Contributions to Finance and Accounting

Zhiqiang Zhang

Fundamental Problems and Solutions in Finance

Springer

Contributions to Finance and Accounting

The book series 'Contributions to Finance and Accounting' features the latest research from research areas like financial management, investment, capital markets, financial institutions, FinTech and financial innovation, accounting methods and standards, reporting, and corporate governance, among others. Books published in this series are primarily monographs and edited volumes that present new research results, both theoretical and empirical, on a clearly defined topic. All books are published in print and digital formats and disseminated globally.

Zhiqiang Zhang

Fundamental Problems and Solutions in Finance

Springer

Zhiqiang Zhang
Business School
Renmin University of China
Beijing, China

ISSN 2730-6038 ISSN 2730-6046 (electronic)
Contributions to Finance and Accounting
ISBN 978-981-19-8268-2 ISBN 978-981-19-8269-9 (eBook)
https://doi.org/10.1007/978-981-19-8269-9

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Everyone makes mistakes, and the world is a mixture of right and wrong; the important thing is that what we pursue is right.
Everyone has irrationality, and reality is a mixture of rationality and irrationality; the important thing is that we pursue rationality.
For the public, the road of rationality is obstructed, but that is their pursuit, an in-blood pursuit.
For professionals, the road to rationality is rocky, but that is their duty, an unshakable duty.

Preface

About 1000 years ago, Su Shi or Su Dongpo (1037–1101) lived in the Song Dynasty of China (960–1279). He is one of the giants of Chinese literature and had high attainments in poetry, prose, calligraphy and painting and even some unique achievements in cooking. A couple of delicious foods in China today are his innovations.

However, due to his bold and forthright words, Su Shi often suffered setbacks and relegation. Su Shi was demoted from Huangzhou to Ruzhou as the deputy envoy of regiment training in May 1084. When he went to Ruzhou, he passed Jiujiang and visited Lushan Mountain with his friends. The magnificent landscape triggered Su Shi's interest to write poetry, so he wrote several poems about Lushan. One of them describes Lushan as follows:

It is peak viewing from one side but a ridge from the other side;
You can see different picture from different point of view;
You cannot find out the true face of Lushan Mountain;
Because you are just in this mountain.

This poem not only describes the wonderful and variable beauty of Lushan Mountain, but also expresses the profound philosophy to cognize the world and enlightens people from generation to generation. It is loved so much by people and is handed hundreds of years down to nowadays. Yes, people are often in some specific situation (profession or affairs), and they cannot help falling into the puzzle of the situation. They should think about problems from the viewpoint of an outsider, and thus draw more objective and comprehensive conclusions.

As the convention in that times, the poem was inscribed (published) on the wall of Xilin Temple in Jiujiang when it was created. Therefore, it was titled "The poem on Xilin wall". The philosophy implied in Su Shi's poem is coincided or consistent with what implied in an ancient Sutra story (sourced from Mahāparinirvāna-sūtra). Here is the story.

Once, several blind people came to the palace to see the king. The king asked them, "what can I do for you?" The blind people replied, "Thanks to your Majesty's kindness, we are born to see nothing. It is said that the elephant is a huge animal, and we are very curious. Please let us touch the elephant with our hands to know what it looks like."

The king was a kind-hearted man, so he readily agreed. He said to his ministers, "go and take an elephant and let these people touch it, so as to meet their wish." The ministers obeyed. After a while, the minister came back with an elephant and said to those blind people, "The elephant is coming, the elephant is coming, come and touch it!"

So these blind people happily walked towards the elephant. The elephant is too big. One of them touched the elephant's legs, one touched the elephant's head, one touched the elephant's nose, one touched the elephant's ears, one touched the elephant's teeth (tusk or ivory), one touched the elephant's body, and one grabbed the elephant's tail. They all thought that what they touched was an elephant. They felt it carefully and thought about it.

After a while, when the king saw that all of them had touched the elephant, he asked, "Now do you realize what an elephant looks like?"

The blind people answered in unison, "Yes!"

The king said, "Tell me all about it."

The person who touched the elephant's leg said, "The elephant is like a big pillar!"

The person who touched the elephant's head said, "The elephant is like a large stone."

The person who touched the elephant's trunk said, "No, no, the elephant is thick and long, just like a huge python."

The person who touched the elephant's ear hurriedly interrupted, "What you said are all wrong. The elephant is smooth, just like a big fan."

The person who touched the elephant's tusk said, "The elephant is like a long, long carrot."

The person who touched the elephant's body said, "The elephant is obviously thick and big, just like a wall."

Finally, the person who caught the elephant's tail said slowly, "You are all wrong! In my opinion, the elephant is thin and long, like a rope."

The blind people did not agree with each other. Everyone thought the right answer was found by himself or herself. So they quarreled endlessly.

The king and his officials burst into laughter. Then the king smiled and said to the blind people, "My dear people, it makes no sense for you to argue now, because none of you have seen the whole elephant and got the whole picture."

Finance is just like the Lushan Mountain or the big elephant, you can get some idea about it from your point of view within a short time, but your perception may not be right because you may just feel or see part of it so far. Even you are not one person, but a group of many persons or people, it is still possible that none in your group get a right understanding about finance. Similarly, the discussions in this book, although I try my best to give readers a whole picture of finance, and the solutions have been closely checked and inspected in numerous situations since they were found, are mainly my understanding, may also be a peephole view. Welcome and thank you to make close examination on them and provide severe criticism and valuable suggestions for improving them.

Finance as an independent science has been exiting for about 70 years since Harry Markowitz published his well-known portfolio theory in 1952. However, partly because it is too profound to be understood, many of the fundamental problems in this subject remain unsolved or even misunderstood, such as how to value a bond, how to value a stock, how to value the bankruptcy risk of a company, how to find the optimal capital structure of a firm, how to estimate a discount rate for debt, equity or total capital based on their respective total risk rather than only systematic risk.

Financial practice has been calling for theoretical solutions to these fundamental problems and the related ones. In the absence of solutions to these problems, more and more practical problems and puzzles have been cumulated, such as what is the fair or reasonable price-earnings ratio, price-book value ratio of a stock or a market in average, what is the appropriate method to measure the bubble of a stock or a market, what is the right approach to calculate a firm's bankruptcy probability and bankruptcy cost arising from debt financing, how much can debt financing add value (tax shield minus the bankruptcy cost) to a company? When should a company stop using debt and turn to equity financing, what is the efficient way to find a firm's optimal leverage or debt ratio under common or simple situations as well as some specific and typical conditions, like the situation with debt guarantees, or with transaction costs, or on book value rather than market value, etc., why does "financial conservatism" in practical financing decisions spread so extensively and persistently, how to measure the risk of equity and debt as well as the whole company and how to derive the discount rate based on the right measured risk? How to make a loan decision or how to determine the interest rate of a business loan, is the discount rate increasing or decreasing or constant over time, how can we estimate the risk-adjusted discount rate for equity capital, debt capital as well as total capital, how can we differentiate and use these discount rates for valuing an asset or evaluating a project, etc.

You may wonder why so many fundamental and important problems left after 70 years' intensive research. As most scholars know, finance has been a very hot research area since last fifties. This implies a lot of research (literature) just does useless work. The above ancient poem and Sutra story have revealed the reason in some sense—because a lot of research just sees or touches part of finance and just reaches a conclusion without knowing the whole picture of finance, and such a conclusion cannot be a right solution to the relevant financial problem. Please note that the whole picture of finance is determined by the right role of finance in social science, rather than the description in past literature, because the past research itself is possible to deviate from the right perception. Anyway, in shortage of sound solutions to those fundamental problems, financial practitioners have to make their decisions relying only on their intuitions or industrial conventions. But obviously, decisions not backed by theoretical understanding or solutions can be right only by chance.

This book records my efforts to deal with these fundamental problems as well as the related problems. Fortunately, most of them are solved, and most of the solutions are closed form models.[1] The book consists of fourteen chapters including some

[1] Most of them are named after ZZ, such as ZZ growth paradox, ZZ growth model, ZZ P/E ratio, P/B ratio and P/S ratio model, ZZ certainty equivalent model, ZZ equivalent coefficient model,

further discussions as special chapters. They are divided into three parts, which is in an order of logic as well as a natural order from easy to hard.

The first part consists of the first five chapters. The main topic of the first part, as expressed by its title, is *Asset Valuation.* Specifically, in addition to the general exploration about finance as the first chapter, we explore the valuation of regular bond and common stocks, respectively, in the second and third chapters. Literally, valuation concerns converting asset risk and return into its value or the consideration of risk and return in valuation. As a matter of fact, the risk is relatively difficult to consider, so the Part I focuses on the consideration of return, that is, the conversion of return into bond or stock value under the assumption that risk can be appropriately represented by the required rate of return or discount rate.

The first chapter explores the fundamental features of finance based on the right role of finance in social science as well as the right role of social science in our society, which is necessary for better understanding and solving financial problems. Naturally, the insights revealed in this chapter are the ideological foundations for the whole book to solve so many tough financial problems. In this sense, this chapter is on the top in importance in this book.

The second chapter explores the valuation of bond, or specifically, the regular corporate bond. Asset valuation is the basic function of finance; regular or common corporate bond is the simplest asset. So regular bond valuation is often supposed to be a solved problem in finance. But this is not true. We discuss on the benchmark for a right solution to this problem and finally work out such a solution, i.e. the valuation models of bond in clean value and dirty value, respectively, factored in the frequency of interest payment and accrued interest.

The third chapter explores the valuation of common stock or equity. For various reasons, common stock valuation is often supposed to be a solved problem in finance. However, a deeper discussion reveals that the prevailing methods or models are neither sound in theory nor feasible in practice. There has been no qualified solution for stock valuation so far in mainstream finance, including the multiple or ratio methods (such as P/E, P/B and P/S methods) as well as the Gordon growth model, though they are well known in every business school and capital market! Even more and bigger surprises are revealed in this chapter, such as no positive perpetual growth rate exists and discounting cash flow (DCF) method does not work for stock valuation.

ZZ capital asset pricing model (ZZ CAPM), ZZ debt or loan pricing model, ZZ equity pricing model, ZZ tax shield model, ZZ bankruptcy cost model, ZZ optimal capital structure model, ZZ optimal capital structure model with various considerations (such as debt guarantee, transaction cost, personal income tax), ZZ overall bankruptcy probability model, ZZ current bankruptcy probability model, ZZ company life expectancy model, etc., ZZ is my initials. This way to name my models is not for self-promoting or self-boasting. It is the need of appellation when exploring the relevant problems, because it is the most convenient way to name and to mention or refer to those models. For example, if you want to compare the Gordon growth model with my model for valuing equity, or if you want to compare the Sharpe model with my model for capital asset pricing, you need a name of my model. In fact, in a sense, ZZ stands for modesty and willingness to be last (alphabetically). If someone does not like the author naming his own model after his own name because of some cultural values, the ZZ in this book can be understood as the acronym of "Zhongguo Zhizao" (means made in China) or "Zhongguo Zhihui"(means Chinese wisdom).

Of course, a real solution to stock or equity valuation is worked out finally which is sound in theory and feasible in practice. This is a brand-new valuation method—valuation based on required payback period, which can overcome most defects of DCF method or the Gordon growth model, and it can improve or rescue the multiple or ratio methods as well—by enhancing both their theoretical soundness and practical valuation power.

One of the reasons why the author can find and solve the problems in bond and stock valuation is that the author correctly understands the basic nature of finance and insists on doing research for the purpose of solving problems. In the past decades, there have been a vast number of studies on bonds and stocks, but a lot of studies deviate from the essence of financial science and do not aim at solving problems. The first chapter elaborates on these basic ideas and understandings. However, surround by the overwhelming number of research literatures that ignoring the decision-making attribute of financial science, it may be difficult for readers to master the decision-making attribute of finance.

Therefore, after the first three chapters, two special chapters on relevant issues are added at the end of this part, namely the two further discussions.

The first discussion explores the relationship between theory and practice (of social sciences) from a perspective, that is, who should take the lead in theory and practice. This problem often causes people's confusion. On the one hand, it is often heard that theory comes from practice, which implies that practice precedes theory; on the other hand, it is often heard that theory guides practice, which implies that theory precedes practice. The discussion reveals some insights. For example, practice should precede theory; theory should solve the problems left by practice; theory should have difficulty and application value; the success of practice may be evaluated by the amount of money made, but the success of theory can only be evaluated by the soundness in theory and convenience and reliability in application.

The second discussion focuses on the classification of science and social sciences and the characteristics of each category. Through discussion, it is found that, comparing with natural science, social science relies more on logical reasoning in research methods, although it does not require high quantitative accuracy. In addition, the more unique aspect of social science is its internal structure. Social science mainly answers the questions about what the world or related objects look like and what we should do. Accordingly, various disciplines in social science can be divided into two categories: descriptive science and decisional science. In fact, making better description is not the ultimate goal; the ultimate goal is to make better decisions. In order to support decision-making, describing the past and the present is not enough, predicting the future is equally important, because decision is future oriented. Therefore, prediction and decision are the two important branches in social science, which is the major difference of social science from natural science in internal composition. However, for a long time, due to various reasons, including blind imitation of natural science, the research on prediction and decision in social science is far from sufficient; but the descriptions of the past and the present are overflowing. On the one hand, many important problems are lack of research and cannot be solved; On the other hand, many specific disciplines in social sciences, such as finance, have lacked theoretical progress for decades.

The second part consists of four chapters (sixth to ninth chapters). The main topic of the second part, as expressed by its title, is *Asset Pricing*. To some extent, asset valuation is the main topic of the whole finance or this book. The right way of valuation is to value an asset based on its risk and return. The firt part focuses mainly on the relationship of return and value, simply assuming the risk is fairly factored into the discount rate. This task of factoring the risk into the discount rate is naturally left to the second part. Based on the right concepts, the discount rate is the appropriate or fair (rate of) return on the relevant asset. As the convention in financial community, the process of finding fair return on an asset (such as stock or equity and bond or debt) is referred to as asset pricing.[2] So the part II is titled as *asset pricing*. This part finds two ways to derive the CAPM incorporating total risk, and solves the problem of asset pricing completely, or provides a total solution to capital asset pricing, that is, the three discount rate models for debt, equity and total capital asset. Some financial puzzles, such as the tendency of discount rate over time, are easily to answer based on those models.

The sixth chapter introduces the concept of option, the principle and method for option valuation or option pricing. Generally, discounting and option pricing are the two complementary valuation tools. Discounting is suitable to value volatile cash flows which are certain in existence, while option pricing is suitable to value contingent cash flows which are uncertain in existence. Further, option as a financial instrument can divide risk from return or isolate risk absolutely. This implies that option pricing is a powerful tool to consider or evaluate risk. In 1997, the Nobel Prize in economics was awarded to the contributors in option pricing achievements, Scholes and Merton, and affirmed the huge application potential of option pricing. Unfortunately, financial research has not fully tapped this potential after that, which is part of the reason for many fundamental and related financial problems remained unsolved so far. Of course, this chapter introduces the option pricing theory and method, especially the Black–Scholes model, in the purpose to pave the way for the subsequent chapters to apply the option pricing method to solve the relevant financial problems.

The seventh chapter explores the fundamental theory and method for determining discount rate. As well known in finance, discount rate can be estimated by a couple of methods, such as the actual or opportunity cost of capital, Sharpe CAPM, and it thus seems that the problem has been solved. Unfortunately, this is not the case. This can be proved by a simple reasoning. Different method leads to different discount rate; however, for an investment or asset, there is only one correct discount rate. This means that only one of these methods can be correct at most. Taking the average of

[2] There are also other different concepts and understandings about asset pricing, which bring a lot of chaos in learning and studying of finance. Similar to the way we understand finance rightly, we would like to define and understand asset pricing based on its right role in finance. That is, asset valuation is to derive the value or fair price of assets; asset pricing is to derive the value or fair price of capitals, which takes the form of rate of return. The asset in Part II is the abbreviation of "capital asset", which represents "capital" rather than "asset".

the results from different methods as the final discount rate used in valuation is a common practice, which is actually wrong because averaging the correct and wrong results inevitably leads to a wrong result.

A little deeper discussion reveals that the discount rate shoulders a mission of accounting for the time (delay) discount and risk (taking) discount; therefore, the only qualified structure for estimating discount rate is "risk-free rate + risk premium". Based on this standard, only the Sharpe CAPM is possible to be right for estimating discount rate, none of the other prevailing methods is right because they cannot account for risk properly. A little further discussion reveals that even the Sharpe CAPM is not sound in theory because it accounts for only systematic risk rather than total risk in determination of the "risk premium". This cannot be right because prudence is the rule of thumb in decision-making. Following the prudence principle, we prefer accounting for more to less risk. Therefore, even portfolio or diversification can eliminate part of the non-systematic risk, it is better or more correct to account for total risk than to account for only systematic risk.

It is not easy to find an effective way to account for total risk in estimation of discount rate. We generalize the problem into a broader context, i.e. how to account for total risk in (investment) decision-making. So the problem is transferred as how to find the certainty equivalent of the future return; but it seems equally difficult as the determination of discount rate. We further transfer the problem into how to find the risk equivalent of the future return and finally solve the problem by resorting to the option pricing model, i.e. the Black–Scholes model. Based on the model of certainty equivalent coefficient, we derive the discount rate model or the new CAPM accounting for total risk, and the model takes the form of "risk-free rate + risk premium". We thus solve the problem, the determination of discount rate, by the new CAPM. In addition, a windfall is the quantification of certainty equivalent and certainty equivalent coefficient, which are consistent in logic with the new CAPM, and can be regarded as a more general solution to account for total risk.

The discount rate model derived in the seventh chapter is a fundamental solution for asset pricing, or for the total (capital) asset pricing. This fundamental solution needs to be extended to asset pricing of debt capital and equity capital. The eighth chapter deals with the discount rate determination for debt capital or debt pricing. Similar to the seventh chapter, the standard to solve the problem is to provide a discount rate model structured as "risk-free rate + (debt) risk premium", and the main task is to quantify the debt risk premium. This is proved to be even more tough. A major obstacle to solve the problem is the bankruptcy cost, which is also a major obstacle to the problem of optimal capital structure (third part or tenth chapter). So we solve the quantification of the bankruptcy cost first and then find the solution to the debt risk premium and then the solution to the discount rate for debt capital or debt pricing as a natural consequence.

A big surprise is that our debt pricing model derived from bankruptcy cost coincides exactly or perfectly with the debt pricing model derived by Robert C. Merton in 1974 via another reasoning path.[3] Put it another way, the two debt pricing models are exact the same in the model form but different in derivation process; hence, they can confirm each other. The confirmation this way is the most convincing proof in the world. The debt pricing model, hence, is undoubtedly correct.

In absent of the transaction cost, the discount rate for debt capital is also the interest charged by the lender. Therefore, the debt pricing or discount rate model can also be used in the loan pricing in commercial bank. However, the model has been given a wide berth since it is published in 1974, because the model was found with two serious and insurmountable drawbacks. One is the interest rate derived based on it is too low, and the other is that the interest rate is negatively related to debt maturity when the leverage of the borrowing company gets high, which seems not consistent with intuition or common perception. Now, since the proof is confirmed after decades, the drawbacks are worth to re-examine closely.

Therefore, we further explore the application of the debt pricing model. We finally get another surprise—the two drawbacks are just the misunderstandings of the model; the model has no serious drawbacks at all, and the model has unexpected functions, that is, it can solve all the problems in loan pricing or loan decision, i.e. the decision on the approvement, the determination of the debt size at most and the interest rate. Put it another way, the model is a very useful tool for loan pricing and decision. We further discuss some specific application issues, such as how to determine the incremental leverage for a specific loan. The debt pricing model is further confirmed to be sound in theory and feasible in practice.

The ninth chapter tries to solve the discount rate for equity or equity pricing. Since the seventh chapter and eighth chapter have solved the discount rate for total capital and debt capital, respectively, based on the two solutions, it is relatively easy to solve the discount rate for equity or equity pricing. The ninth chapter also finds another easier way to work out the discount rate model for total capital and further a unified solution to capital asset pricing, i.e. a model series to debt, equity and total capital asset pricing—all the models have unified structure (risk-free rate + risk premium), based on the same risk measurement (company volatility) as well as the same concept and logic.

Therefore, we can derive the three discount rates in two ways: one is calculating the discount rates based on the same basic ZZ CAPM by input debt, equity and company volatility, respectively; the other is inputting the company volatility, respectively, into debt, equity and total asset pricing models. The former uses one model but three volatilities; the latter uses three models but one volatility. The discount rates are calculated based on the structure of "risk-free rate + risk premium" for sure in both ways. Further discussion reveals that the latter way is sounder in theory, it is more consistent in logic and variables with the solutions to other problems in the subsequent chapters.

A financial puzzle concerning the tendency of discount rate over time has been hotly debated over decades. This puzzle can be solved easily with our solution to

[3] Merton, R. C., 1974, On the pricing of corporate debt: The risk structure of interest rates, *Journal of Finance* 29, 449–470.

discount rates. Based on the three discount rate models, all the three discount rates are decreasing over long time, but the discount rate of debt is increasing over some decades and then decreasing over long run. This further solves a practical problem of how to determine the discount rates in government dominated projects with normally long-life expectancy.

The third part consists of five chapters (tenth to fourteenth chapters). The main topic of this part, as expressed by its title, is leverage and risks. The most serious risk faced by a company should be bankruptcy risk; bankruptcy risk is related to the overall risk and debt ratio of the company. This part discusses the impact of company volatility and debt ratio on the company's value, risk and bankruptcy probability, as well as related issues. Obviously, these problems are very important for operation management, risk management and related equity and debt investment decisions. However, due to the great difficulty, they have not been solved or even rarely discussed for a long time in mainstream finance. After discussion, this part solves the quantification or modelling of tax shield and bankruptcy cost. On this basis, the problem of optimal capital structure, i.e. the optimal debt ratio, is solved; some or most of the related capital structure puzzles are then explained. This part further discusses and solves the capital structure decision-making problems under various specific circumstances, including considering the book value, external investment, external guarantee, abnormal growth, the need for external guarantee, transaction costs, etc. It discusses and solves the calculation problems of the company's current and overall bankruptcy probabilities. These findings promote the progress of financial statement analysis by relating some of the financial ratios to the optimal standard or the bankruptcy probabilities. Based on the previous findings, this part further finds a solution to calculate a company's life expectancy.

The tenth chapter solves the problem of optimal capital structure. It reviews some important findings in past capital structure research and reconfirms the reasonable way to solve the problem of optimal capital structure is to trade off between the tax shield and bankruptcy cost resulted from debt financing. A little deeper discussion reveals that the time horizon is a neglected factor in previous research. Based on the new derived models for valuing tax shield and bankruptcy cost, this chapter finally derives the theoretical solution for the optimal capital structure. The models of the tax shield and bankruptcy cost as well as the optimal capital structure reveal that there is indeed an optimal debt ratio for every company, but the trade-off value or the benefit from the optimal use of debt capital is very small and the loss resulted from over use of debt capital (over leverage) is much larger. It is not worth to adjust the capital structure so long as a company is not over leveraged; and financing is the best chance to adjust capital structure for a company. Based on the solution, various capital structure puzzles can be easily and reasonably explained, like why some firms are inclined to conservatism in financing, or some firms prefer certain pecking order, or some firms have no leverage target, etc.

The eleventh chapter conducts some extensive discussions based on the optimal capital structure model derived in previous chapter. The discussions mainly involve how to adjust the model to obtain the optimal debt or leverage ratio of the company under a variety of specific circumstances. The specific circumstances considered include abnormal growth, bankruptcy expectancy, market value vs book value, guaranteed debt, transaction costs, personal income tax, inter-firm's investments, etc. The relevant solutions are illustrated with a case study of Haier, a home appliance giant in China. Based on the discussions in all the previous chapters, this chapter also discusses an issue with top importance: what is the standard or hallmark of a problem is solved. This issue seems simple and clear but actually neither simple nor clear in finance. This issue belongs to the basic concept of finance and should have been discussed in the first chapter. However, it may not be possible to have a valuable discussion without contacting or discussing some financial issues. Therefore, we choose to discuss it in this chapter and hope it is more convenient to illustrate the relevant opinions or ideas with examples.

Literally, the twelfth chapter explores two issues: bankruptcy probability and company life expectancy. At a deeper level, similar to the optimal capital structure, this chapter is also an extension of the bankruptcy cost model. Bankruptcy probability and bankruptcy cost are the major bankruptcy risk measurement and major concern of the relevant parties of a company. However, neither bankruptcy probability nor bankruptcy cost can be calculated in traditional financial analyses. This chapter solves this calculation problem based on the bankruptcy cost model derived in the eighth chapter. In addition, the bankruptcy probability and cost can be calculated for potential current bankruptcy and overall bankruptcy, respectively. Such new bankruptcy risk analyses are illustrated based on the case of the three home appliances giants in China. Further, the findings in the bankruptcy risks set up the theoretical foundation for us to further predict the company life expectancy. This chapter then explores the estimation of company life expectancy. Based on the queueing theory, the estimation of company life expectancy depends on the long run applicable annual bankruptcy probability. Further discussion reveals the logic from cumulative bankruptcy probability to annual bankruptcy probability further to the long run applicable annual bankruptcy probability. Hence, the problem of company life expectancy estimation is ideally solved in the end.

The third part ends with two special chapters, namely two further discussions, which are, respectively, a financial science enlightenment novel and a more serious discussion on financial issues. The two further discussions seem to be different in themes, contents and styles. In fact, their purposes and functions come down in one continuous line. They help readers understand the context of financial science and understand the deviations and errors in financial science research in recent decades.

In current financial field, on the one hand, most fundamental theoretical problems remain unsolved. On the other hand, problems discussed in published papers or academic seminars are extremely boring and have neither theoretical significance nor application value. However, such studies and discussions have been published in the top journals of finance. This means that there is a misunderstanding of fundamental issues in the mainstream perspective, such as what is a scientific problem in finance. If we cannot find scientific problems, there will be no scientific research at all. In recent decades, it has become very popular in finance to find and study the relationship between any random guessed "factor", such as the length of women skirt, the ratio of the length and width of a company leaders face, the beauty of stock analysts, to stock price or company performance.

In fact, scientific thinking comes before finding scientific problems; only by studying scientific problems and drawing scientific conclusions can we promote the progress of scientific theory. Scientific thinking should belong to logical reasoning at a level higher than common sense. Random guess at or even lower than common sense level is not scientific thinking; the data test or regression for such kind of random guesses cannot be scientific research, and pseudo-scientific conclusion can be drawn at most along this way. Such pseudo-scientific research and conclusions cannot solve financial problems, nor can they promote the progress of financial theory.

This can explain why financial theory has stagnated in recent decades. Unfortunately, the current financial community is not aware of this. Many people even think that financial theory has developed to a very advanced stage. Therefore, the current financial theory research urgently needs the most basic scientific enlightenment. This part reminds people of the importance of scientific thinking and finding scientific problems in scientific research by making up the story of Newton's publishing the law of universal gravitation. It can hopefully enlighten the financial research.

Financial theory has not made substantial progress for decades; financial research in recent decades has failed to produce financial theories that are valuable and qualified to be written into textbooks. This means that there are fundamental deviations and errors in financial research in recent decades.

The first further discussion reminds the existence and severity of the problem in the form of a novel; however, what serious deviations and errors exist in the financial research need to be discussed in detail. Therefore, this is explored in the second further discussion. It turns out that there are serious deviations and errors in the research content, research methods, assessment criteria, evaluation principles and related academic ideas and practices in financial research in recent decades.

The solutions in this book may not be perfect, but they are surely innovative in concepts, sound in theory, simple in the model form and convenient for use. Since there are no reliable solutions to these fundamental problems in current finance books and journals, the solutions in this book can surely benefit students, researchers, analysts and practitioners, especially when they encounter financial problems that are difficult to understand or solve.

The solutions in this book have vast application potential in valuation, debt and equity investment, capital budgeting, risk management, investment and commercial banking and the related business like insurance and debt guarantee, firms' financing and capital structure decision, etc. Specially speaking, venture capitals and investment banks can use the brand-new stock valuation model in the valuation of equity investment, IPO, P&A, etc.; hedge fund and other investment institutions can use the new stock valuation model and theoretical ratio models in stock selections and portfolio decisions. Commercial banks, rating agencies and insurance companies can use the bankruptcy probability and bankruptcy cost models, the debt/loan pricing model as well as the risk and certainty equivalent model to evaluate the risk and make better judgments and decisions. Most non-finance companies can use the new capital asset pricing model series to improve their capital budgeting and project investment decisions and use the optimal capital structure model to improve their financing decisions and risk management.

In the noisy and bustling field of financial research, my research has received little but not no attention. Thank God, I have met many scholars who put the truth before personal interest. Most of them are unknown, but their hearts are like a mirror. Their understanding, criticism and suggestions are full of insight, which is the treasure I gained in my research career, the source of my research inspiration and the source of my research confidence and perseverance. With their presence, my research is no longer a lonely journey.

Special thanks to those academic and practical experts, including K. Thomas Liaw, Aswath Damodaran, Ruqi Wang, Mingxuan Yu, Dongming Liu, Jiwen Song, Shufang Xiao, Wuxiang Zhu, Hua Zhou, Jinghao Ma, Xingbang Liu, Gehong Wang, Xueyun Gu, Xiaoyang Zhuo, etc., for their various helps and insightful comments during my research and writing.

Special thanks to Ms. Yingying Zhang, Ms. Coral Zhou, Ms. Nan Zhang, Ms. Kokila Durairaj, Mrs. Padmavathi Jagadeishkumar, Mr. Saravanan Murugan and other employees in Springer for their enthusiastic help and efforts for publishing this book.

I have been extremely busy year after year for seeking all the solutions in this book and have spared little time to perform my duties in my family. Thanks to my father, my mother, thanks for their consistently support from beginning to end; thanks to my wife and my son, they have to deal with all troubles I left to them and endure various hardships because my research cannot obtain any financial support in current research environment.

Finally, thanks to all people who are kind-hearted to the progress of financial theory, since there has been seldom breakthrough discovery in financial theory over recent 40 years though the financial research has been booming strongly.

All the faults and errors in the book are my own. Your criticisms and comments on my reasoning and models as well as writing are welcome.

Beijing, China
July 2022

Zhiqiang Zhang
jinronglilun@126.com

About This Book

Financial theory is supposed to solve the problems in financial decisions. Unfortunately, none of the problems in financial decisions has been solved in theory so far except the valuation of the European option. Part of the reason is that financial research over recent decades deviated from problem solving to phenomena describing and explaining.

Among those unsolved problems, the fundamental and important problems are selected and solved in this book. They are the valuation of regular bond and stock, the pricing (discount rate) of debt, equity and total capital asset based on total risk rather than systematic risk, the optimal capital structure based on the trade-off between tax shield and bankruptcy cost, the bankruptcy probability calculated based on financial ratios like current ratios and debt ratios as well as the business risk (volatility) of the company, etc. These theoretical solutions or models cover most areas of finance, hence have vast potential for application.

Those solutions are backed by radical innovations in concepts, logic, methods, etc. Of course, most of those solutions are basic or initial solutions rather than final or perfect solutions; further challenges, debates and improvements are needed.

Contents

Asset Pricing

Leverage and Risks

Part I
Asset Valuation

Finance and Its Fundamental Problems

Contents

It needs a right and thorough understanding of finance to find the right and efficient solutions to financial problems. There are actually numerous misunderstandings about finance within and beyond the financial community. Those misunderstandings directly damage our judgment on financial research and financial theory, and are some of the reasons why so many fundamental financial problems remain unsolved after 70 year's intensive research.

1 Financial Theory and Business Practice

Many people regard financial theory as a tool to make money. This is not very correct. As a social science, financial theory has its unique concepts, unique methods as well as unique models for us to understand and solve the relevant problems. If it is regarded as a tool to make money, our attention may be stopped at actual profit from buying and selling financial assets, regardless the underlying financial principles and problems. That is why we may be confused about some financial problems. Actually, those who are successful in making money are not necessarily successful in theoretical contribution, and vice versa

Fischer Black (1938–1995) is a good example. Apart from his famous work on option pricing, his works cover numerous financial problems including CAPM

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_1

and continuous time finance, dividend policy, etc. So many people were struck by the depth of his insight and intuition into economics and finance, even though his manuscripts were rejected often by financial journals. There is a famous story of a presentation by him, to an industry audience, where a smart aleck kid asked him "if you're so smart, how come you're not rich?" Black replied with a smile "if you're so rich, how come you're not smart?"

Making money is actual the main objective pursued by business rather than that pursued by academy. The academic research should aim at solving theoretical problems. This is by no means that the academic research's interest is totally different from that of the business practice. A theoretical problem should be the common or generalized problem representing a group of practical problems. Solutions of theoretical problems should be useful in practical decision-making. Put it another way, financial solutions or models should be helpful for practitioners to make money. However, this is by no means that the inventor of the theoretical solution can also manage to apply the solution to make money, because financial theory is only one of numerous factors in the success to make money. Hence it is not correct to judge a financial theory or model based on how much its inventor makes money.

The importance of practice is over stressed on in most circumstances. Actually, one of the main differences in practice between mankind and animal is just that one is guided by some science or theory, the other is just pure practice. Thus, when we stress on the importance of practice, be careful not to overstate it to the extent that the theory or science is totally unimportant. If we totally abandon science or theory, a foreseeable result is that our society cannot progress anymore; we just repeat the same thing year after year and generation after generation just as what happens in wild animal world. Similarly, if we abandon financial theory or the financial theory stops progress for some reasons, a foreseeable result is that we do not know the value of financial assets, and have to make financial decisions just by guessing blindly and financial crisis will occur more often than not.

In the range of social affairs, the main function of the social science (such as financial theory) is offering support to the relevant decisions making. From a point of view, a theory is an answer to a question or a group of questions. We can find answers to various questions. One feature of a good theory is that the question rather than the answer should come from the relevant practice. This is easy to understand. If the question is not from practice, the theory is not possible to be useful even it is a correct answer. But if the answer is also from practice, the theory is definitely useless, because practice itself has solved the problem already, and the theory cannot offer any additional support or help.

A similar argument is that social theory or science should be consistent with reality. This is not very correct either. Theory can be viewed as a statement with a conclusion based on some premises. The main premises of a theory should be consistent with reality; but the conclusion is not necessarily be consistent with reality. If a theory, from premises to conclusions, is completely consistent with reality, it must be reality itself, and cannot be a theory anymore. Similarly, such a "theory" is impossible to be useful to guide decision-making in practice.

Strictly speaking, besides the conclusions, the premises of a theory should not be completely consistent with reality either. For instance, the extent in rationality of mankind, as one premise, must not be consistent with reality; otherwise, there is no way to derive a correct conclusion. The decision-makers may be somewhat irrational in reality, which implies that practical decision-making may not be right to some extent. The irrationality varies widely across persons, problems and times, which implies the conclusions drawn by different persons at different times may be much different. A man with normal intelligence can get a certain answer to the question like "$3 + 2 = ?$". Yes, it is 5. "5" is the only right answer to this question. However, what is the answer to the same question if the man is irrational? Obviously, the answer may be any number under such an assumption.

Similarly, if we allow some irrationalities existed in our premises when we research or discuss on a problem, we cannot obtain a certain conclusion, do not mention the conclusion is correct or not, because irrationalities are fundamentally uncertain in terms of direction and extent in error or deviation. In addition, a rational answer is useful for guiding the practical decision; the irrational answer is usually useless. For instance, if one day a drunkard worked out that "$3 + 2 = 6$", this information makes no sense because this is uncertain; next time, his answer may be 7 or another drunkard's answer may be 32, etc. Researching on those wrong answers makes no sense. All we need to know is that the right answer is 5, and any answer other than 5 is wrong. That is, academic or theoretical research should be based on the premise of rationality.

We now get some insight. The questions or problems worth researching should come from practice; but the conclusions or solutions are not necessarily from practice. In practice, decisions must be made within a limited time (before deadline); hence the conclusions or solutions may not be correct. However, the theoretical research should pursue the right conclusions or solutions after understanding the question or problem well. The optimal capital structure decision is a good example. Companies in reality should make their capital structure decisions within a given time, regardless whether they really know how to determine an optimal debt ratio. Therefore, the completion of the practical decision-making neither mean an optimal debt ratio is found, nor mean the problem is solved; the real solution of the optimal capital structure as well as other difficult problems relies on the relevant theoretical solutions.

2 Finance in Academic Knowledge Spectrum

Knowledge is the wealth of mankind. In academic circle, most of the knowledge can be divided into two categories: art and science; they are much different from each other. While arts (in various forms) try to give people sensations or feelings, such as music, painting, novel, film, etc.; science try to answer or solve problems directly, such as answer the question about bacteria or aster, etc., solve the problem like environment pollution or traffic congestion, etc. The answer or solution provided by a scientist may be right or wrong; but for art, there is no certain standard about

- Theory
  - Science
    - Natural science: mathmatics, physics, chamistry, etc.
    - Social science
      - Description: history, statistics, accounting, etc.
      - Decision: economics, management, finance, etc.
  - Art: visual art, Auditory art, etc.

Fig. 1 Academic knowledge spectrum

right or wrong. A novel can be ended as a tragedy or as a comedy; the leading role in a movie can be handsome or not so handsome; the main color of a garment can be red or purple. There is no absolute right or wrong for all of them.

Science can be divided further into two categories: natural science and social science. Social science (excluding art) further can be divided into two categories: descriptive science and decisional science. Descriptive science aims at answering questions like "what is it", "what have been done"; decisional science aims at answering questions like "what should it be", "what should be done". Examples of descriptive science like history, statistics, accounting, etc. Examples of decisional science like economics, finance, management, etc. Those classifications of knowledge are shown in Fig. 1.

Finance or financial theory belongs to the decisional science in social science. It offers concepts, theories and models to help and support students, researchers, practitioners, etc. to understand financial issues and to make financial decisions. As a decisional science, finance is different in many aspects from other kinds of science and knowledge.

Finance as a science is different from art. Finance aims at revealing financial principles and solving financial problems. A financial research or theory is much different from an art show or exhibition. We can judge the show or exhibition based on our sensation or feeling, but we cannot judge a financial research or theory based on our sensation or feeling before we really understand the logic or reasoning behind it. It is not necessary to judge a work of art by correctness; but a financial conclusion or theory is possible to be correct or incorrect in most circumstances. The criteria to judge a financial research or theory mainly include: whether it is correct in concepts understanding and logic or quantitative reasoning; how effective and efficient it answers the question or solves the problem; how important and difficult the problem is; and how innovative the research method is, etc.

Finance as a social science is different from natural science. To understand some natural phenomenon or issues, scientists sometime resort to data from observations and experiments. This is often an efficient way in research of natural science, because the conclusion derived from data is usually certain or reliable. However, the data in social science (hereafter referred to as social data), from observation or experiments, is much different. Social data is the results of people's behavior or decision-making. The behavior will surely be affected by people's learning effect, mentality, rationality, accidence as well as other complex factors, hence most (if not all) social data is fundamentally uncertain. This implies that it is often impossible to derive a certain or reliable conclusion from social data.

Things in financial area are similar. For instance, data of stock price are the result of people's behavior of buying and selling in the market. Numerous factors may affect people's buying and selling strategies, such as learning, etc. Therefore, financial theories or solutions should rely mainly on logic reasoning or logic-based research rather than data-based research. Actually, even a social data-based conclusion with consistency can hardly be useful, because no one can tell the premises of the based social data. If we do not know the premises of the conclusion, how can we know when and how to use it? Therefore, we should judge a financial solution or theory based on whether its logic or reasoning is correct and rigorous rather than whether it is consistent with some past data.

For instance, to explore the relationship between interest rate and stock price, we can derive a conclusion based on careful reasoning that they are negative related. However, if we observe the actual data, they are negative related in some cases and positive related in other cases. Our conclusion thus depends on the data we collected. The question is: do you really believe the data-based conclusion? If your instant answer is "yes", which conclusion do you really believe, negative related or positive related? Further, when and how will you apply the two contradicted data-based "conclusions"? The embarrassment aroused from these questions reveals that data-based conclusions in finance are very unreliable; and the basic and final judgment should be based on logic or reasoning rather than unstable past data. This is a very important insight for financial learning, studies and reviews.

Finance as a decisional science is different from descriptive science. The purpose of descriptive science is to describe what our society like or what people have done. So we can judge the descriptive research or theory based on how it conforms to the reality or practice. However, the purpose of decisional science is to guide or improve people's decision. So we definitely cannot judge the conclusion or theory of a decisional science based on its consistence with the reality or practice. Rather, the judgment should be based on its improvement of the prevailing decision making. Most of the descriptive sciences aim at describe the past decisions and their consequences, so they can be past data-based research. However, decision making is always future-oriented. Financial decision is not an exception. Therefore, most (if not all) of the financial research should focus on how or what to do based on foreseeable future. This is very important for understanding and solving financial problems.

As a decisional science, finance is different from other subjects, such as economics and various management subjects. Comparing to each other, economics is a theoretical science, whereas finance is an applied science. As a theoretical science, economics stresses on principles and does not care much about the feasibility of the theory or model. For instance, many economic models use utility as an independent variable; but utility can hardly be estimated. As an applied science, finance must consider feasibility of the model or theory and should avoid incorporating variable like utility in its model. On the other hand, finance is not a pure applied science; it has some branches as pure applied subjects, such as investment, valuation, risk management, investment and commercial banking, etc. Finance offers theoretical supports rather than techniques in detail to these pure applied subjects.

Further, finance is different from other management or decisional subjects. Some management subjects focus on qualitative problems, such as strategy, organization, marketing, etc., whereas finance focuses on quantitative problems. Finance is also different from other quantity-based decisional subjects, such as operation management, management science, etc. Finance is a value-based subject with valuation approach as its main and basic technique, whereas other quantity-based subjects are not. Therefore, finance as a science has unique problem, unique method and unique theory. As a unique subject, it is responsible to deal with and solve special and important financial problems.

3 Fundamental Problems in Finance

3.1 The Objects of Financial Research

Finance as a subject is supposed to deal with problems in financial practice, including those in investment and financing (raising fund or capital). The objects of financial research thus include the objects involved in investment and financing, i.e., assets and capitals.

Investment is buying assets in some forms, such as fixed assets, current assets, the short and long-term investments and the intangible assets. Fixed assets include land, factories, buildings, equipment, machinery, furniture and various vehicles, etc. Intangible assets are another kind of long term assets without physical existence. Goodwill, copyrights, trademarks, and intellectual property are common examples of intangible assets.

Current assets are assets expected to convert their value completely into the finished products or cash within one year or one normal operating cycle. Cash, materials, parts, semifinished products or work-in-process, finished products and various receivables are all typical current assets. Assets are sometime regarded as the opposite to cash. Sales of assets can increase a firm's cash, while purchases of assets will decrease its cash.

The short and long-term investments are the investments on stocks, bonds, real estate, etc. that a company bought to keep for a short or long time. Those assets are not involved in the operation of the company, and are also referred to as financial assets. As a convention in accounting, "current" or "short-term" is normally referred to the maturity within one year. Correspondingly, "noncurrent" or "long-term" is referred to the maturity beyond one year. All the assets or the objects of investment are listed on the left side of a company's balance sheet, as shown in the balance sheet in Table 1.

In contrast, financing is selling assets in some forms, such as issuing stocks, bonds as well as borrowing in various forms. Those capitals raised are listed on the right side

Table 1 A short version of balance sheet of company W

Assets	20 × 7	20 × 6	Liabilities	20 × 7	20 × 6
Current asset			*Current liabilities*		
Cash	220	200	Short-term loans	1800	1500
Inventory	2600	2300	Current portion[a]	5000	4500
Receivables	3000	2500	Payables	7000	6500
Subtotal	5820	5000	Subtotal	13,800	12,500
Fixed asset			*Long-term liabilities*		
Building	15,000	12,000	Bonds	5000	5000
Equipment	20,000	16,000	Bank loan	6000	5000
Less: AD	7000	5000			
Subtotal	28,000	23,000	Subtotal	11,000	10,000
Intangible assets			*Owners' equity*		
Copyrights	1000	800	Common stock	5300	3600
Patents	2000	1800	Retained earnings	9720	6500
Goodwill	3000	2000			
Subtotal	6000	4600	Subtotal	15,020	10,100
Total asset	39,820	32,600		39,820	32,600

[a] Current portion = Current portion of long-term debt

of the balance sheet under the accounting title of liabilities and equity. Total capital in a firm can be classified into two categories: debt and equity. Equity is regarded as the capital comes from the owner of the firm; debt is the capital borrowed from outside lenders.

The debt capitals can be classified into two categories, current debt and long-term debt, depending on their maturities are shorter or longer than one year. Every year, a proportion of long-term debt will be due within one year, which is referred to as current portion of long-term debt. The proper using of debt capital can increase the return of equity capital, hence is also referred to as making use of financial leverage. As the capital borrowed in a levered firm, the debt is mandatory in repayment of the interests and principal, unless in the case that the firm or borrower is insolvent and go into bankruptcy.

Equity capitals include the capitals input by the owners or shareholders and those retained from the past earnings (the other parts of the earnings are dividends paid out). Equity capitals do not require mandatory repayment, and companies can use equity capitals forever. Equity capital represents the shares of ownership in the company. As an owner or shareholder of the firm, one can sell the shares or equity in hand, but cannot require the company to repay the initial capital input. Common equity (stock) is the typical equity capital; preferred equity is also equity capital but has some features of debt capital, such as almost fixed dividends, and may be redeemed at a promised maturity.

The current debt is also referred to as short-term capital, and the long-term debt and the equity are also referred to as long-term capital. The total capital corresponds to the total asset of the firm. This leads to the accounting equation (identity), i.e., total asset = debt + equity. The debt holders have priority in claiming their interest in the firm; the equity holders are second in line. The equity holders have the right to vote for the important decisions in the company but debt holders have not. The priority of payment at the liquidation is debt first, preferred equity second and common equity at last.

In balance sheet, the sum of the left side is referred to as total asset; the sum of the right side is referred to as total liabilities and equity or total capitals. Understandingly, the sum of left side always equals to the sum of the right side, because the right side shows where the capitals come from, and the left side shows where the capitals go. That is why the table is referred to as balance sheet, as shown in Table 1.

The above analyses do not imply the items in balance sheet are the research objects of finance. It is the duty of accountant to prepare financial statements which include mainly the balance sheet, the income statement and the cash flow statement. The accountant is responsible for working out the numbers in balance sheet based on historical costs and balancing the two sides by using accounting knowledge rather than financial knowledge.

Finance is in charge of neither side of the balance sheet. Financial analysts are more interested in the current or fair values rather than the historical costs of those assets, because finance is concerning the decision making in investment and financing, or buying and selling assets. The good decision is to buy asset at an undervalued price and sell asset at an overvalued price. Therefore, to make a good financial decision, it is necessary to know the current or fair value of the relevant assets rather than its historical cost or value.

Specifically, financial analysts are more interested in the aggregate values of those assets rather than the value of every item, such as the fair values of the debt and equity of the company. As a descriptive subject, accounting is backward looking, it focuses on recording what happened in the past; while finance is a decisional subject and need to be forward looking. This is also confirmed by the value determination, because the asset value is determined by the future situation hence featured as forward looking.

Therefore, the duty of a financial analyst is not preparing the balance sheet, but rather, finding or estimating the fair values of the relevant assets, or valuing the assets. Obviously, valuing assets or valuation is much more difficult than finding its historical cost. Hence, finance is much different from accounting in methods or tools package although they are closely related in most companies. Finance as a subject is focusing on developing the theory and methods for valuing asset and making the relevant decision.

The value of an asset is determined by its return and risk. Specifically, the total earnings or profit (earnings before interest and taxes, EBIT) every year are allocated a relative fixed part as debt interest or the return to debt holders. The rest after further deducted as corporate taxes to government is remained to equity holders. What is allocated to debt holders is fixed even when the year the firm generates zero or negative EBIT. Therefore, as a leverage tool, debt financing makes a firm better in a favorable situation, but worse in an unfavorable situation.

Anyway, asset is a type of commodity. Asset value is determined by its return and risk. There are some ways to incorporate the return and risk into the asset value. Hence assets are not too difficult to value, and it is possible to find feasible solutions to problems in finance as well as in valuation. On the other hand, economics takes the whole commodity as its research object. As a decisional science, economics need to value commodity, or is supposed to provide general methods to value commodity. Comparatively, commodity is much more difficult to value. Economists try to value commodity based on labor, utility, the interaction of supply and demand, etc., but have not got a feasible or ideal method so far.

Literally, the total risk of a firm is sharing by debt and equity. However, besides the abnormal situation like bankruptcy, because of the priority pattern in allocation of EBIT, most risk is undertaken by the equity, the debt bears limited risk. Understandingly, on the basis of per unit capital, equity holders bear more risk than average and debt holders bear less risk than average. Average here represents the firm risk. In another word, on the basis of per unit capital, equity risk is higher than firm risk; and the firm risk is higher than the debt risk.

3.2 Fundamental Problems in Finance

As revealed above, Asset value is the key benchmark in financial decision, and asset valuation becomes the theme and basic method of financial theory.

Asset value is determined in a more specific and simpler way than the value of other general items, such as commodity. This distinguishes finance from other value-based subjects, such as economics (especially microeconomics) and makes finance as an independent subject.[1] Financial theory thus gains more feasibility for practical application. Economics deals with values and prices of commodities, and cannot determine the value of commodities merely based on risk and return. Economics aims at revealing more general principle of value determination, hence is perceived as the theoretical foundation of finance.

Asset (financial assets and real assets) value is determined by its future or expected risk and returns (such as earnings and cash flows). Asset value increases as its (expected) returns increase and decreases as its risk increases. This is the basic axiom of asset value determination, and naturally becomes the start point of financial theory and the basic standard to build as well as to judge a financial theory or model.

As an application-oriented subject, finance should offer methods or models to support investment and financing decisions. Investments include security investment (such as stocks, bonds and derivatives, etc.) and real asset investment (such as project,

[1] Finance is just a branch of economics before last 50s. Even today, the difference between economics and finance is not cared much in many situations, scholars specialized in finance are also entitled as economists.

merger and acquisition, etc.). Financing is issuing securities to raise money, which is the opposite of investment. The process of financing is just the process of investment of the counterpart (companies and investors). Finance thus needs to support the relevant decisions by providing methods or models to value securities and projects based on the expected returns and risk.

For valuing securities and projects, the future returns of the asset need being taken into account. There are multiple ways to measure asset returns. Accounting uses some indicators to record a firm's past returns, such as earnings, net income and cash flows. Asset future returns thus can be forecasted in terms of such measures. In fact, forecast (as a subject) rather than finance is responsible for providing the data of future returns (cash flows) of an asset. Finance as a science is mainly responsible for decision method or model, or how to incorporate the asset returns into its valuation. Forecasted data is the input of financial method or model, and the valuation or decision conclusion is the output of financial method or model.

Risk is uncertainty. In finance, risk is specifically referred to as the uncertainty in asset returns (earnings, cash flows, etc.). The return uncertainty can be measured by the distribution (the extent of decentralization) of the possible future returns. Statistics uses standard deviation (σ) or variance (σ^2) to reflect the distribution of a variable. Risk thus can be measured this way. Based on the measurement of risk, finance is mainly responsible for offering efficient methods to incorporate the amount of risk indicator into asset valuation.

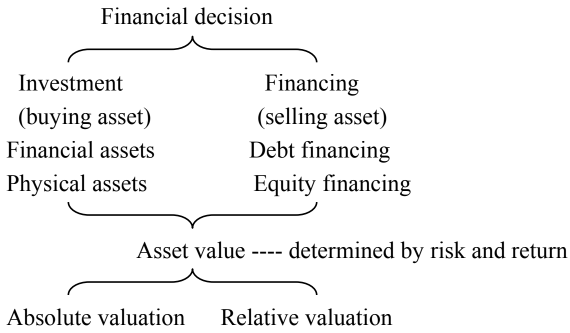
Therefore, given the returns and risk of an asset, such as earnings and their standard deviation of the asset, finance should be able to value the asset and support the relevant decision-making. This implies that financial theory has the following basic tasks:

(1) Modelling the quantitative relationship between an asset value and its returns, or how the asset returns should be considered in valuation.
(2) Modelling the quantitative relationship between an asset value and its risk, or how the asset risk should be considered in valuation.

Absolute valuation and relative valuation are the two basic ways for valuing assets in finance. Absolute valuation values an asset based on its fundamentals, i.e. its expected risk and return. Relative valuation values an asset based on the prices (or values) of other similar (comparable) assets. The philosophy behind absolute valuation is the axiom of "risk and return determine asset value". The philosophy behind relative valuation is a common sense of "similar assets should have similar value (prices)". Figure 2 shows the logic from financial decisions to the financial concepts and methods as well as financial problems.

Currently, absolute valuation is almost equivalent to discounted cash flow method (DCF). According to DCF, value of an asset is the sum of the present value of its future returns (cash flows). Valuing an asset with DCF method, the risk is incorporated into an appropriate discount rate and the returns are measured as the future cash flows of the asset. In addition, contingent claim valuation has become a new and advanced valuation method and applied widely in finance since the breakthrough in option pricing in last seventies.

Fig. 2 Financial problems in financial decisions



Relative valuation is equivalent to ratio method so far, which is mainly used in stock or equity valuation. According to ratio method, value of an asset is close to the product of its value-driver (often earnings, net asset, revenue, etc.) multiplied by a fair ratio.[2] Ratio method obviously depends on the methods to determine the fair ratios, such as a fair P/E ratio or P/B ratio. The fair ratio is often determined at a level around the industrial average or median without much reasoning, which implies that the relative valuation is currently a pure practical valuation method and lack of theoretical soundness (Fig. 2).

Value is the fair or reasonable price which can be used as a benchmark in the decision of investment or buying assets as well as financing or selling assets. On the other hand, the decision of financing or raising capitals is often made based on another benchmark, which is fair or reasonable capital cost. The fair or reasonable capital cost is actually the fair or reasonable price of capital, and also the discount rate. In practice, actual capital cost is often used as discount rate. This is not right; only the fair or reasonable capital cost can be used as discount rate. Anyway, finance is also responsible for providing a fair or reasonable capital cost or discount rate to support the relevant valuation or financial decision.

Transactions in capital market is just a zero-sum game. The investors and issuers (the financing company) are counterparty to each other. Neglecting the transaction cost, the return on investment equals to the capital cost in financing. Therefore, neglecting the transaction cost, the fair return on investment equals to the fair capital cost in financing and further equals to the discount rate. Some important insights about the fair asset return or capital cost were revealed during 1960s. These findings were entitled as "capital asset pricing", and as time passed by, the terminology was simplified as "asset pricing".

Therefore, asset valuation and asset pricing are the two major tasks in finance, which aims at working out the benchmark for asset price and capital price (cost) respectively. The fair cost of capital provided by asset pricing is also referred to as discount rate, which is the key input in asset valuation. Different assets have different valuation issues need to be solved, such as bonds, stocks and derivatives; similarly, different capitals have different asset pricing issues and different discount rate, such

[2] Such as Price to earnings ratio or P/E, Price to book value ratio or P/B, Price to sales ratio or P/S.

as debt capital, equity capital and total capital. The fundamental problems need to be solved in finance consist of those issues.

In addition to the problems in asset valuation and asset pricing, firms have to trade-off between equity capital and debt capital and choose an optimal capital mix. This implies that the problem of optimal capital structure (the ratio of debt in total assets or leverage ratio) is also an important problem and need to be solved in finance. The debt or leverage ratio determine to a large extent the bankruptcy risk of a company. The problem of optimal capital structure is thus close to the risk analyses and risk management of a company.

In summary, finance as a science has to solve the following fundamental problems:

(1) How to value a bond on its pure and dirty price bases.
(2) How to value a stock based on the right measurement and consideration of its future return and risk.
(3) How to determine the fair or theoretical ratios (P/E, P/B, P/S) of a stock.
(4) How to value a contingent claim, such as an option as well option-like opportunities.
(5) How to measure and consider the risk of an investment or an asset.
(6) How to estimate a discount rate (or fair return) for debt capital, equity capital and total capital respectively based on the relevant risk.
(7) How to measure the value added and value loss from debt financing respectively, i.e., the tax shield and the bankruptcy cost.
(8) How to determine the optimal capital structure of a firm.
(9) How to measure the bankruptcy or default risk of a firm.

We shall discuss the first three problems ((1), (2), (3)) in the first part (asset valuation), the second three problems ((4), (5), (6)) in the second part (asset pricing), the last three problems ((7), (8), (9)) in the last part (leverage and risks).

3.3 Fundamental Problems Remain Unsolved

We sort out nine fundamental financial problems in last section. Unfortunately, most of them are unsolved in theory so far, with one exception of the contingent claim valuation, which is solved by the Black-Scholes model. To make things worse, most of the problems are pervasive in finance, since they are all fundamental problems in finance, such as any one problem from (1) to (9). Any one of these problems remain unsolved may hinder us from solving most of other financial (investment and financing) problems in theory and in practice. For instance, all stock and project investment as well as stock or bond issuing need an appropriate discount rate if the relevant decision making is based on absolute valuation.

As DCF method is almost the only valuation method sound in theory, if the future returns (given a best forecast) or its risk cannot be correctly considered, the result derived from the DCF will be unreliable. In this case, how can we rely on the financial analyses to make investment and financing decisions? Ratio method is also widely used in practical stock valuation. Since there are so many firms in an industry, the industrial average of the ratio is impossible to be a good estimation of the fair ratio for a specific firm. Similarly, there is no reliable method so far to determine an optimal capital structure of a firm. This implies that the decisions for a commercial bank to offer loans and the decisions for a non-finance firm to raise debt and so on are all made without solid foundations.

The capital structure is related to bankruptcy or default risk. Literally, companies care risk management, including credit risk management. However, in reality, most companies seem have no risk indicator in risk management. In financial theory, the basic risk index is volatility, that is, the standard deviation of asset return. But this standard deviation or volatility is not an intuitive risk indicator. The most intuitive and convenient indicator is bankruptcy probability. However, how to deduce the bankruptcy probability from the volatility is an unsolved problem in financial theory. In reality, the bankruptcy or default risk is assessed by credit rating agencies. As the final assessment result provided by the rating agencies is the discrete risk grades, rather than the continuous variable measuring probability of bankruptcy or default, this operation fully shows that the quantitative problem of risk assessment has not been solved.

Perhaps the biggest gap in nowadays finance is the absence of a theoretical sound discount rate model. The discount rate represents the consideration of risk in finance. As DCF is an indispensable method in most financial analyses, and discount rate is an indispensable variable in DCF, the discount rate has most widely influence in finance. The current prevailing method is to determine the discount rate based on the Sharpe CAPM, which assumes arbitrarily all the non-systematic risk is eliminated by full diversification. This does not follow the rule of thumb in decision making, i.e., the prudence principle. According to the prudence principle, if we are not sure about the return and risk, we should consider adequately more of the risk and less of the return. Similarly, if we are not sure the proportion of non-systematic risk being eliminated by diversification, we would rather assume the proportion being eliminated is zero. However, it is assumed as 100% in the Sharpe CAPM.

Although millions of related decisions are made every day in the world, most of them have to be made intuitively rather than under the guidance of some reliable financial theories or models. The following chapters reveal that some of the prevailing financial theories or models have fundamental defects in logic. Financial studies after Gordon, Sharpe, Modigliani and Miller gradually turn to empirical research, which focuses on explaning past sample data rather than finding solutions to important financial problems. Therefore, these fundamental problems remain unsolved over decades. Obviously, solving these fundamental problems is urgently needed for strengthening the mansion of financial theory as well as for supporting the decision -makings in practice. This is the intention of this book. Fortunately, we find solutions to most of these fundamental problems in this book.

Discounting and Bond Valuation

Contents

Asset value is the core of financial decision; Ordinary bond is the simplest security in finance. However, for a long time, while the financial theory and practice have become more and more complex and abstruse, the problem of ordinary bond valuation has not been solved, or to be exact, has not been solved correctly.

1 The Methods of Valuation

As indicated in Chapter "Finance and Its Fundamental Problems", asset value is determined by its risk and return. Literally, asset should be valued based on its risk and return. In reality, this is one way to value asset, which is referred to as absolute valuation. Another way to value asset is relative valuation, which values an asset based on the prices of similar or comparable assets.

Option pricing method emerged as a new valuation tool after the breakthrough of option pricing in the 1970s, which is often added as the third valuation method

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_2

in textbooks. Conceptually, there cannot be a valuation method other than absolute valuation or relative valuation; option pricing is actually a kind of absolute valuation.

As a starting point, we introduce here the traditional absolute valuation and relative valuation; the introduction of option pricing method will be left to Chapter "Option Pricing and Valuation of Contingent Cash Flow".

1.1 Absolute Valuation

Absolute valuation attempts to find the intrinsic or true value of an asset based on its future risk and return. According to the mainstream understanding, the absolute valuation is equivalent to discounting or discounting cash flow (DCF) method. DCF method derives asset value by discounting its future returns (cash flows) at a discount rate which is supposed to be determined based on the risk of the future returns.

Absolute valuation as a valuation method is obviously sound in theory, since it accords to the basic axiom or principle of value determination, i.e., the risk and return determine the value of an asset. Take the DCF or discounting method as an example, the valuation formula or model comprehensively considers the risk and return of an asset, with return reflected by the future profit or cash flow, and risk incorporated in the discount rate (k), just as shown in Eq. (1). The valuation result or the value of the asset is positively related with return and negatively related with risk. Therefore, the absolute valuation has sufficient rationality in theory.

$$\text{Value} = \sum_{t=1}^{n} \frac{CF_t}{(1+k)^t} \tag{1}$$

Obviously, for using the absolute valuation or discounting method, it is important that the return and risk of assets be measured reasonably.

1. Measurement of return

In finance, return refers to the net income from operation and investment, normally on yearly basis. Profit and cash flow are two specific indicators of return. Generally speaking, cash flow is more accurate than profit in terms of the time and size of return. Therefore, cash flow is more often used as the return indicator in valuation.

In investment, the annual return normally includes two parts: the cash flow and the capital gain or appreciation which is the change of asset value during the year. The capital gain and cash flow can be measured on absolute and relative bases. Accordingly, there are two forms of return: relative return and absolute return. The relative return is also known as the rate of return or percentage return. Thus:

$$\text{Return(rate)} = \text{cash flow(rate)} + \text{capital gain(rate)} \tag{2}$$

The annual cash return is interest for bond investment and cash dividend for stock investment. The annual capital gain is the change in market price during the year for both. The capital gains may be positive or negative depends on the change of the price; as a result, the total return may be larger or smaller than the cash return. So does the investment in real project, in addition to the profit or cash flow obtained during the year, the project itself may also have some capital gains.

For future oriented decision-makings, return in finance more often refers to the predicted or expected return from investment or operation, that is, the weighted average return of various possible returns during a future year:

$$\mathrm{E(r)} = \sum_{i=1}^{n} r_i p_i \tag{3}$$

where,

E(r) Expected return;
r_i The ith possible return;
p_i Probability of occurrence of the ith return;
n The number of possible returns.

No matter how to estimate or predict a future return, it is the expected return literally, or the weighted average of various possible returns in a future period; Past "objective data" does not represent future "objective expectations". Please note that to estimate the expected return, we must fully consider various possibilities in the future.

2. Measurement of risk

Risk in finance is specified as the uncertainty of future return. In a given period (year), various possible returns are distributed on both sides of the expected return. It is conceivable that the more dispersed the possible returns, the greater the risk.

The risk can be defined as average distance of possible returns. Such an average distance can be measured by the standard deviation (σ), as shown in Eq. (4).

$$\mathrm{s} = \sqrt{\sum\nolimits_{i=1}^{n} p_i [r_i - E(r)]^2} \tag{4}$$

where,

E(r) Expected return;
σ Standard deviation of possible future returns. Other variables are the same as before.

The standard deviation thus turned into a fundamental risk indicator in finance which is also referred to as volatility. The larger the volatility, the greater the risk. The square of the standard deviation is variance. Obviously, variance can also be used to measure the risk.

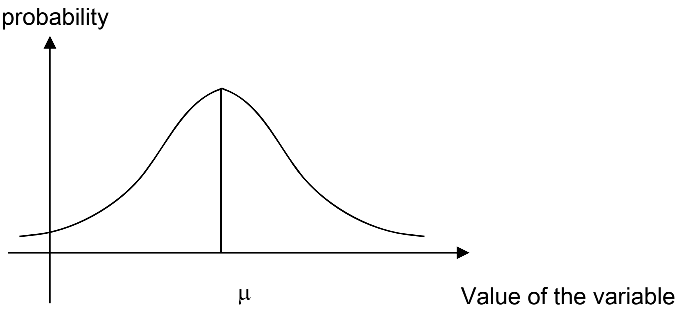


Fig. 1 Mean and standard deviation of normal distribution

In reality, most variables conform to or approximately conform to the normal distribution. Normal distribution variable has possible values centered on the mean (expected value), and distributed symmetrically on both sides of the mean with the probability gradually decreasing as the possible value gets further from the mean, as shown in Fig. 1. The horizontal axis represents the value of the variable, the vertical axis represents the probability of each value. The points representing the probabilities of those possible values form a bell shaped curve.

For continuous variables, the number of possible values is unlimited, and the probability of each specific value approaches zero; what is of practical significance is the probability of a certain value range. In Fig. 1, the probability of a value range equals to the relevant area under the bell curve. The total area under the bell curve represents the sum of the probabilities of all values of the variable, which is equal to 100%; The probability that the value of the variable is greater than or less than the expected value (μ) is 50% respectively. The probability that the value is within the range of one standard deviation (σ) around the mean is about 68.27%; The probability within the range of two standard deviations around the mean is about 95.45%; The probability within the range of three standard deviations around the mean is about 99.73%.

When the mean value is 0 and the standard deviation is 1, the normal distribution is called normal standard distribution. In reality, most variables that conform to the normal distribution have different mean and standard deviation, but the probability calculation can be simplified through standardization. The independent variable value x is subtracted from the mean and then divided by its standard deviation to obtain the new variable Z, and the distribution of this variable Z conforms to the standard normal distribution. In this way, the probability of variable value in a certain range can be obtained by comparing with the standard normal distribution.

The function represented by the bell-shaped curve is called probability density function, i.e. f(x); the area below the bell-shaped curve to the left of a value x of the independent variable is called the cumulative probability, and the function for calculating this cumulative probability is called the cumulative distribution function, i.e. F(x). According to the basic characteristics of standard normal distribution, $F(0) = 0.5$; $F(x) + F(-x) = 1$; When $x < 0$, $0 < F(x) < 0.5$; When $x > 0$, $0.5 < F(x) < 1$.

3. From risk to discount rate

The return, either measured by profit or by cash flow, can provide the data used directly in valuation. However, the risk, either measured by volatility or by variance, cannot provide the data used directly in valuation, because in absolute valuation or discounting method, risk is incorporated via discount rate rather than a risk indicator, neither volatility nor variance.

This implies the volatility should be further turned into discount rate. However, this is proved to be an arduous task. Mainstream of finance has not solved this problem so far. Put it another way, the risk is measured by standard deviation or volatility, and considered by discount rate, but the relationship between the volatility and the discount rate is not found. Therefore, there is a big logic gap here.

Anyway, it is a right way to value an asset or make a financial decision by the trading-off between risk and return. We will deal with the problem of discount rate in part II or Chapters "Certainty Equivalent, Risk Premium and Asset Pricing, Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing and Capital Asset Pricing: An Easy and Unified Solution". In this chapter and next chapter, for bond and stock valuation, we just suppose the discount rate can be determined in a proper way based on the risk of the relevant asset.

1.2 Relative Valuation

Different from absolute valuation, relative valuation gives up the pursuance of true value and no longer value an asset based on its risk and return. From a point of view, relative valuation is only a convenient method rather than a professional method. Simpleness is the only advantage of relative valuation, and over simpleness is its serious disadvantage.

The idea of relative valuation is to determine the value of a target asset based on the market price of similar or comparable assets. According to the current mainstream understanding, relative valuation is basically equivalent to the ratio or multiple method, which involves calculating multiples and ratios, such as the price-to-earnings multiple, and comparing them with the same multiples of similar companies.

The relative valuation believes prices among assets should be relatively reasonable. When the prices of assets are relatively reasonable, there should be a roughly equivalent proportion between the value or price of different assets and their important value drivers or factors. Here, the sizes of the drivers or factors are called comparable indicators, and the similar assets are called comparable assets. The reasonable relationship is:

$$\begin{aligned}&\text{Value of asset under valuation/comparableindicator of asset under valuation}\\&\quad=\text{Price of comparable asset/comparable indicator of comparable asset}\end{aligned} \tag{5}$$

Obviously, the key issues in using relative valuation method are choosing comparable assets and comparable indicators. In practical application, only the most important value driver can be selected as the comparable indicator hence the valuation can be simplified. The annual earnings, book value of assets are the common comparable indicators for firm or stock valuation. The comparable indicators vary across the assets under valuation.

Those assets with prices easy to observe in the market and similar to the asset under valuation are easily chosen as the comparable assets. For valuing a company or its equity, the comparable assets are usually those listed companies similar to the company under valuation in various aspects, such as the industry or sector, the business mix and model, the development history and size or scale, the strategy and prospects.

The valuation ratio or multiple is "the price of comparable assets/comparable indicator of comparable assets". For equity or stock valuation, the three most common ratios are: price earnings ratio (P/E), price book net asset ratio (P/B) and price sales ratio (P/S). You can estimate these multiples either on total amount basis or on per share basis, as shown in Eqs. (6), (7) and (8).

$$\mathrm{P/E} = \text{total stock value/net profit} = \text{price per share/earnings per share} \tag{6}$$

$$\mathrm{P/B} = \text{total stock value/net assets} = \text{price per share/net assets per share} \tag{7}$$

$$\mathrm{P/S} = \text{total stock value/total sales} = \text{price per share/sales per share} \tag{8}$$

Therefore, stock can be valued based on these ratios as:

$$\mathrm{P} = \text{earnings per share} \times \mathrm{P/E} = \mathrm{E(P/E)} \tag{9}$$

$$\mathrm{P} = \text{net assets per share} \times \mathrm{P/B} = \mathrm{B(P/B)} \tag{10}$$

$$\mathrm{P} = \text{sales per share} \times \mathrm{P/S} = \mathrm{S(P/S)} \tag{11}$$

P/E ratio is used most commonly among the three ratios. It is sounder in theory, since it values a stock based on its returns in current year, while the other two ratios value a stock based on neither its returns nor its risks. But the P/E ratio has a disadvantage, it cannot be used when the earnings fluctuate too much or become negative in current year. The actual earnings should be replaced by rationalized earnings to ensure its representativeness when the earnings are too volatile.

P/B ratio can avoid such a problem, because compared with earnings, the book value of net asset is positive most of the time and more stable as well. Therefore, if a company has a large scale of fixed assets and unstable earnings, it may be suitable to be valued by P/B ratio. The P/B ratio is lack of soundness in theory, since asset value is determined directly by its return and risk, rather than its book value; there is no reliable relationship between the size of net asset and its value.

The P/S ratio is mainly applicable to the stock valuation of emerging high-tech or network companies. It is difficult for such companies to apply P/E ratio and P/B ratio, because the company's visible asset is often small in size, and their profits are often negative in current year as well as in a foreseeable future. In this case, the sales may determine the market share as well as the future potential. At the same time, the sales are seldom to be negative and suitable as a comparable indicator.

As well known in finance, the relative valuation is mainly used in stock or equity valuation, and seldom used in bond valuation. Why? The absolute valuation is sounder in theory, hence in priority for method selection. But the absolute valuation is more sophisticated and time consuming, so it is suitable for simple case. Compared with other assets and securities, the future return of ordinary bonds, its interest and principal, are almost sure, so it is the simplest case. Therefore, the ordinary bonds are naturally valued with absolute valuation, and often become the starting point for the application of valuation methods in textbooks and classes of finance.

2 Bond Valuation I

A bond is a certificate of debt used to raise money for the issuer. Large corporations and governments are the major issuers of bonds; and various banks are the major investors or purchasers of bonds. The company or government issuing bonds are borrowing money from the investors; and are supposed to pay interests and principals to the investors during the life of the bond for borrowing' their money.

2.1 Types of Bonds

There are various types of bonds, such as corporate bonds issued by large companies and government bonds issued by governments institutions; those bonds can be further divided into more detailed types.

For instance, the government bonds can be divided into municipal bonds or treasury bonds, which are issued by municipal authorities and federal government agencies respectively. Bonds issued by federal government have maturities greater than 10 years and are normally regarded as default free, this is not the case for municipal bonds, because the municipal governments are possible to go bankruptcy.

The corporate bonds may be guaranteed or asset-backed or just credit-backed. Anyway, corporate bonds may be default hence have credit or default risk. The corporation bonds with higher default risk are also called junk bond. Specifically, a bond rated as lower than BB is said to be of speculative-grade or a junk bond. Most investors especially the risk-averse investors should avoid bond in this category.

The corporate bonds illustrate more characteristics of bonds in dimensions of risk and return. In the dimensions of risk, apart from the default risk, corporate bonds have interest rate risk and reinvest risk as well; in the dimensions of return, corporate bonds normally have both cash return and capital gains.

For example, a bond agreement may have following typical terms.
The par or face value is \$1000;
The interest or coupon rate is 6%;
The time to maturity is 5 years;
The interest is paid semi-annually or two times every year.

If you invest \$5000 into this fixed-rate bonds, you will receive 3% in interest (\$150) per half year and made \$1500 in interest over the 5 years; as the bond expires (at the end of year 5) you will get the original \$5000 back if no default.

As shown in above example, capital raised by issuing bond is debt capital, just as other debts borrowed from banks, etc., the issuer is obliged to pay interest and principal at due time. On the contrary, capital raised by issuing stock is equity capital, the issuing company has no obligation to pay interest or principal. The investors are the counterparty of the issuers in the market. From the investors' point, bonds investment is relatively safe. In most of cases, bonds will pay a fixed rate of interest, which is determined in advance. This means that if you hold on the bonds until the maturity date, then you obtain exactly what you know upon you making the investment decision. When a company goes bankrupt, the bondholders have priorities in claim over the property of the issuer before the stockholders. As the amount of bond interest is normally fixed, bond is a type of fixed-income security.

Sometimes the issuer will add some attractive features to the bond to lower its interest cost. One common feature is the flexibility to convert it into the common stocks of the issuing company. Other special terms like redeemable or callable bond, puttable bond, etc., are also used in bond design. These special terms are actual options and need option pricing method to value. We do not intend to involve those specially designed or complex bonds in this chapter, but focus on ordinary bonds. The ordinary bonds here refer to fixed rate bonds without special terms. The valuation of ordinary bonds is the basis of the valuation of various complex bonds. If the simple ordinary bonds cannot be valued correctly, complex bonds valuation cannot be expected.

2.2 Basic Model for Bond Valuation

Bonds have relative certain interest, face value and expiration date at which the investors can receive their principals, put it another way, bonds have relative certain future cash flows. Hence, bonds can be valued conveniently by the DCF method, and need not to consider relative method or ratio method for valuation.

1. Elements of ordinary bonds

As a general rule, the value of bonds depends on their risks and returns. When using the DCF method to value a bond, the risk of the bond determines the discount rate used in the calculation; the return of the bond depends on its face value, interest rate, maturity and the frequency of interest payment of the bonds.

(1) Bond face value M. The face value of the bond is also referred to as par value. It is the principal that the issuer promises to repay when the bond matures. It is also the basis for calculating the annual interest of the bond.
(2) Interest rate r. Also known as coupon rate, it is the ratio of the annual interest of the bond to the face value. The issuer pays the bond interest according to this interest rate during the life of the bond, that is, the annual interest I = rM. This is usually paid once per year, but some bonds pay interest semiannually or even quarterly.
(3) Bond maturity n. Also known as the time to expiry, which is the time before the issuer repays the principal of bonds as promised. Bond maturity has two specific meanings. First, the life of bond as stated in the bond agreement; Second, the remaining time of the bond life. Note that the remaining time to maturity is shortened over time, and the remaining time to maturity rather than the original time to maturity determine the bond value at any time in the bond life. In another word, other things being equal, the bond value will be changing over time before the expiring date.
(4) Frequency of interest payment m. The annual interest of a bond can be paid in several installments. Paying interest more than one time a year provides more interest appreciation or time value of money; therefore, investors can actually get more return. Paying interest once or twice a year is the common case in reality.

The above four elements together with the risk determine the yield and the value of a bond. As a general rule, the risk of the bond is incorporated into the discount rate, which is used to discount the future returns of the bond.

As a convention in bond issuance, the coupon rate is estimated based on the risk of the bond, so that the coupon rate equals to the discount rate and the bond can be sold at par, that is, the price equals to the par value of the bond.

1. Common valuation model

As a general rule, the value of a bond is equal to the total present value of all future interests (deferred annuity) of the bond plus the present value of the principal (face value). Let k represent the discount rate, I represent the annual interest, then, the value or reasonable price of the bond, B, is:

$$\mathrm{B} = \sum_{t=1}^{n} \frac{I}{(1+k)^t} + \frac{M}{(1+k)^n} \tag{12}$$

Equation (12) is the basic model of bond value in the case of paying interest once a year. Based on this model, it is necessary to consider the bond value when the interest is paid twice or more times a year.

As a convention in bonds design and issuance, no matter how many times the interest is paid in one year, the instalments are equally divided in amount and evenly distributed in time. For example, for a bond with annual interest I in n years, when the interest is paid twice a year, it is equivalent to semiannual interest I/2 in 2n half years.

For the case of semiannual interest payment, the Eq. (12) is changed as: the annual interest is divided by 2 and replaced by the interest of half year, the annual discount rate is divided by 2 and replaced by the discount rate of half year, and the number of years multiply by 2 and replaced by the number of periods in terms of half year as a period. Then, the bond value is:

$$\mathrm{B} = \sum_{t=1}^{2n} \frac{I/2}{(1+k/2)^t} + \frac{M}{(1+k/2)^{2n}} \tag{13}$$

Of course, if the interest is paid m times a year, the bond value is:

$$\mathrm{B} = \sum_{t=1}^{mn} \frac{I/m}{(1+k/m)^t} + \frac{M}{(1+k/m)^{mn}} \tag{14}$$

Assuming the bond value equals to its price, based on the above models, a discount rate can be derived out. Such a discount rate is also called the yield to maturity of the bond. The yield to maturity or simply the yield is actually the total expected return of investment in the bond at current price. The yield equals to the actual return under two conditions: the investor buys the bond at current price and holds it until its maturity; the issuing company pays all the interests and principal on time or no default.

In the case of a government bond issued by central government, which is default free, the yield to maturity derived from the above models is referred to as risk free interest rate or simply risk free rate. Risk free rate is a very important benchmark in finance, especially for consideration of risk, such as the determination of discount rate, etc. Please note that investment in government bond has risks other than default as well, such as interest rate risk, etc. So the risk free rate is just an abbreviation of default risk free interest rate.

The above models are usually the content of bond valuation on many occasions, such as many finance related textbooks and classes. However, it is little noticed that the time variable, t, in these models can only be integer. Put it another way, only limited bond values can be derived out by using these models. Specifically, the bond can only be valued at the beginning of each year or each period rather than at any time during its life span. Therefore, these models are far from enough for bonds valuation.

2.3 Fundamental Problems Remain Unsolved

Obviously, the value of bonds at any time includes not only the principal and interest of all future integer or whole periods, but also the accrued interest of current non integer periods. Accrued interest refers to the interest generated from the previous interest payment date to the transaction date.

It is understandable that this interest is only book return, that is, so much can be counted in bookkeeping; It's not the actual return, because the interest payment date is still days or months away. The bond price (value) excluding accrued interest is called the net or clean price (value) of the bond; The bond price (value) including accrued interest is called the total or dirty price (value) of the bond.

At present, most bond markets in the world quote at the clean price and trade or settle at the dirty price. It is easy to understand that bonds should be traded or settled at the dirty price, since the accrued interest is also belong to the seller in addition to the interest in the integer or whole periods. But why it is quoted at the clean price? That is because the change of the dirty price is misleading as the accrued interest is also included, while the change of the clean price reflects the true change of the bond.

Specifically, the dirty price will go up along with the accrued interest accumulated over time until the next interest payment, and will go down suddenly after every interest payment. These ups and downs are not true change in value of the bond and misleading to investors. On the contrary, the clean price is unaffected by the accrued interest, and the change in clean price truly reflects the change of bond value, which is helpful for investors to analyze and judge the trend of bond price or value.

Therefore, when the risk and return of the bond remains unchanged,[1] its value should remain unchanged, so does the clean price or value of the bond. This is the correct standard of the clean value model; the dirty value of the bond is just sum of the correct clean value and the correct accrued interest. Of course, both the clean value and dirty value need to consider interest payment frequency correctly.

As a benchmark case, consider a company issued bonds on January 1, 2022, with a face value of 1000 dollars and a maturity of 5 years (due on January 1, 2027). The coupon rate is 6%. From year 2022, the interest will be paid (60 dollars) on December 31 once a year. Assuming that the appropriate discount rate is 6%, the bond is sold at par, and both the value and market price on the day of issuance are 1000 dollars.

Assuming that all conditions including the discount rate remain unchanged, regardless of the time value of money in the year, i.e. just make a rough calculation, the value of the bond should be 1030 dollars on June 30, 2022 and 1060 dollars on December 31, 2022. Of course, by January 1, 2022, because the interest of 60 dollars has just been paid, the next interest will have to wait one year and no accrued interest at all, the bond value should be back to the clean level or 1000 dollars. Again, the value of bonds on June 30, 2023, December 31, 2023 should also be 1030 dollars, 1060 dollars respectively.

[1] Assume the risk free rate also remains unchanged.

Here, 30 dollars on June 30 and 60 dollars on December 31 are the accrued interest of bonds. When the risk-free interest rate in the market remains unchanged and the company's own risk and return remain unchanged, the bond value gradually rises from 1000 to 1060 dollars during every year, which is entirely the influence of accrued interest and does not represent the change or trend of the bond value.

Anyway, the above 30 dollars is only a rough calculation. So, how to calculate the accrued interest? Here is a popular formula or model in use:

$$\begin{aligned}\text{Accrued interest} = \text{face value of bonds} \times \text{Coupon rate} \\ \div 365 \times \text{days for accrued interest}\end{aligned} \tag{15}$$

For example, in the above example, by June 30, 2022, if calculated in detail, the accrued interest is not exactly 30 dollars. Days for accrued interest: 31 + 28 + 31 + 30 + 31 + 30 = 181 days. Therefore, the accrued interest of the bond is:

$$1000 \times 6\% \div 365 \times 181 = 29.75 \text{ dollars}$$

Unfortunately, under the prevailing and mainstream method, neither the accrued interest nor the interest payment frequency is incorporated correctly into the bond value or neither of them is solved correctly.

3 Bond Valuation II

This book aims at solving the fundamental problems unsolved in finance. So does this chapter. Now the problem is how to consider the accrued interest and interest payment frequency in bond valuation. We discuss on the accrued interest and interest payment frequency separately first and then put them together for a comprehensive solution, i.e. the correct models for clean and dirty valuation.

3.1 The Accrued Interest

Based on the previous analyses, we know that, when the risk and return of the bond remains unchanged, a correct clean value model should provide a constant clean value. This is the standard that the clean value is calculated correctly.

Taking again the previous base case as an example, a company issues bonds on January 1, 2022, with face value of 1000 dollars, maturity of 5 years, coupon rate of 6%, and pays interest once (60 dollars) on December 31 every year. The appropriate discount rate is assumed to be 6%. The accrued interest, total or dirty value and clean value calculated at the end of each month are shown in Table 1.

Table 1 The accrued interest, dirty value and clean value of the base bond

Date	Days accrued	Days remained	Accrued interest	Dirty value	Clean value
January 1st	1	364	0.164	1000.160	999.995
January 31st	31	334	5.096	1004.961	999.865
February 28th	59	306	9.699	1009.463	999.765
March 31st	90	275	14.795	1014.471	999.677
April 30th	120	245	19.726	1019.342	999.616
May 31st	151	214	24.822	1024.399	999.577
June 30th	181	184	29.753	1029.317	999.563
July 31st	212	153	34.849	1034.423	999.574
August 31st	243	122	39.945	1039.555	999.610
September 30th	273	92	44.877	1044.546	999.669
October 31st	304	61	49.973	1049.728	999.755
November 30th	334	31	54.904	1054.767	999.863
December 31st	365	0	60.000	1060.000	1000.000

Note

(1) Days accrued is number of days for accrued interest or number of days the bond was hold by the seller; Days remained is "365-Days accrued".

(2) The bond value on December 31st, 1060 dollars, will be fully owned by the buyer. The fair transaction price or the dirty value of the bond should be the present value of the 1060 dollars at the transaction date (before December 31st), i.e., 1060/1.06^(Days remained/365). From the transaction date to the date of interest payment, this value goes up to 1060 dollars as the discounting effect becomes less and less, and disappears completely on December 31st.

(3) Conceptually, the clean value should be "dirty value—accrued interest". The data in the last column of Table 1 are derived this way.

As the discount rate equals to the coupon rate and the interest paid once a year, the bond in base case is obviously priced at par when fair valued. Its value is 1000 dollars. This should be its clean value. Of course, other things being equal, as time passed by, the clean value of the bond should remain constant as 1000 dollars.

This implies the clean values in Table 1 are not quite right, since other things being indeed equal, but the clean values move down first and then up, the results are all less than 1000 dollars except that on December 31st.

What is wrong with above calculation?

The biggest deviation happened in the mid of the year, or the June 30th. This implies that the problem is caused by discounting.

It is easy to find that the consideration of accrued interest is incorrect. The interest payment date of the bond is stipulated in advance and agreed by the buyer and the seller. No matter who holds the bond, he or she has no right to claim interest before the interest payment date. Therefore, the accrued interest calculated in Eq. (15) is just a book figure, which should be the interest that the seller can receive on the payment date, rather than on the transaction date. Therefore, the impact of accrued interest on the value of bonds is its "present value" on the transaction date, rather than the accrued interest itself.

Discounting the accrued interests in the way as the calculation of dirty value, i.e., compound discounting, the present values of the accrued interests on the 13 dates in Table 1 are: 0.155, 4.831, 9.236, 14.159, 18.969, 23.988, 28.892, 34.008, 39.175, 44.222, 49.488, 54.633, 60.000 dollars respectively. subtracting those present accrued interests from the dirty values, the clean values from January 1st to December 31st are: 1000.005, 1000.130, 1000.227, 1000.312, 1000.372, 1000.410, 1000.424, 1000.415, 1000.380, 1000.323, 1000.239, 1000.134, 1000.000 dollars respectively. Unfortunately, the results are still not exactly 1000 dollars, though they are no longer smaller than 1000 dollars.

Surprisingly, the biggest deviation again happened in the mid of the year, or the June 30th. This means that there is still problem in discounting.

As the interest is paid only on December 31st every year, the interest has no chance to generate return within a period less than one year, and the discounting should be calculated based on simple interest rather than compound interest. Correct this discounting mistakes in the calculation of dirty value and accrued interest, i.e., dirty value = 1060/1.06^(1 + 0.06 × (Days remained/365)), and accrued interest = 60 × (1-(Days remained/365))/1.06^(1 + 0.06 × (Days remained/365)), the results are shown in Table 2.

Now, we work out the constant clean values, 1000 dollars, which prove all the relevant calculations in Table 2 are correct, and also prove the problem of consideration of accrued interest in bond valuation is solved.The calculation similar to Table 2 was published by Zhiqiang Zhang (the author of this book), Finance and Accounting Monthly, 2021, issue 18, titled as On the valuation of ordinary bonds.

Table 2 The accrued interest, dirty value and clean value of the base bond

Date	Days accrued	Days remained	Accrued interest	Dirty value	Clean value
January 1st	1	364	0.155	1000.155	1000.000
January 31st	31	334	4.831	1004.831	1000.000
February 28th	59	306	9.234	1009.234	1000.000
March 31st	90	275	14.155	1014.155	1000.000
April 30th	120	245	18.962	1018.962	1000.000
May 31st	151	214	23.978	1023.978	1000.000
June 30th	181	184	28.880	1028.880	1000.000
July 31st	212	153	33.994	1033.994	1000.000
August 31st	243	122	39.160	1039.160	1000.000
September 30th	273	92	44.208	1044.208	1000.000
October 31st	304	61	49.476	1049.476	1000.000
November 30th	334	31	54.626	1054.626	1000.000
December 31st	365	0	60.000	1060.000	1000.000

3.2 The Frequency of the Interest Payment

We have known Eqs. (13) and (14), which is the popular formula for considering interest payment frequency. First of all, we test this formula by the base case and to make sure whether this formula is a real solution.

The base case: the bond was issued on January 1, 2021, with face value of 1000 dollars, maturity of 5 years, coupon rate of 6%. Now we try to calculate the bond value on the issuing date (at the very beginning date) if it pays interest annual, semiannual and quarterly respectively. The appropriate discount rate is still 6%.

Based on the Eqs. (12), (13) and (14), the bond value is:

$$B_{\text{annual}} = \sum_{t=1}^{5} \frac{60}{(1+6\%)^t} + \frac{1000}{(1+6\%)^5} = 1000 \text{ dollars}$$

$$B_{\text{semiannual}} = \sum_{t=1}^{10} \frac{30}{(1+3\%)^t} + \frac{1000}{(1+3\%)^{10}} = 1000 \text{ dollars}$$

$$B_{\text{quarterly}} = \sum_{t=1}^{20} \frac{15}{(1+1.5\%)^t} + \frac{1000}{(1+1.5\%)^{20}} = 1000 \text{ dollars}$$

As previous analyses, other things being equal, the higher the interest payment frequency, the earlier the investors receive interest, the more interest appreciation, hence the more attractive of the bond, the more valuable of the bond. Unfortunately, the valuation results based on the prevailing model are unchanged. This proves the prevailing model is not correct and we need to find a real solution.

Please note that this is by no means that the Eqs. (12), (13) and (14) are not precise enough, but rather they are not correct. Perhaps a discount bond can illustrate the mistake more clearly. Take again the base bond as an example, assume everything is the same but the appropriate discount rate is 8% rather than 6%. Then, for the annual, semiannual and quarterly interest payment, the bond value is:

$$B_{\text{annual}} = \sum_{t=1}^{5} \frac{60}{(1+8\%)^t} + \frac{1000}{(1+8\%)^5} = 920.146 \text{ dollars}$$

$$B_{\text{semiannual}} = \sum_{t=1}^{10} \frac{30}{(1+4\%)^t} + \frac{1000}{(1+4\%)^{10}} = 918.891 \text{ dollars}$$

$$B_{\text{quarterly}} = \sum_{t=1}^{20} \frac{15}{(1+2\%)^t} + \frac{1000}{(1+2\%)^{20}} = 918.243 \text{ dollars}$$

Strange things happened; when the interest payment frequency increases, the bond value decreases rather than increases. This is not a problem of precision or not but a problem of correct or not because the bond value changes in the opposite direction.

Obviously, the bond can and should be valued by discounting all its future cash flows, which is just the way of above calculations. But there must be something wrong in above models. The real solution can be found by correcting the mistakes in those models.

But, what is wrong with the prevailing models?

In the case of interest payment more than once a year, there is no problem in converting the annual interest to I/m and changing the number of interest payment periods to mn. This is in line with the real situation. However, there is a problem when the discount rate of each period is changed to k/m. Because $(1 + k/m)^m \neq 1 + k$, it means that the annual discount rate is actually changed if the discount rate of each period is changed to k/m. But when other factors remain unchanged, the increase of interest payment frequency should not affect the annual discount rate. In other words, Eqs. (13) and (14) make a mistake in the discount rate for each period.

Now, how to correct the mistakes of the models? That is, how to calculate the discount correctly? Paying interest m times a year means there are m periods a year. Note that the annual future value factor (1 + k) should remain constant, which implies the future value factor of each period should be $(1 + k)^{1/m}$, so that, $((1 + k)^{1/m})^m = 1 + k$. Then, along with the number of period, n, is adjusted to mn, the $(1+k)^t$ is adjusted to $((1+k)^{1/m})^t = (1+k)^{t/m}$; $((1 + k)^{1/m})^{mn} = (1 + k)^n$. Therefore, the correct bond valuation model that correctly considers the interest payment frequency is:

$$\mathrm{B} = \sum_{t=1}^{mn} \frac{I/m}{(1+k)^{t/m}} + \frac{M}{(1+k)^n} \tag{16}$$

Equation 16 was published in 2012 by Zhiqiang Zhang (the author of this book) in Advanced Finance - theoretical innovation and decision-making application (Peking University Press). Based on Eq. (16), the value of the base bond is:

$$\mathrm{B}_{\text{annual}} = \sum_{t=1}^{5} \frac{60}{(1+6\%)^t} + \frac{1000}{(1+6\%)^5} = 1000.000\ \text{dollars}$$

$$\mathrm{B}_{\text{semiannual}} = \sum_{t=1}^{10} \frac{30}{(1+6\%)^{t/2}} + \frac{1000}{(1+6\%)^5} = 1003.736\ \text{dollars}$$

$$\mathrm{B}_{\text{quarterly}} = \sum_{t=1}^{20} \frac{15}{(1+6\%)^{t/4}} + \frac{1000}{(1+6\%)^5} = 1005.618\ \text{dollars}$$

For the above discount bond, Based on Eq. (16),

$$\mathrm{B}_{\text{annual}} = \sum_{t=1}^{5} \frac{60}{(1+8\%)^t} + \frac{1000}{(1+8\%)^5} = 920.146\ \text{dollars}$$

$$B_{\text{semiannual}} = \sum_{t=1}^{10} \frac{30}{(1+8\%)^{t/2}} + \frac{1000}{(1+8\%)^5} = 924.845 \text{ dollars}$$

$$B_{\text{quarterly}} = \sum_{t=1}^{20} \frac{15}{(1+8\%)^{t/4}} + \frac{1000}{(1+8\%)^5} = 927.217 \text{ dollars}$$

Obviously, based on Eq. (16), bond value definitely increases when the interest payment frequency increases, no matter it is a discount bond or premium bond or a bond at par. The Eq. (16) thus is a correct or real solution for considering the interest payment frequency.

Now we solve both the problems of considering the interest payment frequency and the accrued interest. It seems not difficult to find the mistakes as well as the correct solutions or models. The question is, why have such kind mistakes remained over decades in textbooks, classes, and practices as well as financial calculators, computer software and network platforms without doubt and correction?

In fact, for the mistakes in considering the interest payment frequency, what is even stranger is that as early as the 1980s, some scholars pointed out this mistake and deduced the correct model but got almost no response. That is, in 1983, I. Keong chew and Ronnie J. Clayton (University of Kentucky) published the model on the Financial Review:

$$\mathrm{B} = (\mathrm{I/m})\left[\frac{1-(1+k)^{-n}}{(1+k)^{1/m}-1}\right] + \frac{M}{(1+k)^n} \tag{17}$$

Note that Eqs. (16) and (17) are exactly the same model, though the expression forms are different. In comparison, Eq. (16) is simple in form and convenient in application, which is more suitable for application in computer software, such as Excel; Eq. (17) is more suitable for manual calculation, that is, it is more suitable for the time the above paper was published.

3.3 The Comprehensive Model for Bond Valuation

We have worked out the solutions for considering the interest payment frequency and the accrued interest respectively. Now we can combine them for the bond valuation, or work out the comprehensive models of dirty value and clean value.

1. The Wall Street model

Before start, let us take a look at the current mainstream model, the famous Wall Street model. In case of the semiannual interest, the models of dirty value and clean value are shown as Eqs. (18) and (19) respectively.

$$\mathrm{V_{dirty}} = \frac{CP/2\left[\frac{1-(1+kn/2)^{-2N}}{kn/2}\right] + P/(1+kn/2)^{2N} + CP/2}{(1+kn/2)^{D/182.5}} \tag{18}$$

$$\mathrm{V_{clean}} = \frac{CP/2\left[\frac{1-(1+kn/2)^{-2N}}{kn/2}\right] + P/(1+kn/2)^{2N} + CP/2}{(1+kn/2)^{D/182.5}} - \mathrm{CP/2(1 - D/182.5)} \tag{19}$$

In order to maintain the original appearance of the model, the original symbols in the model are retained and some are different from that in previous models:

V: Bond value;
C: Coupon rate;
P: Face value of bonds;
Kn: Discount rate;
N: Number of years to maturity.

The Wall Street models illustrate the basic relationship that clean value equals the dirty value minus accrued interest. However, based on the previous analyses, it is easy to realize that the model has following mistakes.

Firstly, the consideration of the accrued interest (CP/2(1-D/182.5)) in Eq. (19) is incorrect, it should be discounted or replaced by its present value.

Secondly, the consideration of the interest payment frequency is incorrect. Specifically, the discount rate of each period is incorrect. It is kn/2 in the model, but the correct one is $[(1 + \mathrm{kn})^{1/2} - 1]$, or the future factor of one period is $(1 + \mathrm{kn})^{1/2}$.

Thirdly, the discounting within a non-integer period is incorrect. Specifically, it should use simple interest to discount within a period, because no cash flow (the accrued interest or the bond value) within a period has chance to get compound interest effect.

Obviously, the correct bond valuation models, the dirty value model and clean value model, should avoid all the above mistakes and correctly consider or discount the future interest, including the accrued interest as well as the due principal.

2. The correct dirty value model

As mentioned above, the dirty value is the total present value of all future interest and principal of the bond. It can be calculated in two steps. First, calculate the bond value just before the interest payment; Then discount this value to transaction date.

Assuming that there are still x integer periods left, the bond will have x + 1 interests to pay. Note that the coupon rate is r, the face value is M, the interest payment frequency is m, and the amount of interest each time is rM/m. At the date just before the payment of the current interest, the current interest does not need to be discounted because it will be paid soon, and the remaining x interests will be discounted in an integer period of 1,2, …, x, plus the present value of the face value. The dirty value of the bond just before the payment of the current interest is:

$$B = \sum_{t=1}^{x} \frac{rM/m}{(1+k)^{t/m}} + \frac{M}{(1+k)^{x/m}} + \mathrm{rM/m} \tag{20}$$

An interest period is 365/m days, and its future value factor is $(1 + k)^{1/m}$, that is, the discount rate is $[(1 + k)^{1/m} - 1]$. At the transaction date, assume D is the days remained to the current interest payment, so the proportion of days need to be discounted within this period is, D/(365/m) = mD/365. Therefore, based on simple interest, the interest rate during this period is $[(1 + k)^{1/m} - 1](mD/365)$.

Then, discounting the dirty value from the end of the current period to today or the transaction date, the present dirty value of the bond is:

$$B = \frac{\sum_{t=1}^{x} \frac{rM/m}{(1+k)^{t/m}} + \frac{M}{(1+k)^{x/m}} + rM/m}{1 + [(1 + k)^{1/m} - 1](mD/365)} \tag{21}$$

Equation (21) is a general valuation model that can value the dirty value of an ordinary bond at any time before maturity. Obviously, as a solution of the dirty value, it avoids all the mistakes indicated previously in the Wall Street model.

3. The correct clean value model

Based on the correct dirty value model, it is not difficult to deduce the correct clean value model. The right way is to subtract the present value of accrued interest rather than the accrued interest itself from the dirty value.

The correct clean value model is derived below.

As previous analyses, the accrued interest as a cash flow occurs at the end of the interest period. So the clean value at the end of the interest period can be obtained by subtracting the accrued interest from the dirty value at the end of the interest period; the clean value at the end of the interest period then discounted back to the transaction date is just the clean value at the transaction date.

The days remained to the end of the interest period is D, it means that the days for accrued interest before the transaction date is (365/m) − D; The proportion of days accrued in current period is [(365/m) − D]/(365/m) = 1 − mD/365; The proportion of days remained in current period is then mD/365. The interest of each period is rM/m. Therefore, the accrued interest is: (rM/m)(1 − mD/365) = (rM/m) − (rMD/365). Subtracting the accrued interest from the interest of the period is: (rM/m) − [(rM/m) − (rMD/365)] = rMD/365; or simply, (mD/365)(rM/m) = rMD/365.

Therefore, based on the correct dirty value model, i.e. Eq. (21), the clean value model of bonds can be derived by replacing rM/m with rMD/365:

$$B = \frac{\sum_{t=1}^{x} \frac{rM/m}{(1+k)^{t/m}} + \frac{M}{(1+k)^{x/m}} + rMD/365}{1 + [(1 + k)^{1/m} - 1](mD/365)} \tag{22}$$

Equation (22) is a general valuation model for valuing the clean value of an ordinary bond at any time before maturity. Obviously, as a solution of the clean value, it avoids all the mistakes indicated previously in the Wall Street model.

Now, we derived the models of dirty value and clean value of bonds respectively. It can be seen that it is not too difficult to find and correct the mistakes of popular models such as Wall Street model; At the same time, the final dirty value and clean value models are not more complex than the popular methods or models.

It is interesting and necessary to test our new solution to the dirty value and clean value. Considering the previous base bonds, the face value is 1000 dollars and the maturity is 5 years. The coupon rate and the appropriate discount rate are equal to the expected yield of the bond, that is, 6% and remain unchanged. What are the dirty values and clean values on the end of each month in year 2022 when the interest payment frequency is 1 and 2 respectively?

When the interests payment frequency is 1, i.e., once a year, the dirty values and clean values on the end of each month in year 2022 are shown as previous Table 2. When the interests payment frequency is 2, i.e., twice a year, based on Eqs. (21) and (22), the calculation results are shown in Table 3.

Moreover, this small "premium" is caused by the payment of interest. With the passage of time, especially with the reduction of the number of remaining interests, this effect should gradually weaken, that is, the clean value of bond should get closer to its face value. The calculation results in Table 3 do reflect this trend. This further confirms the rationality of our bond dirty and clean value model.

Therefore, Eqs. (21) and (22), as dirty value and clean value models for bond valuation, correctly consider the impact of interest payment frequency and accrued interest on bond value, are sound in theory and reliable in application, and are real solution to ordinary bond valuation.

The calculation in Tables 2 and 3 shows that the clean value model can completely exclude the influence of accrued interest from the dirty value. This directly confirms the correctness of the dirty value model and clean value model in this chapter. The bond clean value model is a "real clean value model", and the treatment of accrued interest is "cleaner" than the popular calculation and Wall Street model.

Table 3 The accrued interest, dirty value and clean value of the bond (m = 2)

Date	Days accrued	Days remained	Accrued interest	Dirty value	Clean value
January 1st	1	181	0.160	1003.973	1003.813
January 31st	31	151	4.988	1008.735	1003.747
February 28th	59	123	9.535	1013.221	1003.686
March 31st	90	92	14.617	1018.235	1003.617
April 30th	120	62	19.584	1023.134	1003.550
May 31st	151	31	24.766	1028.246	1003.480
June 30th	181	1	29.830	1033.242	1003.412
July 1st	182	0	30.000	1033.409	1003.409
July 2st	1	182	0.159	1003.488	1003.329
July 31st	30	153	4.799	1008.088	1003.289
August 31st	61	122	9.806	1013.053	1003.246
September 30th	91	92	14.699	1017.903	1003.204
October 31st	122	61	19.804	1022.965	1003.161
November 30th	152	31	24.794	1027.911	1003.118
December 31st	183	0	30.000	1033.073	1003.073

Note
(1) In order to evenly distribute the two payments of interest, assume one is paid on July 1st and the other is paid on December 31st. The first half year is ended on July 1st with 182 days, and the second half year is started on July 2st with 183 days.
(2) During the first half year, there are 9 integer period left; during the second half year, there are 8 integer period left. The calculation of dirty value takes this difference into account and discounts the value at the end of the period.
(3) The clean values in Table 2 are exactly 1000 dollars; the clean values in Table 3 are slightly larger than 1000 dollars. This reflects the effect of the higher interest payment frequency, which cannot be captured by the Wall Street model.

Further Readings

1. Chew IK, Clayton RJ. Bond valuation: a clarification. Fin Rev 1983;18(May 1983):234–36.
2. Zhang Z. On the valuation of ordinary bonds. Financ Acc Mon. 2021;18:29–35.
3. Zhang Z, Advanced finance—theoretical innovation and decision-making application. Peking University Press; 2012, pp. 62–83.

Stock and Equity Valuation: Where Discounting Does Not Work

Contents

We have dealt with bond valuation in last chapter; now we turn to the stock or equity valuation. A stock is also a basic security but different in characteristics. The ideas

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_3

in bond valuation are also useful for stock valuation, such as discounting future cash flows. In addition to absolute valuation, stock can also be valued with relative valuation or ratio method.

This is by no means that stock valuation is easier than bond valuation. But rather, stock valuation is much more difficult than bond valuation. As a matter of fact, the use of relative valuation as a sign is more negative rather than positive. For instance, the absolute valuation is too difficult, so we have to resort to relative valuation to value a stock; it is too hard to derive a precise valuation result, so we just want a rough estimation by the relative valuation; we are not so sure about the valuation result from absolute valuation, so we want to see the valuation result from relative valuation as well; etc.

Anyway, stock valuation is more difficult than bond valuation, especially when you want to derive the value of a stock based on its future return and risk, because the future return and risk of a stock is much more difficult to estimate than that of the bond. As an accounting term, equity includes common and preferred stock as well as other items, such as retained earnings, contributed surplus, etc. As a finance or valuation term, stock and equity are the same thing in most of circumstances, because stock value includes also the value of retained earnings and contributed surplus as well. Equity capital is also called share capital or paid-in capital. As a convention, equity is often referred to the owners' stake in a private company, while stock is referred to the owners' stake in a public company. They are similar in valuation. In the following discussions, we focus on the stock valuation, but the findings are suitable to equity valuation as well.

1 The Gordon Growth Model

We have introduced something about the relative valuation for stock valuation in last chapter. Since absolute valuation is more powerful in revealing the mechanism of stock value determination, we now start with absolute valuation.

1.1 Basics of Stock and Equity Valuation

From the issuer point of view, bond represents debt capital, and stock represents equity capital. Debt capital has maturity, at which the company must give the capital (principal) back to investors. Stock as the equity capital has no maturity date; the company can literally use the equity capital forever. Therefore, the lifespan of stock is usually assumed as unlimited.

As a general rule, the value of a stock is also determined by its future return and risk; hence a stock can also be valued by discounting its future cash flows. Specifically, what are the future cash flows of a stock? If you buy a stock and hold it, what you receive are the dividends paid every year in the future, which can be regarded as the cash flows of your investment.

If you hold the stock forever, you receive all the future dividends. This implies that the stock can be valued by discounting all those expected or predicted dividends. Stock valuation in this way is referred to as dividend discount model (DDM), that is,

$$\mathrm{P} = \sum_{t=1}^{\infty} \frac{D_t}{(1+k)^t} \tag{1}$$

where, Dt is the dividend in year t; k is the appropriate discount rate.

Understandably, Eq. 1 is a correct but unfeasible valuation model, because there are unlimited unknown variables which need to be estimated; nobody can predict unlimited future dividends. Equation 1 can be improved as Eq. (2):

$$\mathrm{P} = \sum_{t=1}^{n} \frac{D_t}{(1+k)^t} + \frac{P_n}{(1+k)^n} \tag{2}$$

where, P_n is the (predicted) stock price in year n.

Equation 2 seems better, as the number of unknown variables is limited. In addition, Eq. 2 seems similar to the basic bond valuation model, which is much simpler than the bond valuation model considering accrued interest and interest payment frequency.

However, Eq. 2 is still not so easy for application, because the unknown variables are much difficult to estimate than that in the basic bond valuation model. The dividends in following years are of course difficult to estimate; while the stock price at the end of year n is even more difficult to estimate, and literally infeasible to estimate. If someone in this world can predict the stock price in any near future year, such as year n, he or she will never lose money in the market. This is impossible in theory. In fact, as an investor, he or she may even do not know how many years to hold the stock, i.e. the n in the model is uncertain.

Therefore, DCF or discounting seems not so valid in stock valuation, although it is used everywhere in finance. A tough task in stock valuation is the dividend prediction. In addition, predicting stock price or value at a future time directly is not a good idea either.

1.2 The Gordon Growth Model

Forecast is a tough task, especially the asset returns (earnings or cash flows) forecast in finance. Although forecast is not the theme or core function of finance, finance is responsible for minimizing the forecast workload in setting up its theories and models. This should be one of the requirements or standards for financial theories and models, because finance as an independent science is decision or application oriented.

It is too difficult to forecast asset returns year by year. A wiser choice is to forecast an average annual growth rate extended into the future. Combining this forecasted average annual growth rate and a (normalized) current return, the future returns of the asset then can all be derived out. This has become a convention since the early days of finance. The average annual growth rate is often referred to as a constant growth rate.

For the average annual growth rate, a natural and necessary question is: how long is the time horizon for the constant growth rate to be valid? Within the context of DCF method, an asset value is the sum of the present values of all its future cash flows. As lifespan varies across assets, the infallible choice of the time horizon is naturally the infinite future. Hence the constant growth rate is the average annual growth rate over an infinite time horizon, which is usually called a perpetual growth rate in current financial community.

For stock valuation, Myron J. Gordon (1959, 1962) derives a model[1] with such a perpetual growth rate. Suppose the current (normalized) dividend of a stock is D_0; the perpetual growth rate of the dividend is g, the dividend in year 1, year 2, …, year t, … is $D_1 = D_0(1 + g)^1$, $D_2 = D_0(1 + g)^2$, …, $D_t = D_0(1 + g)^t$… respectively. Then the stock value, P, is,

$$\mathrm{P} = \frac{D_0(1+g)^1}{(1+k)^1} + \frac{D_0(1+g)^2}{(1+k)^2} + \ldots \frac{D_0(1+g)^t}{(1+k)^t} + \ldots \tag{3}$$

Thus,

$$\mathrm{P}\frac{(1+g)}{(1+k)} = \frac{D_0(1+g)^2}{(1+k)^2} + \frac{D_0(1+g)^3}{(1+k)^3} + \ldots \frac{D_0(1+g)^t}{(1+k)^t} + \ldots \tag{4}$$

Equation 3 minus Eq. 4,

$$\mathrm{P}\left[1 - \frac{(1+g)}{(1+k)}\right] = \frac{D_0(1+g)^1}{(1+k)^1} \tag{5}$$

Then,

$$\mathrm{P(k - g) = D_0(1 + g)} \tag{6}$$

Then,

$$\mathrm{P} = \frac{D_0(1+g)}{k-g} = \frac{D_1}{k-g} \tag{7}$$

[1] It was originally published by Myron J. Gordon and other scholars in 1956 based heavily on the book titled “The Theory of Investment Value” by John Burr Williams in 1938.

where D_0 or D_1 is the normalized initial value of dividend per share in current or next year, and $D_1 = D_0(1 + g)$; k is the market (investors) required rate of return matched with the risk of this stock, and g is the estimated constant perpetual growth rate of dividend.

Equation 7 is referred to as Gordon growth model (GGM) or constant growth model in finance community, which was named after Myron J. Gordon of the University of Toronto. The GGM is a special version of the dividend discount model (DDM), and is widely used for stock as well as other asset valuation, because it has at least the following advantages.

(1) Simplicity

Simplicity is necessary for a model to be feasible in application. As an application-oriented science, finance should care about the simplicities of its theory and model under the condition of problem being solved. The GGM is an ideal model in such a sense. The (independent) variables taken into account are not much; the form of the model is very simple; and the derivation process as above is easy to understand.

Consider a base case stock, the current dividend per share, $D_0 = 2$ dollars; its perpetual growth rate, $g = 6\%$; the discount rate, $k = 10\%$. Then, the stock value is:

$$P = 2(1 + 6\%)/(10\% - 6\%) = 53\ \text{dollars}$$

(2) Soundness

For good generality and flexibility, finance as a science or financial theory should be correct exactly in concept. The GGM meets this rule well. For instance, g in the model is a perpetual growth rate rather than a growth rate over a finite time horizon, which is in line with the concept of DCF method. In addition, the asset value increases as the future return (determined by the D and g) goes up and decreases as the future risk (incorporated in the k) goes up, which is in line with the axiom of asset value determination.

As indicated in Chapter "Finance and Its Fundamental Problems", this axiom should be the basic standard to build and to judge a financial or valuation model. Therefore, based on the theoretical soundness and application convenience, it is widely believed that the GGM is an effective solution to stock valuation.

1.3 Variations of the Model

In previous case, if the dividend grows at an annual rate of 7% rather than 6%; other things being equal, the stock value will increase to:

$$P = 2(1 + 7\%)/(10\% - 7\%) = 71.33\ \text{dollars}$$

The one percentage change in the growth rate causes a one third change in stock value (71.33/53–1 = 34.6%). This implies the stock value is very sensitive to the growth rate, and the estimation or forecast of the growth rate should be very careful.

On the other hand, related to the difficulties in the dividend prediction, the estimation of the perpetual growth rate is not easy even at an impreciseness basis. Just consider any company in reality you are familiar with, if you make a prediction of the perpetual growth rate of its future dividend, are you sure your error in estimation is less than one percentage point?

On this account, some variations of the Gordon growth model come up in order to reduce the sensitivity of the stock value to the growth rate. Among those variations, two stages or multiple stages model are the popular one, in which the dividends over near future are predicted year by year, the Gordon growth model is only used for valuing the far future dividends. In this way, the effect of the perpetual growth rate on the stock value is reduced to some extent.

In the above base case, if the dividend grows at an annual rate of 18% over the first 5 years, and then grows at 5% perpetually; the dividends over the following 5 years are 2.36, 2.78, 3.29, 3.88, 4.58 dollars respectively. Then, other things being equal, the stock value is:

$$\begin{aligned} \mathrm{P} &= 2.36/(1+10\%) + 2.78/(1+10\%)^2 + 3.29/(1+10\%)^3 \\ &\quad + 3.88/(1+10\%)^4 + 4.58/(1+10\%)^5 \\ &\quad + 4.58(1+5\%)/(10\% - 5\%)/(1+10\%)^5 \\ &= 12.41 + 59.66 = 72.07\,\text{dollars} \end{aligned}$$

This demonstrates the application of the two-stage model. You may feel that the vertical drop of the growth rate between year 5 and year 6 is too big and not so realistic; then you can add a stage in between as a smooth transition (with growth rate gradually decreasing from 18 to 5%), as such, the model is transformed into a three-stage model.

For instance, other things being equal, but the dividend grows annually at 20% over the first 3 years, and then slows down gradually to 5% in the following 3 years. Specifically, the growth rates in year 4, 5, 6 are 15%, 10%, 5% respectively. Then, the dividends over the following 6 years are 2.40, 2.88, 3.46, 3.97, 4.37, 4.59 dollars respectively. The stock value is:

$$\begin{aligned} \mathrm{P} &= 2.40/(1+10\%) + 2.88/(1+10\%)^2 + 3.46/(1+10\%)^3 \\ &\quad + 3.97/(1+10\%)^4 + 4.37/(1+10\%)^5 + 4.59/(1+10\%)^6 \\ &\quad + 4.59(1+5\%)/(10\% - 5\%)/(1+10\%)^6 \\ &= 15.18 + 54.41 = 69.59\,\text{dollars} \end{aligned}$$

2 Does a Positive Perpetual Growth Rate Exist?

It is taken for granted in class and in practice that the perpetual growth rate of most (if not all) stock dividend is positive; hence a stock can be valued by putting a positive perpetual growth rate into the Gordon growth model. Is that true?

2.1 The Surprises Resulted from Positive Perpetual Growth

It was said that Albert Einstein mentioned, "The most powerful force in the universe is compound interest." Although Einstein was not majored in finance, but his words is a valuable warning for us to use compound interest in financial calculation. Unfortunately, the force of positive perpetual growth rate has not drawn much attention so far.

1. The future dividend per share based on the positive perpetual growth rate

For instance, Haier is a well known refrigerator manufacturer in China. The output of Haier refrigerators is 15.18 million in year 2020, accounting for 17.98% of the total output of refrigerators in China. Haier's earnings per share in 2020 is 1.34 yuan (about 0.209 dollar), and the cash dividends per share is 0.36 yuan (about 0.056 dollar).

Haier's earnings and dividends grow at 16% per year in average during past 20 years. Understandingly, the future growth rate may be somehow less than 16%. As Haier is a good company in the whole economy and the whole economy grows at a rate around 6%. 6% may be a reliable estimation of Haier's perpetual growth rate in future.[2]

Starting from the current dividend, 0.056 dollar, growing at a rate of 6%, after thousands of years (far from infinite horizon), Haier's dividends per share will reach an enormous and untrue level. For example, after 500 or 1000 years' growth, the dividends per share is:

$$0.056 \times 1.06^{500} = 251,837,650,738 \text{ dollars}$$
$$0.056 \times 1.06^{1000} = 1,132,539,327,307,830,000,000,000 \text{ dollars}$$

Dividend per share as such is too large to believe and beyond most people's imagination. Please note that this is only dividend per share in year 500 and 1000. Based on the Gordon growth model, the dividend per share will go on to grow after year 500 or 1000.

At present, Haier's stock is traded at around 26 yuan (about 4.06 dollars). Suppose the stock is fair priced, the growth rate implied in the stock price is bigger than 6%, which means the dividends growing even faster. Based on the Gordon growth model,

[2] The purpose here is to test the effect of the positive perpetual growth rate, rather than to estimate the growth rate; the accuracy of the estimation is not our main concern.

$$k - g = 0.056 * 1.06/4.06 = 1.46\%.$$

Since Haier is a typical blue-chip stock and has moderate risk, suppose the discount rate is estimated at 10% which is about the normal return in stock investment over long run. Then, the perpetual growth rate, g = 10%-1.46% = 8.54%. Growing at such a rate, after only 500 and 1000 years (far from infinity), Haier's dividends per share will be:

$$0.056 \times 1.0854^{500} = 34,921,802,248,659,900 \text{ dollars}$$

$$0.056 \times 1.0854^{1000} = 21,777,362,005,259,$$
$$100,000,000,000,000,000,000 \text{ dollars}$$

In both case, the dividends per share are too large to be true. More surprisingly, even for a very conservative perpetual growth rate, after a long enough period (not as far as infinite horizon), the dividends per share can also reach an untrue amount

For instance, starting from 0.056 dollar, growing at a rate of 1%, after 8000 years (still far from infinite horizon), Haier's dividends per share is:

$$0.056 * 1.01^{10000} = 2,085,346,791,532,870,000,000,000,000,000,000 \text{ dollars}$$

Those enormous dividends per share are obviously insane and unbelievable. Even worse, according to the assumption of the Gordon growth model, the dividends per share will continue to grow after that, and become larger and larger. But, if the estimation of the future dividends is incredible, how can we trust the valuation and the result based on it?

2. The future sales volume based on the positive perpetual growth rate

Take again Haier's fridge as an example. The output and sales is 15.18 million units in 2020. Suppose the unit price and sales margin remain constant after allowing for inflation, and so does the dividend payout ratio. The annual growth rate in dividend thus equals to that of the earnings and sales or output, assuming again those volume or amount grow at 6% annually.

After growing thousands of years in past, the world population has reached about 7.0 billion now. Assuming the population will increase 7.0 billion over every hundred years in future, or 3.5 billion over every fifty years, it is very interesting to estimate the change of Haier fridges purchased per capita worldwide each year in the future.

Table 1 shows the fridge purchased per capita worldwide each year based on the assumption of growth about world population and Haier fridge sales. The fridge purchased per capita starts from less than one (0.002 now) and gets bigger and bigger. As shown in the table, in the year 100, to support the assumed growth of Haier, each person in the world needs to purchase about 0.368 Haier fridges every year; a family with 3 members then needs to purchase about 1 Haier fridges; in the year 150, each person and family every year need to purchase about 5.421 and 16 Haier fridges respectively; and they need to purchase even more in the subsequent years.

Table 1 The forecast of fridge purchased per capita each year

Year	World population[a]	Haier fridge[a]	Per capita[a]	Total weight[a]
0	7.0	0.015	0.002	0.001
50	10.5	0.280	0.027	0.014
100	14.0	5.151	0.368	0.258
150	17.5	94.875	5.421	4.744
200	21.0	1747.61	83.22	87.38
250	24.5	32,191.27	1313.93	1609.56
300	28.0	592,968.13	21,177.43	29,648.41

[a] World Population and Haier Fridge are the corresponding number in billions;per capita is the Haier's fridges purchased per capita of the world's average; total weight is the total weight of Haier fridge output in billion tons

This is obviously untrue. Even more surprisingly, assume the average weight per fridge is 50 kg, after growing 629 years (far from infinite horizon), the total weight of Haier's output will be 6,275,249,928,600 billion tons, which exceeds the earth mass.[3]

These results are so ridiculous and unbelievable that they manifest patently that the positive perpetual growth rate seems not exist in our world.

2.2 Arithmetic or Geometric Average Growth Rate

It seems the perpetual growth rate needs a closer examination. There are two ways to obtain an average growth rate: arithmetic averaging and geometric averaging, which are different in calculation as well as in result. It is necessary to make sure which is better.

Growth means the value change of a variable across times or periods. Let V_t and $V_{t\text{-}1}$ to be the value of a variable in year t and year t-1 respectively, assuming they occur at the end of the year; define $V_t/V_{t\text{-}1}$ as the growth factor of the variable in year t. Hence the growth rate in year t is "$V_t/V_{t\text{-}1}-1$". Use AAG and GAG to represent the arithmetic average growth rate and the geometric average growth rate respectively. Then,

$$\text{AAG} = \frac{V_1/V_0 + V_2/V_1 + V_3/V_2 + \ldots + V_n/V_{n-1}}{n} - 1 \tag{8}$$

$$\text{GAG} = \left(\frac{V_1}{V_0} \times \frac{V_2}{V_1} \times \frac{V_3}{V_2} \times \ldots \times \frac{V_n}{V_{n-1}}\right)^{\left(\frac{1}{n}\right)} - 1 = \left(\frac{V_n}{V_0}\right)^{\left(\frac{1}{n}\right)} - 1 \tag{9}$$

[3] About 5.965 × 1,000,000,000,000,000 billion ton.

Consider an example. Investor W buys a stock today at a price of 100. The price goes up to 150 at the end of year 1 and goes down to 100 at the end of year 2. Then, what is the average annual growth rate of the stock price? Our intuition tells us that the average annual growth rate is 0%, because the stock price is still 100. However, based on Eqs. 8 and 9, we can work out an average annual growth rate 8.33% and 0% respectively.

This demonstrates clearly that the GAG is correct, whereas the AAG is not. Please note that the annual growth rate under GAG is affected only by the beginning value and final value of the variable, and is independent of all the intermediate ups and downs. Obviously, when the growth rate of every year is indeed constant, GAG is equal to AAG; otherwise, GAG is less than AAG. Use SD to represent the standard deviation of the yearly growth rates, based on statistics, the relationship between AAG and GAG is:

$$(1 + \mathrm{AAG})^2 = (1 + \mathrm{GAG})^2 + \mathrm{SD}^2 \tag{10}$$

Actually, the arithmetic averaging is more suitable for measuring the absolute growth, i.e., the annual growth in value; whereas the geometric averaging is more suitable for measuring the relative growth, i.e., the annual growth rate. In other words, using AAG' to represent the average annual absolute growth, Eq. 8 should be rewritten as:

$$\mathrm{AAG}' = \frac{(V_1 - V_0) + (V_2 - V_1) + (V_3 - V_2) + \ldots + (V_n - V_{n-1})}{n} = \frac{V_n - V_0}{n} \tag{11}$$

Based on Eq. 11, the average annual absolute growth of the stock price is 0 dollars, which is in line with our intuition. Similarly, the annual absolute growth under Eq. 11 (AAG$'$) is also affected only by the beginning value and final value of the variable, and independent of all the intermediate ups and downs. Therefore, when we use average growth rate, keep in mind that it should be a geometric average. This is also in line with the convention of compounding calculation of future and present value in finance.

Anyway, based on above discussion, the perpetual growth rate is a geometric average growth rate over an infinite time horizon. This perhaps is a little surprising, which implies that the perpetual growth rate is not well understood in financial community. As a consequence, there may be even more and bigger surprises ahead of us.

2.3 Does a Positive Perpetual Growth Rate Exist?

In fact, no company can live forever. As a constant (average) growth rate extended into infinite future, it can only be negative, because a company will surely go bankrupt or disappear given such a long time! Expected returns in any form (accounting earnings, operating or free cash flows, and dividends, etc.) will be zero after a long enough time.

Bankruptcy or disappearance is the inevitable destination of every company in reality. Even being incorporated into another company via purchase and acquisition, a company finally cannot escape from disappearance together with the buyers. How strong and brilliant are the corporate behemoths like Barings Bank (1762 to 1995), Lemon Brothers (1950 to 2008), Eastman Kodak (1880 to 2012), etc. used to be? But where are they today?

Most companies in the world so far are less than 1000 years old. Unfortunately, the average lifespan of leading US companies listed in the S&P 500 index has decreased by more than 50 years in the last century, from 65 years in the 1920s to just 15 years today, according to Professor Richard Foster from Yale University.[4]

Obviously, it doesn't need an infinite time for a company to devalue to zero. Based on Eq. 6, in a long enough but finite time horizon, i.e., before it extends into an infinite future time, the value of returns will goes from a positive number (V_0) to a number close to zero ($V_n \rightarrow 0$). This implies that $(V_n/V_0) < 1$, hence $(V_n/V_0)^{(1/n)} < 1$. GAG or geometric average growth rate thus can only be negative.

For instance, consider a good company with lifespan or life expectancy of 100 years. The dividend of this good company in year 101 is of course zero. The (rationalized) dividend of this good company in current year is of course a positive number, assume the amount is 1 dollar on per share basis. Based on the geometric average, the annual growth rate will be −100% over the 101 years. To make the calculation more meaningful, assume a (rationalized) dividend in the year 100 of this company's life is something close to zero but not zero, say, 1/10000 dollars, the average annual growth rate over its lifespan is:

$$\text{GAG} = (1/10000)^{(1/100)} - 1 = 0.912 - 1 = -8.8\%$$

Please note that this is an assumed good company. For a bad company, the lifespan is shorter, the average annual growth rate over its lifespan is surely to be negative, no chance to be positive. For instance, other things being equal, but the lifespan is 10 years rather than 100 years. the average annual growth rate of its dividend over its lifespan (10 years) then is:

$$\text{GAG} = (1/10000)^{(1/10)} - 1 = 0.398 - 1 = -60.19\%$$

Therefore, the average annual growth rate for any company over a long time horizon or infinite time horizon will definitely be negative, which means that there is a decreasing trend of return over the company's lifespan. The only difference is that a good company decreases slowly, while a bad company decreases rapidly. Seems

[4] Kim Gittleson, Can a company live forever? BBC News, New York at http://www.bbc.co.uk/news/business-16611040.

too surprising? but it is true and obvious! The logic behind the negative growth rate is virtually invulnerable.

Last section reveals that the positive perpetual growth rate can result in unbelievable and ridiculously large dividends and sales, which manifests that there cannot be positive perpetual growth rate in this world! Now, this section reveals that the perpetual growth rate or long term growth rate can only be negative! These two findings reconfirm each other.

Now the conclusion is clear and for sure: the perpetual growth rate, if exists in this world, can only be negative, no way to be positive.

However, a positive perpetual growth rate is taken for granted in current application of the Gordon growth model. This is a mistake in direction rather than just too arbitrary or too subjective, which implies probably that most (if not all) of the applications of the Gordon growth model are just an affectation or doing it for show. Obviously, when one of the variables in the model are wrong about positive or negative, it is inevitable to result in an intolerant mistake. We thus cannot expect a correct result based on such an application.

An explanation for such a convention is that it can simplify the calculation. The returns of a company are likely to increase for a short or long time and decline thereafter. Within the context of DCF valuation, cash flows in far future are less important because of the discounting. Thus, a mistake in dividend forecast (because of the positive perpetual growth rate) over the future period will not affect the valuation result too much.

This explanation is obviously not convincing. As previously revealed, dividend in far future is likely to be zero because the company is likely disappeared before that time. However, if we blindly use a positive perpetual growth rate, the dividend in far future will become larger and larger, even become an astronomical figure, as revealed in last section.

For example, if it is forecasted Haier no longer exists after 1000 years, and the dividend per share then is 0. However, based on a positive perpetual growth rate, 6%, the dividend per share in year 1000 is 1,132,539,327,307,830,000,000,000 dollars. Can we say that although this is larger than the correct forecast of 0 dollars, since its present value is still close to 0, we can just take it as a correct forecast? Obviously, simplification is not a good excuse for changing the sign of growth rate from negative to positive. In fact, we can simplify the calculation while keeping the growth rate negative rather than changing it to a wrongly positive.

Anyway, the negative growth rate is beyond traditional belief. It is rather difficult for current finance community to accept the negative long term or perpetual growth rate, especially for those normal or good companies. In addition, the valuation with a negative growth rate will lead to a very low result than that with a positive growth rate, which is also unacceptable.

Consider again the base case stock with a perpetual growth rate of dividend of 7% and a discount rate of 10%. Assume the current dividend (year 0) of this stock is 2 dollars, based on the Gordon growth model, value of a share will be:

$$P = \frac{D_0(1+g)}{k-g} = \frac{2 \times (1+7\%)}{10\% - 7\%} = 71.33\,\text{dollars}$$

This is the prevailing valuation.

Now assume this is a typical company listed in the S & P 500 index, and its life expectancy is 40 years ($= 65 \times 50\% + 15 \times 50\%$). Assuming the dividend in the year 40 is something close to zero, say, 1/10000 dollars. Thus, the average annual growth rate is:

$$\text{GAG} = (1/10000)^{(1/40)} - 1 = 0.7943 - 1 = -20.57\%$$

Take the -20.57% (an average annual growth rate over 40 years) approximately as the perpetual growth rate, based on the Gordon growth model, the share value is,

$$P = \frac{2 \times (1 - 20.57\%)}{10\% + 20.57\%} = 5.20\,\text{dollars}$$

Obviously, the valuation difference between 71.33 and 5.20 is too large to be reconciled by prevailing financial wisdom. The most important and most urgent problem, however, may be not to explain the difference, but to judge which one is correct, or which one is more correct.

Unfortunately, it is rather difficult to answer such a "simple" question. On one hand, the valuation result of "5.20", which is 92.71% lower (= 5.20/71.33–1) than the "normal valuation" result of 71.33, is too low to accept. On the other hand, the positive perpetual growth rate 7% seems groundless and indefensible, because no firms will grow positively forever; the average growth rate, based on correct concept and logic, can only be negative.

The perpetual growth rate thus comes into a dilemma: it can only be negative in logic but the negative growth rate can hardly be accepted in reality. As such a dilemma is hard to be explained or solved; we refer to it as the "ZZ growth paradox" rather than a "conclusion", which was first published by Zhiqiang Zhang (2008, 2010). As the long-term or perpetual growth rate is an inevitable variable in many financial and valuation models, the ZZ growth paradox may be a little challenging.

The challenge from the ZZ growth paradox, however, is not as serious or terrible as it seems to be at the first sight. We will further our discussion to reveal the implications of the ZZ growth paradox to valuation and finance, rather than to persuade readers to accept the negative growth rate. Anyway, the ZZ growth paradox pushes us to do fundamental rethinking about some related financial problems as well as financial theories.

3 The Feasibility of Gordon Growth Model

The ZZ growth paradox means that we are not sure whether the perpetual growth rate is positive or negative. This is of course a big challenge to the application of the Gordon growth model.

3.1 The Illusive Positive Perpetual Growth

After understanding the ZZ paradox, a problem perhaps needs to think: why the positive perpetual growth rate, a mistake, is so widely believed and used?

The main reason is that people (investors as well as scholars) are used to infer or deduce future trend by analogy with the past, and infer or deduce the unobservable long run situation by analogy with the observable past short run. However, this time, for the application of the Gordon growth model, the case is extremely different. It is not an ordinary long run, but a very long or an infinite horizon! The trend and growth rate over such a horizon is totally different from that over a past period or over foreseeable horizon!

As shown in Fig. 1, line m and line n illustrate the growth of return (earnings, dividends, etc.) over a foreseeable horizon and over an infinite horizon respectively. The growth rate is possible to be positive over a short foreseeable horizon, like those represented by the points on the increasing line m; but is negative for sure over infinite horizon, like those represented by the points on the decreasing line n, because the positive dividends within the lifespan of the stock are followed by endless zero dividends thereafter. It cannot be correct to derive the perpetual growth rate while neglecting those zeros. It can be imagined that those zeros are much more than the positive dividends in numbers.

Take again Haier as an example, since the average annual growth rate of its earnings and dividends over the past 20 years is about 16%, it seems naturally and reasonably to estimate its perpetual growth rate as a positive number lower than the

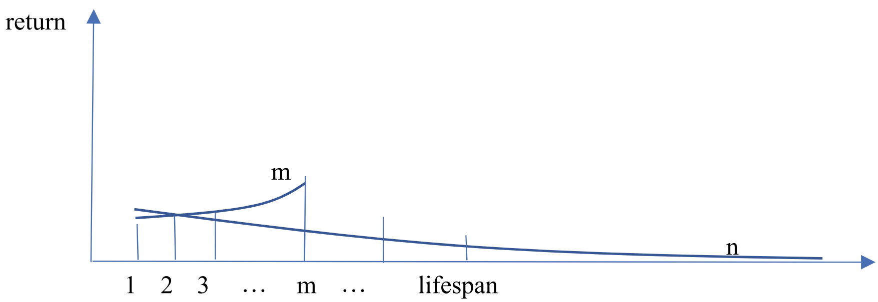


Fig. 1 The possible positive growth rate over a foreseeable horizon within the life expectancy and the surely negative growth rate over infinite horizon

16%, such as 6% or something like that. However, since Haier cannot live forever. The dividend after its end is numerous zeros. From the positive dividends to the far future zeros, the growth rate can only be negative. Just as line n shown in Fig. 1. This implies the positive perpetual growth rate is not correct.

In addition, the perpetual growth rate should not be arbitrarily estimated as zero based on the infinite horizon, because zero growth rate cannot differentiate good companies and bad companies. As revealed before, both good and bad companies grow negatively over very long horizon, the difference is that some decrease slowly some decrease rapidly.

Finding out the reason does not mean the perpetual growth rate can be positive; but rather, this makes it further sure: no positive perpetual growth rate exists in the world!

3.2 *The Feasibility of Gordon Growth Model*

We have mentioned the sensitivity of Gordon growth model to the growth rate and introduced the two stage and multiple stage models. We now explore it more systematically. To focus on the sensitivity, we assume temporarily the perpetual growth rate can be positive.

Consider the following case.

Case 1: Given stock A, B and C, the perpetual growth rates of their dividends are 9%, 9.9% and 9.99% respectively; their dividends per share next year are all 1 dollar; the market or investors required rate of return are all 10%. How much are they in value today?

Case 1 provides the data of risk and return, which is suitable to value the three stocks by applying the Gordon growth model,

$$
\begin{aligned}
P_A &= 1/(10\% - 9\%) = 100\,(\text{dollars}) \\
P_B &= 1/(10\% - 9.9\%) = 1000\,(\text{dollars}) \\
P_C &= 1/(10\% - 9.99\%) = 10000\,(\text{dollars})
\end{aligned}
$$

Value of B is 10 times of A, and value of C is 100 times of A, just because of the less than 1% difference on their growth rates. The stock A, B and C actually can be viewed as just one stock, only that the growth rate is forecasted as 9% with an error of 0.9% or 0.99%. Since the g in the Gordon growth model is a perpetual growth rate, i.e. the average annual growth rate of dividends over infinite time horizon, a forecasting error as lower as 0.99% is obviously inevitable. However, this inevitable and trivial error leads to a valuation bias of 10 times or even 100 times! This is obviously incredible and unacceptable.

Why is the Gordon growth model so sensitive to the growth rate? One reason is that it is a perpetual growth rate rather than an ordinary one. A perpetual growth rate is the average growth rate over infinite future. Believe it or not, it is beyond the ability of human intelligence to forecast such a growth rate, needless to say forecasting it with high precision or accuracy. This implies that it is too difficult to work out a reasonable valuation result based on the Gordon growth model, just because it requires a too precise perpetual growth rate.

Consider another case.

Case 2: Given stock D, E and F, the growth rate of their earnings per share over a foreseeable future period (about 15 years or so) are 10%, 20% and 30% respectively; their current earnings per share are all 1 dollar; the market or investors required rate of return are all 10%. How much are they in value today?

Case 2 also provides the risk and return about three stocks, D, E and F. The information of earnings instead of dividends is clearly more in line with practical investors' interests, because the earnings per share in any given year represent the total return on a share of stock in that year. Unfortunately, things now become even worse. Since the growth rates, 10%, 20% and 30%, are not less than the discount rate, the Gordon growth model does not work,[5] because it can only work in the condition of "k > g".

The valuation result of the first case is too sensitive to the growth rate; there is even no result can be worked out for the second case, which is a more common case in terms of the data given. So the Gordon model seems too weak in dealing with the growth rate hence the valuation of stocks. A major obstacle here is the growth rate. The Gordon model requires a very precise perpetual growth rate, which cannot be larger than the discount rate. This is basically impossible to predict or estimate with human intelligence.

Consider all previous findings, on one hand, the Gordon growth model is very sensitive to the perpetual growth rate, hence requires a very precise estimation of the growth rate; on the other hand, as reflected by the ZZ growth paradox, it is not sure whether the perpetual growth rate is positive or negative. But apparently, nobody in this world can estimate a variable precisely when the variable is not sure to be positive or negative.

Therefore, the Gordon growth model is finally proved no feasibility at all!

4 Where Is a Qualified Model for Stock Valuation?

Since the Gordon growth model is powerless for stock valuation, especially in handling the growth rate, is there other way out for stock valuation? An interesting idea perhaps is: can the relative method or ratio method do better?

[5] Please note that "their earnings per share over foreseeable future period are growing at 10%, 20% and 30% respectively" is a realistic situation or reasonable estimation, because people usually cannot forecast the earnings over too long in future, do not mention over infinite future.

4.1 The Potential of Ratio Method

The Gordon growth model, although fails to climb the mountain of the growth rate, draws out three determinants or variables of stock value: the initial return or dividend, the growth rate of this return, the risk or discount rate of the stock. The three determinants or variables are absolutely right, because they are in line with the fundamental axiom of value determination: return and risk determine asset value.

As we know from last chapter, the ratio method does not pursue the true value, but only a relative reasonable price. Specifically, it values a stock based on variables not directly related to the expected return and risk, or other than the above three variables. But, can stock be valued without factoring in its return and risk?

Just think a simple question. When you intend to choose a stock to invest, which would you prefer, the stock with high P/E ratio or low P/E ratio?

Obviously, no certain answer here, because the information is not enough. This implies the P/E ratio is not qualified to be a standard for stock selection, because P/E ratio being high or low does not necessarily mean overvalued or undervalued. Such being the case, is the P/E qualified to be a valuation ratio? What about the other ratios, such as P/B and P/S?

The answer is also no; that is, P/B and P/S are not helpful in stock selection; they have nothing to do with all the three determinants of stock value. P/E is relatively better; it relates stock value only with the initial return (earnings per share in current year). That is why all the three ratios are not sound in theory. Obviously, none of the three ratios takes the growth rate into account. That is the main reason that they are unhelpful in stock selection.

Therefore, the ratio method is neither sound in theory nor useful in practice, although they are widely used because of their simplicity and easiness. Some efforts have put to improve this method, which is mainly on the P/E because it is a little sounder than the other two ratios.

Previous efforts to improve P/E focus on how to consider the growth of earnings. An improved version is Leading P/E, also named as dynamic P/E, which uses the predicted earnings next year instead of the current earnings in calculating P/E; consequently, the traditional one is referred to as static P/E or trailing P/E.

$$\mathrm{LeadingP/E} = \mathrm{P}/\big[\mathrm{E}(1+\mathrm{g})\big] \tag{12}$$

For example, stock A and B, both have earnings per share 2 dollars in current year, and will grow at an annual rate of 10% and 30% respectively. Suppose both are traded at 30 dollars. Then, their P/Es are both 15 times. Obviously, A is relatively overvalued and B is relatively undervalued; but the trailing P/E tells nothing about the mispricing.

However, the leading P/Es of A and B are different. The earnings per share next year for A and B are 2.2 and 2.6 respectively. Therefore, the leading P/E of A is 13.636 (= 30/2.2); and the leading P/E of B is 11.538 (= 30/2.6). Now, the leading

P/E tells the mispricing to some extent, i.e., B is relatively lower than A, which means B is more undervalued.

Another improved version is PEG, which is the P/E level per unit growth rate,

$$\mathrm{PEG} = (\mathrm{P/E})/(\mathrm{g} \times 100) \tag{13}$$

According to the empirical standard of Wall Street, the standard or fair PEG ratio is around 1; PEG > 1 indicates that the stock is overvalued and PEG < 1 indicates that the stock is undervalued. PEG makes up for the weakness of P/E to a large extent.

Consider the stock A and B again. As both are traded at P/E of 15, and their growth rates are 10% and 30% respectively, their PEGs are:

$$\mathrm{PEG_A} = 15/10 = 1.5$$
$$\mathrm{PEG_B} = 15/30 = 0.5$$

Based on the PEG, stock A is obviously overvalued and B is obviously undervalued. PEG breaks the rule of P/E. The higher (lower) P/E does not necessarily mean the stock is more expensive (cheaper) or more overvalued (undervalued).

Written from subjective intuition, PEG is inevitable unsound in theory, or has conceptual and logical loopholes. For example, if a stock maintains constant earnings hence the growth rate is 0, the PEG will go up to infinite. This makes no sense. Or according to the PEG rule, its P/E should be 0, which means the stock should be worthless. This makes no sense again.

Therefore, the ratio method is even more powerless than the Gordon growth model for stock valuation as well as for handling the growth rate.

4.2 The Potential of Discounting Method

It is well known that the Gordon growth model is almost the only absolute valuation model for stock valuation except some variations. Since the Gordon growth model is proved to be infeasible, is the two-stage or multiple-stage variations feasible?

Obviously, multiple-stage variation model does not change Gordon growth model; in essence, it is the restriction or weakening of Gordon growth model, that is, under multiple-stage valuation, only part of the future cash flows is left to value with Gordon growth model; then add the part valued by the regular discounting to obtain the equity or stock value.

Thus, the multiple-stage model cannot overcome any shortcomings of the Gordon growth model, including its infeasibility. The multiple-stage model is not an improvement of the Gordon growth model, but rather a negation or partial negation of the model. In other words, the multiple-stage model is just the evidence supporting our previous findings.

Thus, the Gordon growth model and its application in those variations is not improved at all, and a positive growth rate is put into the model to value the relevant future dividends. A positive growth rate in such case cannot be right, because the far future (after the first one or two stages) is more likely to grow negatively.

Furthermore, the multiple-stage model is not exactly a model, but a simple composition of multiple models. After composition, the number of unknown variables increases significantly. Theoretically, the application difficulty should not be less than the sum of the difficulties of those component models. Since the Gordon growth model itself is infeasible, and the multiple-stage model cannot change it; then, the multi-stage model is also infeasible.

Within the framework of discounting method in valuation, asset value is the sum of the present values of its future cash flows. For stock valuation, how many years need to be considered in discounting the future cash flows? Because the stock or equity has no maturity, people naturally tend to understand it as infinite horizon. Although the period during which the stock has return (earnings, dividends, etc.) cannot exceed the life of the company, the life expectancy of a company is too difficult to estimate and the lifespan varies across companies, it is unlikely to add a variable of lifespan to the model. Then, the stock value is naturally the sum of the present values of its cash flows over infinite horizon. On the other hand, it is impossible to predict the dividend or dividend growth rate year by year. Naturally, the dividend forecast adopts the combination of initial value and constant or average growth rate. Then, based on strict reasoning, Gordon growth model is bound to be obtained.

Therefore, within the framework of discounting or DCF method, the shortcomings of Gordon growth model, the sensitivity, the dilemma of growth rate, the infeasibility, etc., have no way to be changed or improved. The problems and difficulties encountered by the Gordon growth model are just the problems and difficulties in the application of the discounting or DCF method. This implies that the stock valuation is not a field for the discounting or DCF method to work. It is perhaps not a good news, but it is certainly another surprise!

What's more, since the current relative valuation or the ratio methods are even more powerless in dealing with the growth rate and other difficulties in stock valuation, it is actually proved that there is not a qualified model or method for stock valuation so far. The prevailing stock valuations based on the Gordon growth model and the ratio methods, although widely used, perhaps are just play for show to a large extent.

4.3 The Stage of Finance as a Science

The current finance related journals and works are full of advanced or obscure theories and formulas. However, if the basic concepts are not clear and the fundamental problems are not solved, these advanced theories and formulas can only be castles in the air; imposing unnecessary and unrelated psychology, statistics, mathematics, artificial intelligence, etc. into finance before making clear the basic concepts and finding out the fundamental solutions does not help.

Bonds and stocks are the simplest and most common basic securities. The valuation of these securities is the basic task of finance. However, based on the preliminary discussion of these two chapters, the valuation of bonds and stocks has not been ideally solved. In bond valuation, even the discounting is used incorrectly; in the stock valuation, the basic problems, such as the dilemma of positive and negative growth rate, have not even been found, do not mention a solution.

More and more financial papers and books argue that our financial theory is so advanced that no new discoveries can be found without interdisciplinary studies. However, can a discipline solve its basic problem by the help of other disciplines? For instance, can we expect that psychology, statistics, mathematics, artificial intelligence, etc., to answer the question like whether a perpetual growth rate is positive or negative, or to solve the problem that encountered by the Gordon growth model, such as too sensitivity or infeasibility?

An interesting question is: does finance really reach its advanced stage? which stage is financial theory now on over its life cycle? Obviously, if a subject is characterized as some basic concepts remain unclear and most fundamental problems remain unsolved, it must be on its initial stage or start phase rather than advanced stage. The most urgent task for a subject on initial stage is to solve the fundamental problems within the field by itself rather than to search for outside cooperation or interdisciplinary studies.

We may say that Astronomy as a science now is in its advanced stage; but we have to say that Astronomy was in its initial stage 500 years ago, because most fundamental problems remained unsolved at that time. For instance, it was not sure whether earth goes around sun or sun goes around earth. Now, a similar unsolved problem in finance, the perpetual growth rate, a very basic input in finance, is not sure to be positive or negative!

Therefore, finance as a science (not as a practice) now is similar to Astronomy 500 years ago, or just in its initial stage. Just think that 500 years ago, in the time of Nicolaus Copernicus (1473–1543), how many years had been spent on the research of Astronomy? Comparably, the history of finance until today is much shorter than the Astronomy 500 years ago! This is not blamable, but rather, we know the emphases well when we know the stage of finance. For instance, interdisciplinary research may not be the right way now for financial research.

As a matter of fact, every discipline has its own basic concepts and methods, introduction interdisciplinary research too early, especially in the initial stage can only hurt the normal development of the discipline by messing up the basic concepts and logics. New concepts and theories come up endlessly in nowadays financial research, but most (if not all) of those new theories are helpless and useless except making scholars and practitioners more confused.

On the other hand, even on later stages of a discipline, the concepts and theories of other disciplines can only be introduced when they are really needed; introduction the concepts and theories from other disciplines unnecessarily or endlessly can also hurt the normal development of the discipline; do not mention abusing or misusing the concepts and theories of other disciplines! Therefore, interdisciplinary research is not an elixir; most tough financial problems cannot be solved by help from other disciplines, especially those fundamental problems in finance. Of course, if the problem itself belongs to interdisciplinary problem, it undoubtedly needs interdisciplinary research to solve.

Just as other independent disciplines, finance has become an independent discipline because its own methods can solve financial problems most effectively and uniquely. On the contrary, if the methods of other disciplines can more effectively solve financial problems, especially a single discipline, such as statistics, can really solve all or most of the financial problems, finance does not need to exist as an independent discipline anymore. Finance or financial research can be directly put as a branch of statistics. Unfortunately, this has not been widely noticed and understood in nowadays financial research.

Finance is a science and solving the problems (especially the fundamental problems) in finance will benefit finance as well as our society. What's important is that the research efforts should be put on the right direction and right way, so that the effective and efficient solutions with correct and simple methods can be found. Whether the efforts are put on the right direction and right way is determined by the essential features of finance as a science, i.e. a decision and application—oriented science based on valuation, as revealed in the first chapter.

The urgent task of financial research is Not a mechanized statistical test, Not an advanced math game, Not an agency theory game, Not behavioral research, Not interdisciplinary! But a real understanding of the basic or fundamental problems of this discipline!

Anyway, the problem needs us to solve in this chapter is still stock valuation.

5 A New Absolute Valuation Beyond Discounting

Previous discussions reveal that there is no qualified model in stock valuation within the mainstream of finance. This is not a good news, but is not necessarily a bad news either, because for science, finding a problem may imply a possible breakthrough.

5.1 New Criterion Other Than the Discount Rate

As revealed in previous discussions, for stock or equity valuation, the Gordon growth model has no room for improvement, so does the discounting or DCF methods. Then the problem now is not how to improve Gordon growth model, but how to find another way to solve the problem. However, it seems difficult to find a way other than discounting to calculate stock value.

As the only method so far in absolute valuation within the mainstream of finance, DCF or discounting method is actually based on an investment criterion of required rate of return. There is another investment criterion, i.e. required payback period. A creative idea is that it may be possible to find a new valuation method based on such an investment criterion.

Suppose the current year's earnings per share is E, the average annual growth rate of the earnings in foreseeable future is g, then the annual earnings in the consecutive n years will be $E(1+g)^1$, $E(1+g)^2$, $E(1+g)^3$, ..., $E(1+g)^n$ respectively. Note that the earnings per share in any given year represent the total return on the stock in that year, and the price of the stock represents the initial investment. If the required payback period is n, when this requirement is just satisfied (no more, no less), then:

$$P = E(1+g)^1 + E(1+g)^2 + E(1+g)^3 + \ldots\ldots + E(1+g)^n \tag{14}$$

Hence,

$$P(1+g) = E(1+g)^2 + E(1+g)^3 + \ldots\ldots + E(1+g)^{n+1} \tag{15}$$

3–10 minus 3–9,

$$gP = \left[(1+g)^n - 1\right](1+g)E \tag{16}$$

Hence,

$$P = \left[(1+g)^n - 1\right](1+g)E/g \tag{17}$$

Let the required payback period, n, represents the fair or reasonable requirement, then the stock price based on this payback period is a reasonable or fair price, which is the stock value.[6] Then, Eq. 17 is an absolute valuation model. Similar to Gordon growth model, Eq. 17 contains also a variable of growth rate g; it thus literally can be named as ZZ growth model.[7] Please note that the g in the ZZ growth model is an average growth rate over a finite time horizon, which is different from that in the Gordon growth model. The ZZ growth model thus breaks away and liberate easily from the ZZ growth paradox trouble, because the average growth rate over any foreseeable and finite period can certainly be positive.

Understandingly, in the ZZ growth model, the foreseeable future, though no certainty requirement for its length, is supposed to extend beyond the required payback period, so that the average annual growth rate of earnings, g, remains valid

[6] Here we use the original concept of the payback period rather than the discounted payback period prevailing in many financial books and literature. The discounted payback period is a misleading concept because it incorporates two competing criteria: the required rate of return and the required payback period. In fact, the required payback period is another side of the required rate of return, i.e., the required payback period = 1/ the required rate of return. Thus, the application of the discounted payback period implies that the return (or payback period) requirement is satisfied repeatedly (twice) in the calculation. This makes no sense and is obviously wrong.

[7] The ZZ growth model cannot be abbreviated to "ZZ model", because there are other models invented by the author, as showed in the rest of this chapter and the following chapters.

over the required payback period, n, and the investor can get all his or her money back at the end of the payback period. The model was first published in 2008 by Zhiqiang Zhang.

5.2 Solutions Based on ZZ Growth Model

Since the ZZ growth model is a fundamental innovation in finance and valuation, let us now test its power in solving practical problems like case 1 and 2. Note that in case 2, the required rate of return on D, E and F are all 10% implies the required payback period are all 1/10% = 10 years. The foreseeable period (about 15 years) is longer than the required payback period 10. Based on the ZZ growth model, the value of the three stocks are:

$$\begin{aligned} P_D &= \left[(1+10\%)^{10} - 1\right](1+10\%)/10\% = 17.53\ \text{(dollars)} \\ P_E &= \left[(1+20\%)^{10} - 1\right](1+20\%)/20\% = 31.15\ \text{(dollars)} \\ P_F &= \left[(1+30\%)^{10} - 1\right](1+30\%)/30\% = 55.41\ \text{(dollars)} \end{aligned}$$

Is the above valuation results yielded by the ZZ growth model reasonable or convinced? You can judge it based on your own intuition or experience.

Let us further try the case 1.

The case 1 is designed for applying the Gordon growth model. As revealed in first section of this chapter, it is beyond human intelligence to forecast a perpetual growth rate. However, based on the ZZ paradox, the growth rates of 9%, 9.9% and 9.99%, larger than zero, are actually the growth rate over a foreseeable period, which is in line with the growth rate in ZZ growth model.

When a firm maintains its retention ratio,[8] the growth rate of earnings will be equal to that of the dividends. Although the growth rate of earnings may be higher or lower than the growth rate of dividends in any given year, the average growth rate of earnings over a very long period should be close or equal to the average growth rate of dividends. The growth rates of 9%, 9.9% and 9.99% in case 1 thus can be viewed as the average growth rates of earnings of stock A, B and C over the foreseeable period (15 years or longer).

To derive the data of initial earnings per share, we need further the retention ratios of the three firms. For the convenience of demonstration, assume they maintain the same retention ratio of 30% over that foreseeable period. Note that $E_1 = E(1+g)^1$, the ZZ growth model can be rewritten as:

---

[8] Retention ratio is the percentage that a company retains earnings for funding its further operations and investments. The payout ratio is the percentage of the dividends payout in its net income or earnings.

$$\text{Retention Ratio} = \text{Retained Earnings}/\text{Net Income} = 1 - \text{Dividend Payout Ratio};$$

$$\text{Payout Ratio} = \text{Dividends Payout}/\text{Net Income} = 1 - \text{Retention Ratio}.$$

$$P = \left[(1+g)^n - 1\right]E_1/g \tag{18}$$

As their dividends in next year are all 1 dollar, their earnings per share in next year are:

$$E_{1A} = E_{1B} = E_{1C} = 1/30\% \text{ (dollars)}$$

Again, required rate of return 10% implies that the required payback period is 10 years. Based on the ZZ growth model, the values of A, B and C are:

$$P_A = \left[(1+9\%)^{10} - 1\right] \times 1/30\%/9\% = 50.64 \text{ (dollars)}$$
$$P_B = \left[(1+9.9\%)^{10} - 1\right] \times 1/30\%/9.9\% = 52.87 \text{ (dollars)}$$
$$P_C = \left[(1+9.99\%)^{10} - 1\right] \times 1/30\%/9.99\% = 53.10 \text{ (dollars)}$$

Comparing with the valuation results yielded by the Gordon growth model, that A, B and C are valued as 100 dollars, 1000 dollars and 10,000 dollars respectively, the results above yielded by the ZZ growth model obviously make much more sense.

5.3 Comparison of the Gordon Growth Model and ZZ Growth Model

It seems easy for the ZZ growth model to solve the common and typical valuation problems that the Gordon growth model or DCF method and the ratio approach cannot. Why? Generally speaking, the reason is that the ZZ growth model represents a breakthrough in valuation. Valuation with required payback period, as a valuation method based on risk and return, is another absolute valuation method other than discounting.

Anyway, as the Gordon growth model is used so widely, it is interesting to make a more detailed comparison of these two absolute valuation methods now.

We have known that Gordon growth model has two advantages: theoretical soundness and simplicity. The ZZ growth model possesses both of them too.

It is obvious that the ZZ growth model is simple in equation form and derivation process. The model takes only three key variables into account. Where, the E and g combine together as the consideration for the return of the asset; the n accounts for the risk of the relevant asset and n decreases as the risk increase. It even needs not discounting (see footnote 6 of this chapter), which is necessary for most (if not all) financial models.

As a fundamental axiom in finance, value (of an asset) increases with the increasing of the (expected) return and decreases with the increasing of the (expected) risk. Obviously, the ZZ growth model will result in a high value when the E and g are bigger (high return), and a low value when the n is smaller (high risk). Therefore, the ZZ growth model is completely in line with the fundamental axiom of return and risk determining value, hence the model is theoretically sound.

While the ZZ growth model possesses the same advantages as the Gordon growth model, it overcomes most of the shortcomings of the Gordon growth model.

Firstly, the growth rate in the Gordon growth model is a particular growth rate which needs to be valid over infinite horizon; this is much different from a common growth rate; the growth rate in the ZZ growth model, on the contrary, only needs to be valid over a foreseeable horizon, which is more in line with the common growth rate or what in our mind about growth rate. Therefore, the growth rate in the Gordon growth model is much more difficult to understand and forecast or estimate than that in the ZZ growth model.

Further, the Gordon growth model requires that the growth rate must be smaller than the discount rate, i.e. $g < k$. The condition of "$g < k$" restricts the application of the Gordon growth model seriously. For instance, it is hardly used to value high growth stocks. For the same reason, the valuation result is too sensitive to the growth rate, which implies the estimation of the growth rate must be highly precise. These restrictions deprive the basic feasibility of the model.

The ZZ growth model requires that the forecasting period should extend beyond the required payback period. This is just a reasonable requirement and poses no serious restriction for the application, since analysts should do their best to forecast the future returns of the asset for using any valuation model. As for the estimation of the growth rate, there is almost no restriction;[9] it can be any number greater than -100%. The model thus can be used conveniently to value stocks in any sectors, including traditional sectors and high growth sectors.

Secondly, both models take three variables into account. According to the Gordon growth model, stock value is determined by dividends in current year D_0, the perpetual growth rate of the dividends g, and the investors' required rate of return (risk-adjusted discount rate) k. In theory, the estimations of the three variables should be correct over infinite time horizon. Believe it or not, such requirements are actually beyond human intelligence or prediction ability.

According to the ZZ growth model, stock value is determined by earnings in current year E, the average growth rate of the earnings g, and the investors' required payback period (years) n. In theory, the estimations of the three variables should be correct over a finite time horizon rather than very long time or infinite time horizon. Such requirements are obviously more reasonable and feasible.

Earnings as a variable to represent return are more reliable and feasible to forecast than dividends. Even for the same finite time horizon, forecasting dividends and its growth is much more difficult than forecasting earnings and its growth, because the numbers of determinants on dividends are at least one more—the firm's dividend policy, which is further influenced by numerous factors.

A lot of firms employ residual dividend policy, under which they decide whether and how much to pay dividends in any year based on the earnings left after their

[9] When $g = 0$, the earnings per share will keep constant in the future; the ZZ growth model changes its form as: $P = nE$.

investment demand got satisfied. As such, even the internal managers have no idea about the future dividends, needless to say the outsider analysts or investors.

In addition, the ZZ growth model is more realistic and practical because of the following facts: (1) Firms in reality all have limited life expectancies, rather than grow or live "forever". (2) Investors in reality have limited forecasting capability, and are not willing to base their valuations or decisions on the returns over an infinite period. (3) Comparing with the required rate of return, the required payback period is somehow more certain in intuition as a criterion for investment or decision, especially for immature investors.

After all, comparing with the Gordon growth model, the ZZ growth model has at least the following important advantages:

Firstly, the ZZ growth model avoids the paradox trouble and gains feasibility by adopting the new criterion of required payback period.

Secondly, the ZZ growth model gets rid of the unreasonable restrictions on the variable of growth rate and is flexible enough to value stocks in various sectors.

Thirdly, the result yielded by the ZZ growth model is properly sensitive rather than over sensitive to the growth rate and other variables incorporated, hence is reliable.

Before the ZZ growth model, there is never a valuation or pricing model based on the criterion of "required payback period". The absolute valuation models so far, traditional or modern, simple or sophistic, are all based on the criterion of "required rate of return" unexceptionally. In such a sense, the ZZ growth model is not only a brand new model, but also a brand new way abreast to the discounted cash flow approach. We will see further the problem-solving power of the ZZ growth model in the rest parts of this chapter.

Of course, there is no "perfect model" in this world. The ZZ growth model has its own drawback. It is well known that payback period as an decision criterion cannot account for the cash flows or returns beyond the payback period. So does the ZZ growth model. Because of this, the model is more suitable for the valuation of assets with future cash flows relatively well-distributed over years. For instance, it is more suitable for valuing assets like stocks, but not suitable for bond valuation, because the cash flows of bonds (including a big principal repayment at the maturity) are usually not evenly distributed.

Please note that for stock valuation, the drawback of the payback period as a decision criterion does not matter in most of the cases, because we do not know much about the risk and return of the stock after the payback period. Put it another way, it is not much regretful for the ZZ growth model not to consider the returns of the stock beyond the required payback period because all the information forecasted is incorporated into the growth rate of the earnings, or there is no known information left outside the model or wasted.

6 The Theoretical Ratio Models

Fundamental breakthrough in any discipline implies profound application potential. The ZZ growth model represents a fundamental breakthrough in valuation as well as in finance, its application potential is also worth looking forward to. In this section, we try to improve the relative valuation or ratio method by using of this model.

6.1 ZZ P/E Model

The fatal defect of ratio method is that it derives an asset value just based on the multiples of its comparable assets without a convincing way to judge whether the comparable assets are fairly priced. Obviously, the effective way to remedy such a defect is to replace the actual price by the asset value, which leads to the theoretical ratio models.

While P/E ratio is usually the ordinary P/E ratio defined as the current share price divided by earnings per share in current financial year. Based on the Gordon growth model, replacing the actual price by the asset value, a P/E model can be derived out, i.e.,

$$\mathrm{P/E} = \frac{D_0(1+g)/E}{k-g} = \frac{dr(1+g)}{k-g} \tag{19}$$

We refer to Eq. 19 as Gordon P/E model, where dr represents the average dividend payout ratio over infinite horizon, which is equal to dividend divided by earnings.

Obviously, the Gordon P/E model is similar in form to the original Gordon growth model, hence carries on all the defects of the original Gordon growth model, such as the sensitivity to the perpetual growth rate g, the dilemma of the growth rate to be positive or negative. Thus, this theoretical P/E model is not feasible either.

Similarly, we can derive a theoretical P/E model based on the ZZ growth model via dividing the two sides of Eq. 17 by E:

$$\mathrm{P/E} = \left[(1+\mathrm{g})^{\mathrm{n}} - 1\right](1+\mathrm{g})/\mathrm{g} \tag{20}$$

Equation 20 can be referred to as ZZ P/E model.[10] The corresponding forward theoretical P/E model then is:

$$\mathrm{P}/\left[\mathrm{E}(1+\mathrm{g})\right] = \mathrm{P/E_1} = \left[(1+\mathrm{g})^{\mathrm{n}} - 1\right]/\mathrm{g} \tag{21}$$

Needless to say, the ZZ P/E model possesses all the advantages of the ZZ growth model, such as the easiness of the variable estimation, the feasibility, etc.

[10] Note that when $\mathrm{g} = 0$, the theoretical P/E model becomes: $\mathrm{P/E} = \mathrm{n}$.

The ZZ P/E model has also another obvious advantage: it has only two variables—one less than that of the Gordon P/E model. According to the ZZ P/E model, a fair or theoretical P/E is determined by the growth potential (g) and risk (n) of the asset. The fair P/E increases as the growth potential increases and decreases as the risk increases because the n becomes smaller. These relationships obviously make sense.

Traditionally, investors and analysts believe that the appropriate P/E for most stock range around 10 to 30. While this may make sense in some cases, this may also confuses the investors and analysts when they confront nowadays' high tech or dot.com stocks. When you see in the market that "Tesla" is trading at a P/E around 300, how do you judge the price? Are you sure it is over-valued or under-valued? How can you work out (rather than guess carelessly) a relatively certain answer about such kind of questions?

There are rarely effective tools suitable for solving these questions in current financial theory, although these questions are so common, so typical and so important for finance. That is one of the reasons why the related questions remain on-going debates. Applying the ZZ P/E model, however, is easy to work out reliable answers to the relevant questions based on responsible estimation on the earnings growth and the relevant risk of the stock.

For instance, if we believe that Tesla will growing in earnings at around 50% on average over a foreseeable period about 10–15 years, the investors' required payback period is 9 years based on its risk, its theoretical or fair P/E then is:

$$P/E = \left[(1+50\%)^9 - 1\right](1+50\%)/50\% = 112.33$$

The ZZ P/E model tells us that "Tesla" is not over-valued when its actual P/E is around 110–120 in the market so long as the inputs about the growth rate and required payback period are reliable. To get more intuition, Table 2 illustrates the theoretical or reasonable or fair P/Es based on the ZZ P/E model when the growth rate over a foreseeable horizon varies from -20% to 50% and required payback period varies from 8 to 13 years.

According to Table 2, when the required payback period is about 8–13 years, if the earnings of a stock are expected to grow annually at around 10% over more

Table 2 The theoretical P/Es

n	K (%)	g							
		−20%	−10%	0%	10%	20%	30%	40%	50%
8	12.50	3.33	5.13	8.00	12.58	19.80	31.01	48.15	73.89
9	11.11	3.46	5.51	9.00	14.94	24.96	41.62	68.81	112.33
10	10.00	3.57	5.86	10.00	17.53	31.15	55.41	97.74	170.00
11	9.09	3.66	6.18	11.00	20.38	38.58	73.33	138.23	256.49
12	8.33	3.73	6.46	12.00	23.52	47.50	96.63	194.93	386.24
13	7.69	3.78	6.71	13.00	26.97	58.20	126.91	274.30	580.86

than 8–13 years, the theoretical P/E ranges between 12 and 27. If the earnings are expected to grow at around 20% over more than 8–13 years, the theoretical P/E ranges between 19 and 59. If the earnings are expected to grow at 50% over more than 10 years, buying the stock is a safe investment when its actual P/E moves down below 170. If the earnings are expected to grow at -10% over more than 8–13 years, the theoretical P/E ranges between 5 and 7. These theoretical values obviously make sense.

On the other hand, the theoretical P/Es prove that the reasonable range is not always 10 to 30 for any stocks, especially for those high growth stocks. As indicated by the ZZ P/E model, the fair P/E depends on the expected growth rate and required payback period. For a stock with high growth potential and low risk, a P/E of 100 may not mean overvalued; however, if a stock has gloomy or dangerous future, a P/E of 5 may represent overvalued. It is clearly convenient to consider synthetically both the growth potential and risk in judging the fairness of a stock pricing with the ZZ P/E model.

For instance, based on the PEG ratio, when the growth rate equal to or below zero, PEG cannot provide a standard or fair value of P/E. Fortunately, similar problems can be easily solved based on the ZZ P/E model. As shown in Table 2, if the required payback period is about 8–13 years, when the earnings are expected to remain unchanged in average over foreseeable horizon, the theoretical P/E ranges between 8 and 13 rather than the zero derived from the PEG ratio. Similarly, when the earnings are expected to grow at about -20% in average over foreseeable horizon, the theoretical P/E ranges between 3 and 4. Obviously, the ZZ theoretical P/E model make more sense and is more useful and helpful than the PEG ratio.

Figure 2 shows the fair or theoretical P/E curves (FPC) based on the ZZ P/E model. X axis represents the expected growth rate and Y axis represents the fair P/E

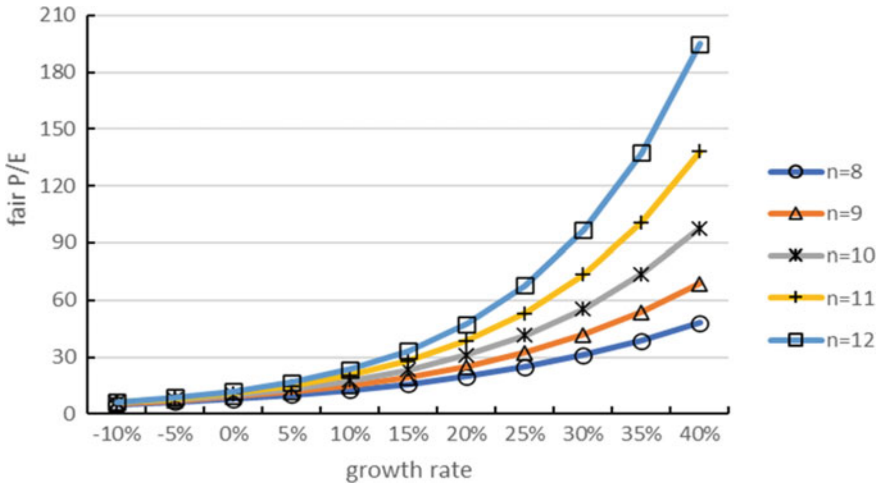


Fig. 2 Fair P/E Curves based on the ZZ P/E model

ratio. The 5 FPCs in Fig. 2 represent the required payback period of 8, 9, 10, 11 and 12 years respectively. Every FPC goes up with the increasing of g, and the position of the FPC moves upward with the increasing of n. Hence, for relative valuation, if investors believe a stock will grow faster in the future, they should use a higher P/E to value it; similarly, if investors believe the stock will experience more uncertainty or risk, they should use a shorter n or lower P/E to value it.

6.2 ZZ P/B Model

The P/B is the ratio of the stock price to the book value of its equity or net assets per share. Where, B is the net assets in the balance sheet divided by the firm's number of shares outstanding. As P/B ratio is also widely used in market and practical valuation, we now try to work out a fair or theoretical P/B model.

The earnings per share, E, is actually the product of the equity per share, B, and the return on equity (ROE). Let r_e to represent the ROE, then,

$$E = Br_e \tag{22}$$

Therefore, the ZZ P/E model can be rewritten as:

$$P/E = P/(Br_e) = \left[(1+g)^n - 1\right](1+g)/g \tag{23}$$

Hence,

$$P/B = \left[(1+g)^n - 1\right](1+g)r_e/g \tag{24}$$

Equation 24 can be referred to as ZZ P/B model. Similarly, a new valuation model can be derived by rearranging the ZZ P/B model as:

$$P = \left[(1+g)^n - 1\right](1+g)Br_e/g \tag{25}$$

Obviously, the theoretical P/B is the product of theoretical P/E and the return on equity. Similar to the ZZ P/E model, the variables in the ZZ P/B model are easy to estimate. Please note that the decrease of the return at a percent within 100% ($g > -100\%$), does not mean a negative return, but rather a smaller positive return; hence the return on equity, r_e, the average rate of return on equity over foreseeable horizon, should be positive in normal case, otherwise, the firm or the stock is not sustainable or not worth to exist anymore.

Comparing with the ZZ P/E model, the ZZ P/B model needs one more independent variable, r. Other things being equal, more variable means more estimated biases. On the other hand, the value of an asset is strongly related to its future profitability and risk, while relative weakly related to its book value. Therefore, if both P/E and

Table 3 The fair P/Bs when the return on equity is 20%

n	k (%)	g							
		−20%	−10%	0%	10%	20%	30%	40%	50%
8	12.50	0.67	1.03	1.60	2.52	3.96	6.20	9.63	14.78
9	11.11	0.69	1.10	1.80	2.99	4.99	8.32	13.76	22.47
10	10.00	0.71	1.17	2.00	3.51	6.23	11.08	19.55	34.00
11	9.09	0.73	1.24	2.20	4.08	7.72	14.67	27.65	51.30
12	8.33	0.75	1.29	2.40	4.70	9.50	19.33	38.99	77.25
13	7.69	0.76	1.34	2.60	5.39	11.64	25.38	54.86	116.17

Table 4 The fair P/Bs when the return on equity varies with growth rate

n	k (%)	g							
		−20%	−10%	0%	10%	20%	30%	40%	50%
8	12.50	0.53	0.92	1.60	2.77	4.75	8.06	13.48	22.17
9	11.11	0.55	0.99	1.80	3.29	5.99	10.82	19.27	33.70
10	10.00	0.57	1.06	2.00	3.86	7.48	14.41	27.37	51.00
11	9.09	0.59	1.11	2.20	4.48	9.26	19.07	38.71	76.95
12	8.33	0.60	1.16	2.40	5.17	11.40	25.12	54.58	115.87
13	7.69	0.60	1.21	2.60	5.93	13.97	33.00	76.80	174.26

P/B theoretical ratios are available, the priority should be given to the theoretical P/E model.

Based on the ZZ P/B model (Eq. 24), Table 3 illustrates the theoretical P/Bs when the return on equity is assumed to be 20%.

Consider the relation between the growth rate and the return on equity, assume the return on equity in average is 20%, but will change in the same direction with the growth rate in a way of: $r_e = 20\% \times (1 + g)$. Then, the theoretical P/Bs are recalculated as Table 4.

From Tables 3 and 4, obviously, when the growth rate is negative, the theoretical P/Bs are getting smaller, and when the growth rate is positive, the theoretical P/Bs are getting bigger; this reflects the double effects of the growth rate and the return on equity.

6.3 *ZZ P/S Model*

It is easy to derive a ZZ P/S model via similar way as that of the ZZ P/B model. Let r_s to be the sales or profit margin, or the ratio of earnings (net income) to sales; as earnings per share (E) equals to sales per share (S) multiplied by the profit margin, i.e., $E = Sr_s$, then, The ZZ P/E model can be rewritten as:

$$P/E = P/(Sr_s) = \left[(1+g)^n - 1\right](1+g)/g \tag{26}$$

Hence,

$$P/S = \left[(1+g)^n - 1\right](1+g)r_s/g \tag{27}$$

Equation 27 can be referred to as ZZ P/S model. Similarly, a valuation model can be derived by rearranging the ZZ P/S model:

$$P = \left[(1+g)^n - 1\right](1+g)Sr_s/g \tag{28}$$

Obviously, the theoretical P/S is the product of the theoretical P/E and the profit margin. The variables in the ZZ P/S model are also easy to estimate or forecast. Similar with the ZZ P/B model, negative growth does not mean the negative profit, and the profit margin r_s should be positive in most cases. Also, the priority should be given to the ZZ P/E model when both the theoretical PE and P/S are available.

Based on the ZZ P/S model, Table 5 illustrates the theoretical P/Ss when the profit margin is assumed as 15%.

Consider the relation between the growth rate and the sales profit margin, assume the profit margin on average is 15%, but will change in the same direction with the growth rate in a way of: $r_s = 15\% \times (1+g)$. Then, the theoretical P/Ss are recalculated as Table 6.

In summary, we derive the theoretical P/E, P/B and P/S model based on the ZZ growth model; hence improve the ratio method fundamentally and bridge the gap between the relative valuation and absolute valuation. The traditional experience-based P/E, P/B and P/S ratios are actually not convincing as a valuation benchmark, because the ratios of the market on average and the individual stocks are fluctuating endlessly and relative valuation itself cannot tell ratios on which day or average ratios over which period are the most correct. The theoretical P/E, P/B and P/S based on

Table 5 The theoretical P/Ss when the profit margin is 15%

n	k (%)	g							
		−20%	−10%	0%	10%	20%	30%	40%	50%
8	12.50	0.50	0.77	1.20	1.89	2.97	4.65	7.22	11.08
9	11.11	0.52	0.83	1.35	2.24	3.74	6.24	10.32	16.85
10	10.00	0.54	0.88	1.50	2.63	4.67	8.31	14.66	25.50
11	9.09	0.55	0.93	1.65	3.06	5.79	11.00	20.74	38.47
12	8.33	0.56	0.97	1.80	3.53	7.12	14.49	29.24	57.94
13	7.69	0.57	1.01	1.95	4.05	8.73	19.04	41.15	87.13

Table 6 The fair P/Ss when the profit margin varies with growth rate

n	k (%)	g							
		−20%	−10%	0%	10%	20%	30%	40%	50%
8	12.50	0.40	0.69	1.20	2.08	3.56	6.05	10.11	16.62
9	11.11	0.42	0.74	1.35	2.46	4.49	8.12	14.45	25.27
10	10.00	0.43	0.79	1.50	2.89	5.61	10.80	20.53	38.25
11	9.09	0.44	0.83	1.65	3.36	6.94	14.30	29.03	57.71
12	8.33	0.45	0.87	1.80	3.88	8.55	18.84	40.94	86.90
13	7.69	0.45	0.91	1.95	4.45	10.48	24.75	57.60	130.69

the ZZ growth model hence can play a vital role in valuation and finance, which is illustrated more clearly in the following section.

7 Some Applications of ZZ Ratio Models

Sound in theory, simple and feasible in practice, as a brand new approach in absolute as well as relative valuation, the ZZ growth model and ZZ ratio models have undoubtedly vast application potentials. The applications demonstrated in this section are by no means to prescribe some limits to the application of these models; rather, they are just some initial or rough examples for inspiring the uses of these models.

7.1 Calculating the Fair-Priced Ratios

Value-based investments are recommended in market. Value-based investors should know first of all the values of the candidate stocks, or at least, know whether the candidate stocks as well as the market as a whole are over-valued or under-valued.

For instance, many countries have two stock markets: one is the market for traditional stocks, such as the New York Stock Exchange (NYSE), often referred to as main board market; the other is for the high tech or high growth stocks, such as the Nasdaq (National Association of Securities Dealers Automated Quotations), often referred to as GEM (Growth Enterprises Market) board or the second board market.

The ratios should vary across markets because stocks in different markets have different fundamentals concerning risk and return. While most investors know that the P/E, P/B and P/S of stocks in Nasdaq should be higher than that in NYSE, seldom among them know a method sound in theory for measuring the ratio differences; the investment decisions and regulation policies in reality thus have to be made based merely on the nebulous experience.

With the guidance of the theoretical ratios based on the ZZ ratio models, such as those used in the calculation of Tables 2, 3, 4, 5 and 6, the practical investment decision-making and regulation policy-making can be improved fundamentally.

For instance, if we believe that the growth rates of earnings on average are expected to be 10% and 30% over the future 10–20 years for the main board market and the second board market respectively; the required payback periods are 11 and 9 years for the investment in the two markets respectively, then the fair P/Es of the main board market and the second board market can be derived based on the ZZ P/E model:

$$P/E_{\text{main board}} = \left[(1 + 10\%)^{11} - 1\right](1 + 10\%)/10\% = 20.38$$
$$P/E_{\text{second board}} = \left[(1 + 30\%)^{9} - 1\right](1 + 30\%)/30\% = 41.62$$

The investors can judge the deviation of the market from fair valuation based on the benchmark derived from the above process and decide their overall investment strategy, such as the mix of cash and stock positions. They can also use the same method and benchmark to judge an individual stock and decide the componential stock selection for their portfolios as well as the timing of buying and selling of the componential stocks.

The regulators, such as the responsible person in the relevant government office and in the exchange, can announce the valuation benchmark of the market as a whole as well as the individual stocks derived from the ZZ P/E model as the guidance on some compulsory or voluntary bases to the market trading, so that reduce the manipulative and speculative behaviors in the market and keep the market more healthy and stable.

Although the theoretical P/E is not over sensible to the determinants of growth rate g and required payback period n, the two determinants themselves do have significant influences on the derived benchmark, the theoretical P/E. The values of the two determinants will vary across persons to do the estimations, so the benchmarks based on the same financial model will be different for different persons.

This is not a problem worthy to worry about; rather, it is necessary for every stock to be traded on the market. Otherwise, as all investors in the market derive exactly the same fair value of a stock, the stock price will go fast toward this value, and then the trading volume of this stock will go down sharply to zero, because when the stock price equals its value, both the buying and selling of this stock are profitless and will be stopped.

Anyway, the investors and regulators as well as the analysts themselves, are responsible to do their best to estimate the inputs of the model in use. This is a general requirement for all the applications of financial models, just as what revealed in Chapter “Finance and Its Fundamental Problems”.

7.2 Measuring the Market Bubble

There are many important financial issues relied on the theoretical ratios. For instance, why has the measurement of market bubble been debated endlessly? The main reason is that no one has a theoretical ratio as a benchmark, whereas actual ratios cannot prove themselves are correct or incorrect. Once the theoretical P/E is known, it is easy to judge whether a market has bubble as well as whether a stock is over-valued.

Let us take China and the US Stock markets as the examples; and use the ZZ P/E model to measure the theoretical P/Es and the bubbles of the two markets.

The overall economy of China as measured by GDP had been growing at a rate around 10%, and down to 6% during recent 10 years. Considering the public companies are the relative good part among the whole economy, assume they on average will grow at 8% annually over a foreseeable period in the future. Based on the average risk over the same period, suppose the investor required rate of return is 10%, which implies the required payback period is 10 years. Then the fair or theoretical P/E is:

$$\mathrm{P/E} = \left[(1+8\%)^{10} - 1\right](1+8\%)/8\% = 15.65$$

Similarly, the growth rate of US GDP is around 5% over a long term in past. Assume the growth rate of all US listed firms is 4% over a foreseeable period in the future, and the investor required payback period is also 10 years. Then the theoretical P/E is:

$$\mathrm{P/E} = \left[(1+4\%)^{10} - 1\right](1+4\%)/4\% = 12.49$$

The economic growth both in China and the U.S. can only support a theoretical P/E ratio around 12–16. The double growth rate of China can only support the P/E in China 1/4 (≈15.65/12.49–1) higher than that in the U.S. These secrets cannot be revealed by historical data based research on P/Es or other ratios.

The fair or theoretical P/E is the bubble-free P/Es. If the bubble-free P/Es in China and in US market are indeed 15.65 and 12.49 respectively, the bubbles of the two markets can be measured easily. If, for example, we find that one day, the actual P/Es are 20 and 15 in China and US market respectively, then the market bubble (MB):

$$\begin{aligned} \mathrm{MB}_{\mathrm{China}} &= 20/15.65 - 1 = 27.80\% \\ \mathrm{MB}_{\mathrm{US}} &= 15/12.49 - 1 = 20.10\% \end{aligned}$$

Therefore, when the reliable theoretical P/E is easy to calculate, the market bubble is easy to be measured effectively and timely, which is the key to monitor and protect against financial crises. That is, if the market bubble is inflated too big, for instance, close to or even beyond 50% or 100%, the responsible institution, such as the exchange, can warn the investors the bubble risk and suggest the prudence principle to the market. This is obviously helpful for avoiding the stock market crash. Thus, the model of the ZZ growth model and the ZZ ratio models solve two issues simultaneously: the valuation of stock and the measurement of market bubbles. As such, the appropriate applications of these models hopefully will enhance our ability greatly to guard against financial crises and stock market disasters.

7.3 Predicting the Change of Stock Price

Stock price fluctuates around its value. New findings or improvements on valuation thus benefit the predicting of stock price change. It is impossible to invent a model capable of predicting the change of stock price precisely, because the stock price goes around rather than equals to its value for most of the time. It makes no sense to test a financial or valuation model based on the precision of stock price prediction in most cases.

As valuation models, the ZZ growth model and the ZZ ratio models can be used to predict stock price change associated with some events or policy. Although we cannot expect a precise prediction, the prediction can undoubtedly help and underpin the relevant investment decision-making and regulation policy-making.

Events and policies affecting stock prices via the value determinants, such as sales, cost, earnings, cash flows and the required rate of return of investors. So we can estimate the influences of the relevant events and policies on these value determinants; then apply the ZZ growth model and the ZZ ratio models to predict the appropriate stock price change.

Estimation of the change in stock value resulted from earnings change

Firms in reality may encounter various events or changes during their operations. Such as fierce competition from other firms, the fluctuation of customers' demand (firms' sales), the change of cost especially the cost of inputs or the variable cost, the change of prevailing product price in the market, etc. These ever-changing conditions will influence the earnings of the relevant firms and their stock value and prices.

Consider that a firm runs into a sales change because of a change of production capability of the sector. According to the principle of the leverage of operating and financing, the change in sales will lead to a larger change in earnings. Suppose the relative change in the estimated earnings (decline or rise) is x over the foreseeable future. Use P and P' to represent the stock value before and after the sales and earnings change being forecasted respectively, based on the ZZ growth model:

$$P' = \left[(1+g)^n - 1\right](1+g)E(1+x)/g \tag{29}$$

Based on Eqs. 29 and 12, the relative change of the stock value X is:

$$X = P'/P - 1 = x \tag{30}$$

Equation 30 implies the relative change of the stock value theoretically equals to the change of the earnings when the change of the earnings is expected to extend over and beyond the required payback period. Most sectors in reality experience a cyclic or periodic change (up and down). So the earnings may not remain rising or falling over a very long period. It thus makes sense to consider the question like how long the change will last.

For example, a firm runs into a sales decline because new competitors move in. The estimated earnings will be 30% less than the original estimation over the next 3 years; then earnings will recover to their original estimated levels (because some competitors move out) over the rest of the foreseeable period. Suppose the investors' required payback period (shorter than the foreseeable period) remains 8 years; the current earnings per share is 2 dollars; the original estimated annual average growth rate of the earnings is 15%.

Stock value before the earnings change

$$\mathrm{P} = \left[(1 + 15\%)^8 - 1\right](1 + 15\%) \times 2/15\% = 31.57$$

Stock value after the earnings change

$$\mathrm{P}' = (1 - 30\%) \times 2 \times \sum_{t=1}^{3} (1 + 15\%)^t + 2 \times \sum_{t=4}^{8} (1 + 15\%)^t = 29.18$$

Hence, the relative change of the stock value X is:

$$\mathrm{X} = 29.18/31.57 - 1 = -7.6\%$$

The decline of 7.6% is much smaller than 30%, which is the decline of the stock value when the estimated earnings decline of 30% will not get recovered over the foreseeable period. This reminds us that when we use the ZZ growth model and ZZ ratio models to predict the value or value change of a stock, be careful about the period over which the influence of the relevant event will be lasting.

Obviously, the decline of the stock value should increase as the number of years before the earnings recover to the normal level increases. As for the above example, the special percentages of the decline in the stock value are shown in Table 7.

Estimation of the change in stock value resulted from interest change

In modern economy, the central bank often adjusts the national economy via the adjustment of interest rate. Both the investors and the central bank want to know

Table 7 The decline of the stock value along with the number of years before the 30% decline on earnings recover to the normal level

Number of years	1	2	3	4	5	6	7	8
Stock value	30.88	30.09	29.18	28.13	26.92	25.53	23.94	22.10
Value decline (%)	−2.2	−4.7	−7.6	−10.9	−14.7	−19.1	−24.2	−30.0

how much the stock market will react to a certain interest rate adjustment. Let us analyze its influence on stock market based on the ZZ P/E model now.

Assume the central bank adjusts the interest rate by a size of x; the original investors' required rate of return is k. Other things (such as earnings) being equal, the investors' required rate of return after the interest adjustment should be k + x. Therefore, the required payback period of investors before and after the interest adjustment should be 1/k and 1/(k + x) respectively. Use P and P' to represent the stock value before and after the interest adjustment respectively, based on the ZZ growth model:

$$\mathrm{P} = \left[(1+\mathrm{g})^{1/\mathrm{k}} - 1\right](1+\mathrm{g})\mathrm{E}/\mathrm{g} \tag{31}$$

$$\mathrm{P}' = \left[(1+\mathrm{g})^{1/(\mathrm{k}+\mathrm{x})} - 1\right](1+\mathrm{g})\mathrm{E}/\mathrm{g} \tag{32}$$

Hence, the relative change of the stock value X is:

$$\begin{aligned} \mathrm{X} &= \mathrm{P}'/\mathrm{P} - 1 \\ &= \left\{\left[(1+\mathrm{g})^{1/(\mathrm{k}+\mathrm{x})} - 1\right](1+\mathrm{g})\mathrm{E}/\mathrm{g}\right\}/\left\{\left[(1+\mathrm{g})^{1/\mathrm{k}} - 1\right](1+\mathrm{g})\mathrm{E}/\mathrm{g}\right\} - 1 \\ &= \left[(1+\mathrm{g})^{1/(\mathrm{k}+\mathrm{x})} - 1\right]/\left[(1+\mathrm{g})^{1/\mathrm{k}} - 1\right] - 1 \end{aligned} \tag{33}$$

For example, consider stock H, the current earnings per share is 2 dollars; the estimated annual average growth rate of the earnings over foreseeable period (15 years or so) is 12%; the original investors' required rate of return is 10%; now the central bank cut down the interest rate by -0.5%. Other things (such as earnings, etc.) being equal, the investors' required rate of return then should be 9.5%. Based on Eq. 33:

$$\mathrm{X} = \left[(1+12\%)^{1/9.5\%} - 1\right]/\left[(1+12\%)^{1/10\%} - 1\right] - 1 = 9.1\%$$

That is, other things being equal, the 0.5% reduction in interest rate will result to a 9.1% rise of the stock value or price. If there is no other important news released, once the interest rate reduction announced, investors can consider buying stock H when its price decline exceeds 9.1%, or consider selling stock H when its price decline is less than 9.1%.

The calculations about stock H can also be used to analyze the market as a whole. If the officers in the central bank want to smooth away the fluctuations of the stock market with the fundamentals like the above stock H; and predict that the change of the various environmental factors will result to the stock prices fall down about 9.1%; then they can adopt a monetary policy to cut down the interest rate by 0.5%.

The theoretical P/B and the theoretical P/S are also bubble-free ratios, hence can also be used to calculate the fair ratios, to measure the bubbles of market and the specific stocks and to predict the change of stock price associated with some events, just like the applications of the ZZ P/E model. As the processes are similar and easy, we do not intend to illustrate their applications one by one further.

8 Summary of the Chapter

The series of valuation models in this chapter, i.e. the theoretical P/E, P/B and P/S models as well as the ZZ growth model represent a brand new way in valuation and an innovation in finance. These models solve the key valuation issues in absolute valuation—avoid the ZZ growth paradox trouble by replacing the required rate of return with the required payback period as the criterion. The valuation models are flexible enough to value an individual stock and a portfolio across sectors; the growth rate as a key input is feasible to be forecasted and considered. These models solve as well the key valuation issues in relative valuation—provide an effective way to find the theoretical ratios in valuation as well as to measure the bubbles of an individual stock and the overall market. Consequently, these models bridge the relative valuation and the absolute valuation. These features betoken the vast potentials in theory and practice of this brand new valuation approach.

Further Readings

1. Gordon MJ, Shapiro E. Capital equipment analysis: The required rate of profit. Manage Sci 1956;3(1):102–10.
2. Gordon MJ, Dividends, Earnings and stock prices. Rev Econ Stat. The MIT Press 1959;41(2):99–105.
3. Gordon MJ. The savings investment and valuation of a corporation. Rev Econ Stat. 1962;44(1):37–51.
4. Zhang Z. Discussion on the relationship between P/E ratio and growth rate - also on issues related to stock pricing, Research on Financial Issues 2008; 1:67–72.
5. Zhang Z, *ZZ financial discovery*. University of International Business and Economics Press, Nov 2008.
6. Zhang Z, Zhao Q. Is there really a positive sustainable growth rate—Fundamental Rethinking on Financial Theory, Contemporary Finance and Economics 2010; 48–58.

Further Discussion: Is Practice Ahead of Theory? —Based on the Case of Finance

Contents

1 Introduction

It is often heard that "practice has gone in front of theory".

This statement or parlance contains at least three meanings. One is that the practice went ahead. The second is that the theory fell behind; The third is that this situation is abnormal, which is worth reminding and calling a upside down change.

In reality, this statement spreads widely and can be heard quite often, which shows its broad and far-reaching impact. Although it may be more often an oral expression, with different intentions and purposes, but it seems welcome in most circumstances and the audience often response with acquiescence, agreement or approval, and there are few doubts and objections. Of course, it is seldom to see the discussion on this statement.

This statement has been widely used in the financial field in recent years with the emergence of fintech. In view of its high-frequency use and wide influence, it is necessary to think about it. Is this statement correct? This problem involves not only the attitude towards theory, but also the relationship between theory and practice. We now try to make a discussion on this topic within the financial field.

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_4

Indeed, due to the application of modern information technologies (IT) such as big data, cloud computing, artificial intelligence and blockchain, the operations of payment, settlement, deposit and loan, investment and financing, credit evaluation, as well as risk management in financial business are changed a lot. Since the relevant practices have not been mentioned in financial theory, some people say that "practice went ahead of theory". Some people even say "practice went ahead of theory" for all the new things. In front of new changes, new situations and new problems in finance, theoretical research began to catch up in a hurry for fear of being "outstripped" more someday in future.

The question is, should "new things" be put forward by theory first? Otherwise, it will be "abnormal" if "practice preceded theory"? With the application of new technical means, does the old financial theory need to be updated? Even need to update in advance? It is conceivable that if these puzzles are not clarified, they will eventually hinder the progress of theory or practice. Therefore, it is necessary to make a serious discussion instead of parrot another's statement without real understanding.

Financial theory is a theory about the principles of financial decision-making and management, not about the technology (machine and tools) used in financial business. Just as business or management theory is not about the technology used in companies. The technology used in finance belongs to the research object or scope of IT application, which is not the same as financial theory.

Obviously, errors in financial models, such as the relationship between bond interest payment frequency and bond value in current prevailing models, will not and cannot be corrected by the application of big data or IT method; Problems not solved in financial theory, such as the optimal capital structure, will not and cannot be solved by the application of big data or IT method. In other words, the mistakes and problems in the financial field can only be corrected and solved by the breakthrough and application of financial theory itself, that is why finance is an independent, unique and irreplaceable subject.

The clarity of the concept is the premise to reach a correct conclusion. Therefore, it should make clear that what is worth discussing here is which in financial theory and financial practice comes first, rather than which comes first in the use of IT methods in finance.

2 Some Basic Facts

Marked by Markowitz portfolio theory, modern financial theory separated from economics and became an independent scientific branch in the 1950s.

Why is Markowitz's portfolio theory perceived as the start point? Because the essence of finance is the value of assets, or how to trade off between the risk and return to value an asset, or the transformation among the risk, return and value. For the first time, portfolio theory clearly shows how to make a quantitative trade-off between risk and return. Therefore, this theory is recognized as the start point of financial theory.

Then, when did financial practice begin?

Of course, the beginning of financial practice is marked by the emergence of financial business. At present, there are many forms of financial business, such as commercial banks, investment banks, securities issuance and investment, venture capital, private equity, securities companies, fund companies, credit guarantee, life insurance, property insurance, commercial insurance and so on. They got start at different time. Taking the commercial banks as an example, early commercial banks appeared about 500 years ago[1]; More than 300 years ago, the first joint-stock bank, the Bank of England, appeared.

The decisions in banking certainly needs to trade off between risk and return. However, it was not until last 50 s, when Markowitz portfolio theory was published, that people found first time the basic theory and quantitative methods to trade off between risk and return. Even today, people have not found a sound quantitative method to trade off between the specific risk and return in bank loan business. For example, how to determine the loan interest rate? Many monetary and banking textbooks introduce the cost markup pricing method for the interest rate determination, which is obviously not sound in theory because it fails to trade off between the loan risk and return.

Something in the current FinTech practice break through the relevant theory to some extent, or a little ahead of theory, comparing with the practice of banking industry, it is nothing. The fintech practice is a head of theory a couple of years; but the banking is ahead of theory hundreds of years. So far, the banking theory has not come up with a powerful answer to the model of determining loan interest rate. As we turn a blind eye to the fact that practice is ahead of theory hundreds of years., why should we call in question when practice is ahead of theory a few years?

The same is true from the perspective of financial products. Option is a good example, according to written records, there were products similar to options in 550 BC[2]; Later in Holland, options were used widely during the tulip foam, then options were used more widely in the early stock markets in Europe and the United States. The practice of option designing and option trading thus has gone through more than 2500 years.

Value is the basis of transaction price. It can be imagined that in the long history, the method or model of option value calculation has always been needed in practice. However, the research on option pricing theory appeared earlier in the nineteenth century, and more in the twentieth century.[3] After near 100 years of research efforts by

[1] As early as 1580 in Venice, Italy, there was the "Bank of Venice". The English word "bank" is transformed from the Italian "Banca". After the establishment of the Bank of Venice, banks were also established in other cities in Italy, Amsterdam in the Netherlands in 1609 and Nuremberg in Germany in 1621.

[2] According to Aristotle's politics, in about 550 BC, Thales, an ancient Greek philosopher, used his money in pocket as a deposit, locked the lease of the olive oil press in the coming year and got success to make a large sum of money.

[3] In the nineteenth century, the research on option pricing mainly focused on at the money option. By 1900, Louis Bachelier, a student of the French mathematician Henri Poincaré, published his doctoral thesis, the theory of speculation, which studied the option pricing problem by using stochastic

generations of scientists, a breakthrough was finally made in 1973 and a convincing model was obtained, i.e., the Black–Scholes option pricing model, which can be perceived as the theory of option.

From the earliest option transaction and pricing practice to Black–Scholes model, the time span is about 2000 years. This implies that option pricing practice is ahead of theory about 2000 years. If we are not surprising to this time gap, why are we surprising about current practice which is only ahead of theory for some years?

So why do people turn a blind eye to the situation that the practice is hundreds or thousands of years ahead, but talk much about the practice only several years ahead? This is probably caused by the limitations of people's own horizons. In other words, it is not easy for people to connect the two things that are hundreds or thousands of years apart, so they will not compare with each other and find the sequence between them; The two things not far apart are easy to connect, compare and draw an "ahead" or "behind" conclusion.

It should be realized that some statements are widely popular but incorrect. For example, seeing is believing. Today, human civilization has lasted for thousands of years, and human footprints have spread all over every corner of the earth, even to the moon, the solar system and outer space. Why should we use our eyesight to limit our understanding of the outside world? We should understand that in terms of eyesight, human beings are far inferior to many low-level animals; The advantage of human compared with other animals is not in the eyes, but in the brain. Why not use our brain to accept and inherit the achievements of civilization for thousands of years, and be willing to limit our own knowledge by "seeing is believing"?

Return to the theme here, that is: is practice really ahead of theory? From the above facts, we can see that practice is indeed ahead of theory, and in the financial field, it is hundreds of years or thousands of years ahead sometime. In contrast, that the practice just a few years ahead is not worth mentioning at all. Throughout history, the practice ahead of theory is inevitable and normal. Because the opposite situation is unimaginable. For example, before the emergence of commercial banks, there were banking theory and textbooks! Early banks were operated according to the textbook of banking! Is that possible? Thus, it is normal to see practice ahead of theory.

So, is there any opposite facts in history?

Take the core function of finance as an example, i.e. value calculation. In the 1960s, the research on option pricing heated up rapidly, because people realized that the value of assets came not only from the future fluctuating cash flows, but also from the future contingent cash flows; The only correct way to value future contingent cash flows is option pricing. The problem of option pricing was basically solved by the Black–Scholes model in 1973. According to reason, many financial problems can be solved by using of option pricing theory and method, i.e. those related to contingent cash flows.

---

processes. His model already contains important ideas and elements to solve the option pricing problem. Later, in the 1960s, Sprenkle formula (1961), Boness formula (1964) and Samuelson formula (1965) appeared respectively, which further approximated the Black Scholes model. In 1973, Black and Scholes published the option value model, which can be considered as basically solving the option pricing problem.

For example, option pricing can be applied to asset valuation, capital budgeting (project investment decision), risk management, pricing of financial products (such as credit guarantee, loan interest rate and various insurance products). More precisely, these financial calculations must use the option pricing method to get the correct results. Whether it is project investment or enterprise development, there will inevitably be future flexibility and uncertain development space, which is called real option; These real options may account for a large proportion of the project and company value. For example, they are common for more than 50% or even more than 100% for the high-tech companies. Conceptually, based on the contingent characteristics of their future cash flows, they can only be correctly evaluated by option pricing method.

Traditional fields also need option model. For example, when some companies suffered from long lasting losses and high bankruptcy risks. They may come to an agreement with banks on converting bank loan into equity. A common computing challenge in the converting is the ratio of bank loan to company equity. Since companies considering debt to equity swap or conversion often have poor management, serious losses and even insolvency, it is impossible to determine the ratio of debt to equity swap by using the traditional equation of "assets = Liabilities + equity", because the book value of liabilities often exceeds the total value of assets; on the other hand, the nature of limited liability determines that the original equity value of the company should not be less than zero. Similarly, as recognized in the practice, such a problem cannot be solved by traditional discount calculation. However, using option pricing to solve this kind of problem is a piece of cake.

This means that without the application of option pricing method, investment analysis and valuation are unlikely to get correct conclusions; Many practical pricing and computing problems cannot be solved. Before the breakthrough of option pricing theory, similar problems can only be decided based on intuition or subjective judgement, because there is no other way. Fortunately, after more than 70 years of research, the option pricing has been solved, and the model has been available for unlimited free use. However, it is a pity that there are few examples of using option pricing model to solve problems or assist decision-making in practice, except for putting on the option pricing formula just for showing the high level of the report. For example, the application of option pricing to determine the debt to equity ratio in the case of debt to equity swap is uncommon. Another example is the valuation of start-ups. Because the current business often suffers serious losses, it cannot turn losses into profits in the foreseeable future. If the future contingent business is not considered, it is almost impossible to obtain a positive value. However, due to various reasons, the option pricing model is put aside, the "ideal" appraisal value is obtained by "adjusting" the parameters in the discounting method or the market method in practice.

This means that the practice is still in the era of discount calculation,[4] or the period before the breakthrough of option pricing, such as the 1940s–1960s. In other words,

[4] In 1938, John Burr Williams proposed in his investment value theory that the investment value of common stock is the discounted value of its future dividends. It is generally believed that this is the earliest elaboration and application of the discount method.

the financial theories and methods used today, are no different from those before the 1970s. So, is practice ahead of theory? From the perspective of the application of option pricing, it is just the opposite; The theory discovered 50 years ago has not been put on the stage of practice. Of course, this refers to the theory that has application value and even must be applied in practice. Some theories really have no application value, and even are only theories in name, which cannot be regarded as theories in essence.

3 The Roles of Theory and Practice

Based on the previous facts review, we are sure that practice ahead of theory, and it can be even hundreds or thousands of years ahead. Of course, the process of history, including the process of theory and practice, is the combination of contingency and inevitability. It is inevitable that some accidental events will occur. The examples like the theory of option pricing established 50 years ago, are still less applied, which cannot deny the conventional or basic fact that practice is ahead of theory.

Therefore, practice is ahead of theory and theory lags behind practice. This is a normal situation and no need to doubt or call in question; If the opposite happens, it is abnormal and worth mentioning and discussing. No matter which is in front in history, there is a question worth thinking and discussing: which "should" in front, theory or practice?

Human society is a society of division of labor. In a sense, the process of human social development is a process of more and more detailed division of labor, including the financial industry. In fact, the refinement of division of labor objectively requires the strengthening of cooperation. Nowadays, almost all kinds of products are the result of multi industry division of labor and cooperation. In all kinds of division of labor, the division of labor between theory and practice is a basic division of labor. It can be imagined that the initial emergence of this division of labor is the symbol of the improvement of human productivity and the emergence of surplus products. This division of labor has lasted for thousands of years, which shows that this division of labor is reasonable and beneficial.

Division of labor means that theory and practice take their own responsibilities and cooperate with each other to complete tasks. The formation of division of labor means that division of labor has higher efficiency than non division of labor, and can complete tasks more efficiently, or achieve common goals. How theory and practice should be divided and how their respective responsibilities are completed should be judged according to whether it is conducive to the realization of this common goal. What is this common goal? Of course, it is the progress of mankind or human society.

How is human society progressing? The historical process shows that human beings make progress by constantly setting foot in new fields and constantly overcoming or solving difficult problems. Among them, getting involved in new fields depends on practice. Practice finds new fields through various attempts, explorations, trial and error. Solving problems depends on science or theory.[5] Without theory, theoretical breakthrough and progress, human beings will repeat the same work and lifestyle for thousands or tens of thousands of years like other animals. Theoretical research makes human cognitive and action ability develop by leaps and bounds by solving difficult problems, which will promote human's ability to explore new fields.

In addition to trying in new fields, practice is repeating in old fields. Whether trying in new fields or repeating in old fields, practitioners have to face large or small decision-making problems. The difficulty of these decision-making problems varies. Most of them rely on the intuition, common sense and experience of decision-makers to make judgments and solutions; A few major problems are solved by team discussion, consultation and even through a certain research process. No matter the size of the problem, whether it is personal decision-making, organizational decision-making or decision-making through research, there is a certain time limit. In other words, a decision should be made within a certain time. It can be called "decision-making within limited time".

Because of the "limited time", it is conceivable that the problems that should be solved may not be solved within the limited time. However, for decision-making in practice, timeliness is often more important than correct results. Therefore, practitioners have to adopt the way of trial and error, and simple or even arbitrary decision-making, or blind decision-making is inevitable. For example, how should the bank loan interest rate be calculated? In practice, even if it is not clear, an interest rate must be determined before deadline, otherwise the business will not be carried out, the customers will not be able to maintain, and the bank will not survive and develop. These unclear or unsolved problems are left to the theoretical research of various professional disciplines as difficult problems.

Different from the "decision-making within limited time" in practice, theoretical research or scientific research aims at pursuing truth and solving problems, and there is no time limit. Tough problems such as the interest rate of bank loan, optimal debt ratio of a company, etc., can be left to scholars as theoretical problems. Theoretical research does not necessarily have time limit. If the problem cannot be solved in one year, the research can go on in the second year; if it cannot be solved in ten years, the research can go on in the following more years. If one generation of scientists cannot solve it, the subsequent generations of scientists can continue the same research, until the problem is finally solved.

Therefore, it can be understood why the study of monetary banking appeared so many years later than the practice of commercial banks, and some key problems in the operation of commercial banks, such as the determination of interest rate, have not been solved so far. It can also be understood why the option pricing problem was solved more than 2000 years after the emergence of options, and it is not completely solved so far. For example, for American option pricing, although there are several methods that can be applied, they are basically expedient measures in the absence

[5] Among them, the theory of natural science plays a leading role and the theory of social science plays an auxiliary role. In any case, social science is also science, and its progress is also the ladder of human social progress.

of appropriate methods. So, it is practice that should be in front. The problem of theory comes from practice. If practice does not walk in front, where do we find the problem or object for theoretical research?

Therefore, the reasonable order is: practice in front, theory behind; Practice provides difficult or theoretical problems, and theory tries to solve those problems. It may be normal for theory to lag behind practice for decades or hundreds of years. Theoretical problem solving takes time; Although the more quickly the problem solving, the better, how long it will take to solve the problem is related to the difficulty of the problem, not to people's will. In any case, there is no need to doubt about the situation that theory lags behind practice for some years or decades, unless the question is too simple, or the answer is easy to think about. Of course, if a problem is so simple, it may not be a theoretical problem; It is not worth pushing to theoretical research, and then returning to practical application. It can be "solved" directly by practice.

4 Difficulty of Theory

Since the significance of theory lies in solving practical problems, theoretical problem and solution (theory)"must" have difficulty, and the degree of difficulty of each theory will be different, because the degree of difficulty of relevant practical problems are different.

Therefore, the difficulty of theories varies among disciplines and problems. For example, in the financial discipline, the discussion of solving the optimal capital structure should be more difficult than the discussion of profit and loss analysis, because the former problem itself is more difficult. You can't expect the difficulty of each question to be the same. On the contrary, if it is found that the solution of optimal capital structure is similar to the difficulty of conventional profit and loss analysis, it indicates that the research may have problems in direction, method or hypothesis. Of course, for the discussion of the same problem, on the premise of solving the problem, the simpler the method or process, the better.

The increase of difficulty in research and solution often means that the difficulty of learning and application increases. On the other hand, there is no certain relationship between the difficulty and importance of the problem, that is, it is not sure whether they are positively or negatively correlated. In other words, in the coordinate system composed of difficulty and importance, such as Fig. 1, there will be problem distribution in all the four quadrants. It is conceivable that the problems solved by practitioners are mainly located in quadrants II and III; The problems solved by scholars are mainly located in quadrants I and IV; The theories that are easy to be adopted in practice are mainly located in the first quadrant; The theories that are not easy to be adopted in practice are mainly located in quadrant IV.

It should be noted that the importance in Fig. 1 is relative importance, and low importance does not mean unimportant. For the questions in quadrants III and IV, if the solution or decision is not right, it will also have a nonnegligible impact on income

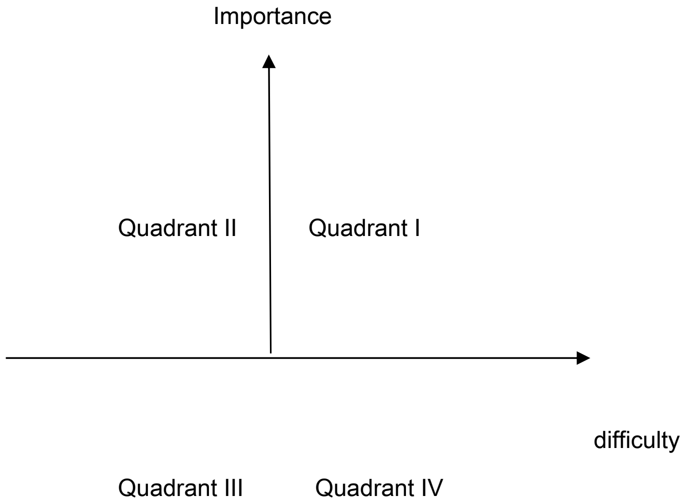


Fig. 1 Distribution of practical and theoretical problems

and value-added of the company. Looking at Fig. 1, it may be easy to understand why some difficult problems have been solved, but the solutions are not adopted in practice. For example, the option pricing theory or model in the figure may be located at the upper right of quadrant I, that is, it is both important and difficult. In the case of limited time, it is not easy to learn, master and apply, so it is delayed to use in practice. In any case, practice has the freedom to adopt or not, but the result of free choice is not necessarily the right choice.

Observing and thinking about Fig. 1 may also help to understand other related issues. For example, the theories that are worth learning and applying may be those that are difficult and important, especially those that are difficult. Because when others can't or don't have time to learn, your team or company will first learn and apply the corresponding theory, which will have a competitive advantage. Therefore, for learning, especially adult or on-the-job learning, we should not have the idea of fearing difficulties, or choose especially the simple and comprehensible courses and learning contents.

On the other hand, the problems worthy of theoretical or academic research should or must be quite difficult; Problems that are not difficult are not worth pushing to theoretical research, and then return to practical application. In a certain field and a certain period, a large number of problems can or should be "solved directly" in practice, and only a small number of problems need to be discussed and solved by academic research. The breakthrough and solution of these problems represent the progress of relevant theories; The application of those theories promotes the progress of practice further.

Of course, this does not mean that the more difficult the financial research is, the better. The significance of scientific existence is to solve problems for practice. Relying on scientific research to solve problems, practical problems should be reduced rather than increased; Or, the difficulty of the problem is degraded rather than upgraded. At the same time, the problem should come from the practice or be found in the practice, rather than assumed or created arbitrarily by the researchers.

5 Usefulness of Theory

Theory is the result of scientific research; The value of theory is determined by its final application value. Most real theories have application value. In reality, theories are not easy to use or have no application value in three cases. First, the theory itself is a pseudo theory, or it is not a real theory; Second, the user has not learned or fully mastered it, and cannot use it correctly; Third, some theories are semi-finished products before the final solution, which is helpful to derive the final solution, but they have no direct application value. The MM model about capital structure is a good example as a semi-finished product or solution.

Generally speaking, the so-called real theory, especially the theory with application value, should have the following characteristics.

First of all, the problem of theory comes from practice. Problems from practice directly represent the needs of practice. Of course, the theories or solutions obtained from research can be applied to practice. On the contrary, if the problem comes from the scholars' own imagination, it is likely to deviate from the needs of practice, and even if the correct solution is obtained, it may have no application value; even worse, if problems are allowed to be created by scholars rather than only found in practice, the problems may increase rather than decrease through the academic research, because scholars can create a large quantity of problems by imagination.

Not all the problems in practice are worth studying. Some problems are not very difficult, ordinary people or people in the industry may know or think of the solutions as soon as they want. Those problems are not worth studying as a theoretical problem. In other words, common sense or problems close to common sense are not worth theoretical research. Theoretical problems should be difficult to some extent, and not easy to solve in practice with time limit. Therefore, the theoretical solutions to such problems naturally have application values in practice.

Secondly, the theoretical problems must have certain commonalities, or exclude or remove the concreteness or particularity, so that the solutions obtained from the research can have wider applicability. Some theories are said to be outdated, or not suitable to the local conditions. But in fact, real theories often have applicability across time and space, and will not be easily outdated or only suitable for specific places. A "theory" easily outdated and ineffective is likely not a real theory, but may be a statistical conclusion based on a specific sample data, or a summary of experience at some specific time and place, which has not yet been raised up to the theoretical level.

Further, on the premise that the problem comes from practice, the theoretical solution must be correct.[6] Incorrect solutions are naturally of no value. In reality, there may be many solutions or theories to a problem, most of which may be incorrect. For the same scientific problem, there is often only one correct solution. Therefore, in learning and application, if there are multiple theoretical solutions, we must first distinguish which are correct solutions and which are incorrect solutions. For instance, to the problem of how to determine the discount rate, there maybe several solutions or methods. In this case, we need to make a judgement and select a correct method or the most correct method. Please note that the average of the right and wrong answers is wrong rather than right.

Finally, the theory that solves the problem has direct application value. The midway research conclusion that not finally solves the problem, may contribute to the follow-up theoretical research, but cannot have direct practical application value. A typical example is the MM model on capital structure. MM models I and II are pioneering achievements in the study of capital structure. However, due to the failure to quantify the bankruptcy cost, the cost and benefit of debt financing cannot be weighed quantitatively, and the optimal capital structure of the company cannot be obtained, so the conclusion cannot be applied.[7] In reality, there are occasional literatures discussing the application of MM model I or II, which only shows that the authors lack understanding of the theory.

6 Evaluation or Assessment

The goal of the company is to make money, or profit or value-added. According to the common understanding, relevant economic or management theories, such as financial theory, are to guide companies to make money or profits. Some people further reasoned according to this logic that experts and scholars who master relevant theories should be easier to make money. Therefore, in business school classes, students often question the teachers who give lectures. How can you teach me to make money if you don’t earn as much as I do? Blake encountered similar questions in his lectures.

The division of labor between theory (research) and practice is the basic social division of labor. Of course, there are great differences in their pursuit, even more than the differences between science and art. In the field of economy or management, it is understandable that the goal pursued by practice is profit; Of course, its success can and should be evaluated by the standard of profit; However, theoretical research, including theoretical researches in economics, management and finance, aims at

[6] This is the basic feature that distinguishes science from art. Strictly speaking, finance and financial theory belong to science, and their solutions may be right or wrong.

[7] The conclusion of MM model I is that the debt ratio has no effect on the company value, so it doesn’t matter how much capital financed by debt. MM model II considers the benefit of debt, that is, the tax shield, but fails to consider the cost, that is, the bankruptcy cost, so it comes to the conclusion that the more debt, the better.

solving problems and seeking truth. How can we evaluate its success according to the money made?

In fact, in terms of making profits or making money, it usually needs the support of a variety of factors, and the relevant economic, management and financial theories are only one of them. These elements may include technology, production, market, capital, contacts, information and theory. Companies or individuals in practice may possess all these elements, but a little deficient in theory; However, experts and scholars may be short in all other factors except a little theoretical expertise. Therefore, of course, people in practice usually make more money; while experts and scholars make less money.

There is another understanding, which seems more logical, that is, the economic theory can be evaluated based on how much money the author made relatively to other scholars. This is not true either. How much a scholar makes is also determined by many factors, and his theoretical level is only one of them. From the current practical considerations, when a theoretical problem has been solved, whether it can be published or not depends on the attitude of the review experts and the professional journal; whether it can be applied depends on the attitude of the practice field. If, for any reason, may be reasonable or unreasonable, the review experts, the professional journals and the practice fields do not understand, or not satisfied or do not recognize, the corresponding theories cannot be published and applied. Then the corresponding scholars will not earn personal income.

Sometimes the profitability may be determined by some accidental factors, and even the practitioners themselves do not know the reasons. For example, a TV station has launched a "stock speculation show by common person", each time inviting two or three "individual" investors who have made money in the stock market to introduce their experience and methods. On one occasion, two profitable investors showed on the same stage and introduced their own theories, methods and performance. Both the host and the audience were puzzled by their introduction, because the two investors' trading principles were almost opposite to each other, but they both succeeded in making money. The two investors were also unconvinced by each other. They even reported their teachers' names and tried to "convince" the other. Of course, they failed in the end, because both sides felt that their "theory" had been proved by practice, "making money is the last word".

In fact, there are many factors influencing whether to make money. In addition to decision-making theories and methods, there are many other important factors, such as size of capital, related resources, information, influence and unknown contingency. "Making money is the last word" may be used to evaluate the actual investment performance, but it cannot be the standard to evaluate the quality of decisional theories and methods.

In any case, in reality, money or wealth one owned is bound to constantly changing. If you have money today, you have the final say, and tomorrow he has money, he has the final say, then people's thinking and the real world are bound to be chaotic. On the contrary, the truth or truth contained in the theory remains unchanged. Therefore, logic or truth is the only correct standard which can be relied to judge who is wise and who is confused, or which theory is right and which is wrong.

It seems that the reason is not difficult to explain clearly, but it seems that how much a scholar makes money has increasingly become the standard for people to evaluate scholar's professional research and theory. This is, of course, the manifestation of some social prejudices in the evaluation of professional theories. Academic research aims at revealing truth and solving problems, which can only be evaluated according to the soundness in theory and the convenience and reliability in application

In any case, whether to make money should not be taken as the standard to evaluate a theory. Otherwise, the more in making money, the more qualified to become business school professors; and further, the richer, the more qualified to win the Nobel Prize in economics. Evaluation like this is obviously wrong and ridiculous.

7 Conclusions

Based on all the previous discussions, we get the following insights.

For the phenomenon of practice being ahead of theory, it is not worth surprising, because practice should go ahead of theory, which is also the normal situation in history. In history, "practices were often ahead of theories" hundreds or thousands of years.

In terms of reasonable division of labor, practice should be responsible for exploring new fields and providing problems to be studied for theory; Theory should solve problems for practice and promote the progress of practice at the same time. The "limited time" of practical decision-making and the "unlimited time" of theoretical research are their distinctive characteristics, and matched well with their roles in the social division of labor. Therefore, practice should be "ahead of theory"; theory and practice cooperate with each other to jointly promote the progress of human society. Based on these findings, we further discuss on the difficulty and usefulness of theory.

Further Discussion: On the Basic Classification of Social Science—Decisional Science and Descriptive Science

Contents

1 Foreword

Science is the ladder of human progress. With the development of human society, science has become a huge and complex system. In history, natural science emerges first, and then social science. At the same time, after the emergence of natural science and social science, they are constantly refining the internal division of labor system. Relatively speaking, social science appeared lately, and some basic problems have not been clearly understood. For example, the problem of basic classification or internal structure is ambiguous in social science.

As we all know, classification is one of the most basic and important ways to know and get familiar with a field. Basic classification hence is the first step for understanding and exploring a profession or a discipline. If the basic classification is not clear or wrong, it will lead to a series of errors. For example, different categories of science have different applicable research methods. If the basic classification is wrong, it is inevitable to use the wrong research methods; Further, it is possible to invest a lot of research time and resources and have little effect.

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_5

Objectively speaking, social science has also developed for a long time, although its history is shorter than that of natural science. There are many versions of the classification of social sciences. However, here we do not intend to review the classification of various versions, because most (if not all) the classification versions are about the segmentation in a specific field, and here we only intend to discuss the basic classification, not classification in a specific field. Strictly speaking, the basic classification of social sciences is seldom discussed so far.

In addition, we do not simply discuss classification, but try to understand better some issues based on the discussion. In recent years, there has been an embarrassing situation in some fields of social sciences, such as large input in research funds but stagnation in theory progress. Finance is a good example, although it seems very hot in academic and practical activities, in fact, there has been no substantial progress in the financial theory over decades, so has been the financial textbook. This problem has been extremely serious and urgent. In the long run, some fields of social science seem hard to avoid slipping into the quagmire of pseudoscience. Therefore, the basic classification of social sciences is an urgent problem involving the progress of social sciences. The discussion of classification can hopefully reveal the crux of the stagnation of scientific development.

The following contents are arranged as follows: the second part compares science and art to distinguish their characteristics and differences; the third part compares natural science and social science, and distinguishes the differences in research objects and methods between them; the fourth part discusses the basic classification of social science, that is, the division of decisional science and descriptive science in social science, including their characteristics and differences; the fifth part discusses the research methods of social science, and try to reveal some important reasons for the stagnation of some disciplines in social science. The sixth part summarizes the discussion.

At present, the theories of social science are numerous and miscellaneous, and contradictory theories and theories with loose logic are mixed with those good or correct theories. However, in fact, people's time and energy are limited. No matter learning or researching, it is not worth to waste time in wrong or bad theories. The basic classification is a very useful tool to help to judge whether a theory or a research method is right or wrong, good or bad.

2 Science and Art: Similarities and Differences

Science and art are two treasures of human society. To a large extent, the development of science and art has made human beings different from animals and become the spirit of all things. Today, with the development of human history, both science and art have achieved high achievements. A comprehensive summary and study of science and art is beyond the ability of any individual, including the author. Therefore, under this topic, we just discussed in the simplest and superficial way. Please forgive me for any deviation.

In the simplest and most direct sense, science answers or solves problems; art gives people feelings, including visual and auditory feelings. For instance, whether the earth is flat or round, whether the sun revolves around the earth or the earth revolves around the sun, are there any higher animals similar to human beings in and outside the solar system? Such questions need to be answered by science. Similarly, how to help people with machines to complete heavy physical and mental work, how to produce material products that meet people's needs at low cost and high efficiency, how to deal with traffic congestion, how to deal with air pollution, and how to maintain social order and world peace? Such problems also need to be solved by science, including natural science and social science.

As a high-level animal, human beings need not only material satisfaction, but also sensory and spiritual enjoyment. Therefore, with the development of human society, with more leisure time and more material resources, arts in various forms have been developed, including literature, painting, music, dance, etc. In modern times, ancient arts have developed more dazzling and diversified art varieties with the help of material means brought by science and technology, such as film, television, etc.

Common ground: innovation

What science and art have in common is that they both stress innovation and do not stick to one pattern. According to reason, every scientific paper should have new discoveries; every work of art should give people a new feeling. Sameness, plagiarism and imitation are taboos in the field of science and art, and represent the low level of the works. This is significantly different from various industries in society, such as manufacturing and service industry, which pay attention to standardization. Although manufacturing and service industry also emphasize innovation, it is necessary to emphasize the standardization and normalization of manufacturing and service processes in order to improve efficiency, reduce costs and ensure quality. However, scientific research cannot emphasize quantitative efficiency, nor can it improve its quality by emphasizing standardization. It is understandable that standardization or normalization are the opposite to innovation.

Therefore, for scientific research, it is not right to set a unified standards for the article structure and research methods, and turning scientific research into standardized product manufacturing, which will reduce the quality of research rather than improve it, and will stifle innovation rather than encourage innovation. The development of human society depends on scientific research to solve tough issues in various fields. In this case, there must be creative ideas and innovative methods to solve the problems that cannot be solved before. If the ideas and methods as well as the pattern of articles are standardized to be the same, what major scientific discoveries and social progress can be expected?

In the field of science, apart from standardization, emphasizing data verification is another factor that restrict and threaten innovation. In nowadays academic research, the worship of data and statistics has reached an unprecedented level. Regardless accounting, finance, economics or management, the main process of a research must have data and statistics. A manuscript without statistical analysis (hypothesis test or regression analysis) based on data can hardly pass the peer review. Whether the logic or reasoning is rigorous is unimportant. As a matter of fact, throughout the history of scientific development, most (if not all) major scientific discoveries came from scientists' innovative and rigorous logical reasoning; data have proved to be the usual means for pseudoscience to encircle and suppress real science. The reason is easy to understand, because pseudoscience often appears first and accumulates large number of favorable supporting data over time; as an innovative or new theory, true science is bound to lack sufficient data accumulation.

One of the reasons why Copernicus' heliocentric theory was attacked and criticized was that the geocentric theory was consistent with the observed data at that time; the heliocentric theory could not provide enough data to prove itself. Similarly, Einstein's theory of relativity has been attacked and criticized by peer scholars, for the reasons of no evidence of observational data. However, an obvious question is, should scientific discovery wait for the emergence of data, or should it lag behind the data? Einstein deduced the gravitational wave based on the theory of relativity in 1916; and the relevant data of gravitational wave was captured a little bit 100 years later by scientists using huge and expensive astronomical monitoring equipment. So, should Einstein wait 100 years and announce or publish his findings of gravitational wave until 2016? Likewise, should Copernicus wait for 1000 years to announce his heliocentric theory? How ridiculous it is!

How simple and deficient Einstein's research conditions were 100 years ago; this shows how hard and great it was to deduce the gravitational wave at that time. But the result was that he could not provide data to prove it, and then he could only be blamed or laughed at. Of course, this is a common thing in the history of science. But the logic behind those historic events seems to be: once anyone contributes scientific discoveries, he must come up with data, otherwise he should be blamed and mocked; On the contrary, because others have no discoveries, they do not need to provide data, and they are not criticized and mocked. Is this custom helpful or harmful to the development of science?

However, the most basic truth is that data represents phenomena, and in many cases represents the range of people's sight and hearing; how can science be limited by people's sight and hearing when it wants to reveal the essence of various phenomena? It is conceivable that if we know the world by sight and hearing, human beings are far inferior to many animals and can never become the spirit of all things. In fact, data processing lacking professional depth cannot belong to scientific research, and it is easy to draw wrong conclusions. For example, simply observing that the sun rises in the east and sets in the west every day, it seems that the sun revolves around the earth. Observing day after day, year after year from more points all over the world, you can collect a "big data" about this topic. According to the "big data", you can draw a conclusion that "the sun revolves around the earth". This is similar to the current research ideas and standards in many social sciences. If we carry out scientific research according to such ideas and standards, human beings will surely return to the era of geocentric theory.

Strictly speaking, few of the major scientific discoveries in history came from observation or data processing. The heliocentric theory is finally accepted by people because it got support from Newton's law of universal gravitation. Did Newton's law of universal gravitation come from observation? Of course not. People like to talk the story of the fall of apples. It seems that the law of gravitation has a great relationship with the fall of apples. But obviously, before Newton, countless people had seen the apple falling countless times, but no one had discovered the law of universal gravitation. The discovery of the law of universal gravitation is the result of Newton's unique and effective logical reasoning! The greatest scientific discovery since Newton was Einstein's theory of relativity. Did the theory of relativity come from observation? Of course not. Did Einstein use big data processing or regression to get it? Of course not!

It is the same in the area of finance. None of the major scientific discoveries in the financial field, such as Markowitz's portfolio theory and Sharpe's CAPM, comes from observations or statistical data on the stock market. Therefore, with the development of science and society, we can no longer ignore or even belittle the power of scientists' logical reasoning and the role of logical reasoning in promoting scientific progress. Copernicus' heliocentric theory has not been recognized for a long time. This is the tragedy of science and the tragedy of the whole human race. Einstein's theory of relativity was published by chance, and he could not win the Nobel Prize relying on the theory of relativity. This is a disgrace to the Nobel Prize and the whole mankind! From one perspective, such tragedies and humiliations becoming less and less (rather than the opposite) is a manifestation of the progress of human society!

Another factor that hinders innovation is the adoption of decision-making or judgment rules in which the minority is subordinate to the majority. In industrial and commercial companies, investment and management decisions may be made by voting, and ultimately the minority is subordinate to the majority. Purchasing decisions in consumer fields can be made based on the sales volume of similar products in the market, which is equivalent to the minority subordinate to the majority. But in science or scientific research, theories or papers cannot be evaluated or selected in this way. Any new theory, especially a new theory with major breakthroughs, is bound to be accepted by the minority in a considerable long time because it challenges people's accustomed authority and traditional cognition and is often not understood and recognized by people. Similarly, the quality of a scientific research paper cannot be evaluated by the way of citation rate and other voting-like methods; In particular, it cannot be evaluated according to the citation rate in a short time. It is obvious that the papers with high citation rate in a short time must be simple and easy to understand, which hard pose a challenge to the authority or tradition, and may even be something like the literature review of previous research.

Differences: logical thinking and divergent thinking

Many people like both science and art. For example, Einstein, the greatest scientist since the twentieth century, could playing the violin very well. Many of the things we consume and enjoy are also the combination or synthesis of science and art, such as houses, cars, instrument, apparel and so on. However, these are not enough to deny that science and art are two categories of knowledge, and there are great differences between them.

As mentioned above, science focuses on answering and solving problem; art, on the other hand, focuses on giving people feelings, which is, of course, a major difference between science and art. In addition, there are at least two important differences between science and art.

1. Science stresses logic and logic reasoning, pursues to discover essence through appearance or illusion. Art stresses the trend and fashion, pursues to get creative inspiration by divergent thinking. Art conveys or passes on feelings with various visual and auditory forms. Such as film, magic, etc., which gives people specific feelings and enjoyment with illusion.

   With the power of extraordinary and rigorous logical reasoning, scientists can break through the limits of human vision and hearing, and find out what people can't find by conventional observation. For example, the discovery of many new stars is often based on quantitative derivation and then confirmed by observation. In this sense, as previously revealed, scientific research should not be limited by "seeing is believing".
2. People evaluate or assess a scientific research in terms of right or wrong; while they evaluate or assess a piece of art in terms of good or bad rather than right or wrong.

For example, Mo Yan won the Nobel Prize in literature, which perhaps implies that Mo Yan's novels are better than others; but does not mean that other people's novels are incorrect; Mo Yan's novels are also difficult to replace other novels. But the situation is different in scientific research. For example, there used to be a stock valuation model. If a new model is found at the same level of detail,[1] the valuation results of the old and new models are naturally different. In this regard, it is necessary to make clear which one is right or wrong, unless in some cases the old and new models do have their own advantages and disadvantages. Finally, the correct model replaces the wrong model, or the model with good effect replaces the model with poor effect. This is the routine of scientific development. Therefore, works of art can be evaluated according to personal audio-visual feelings, but the scientific research cannot be evaluated like this. Neither social science nor natural science can be judged by the reader's own intuitive feelings or habitual knowledge without understanding.

Accordingly, scientific progress may be subject to greater resistance. The original theory may bring vested interests to some people. Overthrowing and replacing the old theory may affect their reputation, status and interests. Therefore, even if more correct and scientific theories appear, the process of replacing the old theories will not be plain sailing. It is common for the authorities of the old theories to suppress the new theories with the help of their superior position, that is, to suppress or obstruct the publication, dissemination and award evaluation. For example, Einstein's theory of relativity, although published by chance, and ultimately considered as the most important and prominent scientific discovery since the twentieth century, was thwarted and criticized by physics peers for at least 10–20 years after its publication,

[1] It should be emphasized here that on the premise of the same level of detail, because the evaluation models with different levels of detail should be regarded as different in function, that is, the models with different functions are naturally irreplaceable.

and finally did not win the Nobel Prize.[2] In the extreme case, the scientists with innovations might also be unprovoked charges and physical persecutions. For example, the process of replacing geocentric theory with heliocentric theory is a profoundly affecting example.[3]

Although they all emphasize innovation, scientific innovation is innovation under strict logic; the innovation of art is the innovation under divergent thinking. Science fiction as an art can be created without logic or data; But scientific breakthroughs must have rigorous logic. It is conceivable that since science and art are so different, the evaluation of science and art cannot be confused. For example, the evaluation of a scientific paper (or a Scientific Lecture) should be based on whether there is unique innovative discovery, whether the relevant professional problems are solved and whether the solutions are more reasonable and effective. In addition, it should pay attention to whether the concept is correct, whether the idea is clear and whether the logic reasoning is rigorous. On the other hand, we should not judge a scientific research by whether the language is gorgeous, wonderful and whether the writing conforms to a fashion.

Relatively speaking, the differences between scientific conclusions are often right and wrong, which cannot be reconcilable, because for the same problem, there is usually one correct answer; however, there are only differences in quality and even characteristics between artistic works. An infinite number of works can be created under the same theme or even the same topic, but they are not mutually exclusive and can coexist harmoniously. Therefore, if there is competition among works of art, it is not worth mentioning compared with the degree of competition among scientific works. Because of this, scientists in history were often excluded, suppressed, beaten, persecuted and even killed because of their scientific discoveries. However, there are far fewer cases in which artists' lives are harmed because their works are too good.

If there is not enough punishment for doing bad things, the cost of doing bad things is too low, and doing bad things will be encouraged and intensified. This seems to be a well-known truth. Therefore, people will formulate various laws and regulations to punish those who abuse their power for personal gain. Even bad people and bad deeds will be punished in terms of social and cultural customs. For example, Qin

---

[2] Einstein won the Nobel Prize in physics in 1921, but when the Nobel Prize Committee announced the prize, it stressed that this was not because Einstein discovered the theory of relativity, but because of his theory of photoelectric effect. It can be said that Einstein could not win the Nobel Prize by relying on the theory of relativity alone, because this scientific discovery was too significant, too innovative and too subversive to be accepted by the authorities of his time.

[3] After decades of astronomical observation and research, Copernicus (1473 ~ 1543) speculated that the earth was not the center of the universe, wrote the theory of celestial bodies' motion, and put forward the "theory of Solar Center". But for fear of being punished by the Church (equivalent to the school representing geocentrism), Copernicus did not dare to make his discovery public until 1543, when he was dying. In more than half a century after the publication of the theory of celestial bodies, geocentric theory still occupied an absolute dominant position, and there were few supporters of Heliocentric theory. Scientists who believed in and spread heliocentric theory were also excluded and persecuted in various ways.

Hui[4] in Chinese history has been reviled from generation to generation. But it is strange that the cost of doing bad things in the field of science is very low, because there is almost no relevant punishment. Therefore, major scientific discoveries in the history of scientific development have been repeatedly blocked and besieged; Scientists who have made breakthroughs in research have been repeatedly excluded and persecuted.

But so far, those "unfair" incidents have not been formally and properly condemned, the "injurer" of these incidents have not been properly punished, and scientists who have been persecuted have not been properly "compensated or made up". In any case, past experience, if not forgotten, is a guide for the future. When we enjoy the material and spiritual wealth brought about by scientific progress, we should not forget the scientists who made contributions at the expense of personal interests, and the hard earned scientific discoveries, especially those did not win awards. For example, Black, who has made outstanding contributions to option pricing, and Treyno, who has made outstanding contributions to capital asset pricing.[5] After all, scientific progress is the ladder of human progress!

3 Natural Science and Social Science: Differences in Objects and Methods

Both social science and natural science belong to science. As science, they have common characteristics, including transcending common sense, revealing the essence of things, answering or solving problems, and so on. On the other hand, scientific research also pays attention to revealing the essence of things in the simplest and effective way to better answer or solve problems. In the same effective case in solving problems, the simpler the method, the better.

Just as science and art are often combined, the solution of many problems in reality requires the combined application of natural science and social science, such as the design and construction of roads, the treatment of traffic congestion and air pollution, the design and operation of production lines, and the site selection, investment and operation of companies, banks, schools and hospitals. Nevertheless, there are significant differences between social science and natural science. The most obvious difference is the difference in research objects. In addition, there are obvious differences in the purpose of the research, the accuracy and stability of the conclusions.

---

[4] Qin Hui (1090–1155) was a famous prime minister in the early Southern Song Dynasty. According to historical records, Qin Hui betrayed the country for prosperity, mutilated loyal and good people, and executed patriotic general Yue Fei for "unnecessary" charges. For a long time, Qin Hui has been regarded as a traitor by the world. However, the evaluation of Qin Hui in history is also controversial.

[5] Option pricing is the most important financial breakthrough in history. The Black Scholes model of option pricing has been perfectly confirmed by the binomial tree model in 1979. It is obviously too late to issue the Nobel Prize in 1997. Treyno's CAPM manuscript could not be published after several rounds. Unfortunately, these regrets in the development of science have not attracted enough attention, let alone relevant improvements.

In the aspect of research object, social science takes people with social attributes or human behavior as the research object, and social science, including economic management, studies and solves social problems; Natural science studies substances from celestial bodies to basic particles. Natural science, including science, engineering, agriculture and medicine, studies and solves natural problems, including human problems of natural attributes, such as the treatment of various diseases. The contact between man and nature is inborn. Human beings live in the natural environment and obtain all the things necessary for production and life from the nature. Therefore, natural science appeared earlier and developed more fully. The social attributes of human beings are constantly enriched with the increase of human population and social interaction; Social science is produced and developed with the improvement of human productivity and the evolution of production relations, that is, the continuous enrichment of human social attributes.

Related to the differences of research objects, social sciences and natural sciences are obviously different in the accuracy and stability of research conclusions. All kinds of substances from elementary particles to celestial bodies are completely affected by natural forces, and their movements have accurate and stable trajectories or laws. However, human behavior is dominated by human brain and often acts according to intuition or experience. There are ideas “on a sudden inspiration” and “on the spur of the moment” and other contingencies, resulting in great imprecision and instability of human behavior and its results. At the same time, human behavior also has the characteristics of learning and improvement, which further strengthens its instability. Therefore, natural science research often requires and can obtain accurate conclusions, especially accurate quantitative conclusions, which are often stable for a long time; however, social science research can only get qualitative conclusions, or inaccurate and unstable quantitative conclusions. The accuracy and stability of natural science research conclusions are repeatable in tests or observations; however, the research conclusions of social sciences often vary across time, place and conditions, and can hardly repeat in different cases, which shows that the conclusions have the characteristics of the times and places; and may be “not applicable” due to the change of location, or “obsolete” due to the change of time.

For example, according to Archimedes’ law discovered in ancient Greece,[6] an object immersed in a liquid (or gas) is subject to upward buoyancy, and the magnitude of buoyancy is equal to the weight of the object to displace the liquid. Such a conclusion will not be invalidated because decades, hundreds or thousands of years have passed, nor will it be invalidated in the East because it was discovered in the West.

From another perspective, the reason behind the above differences also lies in the good quality of data (observation data and experimental data) in natural science; however, the quality of data (often observational data) in social sciences is not so

[6] Archimedes was an ancient Greek philosopher, mathematician, physicist and founder of static mechanics and hydrostatics from 287 to 212 BC. Archimedes’ methodology contains the advanced research of "infinity" in mathematics and contains the thinking of calculus. He used the "approximation method" to calculate the sphere area, sphere volume, parabola and ellipse area. Archimedes, Gauss and Newton are listed as the three largest mathematicians in the world.

good. Accordingly, the research results of natural science may be verified by the fact that the theory conforms to the actual data, because within the range of individual life span or the life span of several generations and even the whole human life span, the actual data at one time and one place are likely to represent the actual data at all times and all places. However, even if the research results of social sciences conform to a set of actual data, they cannot prove their correctness, because there are far more than thousands sets of such data, especially when people's behavior is affected by a variety of accidental and inevitable factors in changes, and there is a lack of consistency or stability between the data in different time and places.[7] Being verified here and now does not mean that the conclusion will also be verified at that time and that place.

The reason why the data quality in social sciences is not good is that the data contains complex factors that people recognize but cannot consider or do not recognize. For example, will higher interest rates lead to higher or lower stock prices? Referring to the actual data of China in recent decades, we can find that in over half cases, the stock price rises after the interest rate rises, but the stock price drops after the interest rate rises nearly half of the time. Then, based on such data, can we conclude that the interest rate and stock price are positively related? Or they are negatively related? Obviously, neither of these two conclusions is tenable. Because there are many factors that affect the stock price besides interest rate, sometimes more and sometimes less. Sometimes this factor is important and sometimes that factor is important. Based on the actual stock price fluctuation data, it is impossible to understand and exclude the influence of other factors, so it cannot reveal the certain relationship between interest rate and stock price this way. In other words, this research method cannot answer this question at all.

Therefore, in the social sciences, research based on actual data, even in the case of a large amount of data, can usually obtain only temporary and effective conclusions applicable to one place at a time. It needs rigorous logic reasoning to get a stable or certain conclusion. Rigorous logic reasoning is the only way for social science to get a certain conclusion. Take the above problem as an example, based on logic reasoning, other things being equal, stock price is negatively related to interest rate. In contrast, natural science has two research approaches to get certain conclusions. It can draw conclusions through logic reasoning; when logical reasoning is difficult or impossible, it can also draw conclusions based on experimental or observed data. This means that learning and researching social science need stronger logical reasoning ability than learning and researching natural science, because logical reasoning is the main research method in social science. This inference is contrary to the general understanding.

Note that this does not mean that data is dispensable in the social sciences. Generally speaking, the so-called science or theory is actually a logical relationship beyond common sense. For example, Archimedes, the ancient Greek scientist mentioned

[7] At present, the so-called big data is just one of thousands of sets of data applicable to the research of a certain problem, and as the data of the past, it is still a question how long it will remain valid in the future.

above, also invented the lever principle, which is expressed by the formula: power $\times$ power arm $=$ resistance $\times$ resistance arm. This is a scientific theory. To use this theory to draw a conclusion, we also need the value of the basic variable. For example, if a person has only 30 kg of force and wants to pull up a 60 kg weight object with a lever, the length of the power arm should be twice that of the resistance arm according to the lever principle. No conclusion can be drawn without the data of 30 and 60 kg; if the two data are not correct, we cannot draw a correct conclusion either.

In social sciences, the theory of quantitative disciplines such as finance is expressed as formulas or models. Just as illustrated by the above example, models and data support each other to draw conclusions, and both are indispensable. The process of collecting and processing data to obtain (independent) variable values may require statistics or computer assistance, but this does not mean that formulas or models also depend on statistics or computer. If it is necessary for any professional discipline to exist independently, its theory and model must be based on professional logical reasoning. It is conceivable that if the theory and model of a major and the value of variables in the model are completed by statistics, this discipline will not be an independent discipline, but can be regarded as a branch of statistics at most.

Of course, it is not easy to draw a theory or model through (professional) logical reasoning. Some problems may not be solved by a logic-based theory or model for a long time. If you think that a temporarily conclusion valid for specific place and time is better than no conclusion, you can obtain conclusions based on a set of data. But do not forget that in the social science research, the data quality is not high and the conclusion is not reliable. On the other hand, if you can draw conclusions based on logic, there is no need to draw unreliable conclusions based on data. In other words, the method of using statistics to obtain models should be the second choice in professional disciplines, which is similar in natural science.

In addition to the above differences in research objects, research conclusions, data quality and research methods, social science and natural science are also different in their research purposes. The ultimate purpose of social science research is not only to reveal the operation law of the behavior subject in society, but also to guide the operation of the behavior subject, or to support the decision-making of individuals or organizations. However, the ultimate purpose of natural science research is only to reveal the operation laws of the behavioral subjects, including the operation of large celestial bodies and small basic particles, but it is neither responsible nor possible to guide the operation of these subjects. This is one of the reasons why social science research needs more logical reasoning, because correct or reasonable actions cannot be directly counted from the existing data. The next section will further explore the characteristics of social science in this regard.

4 Basic Classification of Social Sciences: Descriptive Science and Decisional Science

As mentioned above, due to different research objects, social science and natural science are much different in research purposes. The significance of social science research lies in guiding the decision-making and operation of relevant parties in society; however, the purpose of natural science is mainly to reveal the mysteries or operating laws of nature, rather than to guide the operation of behavior subjects in nature.

From a certain point of view, social science actually studies the interaction between behavior subjects (individuals, families, groups, organizations, industries, regions, countries, etc.) and the external world (natural environment, human environment, other subjects, etc.). The humanistic environment here mainly refers to the unnatural environment related to human activities, which roughly includes social structure, institutional system, laws and regulations, administrative intervention, social customs, value orientation, moral standards, etc.

Therefore, social science needs to answer or solve two kinds of questions:

The first is: what is the world (or research object) like?
The second is: what should we do?

On the first kind of problems, the research of social science is similar to natural science, which is to describe the relevant world or object. That is, natural science describes the matter movement of the material world, while social science describes the behavior of individuals or groups in human society. The matter movement in nature is usually stable or constant, while the behavior of individuals or organizations in human society has great uncertainty. Therefore, natural science describes the laws of matter movement in nature with long lasting and more accurate conclusions, but social science describes the laws of individual or organizational behavior with tentative and more rough conclusions.

The discipline responsible to solve the first kind of problems can be called as descriptive science. The discipline responsible to solve the second kind of problems is supposed to tell the behavior subject what to do, which can be called as decisional science. This is a unique research category of social science, which is significantly different from natural science, because natural science cannot tell the celestial bodies or particles what to do, and do not have such a category. The ultimate purpose of decisional science research is to guide the decision or action of the relevant behavior subject.

Therefore, as the most basic and simplest classification, social science is divided into two categories: descriptive science and decisional science. Descriptive science mainly discusses the description of the past, present and future; a description of the future situation is commonly referred to as a forecast. History, news, forecasting, accounting, statistics, etc. are all intended to describe a certain social object, so they all belong to descriptive science. The decisional science mainly discusses what or how to do when facing a choice. It is usually divided based on the types of problems, such as the qualitative problems and the quantitative problems, the problems of individuals, companies and countries, and the problems of investment, financing and operation in companies. Economics, finance, management, etc. are all meant to discuss what to do about a certain kind of problem. Therefore, they all belong to decisional science. Decisional science is a unique part of social science that natural science does not have. Moreover, in social science, describing the world or object is often not the end, and the ultimate purpose is to guide people's behavior. Therefore, the decisional science is at least as important as the descriptive science.

4.1 The Difference Between Descriptive Science and Decisional Science

It is often said that natural science studies laws in nature and social science studies laws in society. This statement is actually bias or unfair. It can be said that in social science, descriptive science is to research laws in society; but the decisional science is not. The decisional science tries to find the truth or principle of decision making. Here, the law represents the connection between phenomena, including the law in the short term and in the long term. Descriptive science obtains such kind of laws by summarizing a lot of phenomena. And behind the truth or principle is logic. Decisional science obtains the principle of what to do through logical reasoning. If social science needs logic more, then decisional science is especially so.

Therefore, social science needs to answer the question of "what or how should we do". Without answers to such questions, social science research is incomplete. In reality, there is a view that objective conclusions can be drawn from data-based descriptive research; the research of "what or how should we do" is to get the subjective conclusion based on the subjective value standard; Moreover, the "objective" and "subjective" here have the meaning of commendation and derogation respectively. For a long time, misled by similar views, the research of decisional science has been congenitally deficient because of being despised. There are also a series of related misunderstandings in the research of the whole social science, such as the more data, the better, the conclusion obtained by clear and strict logical reasoning is not tenable, and the corresponding papers do not conform to the research norms and cannot be published, etc. It is conceivable that these understandings have affected and even hindered the development of decisional disciplines as well as social sciences as a whole.

Take economics and finance as an example, without knowing whether they belong to descriptive science or decisional science, the relevant research tries to get "objective" conclusions based on actual data, that is, to pursue reality. Unfortunately, data is the result of past decisions. The real economic and financial problems need to be decided based on the truth of cost and benefit analysis, rather than the previous decisions. If decisions of economic and financial problems are made based on past decisions, how to understand the role of economics and finance? Of course, people may realize that such research is of little significance. As a result, most economic and financial studies turn to interpreting the data now. However, the actual decision-making situation is complex, which is the result of the disorderly combination of various accidental and inevitable factors, and even some undisclosed and unknown factors among them. What valuable or useful discoveries can be expected by explaining the results without clear reasons?

Therefore, the division of Decisional science and descriptive science is of fundamental significance to social science. The research purpose or pursuit (at least) must conform to the discipline category, and the input of research resources thus can be reaped. Otherwise, if we don't understand or make a mistake in the subject category, we may get half the result with twice the effort, or even the research may be not helpful but harmful to the theory and practice. Knowing the scientific attribute of decision-making in economics and finance, we can make it clear that the direction of our efforts is to research "what or how should we do", that is, how to solve economic or financial problems, rather than explain the decision-making results that have been completed, we can and should righteously research and answer the "what or how should" problems.

Many decision-making problems in economics and finance have not been solved, and theories are needed to answer those "what or how should we do" problems. For example, what is the reasonable or optimal debt ratio of the company? Is it too much or too less debt used at present, so the next financing "should" mainly use equity financing or debt financing? What is the "should" value of a bond or stock? Is the current market price overvalued or undervalued, so "should" we buy or sell the corresponding bond or stock? What should be the discount rates of the company's equity capital, debt capital and total capital, so what is the discount rates at which to evaluate the investment projects? and whether the capital costs of the company's equity and debt financing are high or low respectively, so the company should choose equity or debt financing? How much should the risk cost of a debt or loan guarantee or insurance, is it higher or lower than the current industry conventions? Based on the customer risk, what is the "should" interest rate of a loan, whether it meets the acceptable scope of the bank, whether or how large the loan can be approved? For a long time, the research of economics and finance has been busy describing and explaining the phenomena of corresponding problems, that is, the past decision-making results. As a result, the above problems have been no positive and convincing answers or solutions. However, if economics and finance do not answer or solve these problems, what other disciplines are expected to answer these problems?

In fact, decisional science provides decision standards, including cost and benefit standard and optimal standard. Only based on these standards can we answer the problem of "what or how should we do". Obviously, these standards cannot be "mined" from the existing data or phenomena, because the data or phenomena can only represent what did in the past. Based on the data, we can make statistics on the decision-making results, including individual case results and average results. However, neither individual case results nor average results may represent correct or optimal decisions, because actual decisions cannot be completely correct within limited time and theoretical solutions. For example, the stock price data is the result of the decision-making of both the buyer and the seller, behind which are the different valuation models or equations and judgments on the future of both parties. Obviously, it is impossible for both parties to make correct judgments. And even if a decision represents a correct or optimal decision, it can only represent the correct or optimal decision under the past conditions; It is absolutely impossible to represent the correctness or optimality of a future decision, because it is impossible to keep all aspects of the conditions unchanged until a future decision. Therefore, data processing or empirical research cannot solve the problems of decisional science, nor should it become the mainstream research method of decisional science.

It is understandable that the above “should” problems are the problems that must be answered or solved in reality; even there is no corresponding social science to research, we need to answer these problems in practice too. Without professional guidance of decisional sciences, these decisions have to be made in more subjective way. The decisional sciences can provide objective basis and correct standards for these decisions through professional research. In another word, the role of decisional science is just to reduce the subjective randomness and increase the objective rationality in decision-making, that is, to improve the scientificity of decision-making. We cannot to mistakenly think that the decisional science is subjective, because there is “should” in the decision-making problem to be studied, or even indiscriminately exclude the research on “should” problems. Such understandings hinder the development of decisional science and will just increase the subjectivity of decision-making.

If the conclusion of the descriptive science research is “objective”, the conclusion of decisional science research is also “objective”. The difference between the two conclusions is not one is objective, another is subjective; but one pursues to conform to the reality, so it does not necessarily represent the correct decision, and the other pursues to be correct, so it does not necessarily conform to the actual decision. For example, stock prices and their changes are the objects of the descriptive research, which are objective, but they are likely to be overestimated or underestimated, which are not correct; however, as the object of decisional research, stock valuation pursues objective or correct value results that are not subject to subjective influence. Although the goal may not be achieved due to the limitations of theory, calculation and relevant information, and the valuation results may not be correct, comparing with the stock price formed under the interactions of various factors and forces in the stock market, it will be more rational and closer to the stock value. Taking the stock value (result of decisional research) as the standard, the stock price is often either overvalued or undervalued. In other words, the stock value derived from the decisional research is an objective standard, which can be used to measure whether the actual decision-making results are correct. In addition, if more investors believe in the same value standard, it will help to promote the buy low and sell high based on the standard, thus helping to reduce the subjective randomness and speculation in the stock market decision-making, thus contributing to the stability of the stock market.

4.2 The Relationship Between Descriptive Science and Decisional Science

The description of the world or related objects is necessary, but the description itself is not an end. The purpose of social science research is making a better decision, so the decision is the ultimate goal. This reflects the relationship between descriptive science and decisional science, that is, descriptive science focuses on knowing the world or research object, and provides basic information needed for decision-making, while Decisional science provides decision methods or tools. The two sciences cooperate with each other to jointly fulfill the responsibilities that social sciences should shoulder. It is necessary to emphasize that first, both descriptive science and decisional science are indispensable; Second, the two are a relationship of division of labor and cooperation, not a relationship of mutual substitution.

The data quality of social sciences was discussed earlier. Of course, improving data quality depends on descriptive science. For example, if the research object is a company, there may be two main sources of data about the company, one is accounting reports, the other is news reports. Therefore, accountants, journalists and their corresponding professional theories should research how to provide and express information to better reflect the actual situation of the company. Similarly, if the research object is the population status of a city, it may need the application of statistical theory, and it is necessary to explore how to obtain the most accurate population data with the minimum research workload. Apparently, the problem of improving data quality, including data work efficiency, is the responsibilities of descriptive science. Economics, finance, management, etc. do not focus on this topic, nor are they responsible for solving this kind of problems.

A further question is what role the descriptive science should play in social science. Based on the previous discussion, this is easy to understand. the descriptive science is responsible to collect and process of data to get the description of the research object. This description becomes the basis for decisional science. Specifically, decisional science establishes methods or models, and the descriptive science provides the data needed by the decision methods or models. By inputting the correct or reliable data into the correct method or model, we can get the correct decision conclusion. Therefore, to get the correct decision conclusion, both the correct data and the correct method or model are indispensable. The descriptive science should focus on providing the data needed for decision, rather than drawing decision models or final conclusions. The decisional science is responsible to solve the decisional problems and provide decision models.

For financial science, the decision methods it provides are various financial models. For example, net present value is a kind of financial decision model. The basic data required by the decision model or method depends on the descriptive science, mainly the prediction science, such as the annual return and risk data required in the calculation of net present value. How to estimate the basic data more accurately and effectively is the responsibility of forecasting science; how to get the correct financial decision conclusions more quickly and effectively based on the forecast information is the responsibility of financial science.

Please note that forecast or prediction is also future oriented, but it is different from decision-making. Prediction is a description of the future and belongs to descriptive science. Similar to other descriptive sciences, prediction also emphasizes that the conclusion should conform to the reality, only that to the future reality; the more the forecast results are in line with the actual situation in the future, the better the forecast will be. Different from prediction, decisional science is not to conform to reality, but to establish reasonable or optimal decision standards, so as to guide actual decision-making and achieve the results that are most conducive to decision-makers. For example, investment and financing decisions are based on value, so valuation is emphasized. The value calculation or valuation pursues correctness, which is not to conform to the current price or the future price of an asset, but to draw the standard for judging the price bias.

On the other hand, correct prediction is conducive to correct decision-making, but prediction cannot replace decision-making. Correct prediction does not necessarily mean correct decision. Consider a stock selection decision, the current prices of A and B are 20 yuan and 60 yuan respectively. The normalized (excluding the impact of accidental factors) earnings per share during this year (the past 12 months) are 1 yuan and 2 yuan respectively. Assuming the average annual growth rate of earnings per share in the foreseeable 15 years (the situation after 15 years is completely unknown) is predicted as 10% and 20% respectively. How should I make a decision, that is, should I choose A or B? Obviously, even if the average annual growth rate of 10% and 20% is predicted correctly or accurately, it cannot replace the decision of choosing A or B.

Prediction is a description of the future and belongs to descriptive science. If prediction science is regarded as a separate category and the description of the past and present is called descriptive science, then description is the basis for prediction and prediction is the basis for decision-making. Prediction is the bridge between description and decision. As shown in Table 1. Prediction is very important for decision-making because decisions are always future oriented or based on future considerations. The data required in most decision methods or models are actually forecast data. Therefore, whether the decision is correct or not largely depends on the correctness of the prediction. However, it is often difficult to predict, because the future may not be necessarily related to the past. For example, a person who learnt well in primary school might not learnt well in secondary school; a successful startups might not become bigger and stronger in the future. Therefore, in the long history, most of the prediction methods are metaphysics, such as divination, astrology, physiognomy, Zhouyi and so on.

Just as decisional science is neglected, prediction, as an important branch of social science, is also neglected in long history. In many fields at present, prediction often relies on the simple judgment of decision makers based on experience, in which there are few scientific elements beyond common sense. Although most people don't know and can't judge the authenticity of divination, physiognomy, geomantic omen and Zhouyi, they are still the prediction theories and methods that many people rely

Table 1 Descriptive science, predictive science and decisional science

Descriptive science	Predictive science	Decisional science
History	Divination	Economics
Statistics	Astrology	Finance
Accounting	Physiognomy	Operation
Journalism	Zhouyi	Marketing
Mathematics[a]	Mathematics[a]	Mathematics[a]
…	…	…

[a] Mathematics can be used to help to describe past, present and to predict future as well as to make decision. The mathematics branches used in descriptive science are mainly probability theory and mathematical statistics, etc. The mathematics branches used in decisional science are mainly calculus and mathematical programming, etc.

on when making vital decisions. There is no modern prediction science to replace these methods. These theories and methods are the wisdom of ancient sages, which may indicate that people paid more attention to prediction in ancient times than today. So far, it can be understood that descriptive science, prediction science and decisional science are equally important. However, in the current social science research, due to the excessive emphasis on using data to speak, the research on the description of the past has been excessively expanded, and the research on the prediction of the future and the current decision-making has been ignored or underestimated, which is a serious deviation and distortion in the current social science research.

This kind of deviation and distortion has resulted in the lack of research on prediction and decisional science as well as the confusion of decisional science with descriptive science, resulting in large number of misunderstandings and puzzles in theory and practice, and the lack of prediction and decision theories and methods. The prediction is, after all, a description of the future. Some features of the future prediction are similar to the past description. However, in terms of basic characteristics, decisional science is quite different from the descriptive science. The following is a discussion on the decisional science.

In the simplest sense, decision and selection are synonymous, that is, decision is to select the feasible or optimal scheme from the alternatives. Decisional science should improve and guide actual decision, and answer how to make a reasonable or best decision. Therefore, it is necessary to establish reasonable and optimal criteria for decision. For example, in the financial fields, the reasonable standard for investment decision is that the net present value is greater than zero, that is, the investment project is expected to bring value-added. The reasonable standard of financing decision is that debt can bring additional value appreciation. Decisional science does not research how people did in the past, but focus on what should be done now to be reasonable, effective or optimal.

Since it is necessary to improve and guide the actual decision, whether the research conclusion of decisional science is correct cannot be judged according to whether it conforms to the past or existing decision results. As a conclusion of decisional science, being in line with reality means that there is no practical guiding significance, because the actual decision itself has been in line with reality, and it is not meaningful to explain or affirm it again after scientific research. This is a distinction between decisional science and descriptive science. Furthermore, according to the reasonable and optimal criteria established by decisional science, we can test whether the past decisions and practices are reasonable, and we can also know how to improve them. Therefore, the actual decision can be judged right or wrong according to to the conclusion of decisional science, not on the contrary, that is, actual decision results (such as the actual stock price) cannot be a standard to test the correctness of the decision or valuation model.

The purpose and significance of decision is to seek advantages and avoid disadvantages and improve the situation in relevant aspects. The past and current conditions have been determined, and it is only the future conditions that can be used to improve or seek advantages and avoid disadvantages. Therefore, decision makers should make decisions with forward looking. Facing the future is an important feature of decision and decisional science. This means that the decisional science is to research in advance, rather than the statistics or evaluation afterwards. For example, when there is no obviously overestimated or underestimated in the current market, the stock with good future prospects should be selected or invested; On the contrary, if the future prospect of a stock becomes worse, it should be sold.

The contribution of decisional science is not the statistics or summary of past decisions, but the reasonable or optimal solution obtained by logical reasoning. The reasonable and optimal solution has strict logic support, will not be invalid due to the change of time and place. Some people may think that if the time, place and conditions change, shouldn't the decision conclusion change with them? The outcome of the decision should certainly change. It should be noted here that the decision conclusions obtained by applying the methods of decisional science are not the same as the decisional science. The decisional science are actually decision models or methods. When the time, place and conditions change, the basic data input the model should change; and the decision conclusion will naturally change based on the same decision model or method. It is not necessary to change the model or method to obtain a new decision conclusion. Of course, the existing methods or models may not be perfect, and may be deficiencies or have room for improvement in concept, logic and application feasibility. The decisional science will also make progress with the research breakthroughs and innovations.

The division of labor of descriptive science and decisional science determines they are different in responsibilities, so their research and evaluation should follow different standards. The descriptive science should be evaluated based on the extent of the conclusion conforms to the reality; the decisional science should be evaluated based on the criterions that the premises conform to reality, the process is sound in logic and the conclusion is correct. However, due to the lack of discussion on the basic classification of social science, the two categories of research have been confused for a long time, and both of them are evaluated as descriptive science. The features and principles of descriptive science have been imposed on decisional science for a long time, resulting in the wrong and unfair evaluation and screening of decisional science research. Even worse, various disciplines in social science now indiscriminately adopt statistical test and statistical regression instead of professional research. Many decisional sciences have not made substantial progress in theory and application over decades.[8] For example, the capital asset pricing model (CAPM), which is very important in finance, was supposed to solve the reasonable rate of return matching the asset risk, that is, the discount rate. However, it was confused with a descriptive model, and later it was misunderstood as a model to describe and explain the past asset return rate, which further led to the no substantive progress in this field over decades.

5 Research Methods of Social Sciences: Amateur and Professional Methods

As for the research methods of social sciences, the current academic field is divided into "popular" empirical research and normative research. The former is actually statistical or econometric method; the latter is also called theoretical research. In economics, as early as in the 1930s, there was a debate about whether empirical research or normative research should be used. For example, Lionel Robbins advocates that economics is limited to scientific explanation, which can neither answer value judgments nor deduce countermeasures. Samuelson clearly pointed out: "economics is essentially a science based on experience. It takes explaining the world around us as its primary goal, and then helps us design economic policies based on correct and reasonable economic principles to improve people's living standards." This means that value judgments and countermeasures are an integral part of economics. So far, it can be said that empirical research has undoubtedly become the overwhelming method of academic research. the obvious evidence is that many core magazines have been controlled and monopolized by empirical research, and there are almost no theoretical research papers.

However, theory should develop and society should progress. The debates and propositions of predecessors should serve as inspiration for future generations, not as a constraint. For both the development of theory and the progress of society, social science should not sit on the sidelines, but should actively promote and shoulder its responsibilities. If social science or any of its branches only play the role of post event explanation, then such a discipline is not necessary. In any case, the role that social science should play is to answer the two kinds of questions mentioned above. That is, descriptive science answers what the world is like; decisional science answers what we should do. For example, how to make decisions on the capital structure of a company and how to make decisions on bank loans, these questions need

[8] Note that computer application itself is not a progress in the relevant decisional science. On the contrary, to "solidify" the errors in the theory and understanding of some decisional science disciplines into computer software means that it is more difficult to correct the errors.

answers sound in theory; and social science cannot just do the post event statistics and explanations.

Practically speaking, the so-called empirical research, as a method, is a statistical method or econometric method. It is conceivable that one method cannot guarantee the world to solve all social problems. From another perspective, the so-called empirical research is to draw conclusions based on data, which is more suitable for describing the world or objects. Therefore, empirical research can be used by some descriptive sciences to draw corresponding descriptive conclusions. Even, statistical methods should not be used uniformly in all descriptive sciences. For example, when describing the financial status and profitability of a company, the reporting method in accounting is more standardized and effective. There is no need to use statistics instead of accounting to reflect the financial situation of a company.

Of course, the decision problem is even less likely to be solved by statistical methods. Under normal or specific conditions, decisional science research needs to provide methods or models for weighing risks and benefits in the future, rather than statistics or description of how people did in the past. For example, the problem of capital structure has not been solved in reality. Therefore, no company know what its optimal debt ratio is. In this case, can we calculate the optimal debt ratio of a company by counting the debt ratio of some similar companies? This is apparently not the way to solve the problem. In other words, empirical research cannot solve decision problems. In fact, one of the fundamental reasons for the formation of various disciplines is that these disciplines have adopted different methods; and their own methods are the most effective one for solving problems in their own fields.

As mentioned above, descriptive science describes the past, present and future, while decisional science needs to balance risks and benefits in the future. Of course, the methods between these two categories of science should not be the same. The methods are also different among different descriptive disciplines. For example, history mainly uses qualitative methods to describe historical events and historical figures in the past, but statistics uses the quantitative estimation method of extrapolating the population based on samples, while accounting is an accounting method that pays attention to cross checking relationships. Similarly, the methods of different disciplines in decisional science are also different. For example, microeconomics mainly adopts the marginal analysis method to compare cost and revenue, finance adopts the quantitative valuation method, that is, the method of quantitatively trading-off between risk and return, and many management disciplines adopt some other quantitative or qualitative analysis methods. In short, different methods, so to speak, are just a manifestation of different disciplines.

Therefore, it is not a problem what methods should be adopted by each discipline. From the day of its emergence or independence, the problem of what method a discipline should adopt has been solved. The problem now is that in current academic research, no matter what discipline, it is required to adopt empirical research methods. This is obviously to erase the differences between disciplines, including the differences in research purposes and research methods between disciplines. This is a big mistake. It is a big setback in the history of scientific development. At least, social science should have the division of labor between descriptive science and decisional science; With the progress of social science, the internal division of labor of descriptive science and decisional science should be more and more detailed, not on the contrary. It should not to unify all the different disciplines into statistical discipline. Unfortunately, this trend and situation have taken shape and become mandatory. No matter what field, whether descriptive science or decisional science, the empirical research, or statistical test or statistical regression became a must. Otherwise, it is impossible to pass the anonymous review or thesis defense. This practice is extremely wrong, which is an important reason why many decisional science theories have not made progress in recent decades.

From one point of view, the current academic research has gone to the extent that statistics methods determine all fields and problems. This is extreme method oriented or method determined research. Under such a prejudice, whether a problem can be studied in any discipline depends on whether there is relevant data to show the statistical methods. The research purpose is not to reveal the essence of something, nor to solve a problem, but to show this analytical method or process. Such a study is equivalent to students doing a statistical class exercise to show that they have learned or mastered this method. This implies that such a study may have little practical significance. This is also the reason why the top journals continue to have ironic paper titles in recent years. In fact, the research of any discipline should take solving the problems that the discipline should solve as its own responsibility. For the problems that need to be solved, we can use whatever method we should use, including various mathematical and statistical methods. But generally speaking, we should first consider the professional methods of the discipline, because the own method of a discipline is usually the most effective method to solve the problems in the discipline. Otherwise, there is no reason for the discipline to become an independent discipline. Obviously, for non statistical disciplines, empirical research methods, including statistical testing and regression analysis, are amateur or non professional methods, which should be considered after the professional method.

Some descriptive sciences may be probably to adopt the method of empirical research, but the difference between the major categories of disciplines determines that decision science cannot adopt the method of empirical research, but can only adopt the method of logical reasoning, because the purpose of decisional science is to draw correct conclusions. However, descriptive science, whether describing the past or the future, it includes correctness and error, contingency and inevitability. Processing of these data cannot draw necessarily correct conclusions, especially the correct decision conclusions at a certain time, place and condition in the future. For example, finance should solve the problems of stock valuation, capital budgeting, capital asset pricing, optimal capital structure and so on. These questions cannot be correctly answered through processing market data. The relationship between descriptive science and decisional science determines that empirical research and theoretical research are not the relationship between the better and the worse. There is no need to discuss which will replace which. Instead, we need to discuss how to better divide work and cooperate to jointly complete the mission of social science, so as to completely tell people what the world is like and what we should do.

Data in social science includes sample data, large sample data and big data, but in fact, to borrow the concept of data, logic represents total or full correct data. In terms of universality, the conclusions drawn by logical reasoning are much better than those of any large sample and big data statistics. Because of this, the really convincing proof is not the data proof, but the same conclusion drawn from different reasoning processes, that is, the two full data processes confirm each other, which can basically represent a certain conclusion out of doubt.

6 Conclusion

To sum up, social science falls into two basic categories: descriptive science and decisional science. People have long been used to understanding that social science studies laws in society, which actually ignores the general understanding of decisional science.

At present, empirical research is widely used in social science research. As far as decisional science is concerned, it actually misunderstands the discipline category and chooses the wrong research method. The empirical research may not be appropriate to all descriptive sciences which are different in original methods, do not mention the decisional sciences. This is also an important reason why many disciplines, especially many decisional sciences, have made almost no progress with a lot of human and financial resources invested in over decades.

However, in many research evaluations, the so-called unified "paradigm" is forced to go further. Scientific research is required by a unified paradigm, which equates scientific research with the production of industrial products. It runs counter to the "nature" of scientific research, such as "originality" and "innovation", and seriously stifles the natural vitality and innovation ability of scientific research.

Part II
Asset Pricing

Option Pricing and Valuation of Contingent Cash Flow

Contents

In 1973, Black and Scholes published the well known Black–Scholes option pricing model. Robert Merton also contributed a lot in option pricing method and its applications in multiple directions. In 1997, the Royal Swedish Academy of Sciences announced the 1997 Nobel Prize in economics. Robert Merton and Myron Scholes won the reward. Unfortunately, Fischer Black, who died in his mid-fifties in 1995, brushed against such an honor.

As the prize citation said, "for a new method to determine the value of derivatives. Robert C. Merton and Myron S. Scholes have, in collaboration with the late Fischer Black, developed a pioneering formula for the valuation of stock options. Their methodology has paved the way for economic valuations in many areas. It has also generated new types of financial instruments and facilitated more efficient risk management in society."

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_6

1 Option and Its Value

Option is first of all a wise design carrying flexibilities that are needed by the investors. Let's begin from the basics of options.

1.1 Concepts About Option

An option is a right but not an obligation to buy or sell an underlying asset at a predetermined price or exercise price or strike price and predetermined date or maturity date. So the holder of an option is allowed but not required to (can but does not have to) buy or sell the underlying asset at the predetermined price and date.

Options have actually developed over a long period. The underlying assets now include a large variety of financial assets and commodities. Such as stocks, bonds, foreign exchanges, energy like crude oil, gasoline, natural gas, and heating oil, and metals like gold, copper, silver, and platinum, as well as agricultural commodities or grains like sugar, cocoa, coffee, wheat, soybeans, soybean oil, rice, oats, corn, cotton and lumber, even live cattle and pork.

Options on those assets and items have similar features in value. For the convenience in discussion and explanation, we will focus on stock option in this chapter, the concepts introduced and the features or relationships revealed based on stock option are also true to options on other underlying assets.

Class of option: call versus put

If the holder has the right to buy the underlying asset, the option is referred to as call option, or simply call; if the holder has the right to sell the underlying asset, the option is referred to as put option, or simply put. The option gives the holder a flexibility or freedom, he/she is free to choose whether or not to exercise the right with option.

Obviously, the holder will use that right to maximize his/her own interest. For instance, the holder will exercise a call only when the price of the underlying asset is higher than the exercise price or strike price. Similarly, the holder will exercise a put only when the price of the underlying asset is lower than the exercise price or strike price.

Therefore, the holder or buyer of a call expects or looks forward to the price rising of the underlying asset; the holder of a put expects the price falling. In another words, if you predict the price of the underlying asset will go up in the future (within the life of the option), buying a call may be a wise choice; otherwise, buying a put may be a right choice.

Time to exercise: American versus European

Options can be classified into two categories based on the time allowed to exercise: American options and European options. An American option is allowed to exercise in all the trading days within its life; and a European option is only allowed to exercise at its maturity date or the last day of its life. An American option obviously gives its holder more flexibility than a European option, because it can be exercised early.

Based on the concepts, it is easy to conclude that other things being equal, the value of an American option is likely larger than that of a European option. An interesting but tough question is how much is the American option larger than its corresponding European option in value or percentage? It is not easy to answer such a question, but we can get better understanding about options by thinking it over and over again.

Please note that the difference between the two categories may be not as large as it seems to be. Like a European option, an American option is allowed to exercise only one time during its entire life. The flexibility in exercise date just means it may end up early. It is possible that the American option is equal to or not larger than its corresponding European option in value. When or under what conditions or circumstances will this happen? Thinking over this question can also benefit our understanding about options.

Option position: long option versus short option

As a primary concept, an option gives its holder only right. This is actually only true to one party in the deal. From the counterparty's point of view, an option gives its trader only obligation. The party owning right has the long position of the option; its counterparty owning obligation has the short position. There are totally four option positions in the market: long call, short call, long put and short put. At any time, the total long positions equal the total short positions, because every party has its counterparty in the market.

At a same trading time, for an option, some traders are willing to long, others are willing to short. The difference in position choice depends on the different opinions about the future of the underlying asset. Or more specifically, the price trend or movement of the underlying asset. For instance, when investor C believes Google stock will move up in the near future, investor D believes it will move down, no matter by guess or by estimation. Based on the judgement, investor C will long call or short put, investor D will short call or long put.

Since the holder of the short position has only obligation, a question comes up: who is willing to hold the short position of an option? First of all, the trading of option, like other transactions in the market, is a fair game. As an example, consider the trading of futures in the market, each party involved has the same or symmetric right and obligation. The right equals obligation for each party, so it is a fair game. In the case of option trading, the two parties are not equal because of the asymmetric right and obligation, but it is still a fair game, because the party with the short position receives money or proceeds from the counterparty as a price of the right being given away.

Margin and the cover position

Option margin is the cash that traders must submit to the broker as collateral when going into an option contract. Understandingly, the purpose of option margin is to secure the deal or to avoid default risk. Based on the concepts, the default risk in option trading comes only from the party with the short position. The party with the long position has only right but not obligation, so he or she has no obligation to escape or no promise to break. Whether the contract being kept to the end depends totally on the party with the short position.

Therefore, unlike other securities trading, in which both sides in the deal have to submit margin, in option trading, only the trader with short position needs to submit margin. Normally, the short position trader has to submit margin double in amount. This is the situation for the normal case or for the case of a naked option. In the case of a covered option, the short position holder need not to submit margin.

A naked option is referred to as the short position of an option exposed to default risk without protection from a position in the underlying asset. On the contrary, if a short position in option is protected by a position in the underlying asset, it is referred to as a covered option. For instance, if you short a call option and buy the underlying stock; your call in short is a covered position; it is covered by the underlying stock in long. When the counterparty wants to exercise the call, you are ready to satisfy him/her, because you have already the stock in hand for the settlement.

1.2 The Intrinsic Value of Option

The intrinsic value of options

In general, intrinsic value often refers to the theoretical value or fair value of an asset, but it is not the case for an option. The intrinsic value of an option refers to as the value that can be obtained by exercising the option, so it can also be called the exercise value of an option. For a call option, when the market price of the underlying stock (S) is lower than or equal to the option exercise price (X) before maturity, it makes no sense to exercise the option, and the intrinsic value of the call is zero; when the market price of the underlying stock (S) is higher than the exercise price (X), the holder can get the difference by exercising the option, and the intrinsic value of the call is $S - X$.

Hence, the intrinsic value of a call is:

$$\mathrm{IVLC} = \max(S - X, 0) \tag{1}$$

where, IVLC means intrinsic value of a long call; max(p, q) means to choose the bigger one from p and q. Similarly, the intrinsic value of a put is:

$$\mathrm{IVLP} = \max(X - S, 0) \tag{2}$$

where, IVLP means intrinsic value of a long put. Correspondingly, for the short positions in a call and put, the intrinsic values are calculated as Eqs. 3 and 4 respectively:

$$\mathrm{IVSC} = \min(\mathrm{X} - \mathrm{S}, 0) \tag{3}$$

$$\mathrm{IVSP} = \min(\mathrm{S} - \mathrm{X}, 0) \tag{4}$$

where, IVSC and IVSP mean intrinsic value of a short call and put respectively; min(p, q) means to choose the smaller one from p and q. Based on Eqs. 1–4, obviously, the intrinsic value of long positions is not less than zero; while the intrinsic value of short positions is not more than zero.

The intrinsic value curve of option

The option intrinsic value is also referred to as option payoff. Figure 1 depicts the curve of intrinsic value or payoff of the four option positions.

Figure 1 is the geometry expression about option intrinsic values or payoffs, while Eqs. 1–4 are their algebra expressions.

The intrinsic value of a portfolio can be shown more clearly in the rectangular coordinates as the sum of the intrinsic values of the securities within the portfolio. For instance, the naked call and covered call are shown as Fig. 2.

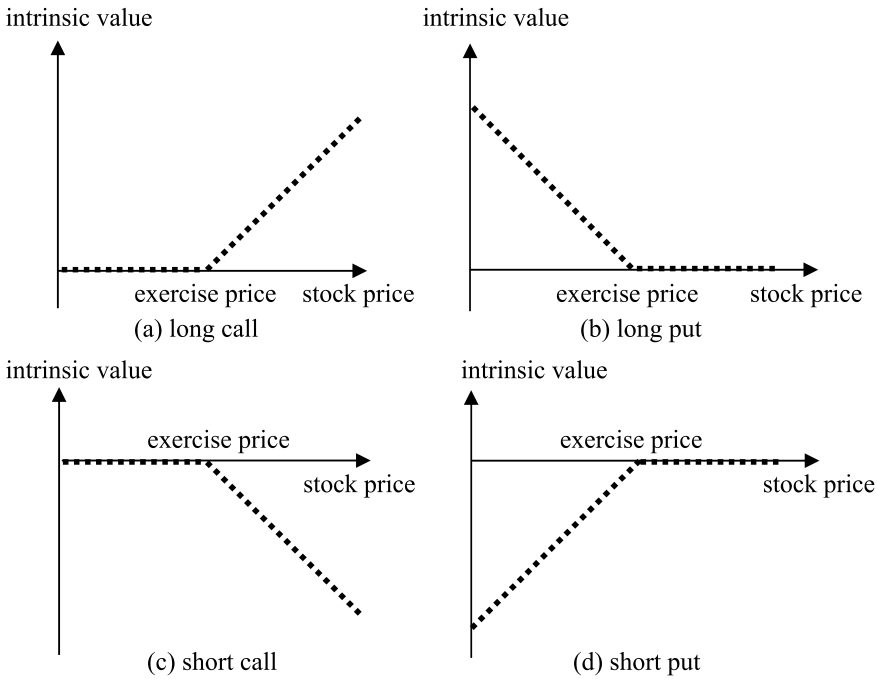


Fig. 1 The intrinsic values or payoffs of option positions

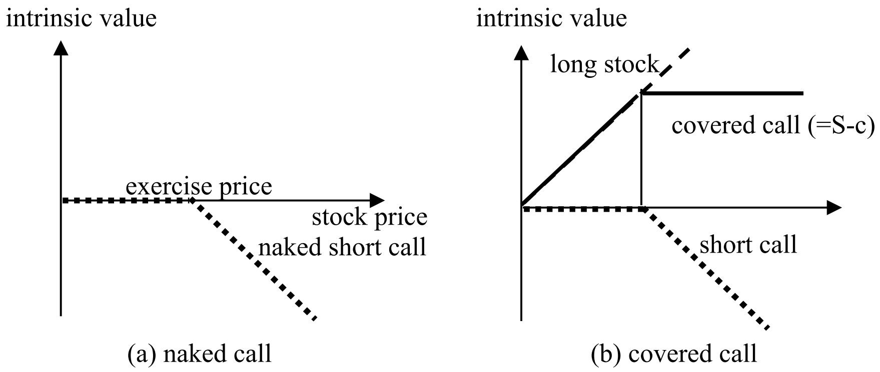


Fig. 2 The naked and covered options

The covered call is the sum of a short call and a long underlying stock (S − c), as shown in Fig. 2, the intrinsic value of the covered call is always larger than zero, which means the holder of a covered call has no further default risk exposure.

The profit of option positions

Note that the profit from option investment or trading is different from the option payoff or intrinsic value. But profit is much easier to measure once the intrinsic value is known, which is simply the intrinsic value minuses the cost. For the option trading, the cost is the initial price at which the option buyer pays to the option seller or writer.

Equations 5–8 are the profits of the four option positions.

$$\mathrm{Ec(long)} = \max(\mathrm{S} - \mathrm{X}, 0) - \mathrm{c} \tag{5}$$

$$\mathrm{Ep(long)} = \max(\mathrm{X} - \mathrm{S}, 0) - \mathrm{p} \tag{6}$$

$$\mathrm{Ec(short)} = \min(\mathrm{X} - \mathrm{S}, 0) + \mathrm{c} \tag{7}$$

$$\mathrm{Ep(short)} = \min(\mathrm{S} - \mathrm{X}, 0) + \mathrm{p} \tag{8}$$

For a call, the breakeven point for its buyer and seller can be derived as:

$$\begin{aligned}
&\max(\mathrm{S} - \mathrm{X}, 0) - \mathrm{c} = 0\\
&\text{since } 0 - \mathrm{c} \neq 0\\
&\text{then } \mathrm{S} - \mathrm{X} - \mathrm{c} = 0\\
&\text{then } \mathrm{S} = \mathrm{X} + \mathrm{c}
\end{aligned} \tag{9}$$

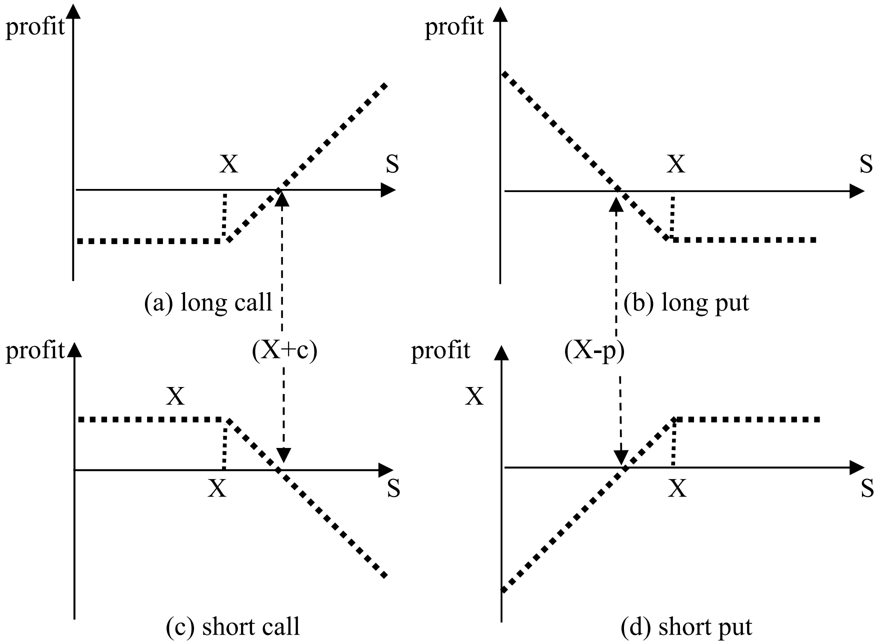


Fig. 3 The profit of option positions and their breakeven points

That is, when the price of the underlying stock goes up to (X + c), the long and short positions of a call reach their breakeven point. Similarly, the breakeven point for put can be derived as: $X - S - p = 0$ or $S = X - p$. That is, when the price of the underlying stock goes down to (X − p), the long and short positions of a put reach their breakeven point (Fig. 3), as shown in Fig. 3.

The moneyness of options

The moneyness is referred to whether or not an option worth exercising, which depends on the relation between the price of the underlying asset and the strike price of the option. Specifically, when the underlying asset price equals the option exercise price, the option is at-the-money; when an option is worth exercising, it is in-the-money; when an option is not worth exercising, it is out-of-the-money. The moneynesses of options are shown as Fig. 4.

The concept of the moneyness, i.e., at the money, in the money and out of the money, are helpful for describing qualitatively the various situations of options in terms of payoff or intrinsic value. For instance, deep or shallow in the money or out of the money, etc.

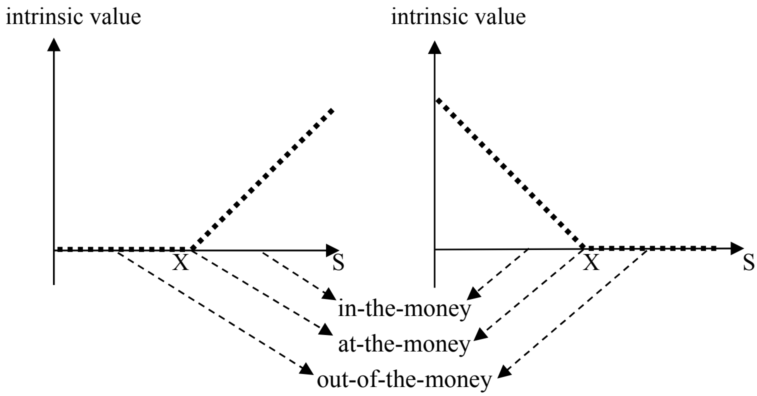


Fig. 4 The moneyness of options

1.3 The Value of an Option

The intrinsic value versus value of an option

For most of assets, intrinsic value and value are the same; they are equal to each other. But that is not true for options. Imagine a call option on Google, which will mature 6 months from now. Suppose the strike price is \$500. If Google is trading now at \$500 a share, the call is at the money, or its intrinsic value is \$0. But its current value or reasonable price is obviously larger than \$0, because its future value is possible to be either \$0 or positive, hence its expected value at maturity date or exercise date is positive, and the present value of such a positive expected value is definitely positive.

Therefore, the value of an option is larger than its intrinsic value prior to maturity date. The difference between the option value and its intrinsic value is referred to as the time value of the option. The option value then is the sum of its intrinsic value and time value, just as shown in Fig. 5. As it goes towards the maturity date, the time value of option moves down towards zero; the value of the option approaches its intrinsic value (Fig. 5).

The determinants of the option value

Since option value consists of intrinsic value and time value, the determinants of option value thus include those determine its intrinsic value and its time value.

Obviously, the further of the maturity date, the larger the time value. So time to maturity is a determinant of the option time value. This is actually one of the two determinants. Another determinant is the volatility of the underlying asset, which is measured by the standard deviation of the annual rate of return on that asset and reflects the risk of the asset.

The determinants of the option intrinsic value include the current value of the underlying asset and the strike price of the option. The strike price as a cash flow normally happens at the maturity or exercise date. We need a discount rate to discount

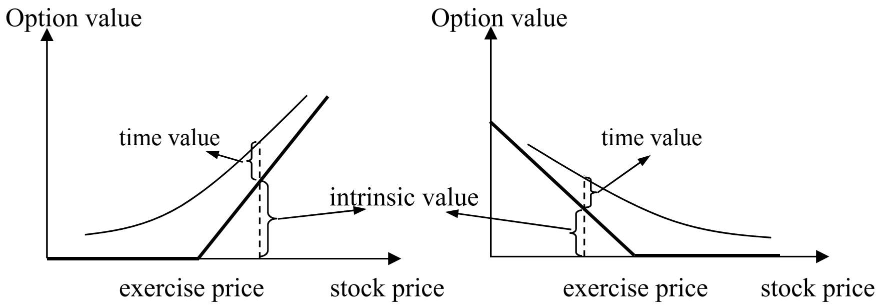


Fig. 5 Option value equals its intrinsic value plus its time value

the strike price back to time zero, so that we can calculate the difference between the underlying stock price and its strike price. Option pricing research reveals that the expected future values in option pricing can be discounted at risk free rate based on a risk free portfolio consists of option.

In summary, there are five factors determining the option value, they are: the current value of the underlying asset, the strike price of the option, a discount rate or risk free rate as well as the time to maturity and the volatility of the return on the underlying asset. For instance, an increase in the underlying price increases the value of a call option and decreases the value of a put option. Reverse is true when the strike price increases. The specific influences of these five factors on option value are shown as Table 1.

Therefore, option value can be derived based on the value of these five factors. But please note that, the specific relations between option value and these factors cannot be derived via statistical process based on market data as the current prevailing research. For instance, we cannot derive option pricing model by finding the coefficients in the regression equation, such as, option value $= aS + bX + cT + d\sigma + er$, because that is not correct, although we can definitely obtain the coefficients from data as well as something like a model. If this can be the right way for modelling, we can model anything or any relation under the sun.

Table 1 The relations between option value and the influential factors

Variables	Denotes	Call	Put
Stock price	S	Positive	Negative
Exercise price	X	Negative	Positive
Time to maturity	T	Positive	Positive
Volatility	σ	Positive	Positive
Risk free rate	r	Positive	Negative

Note positive = option value is positive-related with the factor value; negative = option value is negative-related with the factor value

As a matter of fact, option pricing or option valuation is one of the most tough problems in finance; it cost more than one hundred years for mankind to solve this problem. We shall describe the process in more detail in the following parts.

2 The History of Options

Option is a clever design dated back to the ancient time of human society; it has played important roles in long human history from time to time.

2.1 Options Before 1973

Option is created by our ancestors. Here are some interesting stories about the use of options in history, and hopefully we can get more or deeper understandings and insights about options from those stories.

Thales used option to lock in the leasehold of olive press

The first story took place in ancient Greek around 550 B.C. A scholar at that time named Thales, was good at philosophy and astrology. One night in a winter, he predicted that the weather would be nice in the following year, and the harvest of olive would be good.

Thales then made a reasoning, during the next harvest season, the new harvested olives needed to be pressed to produce olive oil. Good harvest implied the demand for the olive presses would outrun supply and their rent would go up. So Thales made a bold decision.

During that winter, Thales visited the nearby owners of the presses one by one, and persuaded them to accept his upfront deposit, in exchange to give him the right (but not the obligation) to use the olive presses during the next harvest season at an agreed rent (price). Thales thus used option to lock in the leasehold of olive presses.

The owners didn't oppose his suggestion, because they knew nothing about the weather prediction and the ideas in Thales's mind. Moreover, they would get an upfront payment in advance, and if Thales walked away from the deal in the harvest season, they could still be able to rent out the presses as usual. It turned out that Thales' prediction was correct in the next year. Then, everything happened as expected, Thales made a huge fortune by subletting all the olive presses out at higher rents.

Options during the bulb bubble

In the early 1600 s, tulips were regarded as a status symbol among the Dutch aristocracy. As their popularity increased, prices went up dramatically. Before the bubble collapsed in 1637, the unit price of tulip bulb went up as high as annual income of a blue-collar worker.

The price of tulip bulbs fluctuated a lot depending on the good or bad harvest as well as the following supply and demand. Tulip wholesalers and dealers began to buy call options to hedge the risk of price going up, and tulip growers began to buy put options to protect their profits against price going down. As the price of tulip bulbs continued to rise, the value of existing option contracts increased dramatically. So an independent market for those option contracts emerged among the wholesalers and dealers. People used options for hedging as well as for leverage; options thus became a popular tool during the tulip bulb bubble.

The use of options helped inflate the prices of the bulbs. It turned out the high prices of the bulbs were not sustainable; and it collapsed finally. The option transactions were not well regulated, many of the holders of short positions walked away from their obligations.

Stock options in early years

Stock options first began in the Amsterdam bourse in the seventeenth century. Stock and option trading gradually moved to London in eighteenth century after the London Stock Exchange was formally established. London Stock Exchange then became the largest options trading market in the world at that time. The underlying assets then including British and American stocks, British government bonds and Spanish government bonds.

By the nineteenth century, options trading had spread from London to France, and then from France to Germany. The over-the-counter stock options became popular and got flourished in New York in the eighteenth and nineteenth Century. Stock brokers had used options to promote their sales before and after the opening of New York Stock Exchange (1792).

In those early days, the stock options were 100% traded "over the counter," or they were OTC options. Contrary to nowadays standardized options, OTC options were set via negotiation between buyer and seller on each of the option terms for every trading, such as the strike price, expiration date as well as the price of the option.

Such kind of option might meet the demands of the traders better in various aspects, like the date and price for hedging. But the drawbacks were obvious, such as the low transaction efficiency and low liquidity as well as the high transaction cost.

In such a case, it is hard for option holders to resell it. Even worse, the option holder with an in the money position may be hard to make money from the position, because the counterparty might not have the means to fulfill his or her obligation.

2.2 Options After 1973

1973 or April 1973 is a special time in option history; from then on, the standard option came out and the option on exchange as well as the OTC option grew rapidly.

In the early 1970s, to rally the market and expand their business, the Chicago Board of Trade (CBOT) acted jointly with some other exchanges to set up an exchange for stock options—the Chicago Board Options Exchange (CBOE). To ensure the options trading timely and reliably, they set up the Options Clearing Corporation (OCC) in 1973 as well all the trading rules including the rules for margins.

On April 26 of that year, CBOE launched call options on 16 stocks at first. As opposed to the over-the-counter options market, CBOE set terms for those options, hence the options had standardized contract size, strike price and expiration dates.

At the end of May 1973, one month after the launch of the stock options, the CBOE's daily trading volume exceeded that of the over-the-counter option market. The success in standard option attracted more exchanges to join in the development of new option products, bringing options to a wider marketplace. In 1975, the Philadelphia Stock Exchange and American Stock Exchange opened their own option trading floors. In 1977, the CBOE increased the number of underlying stocks to 43 and introduce put options as well.

The first decade thereafter saw the explosive growth of the options market, and it turned out that this was just the first round of growth. In 1983, the index option (the CBOE 100 index) was introduced to market, which helped to further fuel another round growth of the options industry. Up to now, there are at least 50 different index options trading in the market, such as the S&P 500 index (SPX), Dow Jones Index (DJX), iShares Russell 1000® Indx Fund (IWB), Nasdaq-100 Indx Options (NDX), etc.

Another new product was introduced in 1990s, that was options on Long-term Equity AnticiPation Securities (LEAPS). These options have longer life up to three years, enabling investors to take advantage of long-term trends of more different underlying securities in the market. The new millennium saw more innovations in option products and trading. The computerized trading systems and the web-based online trading created a far more viable and liquid options market than ever before. It is a brand-new era showing us more potential and brighter future for various option markets and applications.

2.3 The Role of Options

Options have played an important role in history. The option now is one of the two biggest financial derivatives in most financial markets around world. In history, the option is one of the oldest financial products, which even comes up earlier than most of other financial products, such as stocks, futures, do not mention other far younger products, such as swaps, etc. The Option as a product has facilitated the trading and development of other products, such as stocks, futures, etc.

Along with the breakthrough in option pricing (the following section) and the coming up of the standardized options, the option plays an even more important role in financial market. In addition to expanding traditional services such as deposit and loan business, exchange settlement, check bookkeeping and market operation, the financial service industry introduced financial products, contract products and index products that did not exist before. In the late 1980s, currency Futures (forward currency contracts) and stock index futures in early years (S&P 500 index futures) formed an industrial scale. The financial service industry is no longer indirectly creating value for the whole economy, but directly providing new opportunities by offering their own products to the market, and enlarge its share in the whole economy rapidly.

The success of financial innovation based on the option application shows that futures and options products can be derived from any underlying assets. In fact, following the financial futures and options on currency, futures and options on various financial index, insurance and reinsurance contracts, installment loan contracts and interest rates etc. have developed. Financial industry seems to have more potential to develop new products. The financial industry segments gradually into high-end products and low-end services, and gain more power of guidance in resource allocation for the whole economy.

The rapid growth of derivatives markets over decades based on options as well as on option pricing theory. The option pricing laid a foundation for option related financial innovation since 1980s. As a matter of fact, the option pricing also helps to bring about the revolution in financial theory. On one hand, the core nature of character of options is "can but not obligated" to do something, which represents a flexibility or freedom; similar flexibilities or freedoms are very common in reality, especially in the area of financial and management decision. On the other hand, because the owner of option has only right but no obligation, it divides the return and risk absolutely; so the method from option is an efficient tool for analyzing risk.

In fact, most tough issues in financial theory are related to the neglection or failure to consider the relevant flexibilities or risks. Indeed, before 1973, those flexibilities and risks cannot be properly analyzed because the option pricing is not solved. However, after the Black–Scholes model published in 1973, it became possible or feasible to analyze those flexibilities or risks. This is why we introduce option and the related pricing method this chapter. Readers can further understand the major and specific roles of option pricing method in various parts of finance, especially when you go through the following chapters where the option concept and pricing method are used to deal with and solve the relevant tough and fundamental financial problems.

3 Option Pricing Method

Option pricing is one of the toughest problems in finance. Just imagine that how to value an at the money stock option with a certain time to go before its expiration, you can get to know the difficulty of the problem. The research on option pricing thus experienced a long history before finding the final solution around the corner.

3.1 The Early Research on Option Pricing

From the eighteenth to nineteenth century, the active stock option trading facilitated the development of option pricing techniques. As recorded in Put-and-call by Leonard R. Higgins (1906), in nineteenth century, people had some methods for determining the values of at-the-money, slightly out-of-the-money and in-the-money short-term calls and puts, and used routinely the put-call parity[1] for option conversion and static replication of option positions.

The first significant breakthrough in option valuation is the model worked out by Louis Bachelier in his doctoral thesis (the theory of speculation, published in 1900). Bachelier was a student of the famous Henri Poincaré at that time; the doctoral thesis is considered to be the first paper to value options by using stochastic processes, as shown in Eqs. 10 and 11.

$$\mathrm{C(S,T)} = \mathrm{SN}\left(\frac{S-X}{\sigma\sqrt{T}}\right) - \mathrm{XN}\left(\frac{S-X}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}N\left(\frac{S-X}{\sigma\sqrt{T}}\right) \tag{10}$$

$$\mathrm{P(S,T)} = \mathrm{XN}\left(\frac{X-S}{\sigma\sqrt{T}}\right) - \mathrm{SN}\left(\frac{X-S}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}N\left(\frac{S-X}{\sigma\sqrt{T}}\right) \tag{11}$$

As early as in 1900, the Bachelier model had revealed most important factors in the modern option pricing model, namely, the stock price (the underlying asset price) S, the strike price X, the time to maturity T and the volatility σ. Further, the model used the cumulative function of normal distribution as well as the idea of fair game in the derivation process, which is, to some extent, the same as the no-arbitrage principle.

The model neglected the time value of money. Hence you can derive $\mathrm{C + X = S + P}$ based on the model, instead of the right put-call parity, $\mathrm{C + Xe^{-rT} = S + P}$. In addition, it assumes that the stock price follows the normal distribution, which deviates significantly from the right assumption of lognormal distribution, because it is the rate of return rather than the price of the stock that follows the normal distribution.

Like many other scientific discoveries, Bachelier's work received rare responses and citations at that time, and his model didn't have much influence over several decades although he kept working hard on the research in subsequent years. The value of Bachelier's work was rediscovered by the financial community in the 1950's. Some papers based on his model got published in the 1960's. Among them, the most well-known ones are the models put forward by Case Sprenkle (1961), James

[1] The put-call parity reveals the relationship between a call option and a put option in terms of value. As its modern version states, $\mathrm{c + Xe^{-rT} = S + p}$, where, c and p represent a call option and a put option with the same expiration date T and the same exercise price X; S is the value or price of the underlying asset. The call and put with prices satisfying the put-call parity preclude arbitrage opportunities.

Boness (1964) and Paul Samuelson (1965), as shown by Eqs. 12–14, 15–17, 18–20 respectively.

$$\mathrm{C(S, T)} = \mathrm{e}^{\rho T}\mathrm{SN(d_1)} - (1 - \mathrm{A})\mathrm{XN(d_2)} \tag{12}$$

$$\mathrm{d}_1 = \frac{\ln\left(\frac{e^{\rho T} S}{X}\right) + \sigma^2 T/2}{\sigma\sqrt{T}} \tag{13}$$

$$\mathrm{d}_2 = \mathrm{d}_1 - \sigma\sqrt{\mathrm{T}} \tag{14}$$

$$\mathrm{C(S, T)} = \mathrm{SN(d_1)} - \mathrm{Xe}^{-\rho T}\mathrm{N(d_2)} \tag{15}$$

$$\mathrm{d}_1 = \frac{\ln(S/X) + (\rho + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{16}$$

$$\mathrm{d}_2 = \mathrm{d}_1 - \sigma\sqrt{\mathrm{T}} \tag{17}$$

$$\mathrm{C(S, T)} = \mathrm{Se}^{-(\rho-\alpha)\mathrm{T}}\mathrm{N(d_1)} - \mathrm{Xe}^{-\alpha \mathrm{T}}\mathrm{N(d_2)} \tag{18}$$

$$\mathrm{d}_1 = \frac{\ln(S/X) + (\rho + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{19}$$

$$\mathrm{d}_2 = \mathrm{d}_1 - \sigma\sqrt{\mathrm{T}} \tag{20}$$

These models are similar in terms of the forms and variables. Comparing to previous model, we can find at least two improvements in these models. Firstly, the time value of money is factored into the model. Secondly, the return rather than the price of stock is assumed to follow the normal distribution, and the stock price hence follows the lognormal distribution.

There is still significant deficiency in these models, for instance, they tried to solve the problem by using utility. But utility and the related variables are unmeasurable, such as A in the formula, which is an index of risk aversion, and ρ (rho) is the average growth rate of the stock. Because of these variables, these models are not easy to use in practice.

Anyway, these models are very close to the final solution of option pricing, i.e. the Black–Scholes Model. The economists with interests in option pricing during this period reached a consensus that Bachelier was the originator or founder of option pricing.

3.2 Black–Scholes Model

The research enthusiasm in 1960s lasted into 1970s and gave the birth of the real solution of option pricing, i.e. the well-known Black–Scholes model.

Fisher Black (1938–1995), received Ph.D. in applied mathematics from Harvard in 1964. A year later, he joined the consulting firm Arthur D. Little and met Jack Treynor, who had his own deep understanding on asset pricing. The mentorship of Treynor helped Black go in for financial research, specialized in warrant pricing and capital asset pricing.

Myron Scholes (born in 1941), received his Ph.D. from the University of Chicago in 1969. He was writing his dissertation under Eugene Fama (known for the Efficient Market Theory and the Fama–French model) and Merton Miller (known for the Modigliani–Miller theorem). In 1968, Scholes joined MIT Sloan School of Management, where he met Black. Black and Scholes were both interested in the option pricing. They intensely worked on this issue since 1968 and derived a solution to a European option in about 1971, as shown in Eqs. 21–24.

$$c = SN(d_1) - Xe^{-rT}N(d_2) \tag{21}$$

$$p = Xe^{-rT}N(-d_2) - SN(-d_1) \tag{22}$$

$$d_1 = \frac{\ln(S/X) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{23}$$

$$d_2 = \frac{\ln(S/X) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T} \tag{24}$$

where, r is risk free rate; other symbols are as same as in previous models.

As most real important scientific achievements, the publish of their paper was far from a smooth process, because it was hard to find qualified reviewers. As recalled by Black in 1987, the paper got rejected by Journal of Political Economy and then rejected by the Review of Economics and Statistics. Fortunately, the Journal of Political Economy agreed to reconsider it because Eugene Fama and Merton Miller suggested a second look. After several adjustments, the paper eventually got published in the May–June 1973 issue by the journal. In response to the reviewers' requirement, the paper was retitled as "The Pricing of Options and Corporate Liabilities" in the final version. It turned out that their model is a real solution to option pricing as well as a milestone in modern finance.

It is worth to note that as a real solution, the model is the simplest in form and about a simplest option. Simplest model here means the easiest way to value an option. The model seems not easy or simple, but this is the simplest formula to value an option, because the problem itself is too tough and too complicated. Simplest option here means the model is about the value of a European option on an underlying stock without dividend payment over the option life span. Then only five variables need to be considered in the model, i.e., current stock price, S, the option strike price, X, the option life span, T, the volatility of the stock or the standard deviation of the stock return, σ, and the risk free rate in capital market, r.

The Black–Scholes model is a real solution to the problem of option pricing. A real solution is a general solution or basic solution, which is independent of special cases or scenarios. Such a solution is often featured as simple logic and less variables. The remarkable thing is that the Black–Scholes model is a closed form, analytic model hence easy to use. Further, based on this basic solution, some other factors can be taken into consideration when needed. For instance, based on the model, stock dividends can be incorporated; the early exercise of an American option can also be taken into account under some circumstances; other factors or characters can also be considered whenever needed and possible.

The Black–Scholes model seems very similar to the previous model, especially the Boness (1964) model (Eqs. 15–17). The main improvement is that the average growth rate of the stock ρ is replaced by the risk free rate, r, in the Black–Scholes model. This implies that models in 1960s made progress by considering the time value of money, but failed to choose a right discount rate. The revolutionary idea behind Black–Scholes is that it is not necessary to consider risk premium when valuing an option. Actually, Black and Scholes solved this problem by constructing a risk-free portfolio, which consists of an option and a part of the underlying stock, the profit from the stock position is just always offset by the (negative) profit from the option position, so the portfolio arrives a certain value for sure at the end of the game. Because the portfolio is risk free, in the case of no arbitrage, it can only grow or be discounted at the risk-free rate.

Taking the call option in Black–Scholes model as an example. Since $c = SN(d_1) - Xe^{-rT}N(d_2)$ means $Xe^{-rT}N(d_2) = SN(d_1) - c$, this implies that when we construct a portfolio with a short position of one call and a long position of $N(d_1)$ shares (less than one share) of the underlying stock, we can get a certain or risk free portfolio value of $XN(d_2)$ at time T. Therefore, this portfolio should be discounted at the risk free rate, and its current value is $Xe^{-rT}N(d_2)$. This is the key for working out as well as understanding the Black–Scholes model.

When the CBOE opened in April, 1973, people questioned the wisdom of opening such an exchange. Indeed, it seemed not wise to open a new securities exchange in the midst of the serious bear markets. People also doubted that options were too complicated for most investors to understand and to participate in. Fortunately, in the same month in 1973, the Black–Scholes model got published and was immediately adopted in the marketplace as the standard for valuing the options based on the relevant factors. On the other hand, the Black–Scholes model as an academic or theoretical discovery also benefited from the practical trading generated by the option exchange, which showed to public the viability of the model.

Starting from the application in CBOE, Black–Scholes model played an important role in the follow-up evolution and innovation in the global options market as well as financial market. In 1997, the Royal Swedish Academy of Sciences awarded the Nobel Prize in economic sciences to Merton and Scholes for their groundbreaking work in option pricing. (Black didn't share in the prize because he died unfortunately in 1995).

3.3 Other Methods

The Black–Scholes model is not so ideal from some aspects. The research on option pricing continued. The most valuable new discovery after that is the binomial options pricing model. Up to now, The Black–Scholes model, the binomial model and Monte Carlo simulation are the three most mentioned and used option pricing methods.

Here is a little more information about the two later methods.

Binominal model

Cox, J., S. Ross, and M. Rubinstein published their paper in 1979, titled "Option Pricing: A Simplified Approach". They provided a numerical method for valuing options in the paper named binomial options pricing model.

They valued options based on the future price distribution of the underlying stock. Specifically, they do the valuation by three steps. (1) Build up the binomial price tree of the underlying stock; (2) Find option value at the final nodes of the tree; (3) Find option value at earlier nodes step by step (period by period) until the option value at the start point.

Step 1: Build up the binomial price tree

The binomial price tree method assumes that the underlying stock price will move up or down by a specific factor per step or per period along the tree. To work out the binomial price tree, firstly, to divide the remaining time to maturity of the option into some equal length steps or periods; and then, to decide the values of the up factor (u) and the down factor (d); finally, to work out step by step the binomial price tree of the underlying stock. For instance, let S denote the current stock price, then at the end of the first period, the price will be Su or Sd; at the end of the second period, the price will be Suu or Sud or Sdd, etc.

Normally, the up and down factors are determined based on the stock volatility, and the time duration of a step, t. Please note that starting from a same node, after a two-step move, along the different path, such as moving up and then down vs. moving down and then up, the stock price will reach the same node. This implies that the number of final nodes (possible stock prices) is always equal to the number of steps plus 1. This property makes it easy to decide how many steps to divide the option life span into.

Step 2: Find option value at final nodes

The final nodes are the nodes at the maturity date. The time value of the option reduces to zero at that date. So at each final node of the tree, the option value is simply its intrinsic value. That is, S–X for in the money call and X–S for in the money put, otherwise the value is zero. Understandingly, for most cases, option values at about half of the final nodes are zero.

Step 3: Find option value at earlier nodes

Once the option value at end is found, then the option value one step earlier come into consideration. Since each price at this step moves into two final prices along the up and down paths, the option value at this earlier step is the present value of its expected final value. The expected final value is the up and down values, weighted by their respective probabilities.

Please note that the probabilities are not estimated subjectively, but is determined with objective formula based on the u, d, length of one step and risk free rate.

The option value at each node can be derived by discounting the expected option values at the later step at the risk free discount rate. Repeating this calculation and working back, the option value at the present (first node of the tree) can be worked out. This is the value of the European option at the valuation date.

For an American option, since exercise is permitted at those earlier nodes, it takes the greater of discounted expected value and immediate exercise value or intrinsic value at each node. After a similar step by step back-wards calculation, the present value of the American option can be worked out.

Relationship with Black–Scholes

Binomial option pricing model appears much different from Black–Scholes option pricing model. Nevertheless, they converge in many aspects. They can derive the same valuation result based on the same variables, such as the current stock price, S, the option exercise price, X, time to expiration, T, the stock volatility, σ, and the risk free rate, r.

As a matter of fact, the binomial model provides a discrete time approximation to the continuous process under the Black–Scholes model. As the number of steps increase, the duration of each step becomes shorter, the stock price distribution built in the binomial model approaches the lognormal distribution, which is the same as assumed by Black–Scholes model. Moreover, the valuation result is also approaching that of the Black–Scholes model.

Therefore, same assumptions underpin both the binomial model and the Black–Scholes model. The result of the two models can be used to verify each other. You can say the binomial model is the discrete time version of option pricing model, and the Black–Scholes model is the continuous time version of option pricing model.

In practice, both models are widely used because they have different advantages and disadvantages. Binomial model can consider more conditions or properties, such as valuing American options and Bermudan options. Black–Scholes model is a close form solution, easy to use, fast for simple option valuation, and more suitable for European options.

Monte Carlo simulation

With the aid of computer technology, the option can also be valued by using of computer simulation. Among them, the most common used method is Monte Carlo simulation.

For options with complicated features and more sources of uncertainty (such as exotic options, real options, etc.), valuation by Monte Carlo simulation has its advantages.

As the purpose of our discussion in this book is to reveal the mechanism of the asset valuation or asset pricing, we do not intend to amplify on this approach for now. Readers interested in this approach can consult the related literature.

4 Real Options and Their Value

Option pricing is not only useful in financial market for option trading, but also useful in the areas other than the financial market for common investment and operation decisions, because there are real options everywhere in the real world.

4.1 Source of Value

As mentioned before, the key for financial decisions is to know the value of the relevant assets. As the basic axiom states, asset value is determined by its (future or expected) risk and return. Risk is the uncertainty of asset return (such as ROA). Therefore, the return or cash flows are the source of asset value, or the asset sources its value from expected return.

The expected or future returns usually have more or less uncertainties, and thus fall into two categories in terms of uncertainty: one is fluctuated or volatile returns and the other is the contingent returns. For a company (or its total assets), volatile returns are those expected profits or cash flows from the current businesses; contingent returns are those profits or cash flows from the businesses which may be set up (or not) in near or far future.

An obvious question is: do the future businesses influence the current value of the company? The answer is yes. Why companies in sunrise or emerging industries are more valuable and those in sunset or dying industries are less valuable? As a matter of fact, a lot of startups or companies in high-tech business make no profit from their current businesses. The reason they are so valuable and even become unicorn is that they have abundant business opportunities in the future, or they have more chances to set up new businesses.

Just think two companies with identical returns or cash flows from current businesses. One has opportunities to set up new businesses in the future and the other does not. Obviously, the one with future opportunities will be valued more than the other one in capital market. Therefore, the future business opportunities, although may not bring the new businesses for sure, are definitely a source of the company value.

Anyway, both the future volatile returns and the contingent returns are the source of the firm value. However, it needs different methods to value the two kinds of returns. The discounting method or DCF is suitable to value a volatile return; but it is not a proper choice to value a contingent return, because the return does not come up for sure. The contingent return can only be valued by the option pricing method.

As a matter of fact, those contingent returns often imply future opportunities or flexibilities, which are also referred to as real options because they have the similar features with the option, i.e., the owner can but not obligated to do something.

4.2 Types of Real Options

There are plenty real options in reality. As mentioned before, real options are various flexibilities naturally exist in the real world. Financial options are financial products or instruments designed artificially, which may exist in some financial markets. Comparably, the real options are much more common and exist everywhere, since they need not being designed in purpose, they are here or there naturally.

Focusing on the area of business or finance, real options are those flexibilities, chances and spaces or freedom for adjustment in strategic or operating decisions. Those real options can be classified into three categories: the operation-related options, the investment-related options and financing-related options. The investment-related and financing-related options are normally valued by the option pricing model, such as the Black–Scholes model; the operation-related options may be valued with more flexible or convenient methods. All the three kinds need to be identified before they can be analyzed or valued.

4.2.1 Operation-Related Options

Companies have plentiful flexibilities in their operation. For instance, a factory may use different materials or different mix of certain materials to produce the same product, a dealer has freedom to resell different commodities, a shop has freedom to open or not open in some days, depends on whether the incremental income can cover the incremental cost.

Those flexibilities give the owner or manager of the business the right but not the obligation to make the relevant adjustment, which add value to the business.

Here is an example.

There is a watermelon field (allowed only to grow watermelon), is bidding for the lease in next year, the rental in one year is 250,000 dollars. According to a reliable estimation, the output will be 1000 tons next year.

The operating cost is 1000 dollars per ton of Watermelon (mainly variable cost due to the perfect infrastructure). After successful bidding, the investor can sign supply contract with wholesalers to agree on wholesale price in the harvest season and avoid market price uncertainty.

According to the present forecasting, the per ton wholesale price resulted from the negotiation with the wholesalers is expected evenly distributed between 500 and 2000 dollars (all prices in between are equally possible). If the appropriate discount rate is 10%, is it worthy to tender for the lease of the plot in next year?

Let's work on this problem.

Solution I

Wholesale price = (500 + 2000)/2 = 1250 dollars
Profit per ton = 1250 − 1000 = 250 dollars
Total Profit = 250 × 1000 = 250,000 dollars
Therefore, the net present value,
NPV = 250,000/1.1–250,000 = −22,727.3 dollars
The negative NPV means this investment is not feasible and should be rejected.

This is a solution totally in line with the traditional or prevailing financial analyses. An important problem is: is the solution correct or not?

Solution II

If the wholesale price is below 1000 dollars, the investor can choose not move into the production (irrigation, fertilization and pesticide spraying), and the profit is obviously 0.

The probability of this situation coming up is 1/3.

In case of the wholesale price above 1000 dollars, it is worth to start the production; in this case, the cost per ton is 1000 dollars, the expected income is:

(1000 + 2000)/2 = 1500 dollars
Profit per ton = 1500 − 1000 = 500 dollars
Total Profit = 500 × 1000 = 500,000 dollars
Then, the expected profit based on the two situations is:
0 × 1/3 + 500,000 × 2/3 = 333,333 dollars
Therefore, the net present value is,
NPV = 333,333/1.1 − 250,000
= 303,000–250,000 = 53,000 dollars
The positive NPV means the investment is feasible and can be accepted.

Solution I and II get different answer to the same problem. So there is at most one of them is correct, or at least one of them is incorrect.

It is easy to judge that Solution II is correct and Solution I is not correct, because no rational person will produce watermelon even when the price is less than the cost. The investor has an option not to move into the production process in such a situation. Solution I neglect such an option, hence leads to a wrong conclusion.

It is worth to notice that similar examples are everywhere in reality. For instance, a factory produces a product A. If the price of product A move down to less than its cost, they can adjust to produce product B. So their profit is the greater of profit A and profit B rather than sometime positive and sometime negative. Similarly, a shop sells daily used products, generating profit as well as cost, including fixed cost and variable cost. If they expect in the near future the everyday sales decline too much and the contribution cannot recover the fixed cost, they can close the shop for some days. So the everyday loss is not greater than the fixed cost.

Obviously, such or similar examples are too many to enumerate.

4.2.2 Investment-Related Options

Investment-related options are those flexibilities, chances and spaces or freedom for adjustment in real assets investment. For instance, an investment may enlarge the space for future growth, or for going into the related business area; the assets invested may change the purpose or usage when unfavorable changes happened in environments or conditions.

Previous research reveals that growth options, abandoned options and waiting or timing options are the three main real options concerning investment decisions.

Here is an example about growth options.

Company F has been engaged in computer related products. Since the computer are not very profitable, the company intends to enter the field of home appliances. Depending on the technique advantage, the company expects to move fast in the new industry.

Fridge (Refrigerator) is one of the hot products in the market now, the company plan to acquire a fridge factory, so they can enter this area rapidly. The fridge business is only the first step of the company's whole strategic blueprint. Several other businesses are expected as the later steps to follow the fridge business. One of them is the air conditioner; the others like washing machine, dishwasher, etc.

Anyway, fridge is the pioneer product and trailblazer. The products following it, no matter what it was and when would it launch, depend on the foundation laid by the fridge business, such as the distribution network, the awareness and experience accumulated from the fridge operation, etc. Once their fridge gets a foothold in the market, the other products will follow as a matter of course, and the company's whole strategy will unfold as expected.

The company need to make a decision on whether to purchase the target fridge factory in the year of 2016. The appropriate discount rate is estimated to be 15%; the investment and operating cash flows are estimated carefully as shown in Table 2.

The negative net investment in 2026 means the price at which the business is expected to be sold. The negative $\triangle$WC in the last three years means the current asset is reduced and the relevant capital or fund is expected to come back.

As the company estimated, the most possible time to invest in air conditioner is 2019. The appropriate discount rate for the air conditioner investment is 12%; the

Table 2 Cash flows of the fridge investment (in millions)

FR	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026
OCF		300	400	500	500	600	600	500	500	400	300
NI	3000	0	0	0	0	0	0	0	0	0	−3000
△WC		200	200	100	100	0	0	0	−200	−200	−200
NCF	−3000	100	200	400	400	600	600	500	700	600	3500

FR fridge; *OCF* operating cash flows; *NI* net investment; △*WC* increase in working capital; *NCF* net cash flows; NCF = OCF + NI + △WC

Table 3 Cash flows of the air conditioner investment (in million dollars)

AC	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029
OCF		500	500	600	600	700	700	600	600	500	500
NI	5000	0	0	0	0	0	0	0	0	0	−5000
△WC		200	200	200	100	0	0	−100	−200	−200	−200
NCF	−5000	300	300	400	500	700	700	700	800	700	5700

AC air conditioner

relevant investment and operating cash flows are estimated carefully as shown in Table 3.

The meanings of the OCF, NI, △WC, NCF as well as the negative numbers in the table are as same as they are in Table 2.

The company believes that both the fridge and the air conditioner have life span longer than 20 years. But they will not consider the cash flows beyond 10 years for two reasons. On the one hand, cash flows beyond 10 years are difficult to estimate; on the other hand, omitting the cash flows beyond 10 years or assuming they are zero can reduce workload as well as keep in conformity with the principle of prudence in decision making.

Discounting at a discount rate of 15%, the NPV of the investment in fridge at 2016 is:

$$\text{NPV(FR, 2016)} = 2740.10 - 3000 = -259.90 \text{ (million dollars)}$$

Discounting at a discount rate of 12%, the NPV of the investment in air conditioner at 2019 is:

$$\text{NPV(AC, 2019)} = 4588.75 - 5000 = -411.25 \text{ (million dollars)}$$

Further discounting at a discount rate of 12%, the NPV of the investment in air conditioner at 2016 is:

$$\text{NPV(AC, 2016)} = 3266.18 - 3558.90 = -292.72 \text{ (million dollars)}$$

Anyway, the NPV analyses tell us the new strategy is not feasible, because the NPV is negative for both the investments in fridge and in air conditioner.

The above calculation is in line with tradition in financial practice. However, an important problem is still: is the calculation or analysis correct or not?

As a matter of fact, the investment in fridge paves the way for the company to develop in the new area (home appliances) in future, and the air conditioner is just one of the opportunities generated by the fridge investment. Otherwise, the production and sales of the air conditioner and other related products will be much more difficult.

In short, the investment in fridge produces not only the cash flows from its operation in future, but also the business opportunities for other products.

As an opportunity, a common understanding is that it is something good rather than bad. People like or love opportunities rather than dislike or hate them; people go toward them rather than go away from them. Based on such a common concept, the NPV of an opportunity is by no means to be negative, because you can make no use of it if it seems no helpful to you, and it will not hurt you except benefit you. Put it another way, The NPV of an opportunity is either equal to zero or greater than zero, and can never be negative.

This implies that there is something wrong in above calculation.

As to the decision in 2016, no one knows the situation exactly in 2019. But one thing for sure, once they invest in fridge, they will have the opportunity in 2019 to further invest in air conditioner. As an opportunity, they are not obligated to invest. They have known something about the investment now (in 2016), and will know more about it when the year of 2019 coming. So they can recalculate the NPV of the investment and get a more precise result.

If the 2019 version of the NPV from the air conditioner larger than zero, they can implement the investment, otherwise, they can reject it. In another word, they will make the decision on investing or not investing in air conditioner until the coming of 2019 based on the 2019 version NPV rather than the 2016 version NPV. View from year 2016, the NPV of the investment is either larger than zero (implement) or zero (reject). It is impossible that the investment in the air conditioner has a negative NPV.

The 2019 version NPV is the positive and the zero result, weighted by the probabilities of the two cases. So viewing from 2016, the expected NPV in 2019 should definitely be positive, and so does its value in 2016. But how can we work out a positive NPV of the investment in air conditioner in 2016? The above analysis tells us the cash flows of the air conditioner are contingent returns rather than fluctuated returns, and their value or net present value cannot be calculated simply by discounting the forecasted cash flows.

So the opportunity of investing in air conditioner is a real option, its value thus can be calculated by using Black–Scholes model. This is a growth option, or a long European call option. Although it may be not exactly a European option, since the maturity date is not so definite, we can simply value it as a European option, by which we can reduce the work load as well as keep in line with the principle of prudence in decision making.

As a European call option, its underlying asset is the air conditioner investment. To valued it by Black–Scholes model, we need five variables. (1) the value of the air conditioner investment, S; (2) the exercise price of the option is the amount needed for implementing the investment, X; (3) the time to maturity is the time to go until the investment opportunity expires, T; (4) the volatility of the return of the investment, σ; (5) the risk free rate, r.

Where the X, T and r are easy to estimate. For this air conditioner investment, X = 5000, T = 3. The risk free rate r can be estimated as usual, i.e., it equals to the yield of the government bond with the same life span left. Suppose it is estimated to be 5% based on some appropriate calculations. The S and σ are not difficult to estimate either. We actually know the S in the above calculation of NPV. It is the current value of the investment, i.e., S = 3266.18.

The volatility σ is not as difficult to estimate as it seems to be. Actually, volatility σ has an empiric value range. The common empiric value range is 0.2–0.4 for most equity or stocks. When the returns of a similar or comparable asset over sequent time intervals are available, we can calculate the standard deviation of the returns and get the volatility. If no relevant data available, we can guesstimates or rough estimate the volatility based on the empiric value range and the judgement about the investment risk. That is, if the investment is about average in terms of risk, the volatility is about 0.3; if the investment is relatively risky, the volatility is near 0.4; if the investment is relatively safe, the volatility is near 0.2.

Moreover, when the volatility is really difficult to estimate, we can estimate a rough range. For instance, 0.2–0.25. Because the option value is positively related with the volatility of the underlying asset, according to the prudence principle, when it is a long position option, we can value it by using the lower bound value, i.e. the 0.20; when it is a short position option, we can value it by using the up bound value, i.e. the 0.25.

Suppose we estimate the volatility of the air conditioner investment is around 0.3–0.35, we can simply use 3.0 as the volatility in this case, because we can get the lower bounder of the opportunity value in this way, so that follow the principle of prudence. Then, we have:

$S = 3266.18$, $X = 5000$, $T = 3$, $r = 5\%$, $\sigma = 0.3$

$d_1 = -0.2710$; $N(d_1) = 0.3932$; $d_2 = -0.7906$; $N(d_2) = 0.2146$; $e^{-rT} = 0.8607$

Based on Black–Scholes model (Eq. 21),

Value of the growth option $= c = SN(d_1) - Xe^{-rT}N(d_2)$

$= 3266.18 \times 0.3932 - 5000 \times 0.8607 \times 0.2146 = 360.78$ (million dollars)

The net present value of the fridge investment or the strategy:

NPV(FR, 2016) $= 2740.10 + 360.78 - 3000 = -259.90 + 360.78 = 100.88$ (million dollars)

NPV is larger than zero means the fridge investment or the new strategy is feasible.

Please note again that the real option here in the case is existing naturally without any artificial design or manmade arrangement, and most (if not all) investment implies further growth opportunities in the future.

4.2.3 Financing-Related Options

Financing-related options are those flexibilities, chances and spaces or freedom for adjustment in financing or the contingent cash flows in financing. For instance, the equity of a company is a call option to buy the company at the debt maturity date with an exercise price equals to the debt maturity value. The guarantee of a company (total) debt is a put option to sell the company at the debt maturity date with an exercise price equals to the debt maturity value.

Figures 6 and 7 depict the relationships among the values of a company and its debt and equity holders viewing from the point of real options.

In the rectangular plane coordinate system or plane coordinate system in Fig. 5, the lateral or horizontal axis is company value, and the vertical axis can be any value being analyzed. That is, both the horizontal coordinate and vertical coordinate can be the company value. The line represents the company value thus is the straight

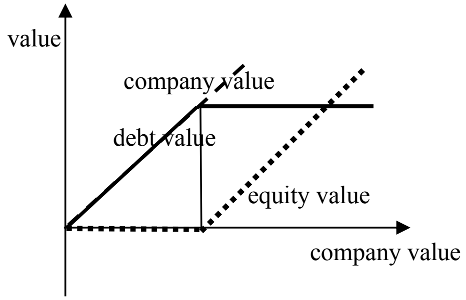


Fig. 6 The values of a company and its debt and equity

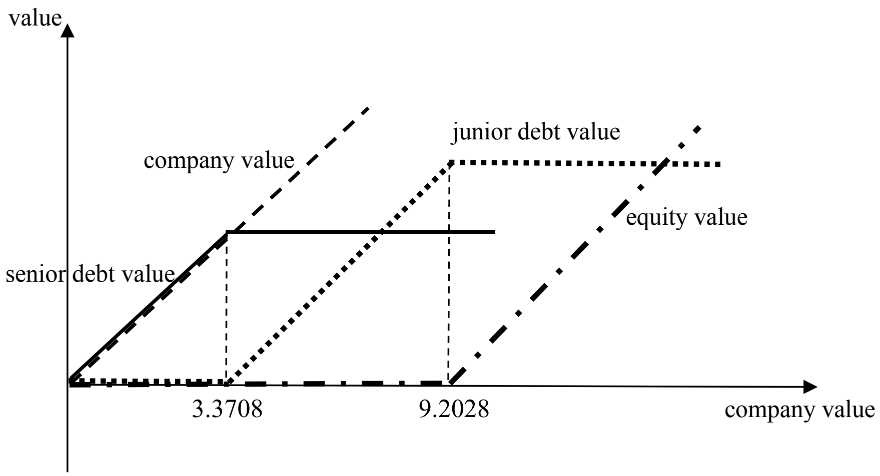


Fig. 7 The more complicated financing mix

line starting from the origin and moving up at a 45° angle. Both the equity holders and debt holders are qualified to claim their own parts from this value. But the debt holders are first in line and equity holders are second in line. This implies that at the debt maturity date, the equity holders can only get the residual part of the company value after the debt value. If the company value is not larger than the debt value, the equity holders get nothing but zero; if the company value increases beyond the debt value, the equity holders get every dollar beyond the debt value.

Hence, the line represents the equity value overlapping with the horizontal axis first and then move up at a 45° angle after the company value increases beyond the debt value. Obviously, such kinked line is just the same as the intrinsic value line of a call option. That implies the equity of a company is a call option of the company value. The debt of the company thus is the company value minus this call option.

The understanding about the equity and debt value from the real option is essential for some financing arrangements. For instance, the conversion of equity and debt in some questionable company. Here is an example.

Bank B lent 8 million dollars to company C eight years ago. The main business of company C has kept declining for years. The loan has two years to go. But it seems difficult for company C to repay the remaining principal and interests of the loan. So the bank and the company sit down to negotiate and reach an agreement. They agree to convert the bank loan from the debt of the company to its equity at a reasonable or equitable price, so that they can manage the business of the company together and revitalize it. By a reasonable or equitable price, they mean a price equal to or close to its fair value.

The book value of the company is 20 million dollars, including current debt 5 million dollars, long term debt 8 million dollars (the bank loan), and equity book value 7 million dollars. The situation of the company is well known in capital market, so the market value of its total asset is below its book value, around 10 million dollars. The interest rate of the bank loan is 8%, and the risk free rate in capital market is 5%. Suppose the company market value can represent its fair value and the current debt needs to be repaid immediately in full amount. So how much should the bank loan convert into the company equity in percentage?

After the deduction of the current debt, the company (fair) value remains:

$10 - 5 = 5$ (million dollars)

This means the total fair value of the bank loan and original equity is 5 million dollars, although their book value is 15 ($= 8 + 7$) million dollars.

Anyway, the problem is: how much should the bank loan convert into the company equity in percentage? Or how much should the original equity be now in percentage?

Obviously, the answer depends on the fair value of the bank loan and the original equity respectively. A well know fundamental accounting equation or identity is that total asset equal to debt plus equity. But now, the total asset is less than debt in value; it seems that we cannot apply the simple accounting identity to calculate the debt and equity values and the problem is obviously not easy to solve within traditional accounting and financial knowledge.

However, the problem is actually quite easy to solve if we use option pricing technique.

As pointed out in above, the company value remains to the bank and the original equity holders is 5 million dollars. If we regard the company equity as a call option of the whole company, this is the value of the underlying asset, S.

The exercise price of the option, X, is the maturity value of the option, i.e.,

$X = 8 \times (1 + 8\%)^2 = 9.3312$ (million dollars)

Obviously, the maturity date of the option, T, is 2 years, and it is given that the risk free rate in the market, r, is 5%. Suppose the volatility (risk) of the company is 30%. Then, based on Black–Scholes model,

$d_1 = -1.0228, N(d_1) = 0.1532;$
$d_2 = -1.4470, N(d_2) = 0.0739;$

The value of the call option then is 0.1417 million dollars. That is, the value of the original equity is 0.1417 million dollars. Therefore, the value of the loan is:

$5 - 0.1417 = 4.8583$ million dollars.
$0.1417/5 = 2.8\%$
$4.8583/5 = 97.2\%$

Then, after the conversion, the company's equity is shared by the bank and the original equity holders 2.8% and 97.2% respectively.

In reality, the financing mix may be more complicated. There are more kinds of debts and equities with different priorities in their claims, such as senior and junior debts,[2] prime and subprime loans, junk bonds,[3] as well as preferred stock and common stock, etc. The option pricing technique can be used to deal with more or any kind of complicated cases.

In the above case, the company may be financed with senior and junior debts rather than only one long term debt, as well as preferred and common stocks rather than only one kind of equity. In such a more realistic situation, we can still use options to depict the financing mix and value each debt and each equity, as long as we can distinguish the claim priorities of those debts and equities.

Specifically, suppose in above Company C, besides the equity capital, the long term capital consists of a senior loan 3 million dollars borrowing from Bank A at an interest rate 6%, a junior loan 2 million dollars borrowing from Bank B at an interest rate 8%, and another junior loan 3 million dollars borrowing from Bank C at an interest rate 8%. Those debts are all two years away from maturity, and now are waiting in line to convert into the company's equity. Obviously, the junior loans borrowing from Bank B and Bank C belong to the same claim priority class. The book values of the debts at maturity can be calculated as:

---

[2] A type of subordinated debt, junior debt has a lower priority for repayment than other debt claims in the case of default.

[3] Junk bonds are mainly issued by some small companies to raise funds for business expansion. Due to the unstable business, their credit rating is often below investment grade. Generally, junk bonds are high-risk bonds that offers a high yield.

senior debt: $3 \times (1 + 6\%)^2 = 3.3708$ (million dollars)
junior debt: $5 \times (1 + 8\%)^2 = 5.8320$ (million dollars)
total debt: $3.3708 + 5.8320 = 9.2028$ (million dollars)

In Fig. 7, there are two call options with the same underlying asset, which is the company value; the only difference between them is the exercise price. The first call option has an exercise price 3.3708 million dollars, and the second call option has an exercise price 9.2028 million dollars. Obviously, the second call option represents the equity of the company; the senior debt equals to the company value minus the first call option; the junior debt equals to the first call option minus the second call option.

For the first call option:
$S = 5, X = 3.3708, T = 2, \sigma = 30\%, r = 5\%$
Based on Black–Scholes model,
$d_1 = 1.3772, N(d_1) = 0.9158;$
$d_2 = 0.9529, N(d_2) = 0.8297;$
$c = 2.0483$
For the second call option:
$S = 5, X = 9.2028, T = 2, \sigma = 30\%, r = 5\%$
Based on Black–Scholes model,
$d_1 = -0.9901, N(d_1) = 0.1611;$
$d_2 = -1.4144, N(d_2) = 0.0786;$
$c = 0.1506$
Therefore:
The original equity $= 0.1506$ (million dollars)
The junior debt $= 2.0483 - 0.1506 = 1.8977$ (million dollars)
The senior debt $= 5 - 2.0483 = 2.9517$ (million dollars)
After the conversion, the equity is divided as:
The original equity $= 0.1506/5 = 3.01\%$
Bank A $= 2.9517/5 = 59.03\%$
Bank B $= 1.8977/5 \times (2/5) = 15.18\%$
Bank C $= 1.8977/5 \times (3/5) = 22.77\%$

There are actually a lot of problems remained in finance and need to be dealt with the option pricing technique. For instance, the pricing or valuation of debt or loan guarantee, etc. You can see more cases that tough problems are solved with the option pricing methods in the following chapters. As you have already understood now, the first step to solve such kind of tough problems is to identify the real options embedded in the relevant issues.

Undoubtedly, option pricing theory is a cornerstone of finance. Just as said by Stephen Ross, “it is the most successful theory not only in finance, but in all of economics.”

4.3 Map of Valuation Method

Asset value is the core in finance. Valuation method is most important in financial theory. To some extent, financial analyses are various applications of valuation methods.

As a common classification, valuation methods are often classified into three categories: absolute valuation, relative valuation and option pricing method.

Absolute valuation derives asset value based on its properties in return and risk. Since risk and return determine the asset value, absolute valuation is absolutely right in concept or in theory. Put it another way, this is the right way to value asset.

Relative valuation derives asset value based on the prices of comparable assets. Relative valuation does not care much about the risk and return of the asset, but believes assets with similar value should be closer to each other in their prices.

As mentioned above, option pricing as a valuation method come up later and has only a short history. It thus became the third valuation method in some textbooks. This is a common opinion in the academic area as well as in practical area.

However, conceptually, absolute valuation and relative valuation should be the whole of valuation. In another word, a valuation method should be either absolute valuation or relative valuation. There should be no the third valuation method.

Then, does option pricing as a valuation method belong to absolute valuation or relative valuation? Obviously, it belongs to absolute valuation, because it cares about the risk and return of the option as well as that of the underlying asset.

Therefore, discounting and option pricing both are absolute valuation. The difference is that discounting method is good at (only able to) valuing fluctuated returns; whereas option pricing is good at (only able to) valuing contingent returns.

Besides the discounting method, we have developed a new absolute valuation method based on the required payback period in last chapter, such as the ZZ growth model, which is also deriving asset value based on the properties of its risk and return.

So there are three methods for absolute valuation and some valuation multiples for relative valuation. The panorama or map of valuation method is shown as Fig. 8.

An important question here is that what is the relationships among those methods? Are they replaceable to each other or complementary to each other?

Valuation
- Absolute valuation
  - Valuing fluctuated returns
    - (1) Discount method ---- valuation based on required rate of return
    - (2) New method ---- valuation based on required payback period
  - Valuing contingent returns ---- Option pricing method
- Relative valuation ---- based on Ratios like P/E, P/B, P/S, etc.
  Improved by ZZ theoretical ratios models

Fig. 8 Map of valuation method

Absolute valuation and relative valuation are replaceable to each other. You can use absolute valuation or relative valuation to value an asset, what ever you like. You can also use one of them to revise the valuation result from the other one.

The relationship between option pricing and the traditional absolute method, such as discounting method, is not so clear. For instance, an equity can be valued with discount method and option pricing method as well, depending on the data available.

Based on the above analyses, as for a company, fluctuated returns come from the current business; whereas contingent returns come from future businesses, which may set up or may not set up in accordance with the supposed time schedule.

If the data available are future cash flows of current businesses and that of the future possible businesses, we can value the current businesses with discount method or ZZ growth model; and value the future businesses with option pricing method. Then the total value of the equity is just the sum of the two results. The option pricing and the traditional absolute method in such a case is obviously complementary to each other.

On the other hand, if the data available are the whole value of the company and its debt, we had better use option pricing to value the equity. So option pricing and other absolute valuation methods are replaceable to each other sometime, whereas more often, they are complementary to each other. Especially, option pricing can be used to solve the unsolvable problem by other absolute valuation methods.

As to the two methods for valuing the fluctuated returns, the discount method and the payback method, seem as if replaceable to each other, they are actually complimentary to each other. The discount method is good at valuing asset with clear life span, such as various kind of bonds, while the payback method is good at valuing the asset with unknown life span, such as equities and stocks.

Further Readings

1. Bachelier L. Theory of speculation (théorie de la spéculation). Annales Scientifiques de l'École Normale Supérieure, Sér 3, 17 (1900), 21–86, translated by D. May.
2. Sprenkle CM. Warrant prices as indicators of expectations and preferences. Yale Econ Essays. 1961;1(2):179–231.
3. Boness JA. Elements of a theory of stock-option value. J Polit Econ. 1964;72(2):163–75.
4. Samuelson PA. Rational theory of warrant pricing. Ind Manage Rev. 1965;6(2):13–31.
5. Samuelson PA, Merton RC. A complete model of warrant pricing that maximizes utility. Ind Manage Rev. 1969;10(2):17–46.
6. Black FS, Scholes MS. The valuation of option contracts and a test of market efficiency. J Financ. 1972;27(2):399–417.
7. Black FS, Scholes MS. The pricing of options and corporate liabilities. J Polit Econ. 1973;81(3):637–59.
8. Merton RC. Theory of rational option pricing. Bell J Econ Manage Sci. 1973;4(1):141–83.

Certainty Equivalent, Risk Premium and Asset Pricing

Contents

Financial decisions rely on the quantitative trade-off between risk and return. It is important to incorporate the relevant risk well into the decision. The relation between risk and return then becomes a core topic in finance. This is the topic of this chapter.

The risk-return relationship is often discussed under the topics like "asset pricing", "risk pricing", etc., which aims at finding the fair rate of return on a capital asset. The capital asset is the other side of real asset. The real assets are those have physical forms, such as fixed assets like factories, equipment, machine, etc., as well as current

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_7

assets like materials, components and parts, etc. Those real assets are recorded at the left side of the balance sheet. The capital assets are recorded at the right side of the balance sheet, which are actually the capitals backing those real assets. The capital assets normally classified into two categories, the equity capital and the debt capital.

Therefore, the asset pricing or capital asset pricing aims at finding the three fair rates of return, that is, the rate of return on equity, debt and total capital respectively. The fair rate of return can also be referred to as risk-matched rate of return or (fairly) required rate of return, which just coincides with the concept of discount rate. This chapter thus aims at finding the right way to determine the discount rate.

1 The Relation Between Risk and Return

All financial analyses and decisions, including investment, financing, risk assessment, valuation and so on, should be based on the future cash flows. Each of those predicted future cash flows is the expected value in the relevant time unit in future; many possible values are supposed to lie around each expected value and often follow a normal distribution. Therefore, such predicted or expected values contain uncertainty or risk.

1.1 Basic Concepts

As a matter of fact, most (if not all) decisions should consider both risk and return, or trade-off between risk and return. But most of the time, they need only to consider or trade-off qualitatively. Finance as a branch of science is a quantitative subject; the relevant risk and return need to be considered or traded-off quantitatively in finance. Or more specifically, finance supplies the theory and methods to support quantitatively trade-off between risk and return. This is one of the unique features of finance.

The discussion about the relation between risk and return in finance is also referred to as capital asset pricing, which is actually aimed to find the fair return of a certain investment or a certain capital, such as debt capital or equity capital. Capital asset pricing is also helpful to determine a fair interest rate of a loan, or a fair return of a debt capital.

As the DCF method becomes the basic or even the only approach for financial analyses and valuations, analysts are used to consider risk via the discount rate. Conceptually, the discount rate should be the fair return of a certain investment or a certain capital. Therefore, in finance, the relation between risk and return, the capital asset pricing, the determining of discount rate are similar to each other in concepts; consequently, discount rate, capital asset price, and required rate of return are also similar concepts as well.

Required rate of return and risk adjusted rate of return are two of the most used other names of discount rate. Required rate of return can be regarded as an abbreviation or abbreviated expression of "reasonably required rate of return" or "fairly required rate of return". By "reasonably" or "fairly" we mean "matching with the relevant risk". These two names indicate the discount rate is related with the risk, and are also the right appellations of discount rate. Capital cost is also another prevailing name of discount rate; but this name contains serious misunderstanding, although it is very popular in academic and in practical area. We'll discuss on it in more detail in the following section.

Therefore, the main topic of this chapter is how to estimate or determine discount rate, more specifically, how to determine a fair rate of return for a capital invested based on its risk. To trade-off or consider risk and return quantitatively, we need first to measure the risk and return quantitatively. Understandingly, most (if not all) of the time, considering something quantitatively is much more difficult than considering it qualitatively.

1.2 Return in Finance

A return is what you get from effort or money you paid. A return in finance is often referred to the money made or lost on an investment over a period of time. Financial returns are often named by various terms like profit, income, earnings, gains, as well as cash flows, etc. A positive return represents a profit while a negative return marks a loss.

Returns sometimes are expressed as the absolute value (dollar value) of return over a period, such as earnings on a certain equity or stock investment, interests on a certain bank loan, cash flows from a business investment, etc.; sometimes are expressed as the relative value (percentage) of return, such as the risk premium on an equity investment, yield on a bond investment, accounting rate of return from a business, etc.

Returns may have different specific names for various investments or capitals. For instance, for loan or debt investment or capital, the return is called as interest (absolute value) or interest rate (relative value); for equity or stock investment or capital, the return is called as return (dividend and price change or capital appreciation) or rate of return (dividend yield and percentage of capital gain); for some other long term assets, such as machines, factories, buildings, and so on, the return may be called as rents.

The total return from a stock includes both capital gains/losses and dividend income. But in practice, for simplicity, it is often taking price change (capital gains/losses) as the total return on a stock or taking index change (average capital gains/losses) as the average gross return on all the stocks in the market, neglecting the dividend payments.

As to a business or a company or a project, the return is often expressed as profits or cash flows over a series of years. It is supposed that the cash flows are more accurate than the profits for recording the returns generated in terms of the size and time. Those profits or cash flows are equivalent to the dividend payments in the case of a stock. In valuing a business or a company or a project, the capital gain is often considered once in the end rather than considered every period or year over the life span or time span in consideration.

Returns are often measured on an after-tax basis. Taxes come into consideration within the context of finance are usually only income taxes, like corporate income tax for valuing a business or a project. The income tax is usually the last step or final line in taxation. So the after-income-tax return is actually the net return after deducting all kinds of taxes. For an investment, it is the difference between beginning value and ending value plus any dividends, interest, or other income received and minus any costs or taxes paid.

Returns are often annualized on percentage basis to facilitate comparisons, because annual rate of return is the most prevailing basis for measure returns. A rate of return is a simple expression of an annual rate of return; in contrary, a rate of return over a period other than one year should be specified the length of the period. Examples like weekly, monthly or semiannual rate of return. We may work out the average daily rate of return of a stock based on its prices during one month, we then need to annualize it. For instance, if the daily rate of return is 0.1%, suppose one year has 360 days, the annualized daily rate of return is 36%.

It is worth to mention or stress that the returns to be considered in finance are often the future returns rather than the past or actual returns, because finance is a subject concerning financial decisions, and decisions should be made based on some forward-looking or expected situations; also, risk and return in future rather in past determine the current asset value. Future returns are simply the forecasted or predicted returns.

For a specific investment, the asset may be held over a period more or less than a year. As a backward-looking, the actual return over the holding period may be expressed as an absolute value or as a percentage; as an annual return or as a periodical return.

Nevertheless, the asset value is irrelevant with those past or actual returns, even irrelevant with the length of the (future) holding period. The value of the asset is determined by all its future returns rather than only the returns over the holding period. Hence, asset valuation model or capital pricing model usually does not consider the length of the holding period. The length of the holding period perhaps finally influences a specific investment's return, but that has nothing to do with the fair return and value of the asset.

Measuring or forecasting returns is one thing, considering them is another. Finance usually considers returns by using discounting. Discounting is the bridge between return and value. People know return and value are positively related to each other based on common knowledge; but the exact quantitative relation is hard to tell without discounting method. The future returns can be added back and compared with current investment only after being discounted.

Discounting is a reasonable method which can incorporate risk and return into asset value. Based on discounting, the influences of inflation need not to be considered, because the inflation effects on return and on discount rate can just cancel each other out. More often than not, the estimated or predicted future returns need not to be adjusted for changes in prices due to inflation. We need not real dividend, real interest rate, real risk free rate, etc., the corresponding nominal returns are just fine for financial valuation.

Here are some common return ratios. ROI is return on investment, which is a percentage return derived from dividing the dollar return by the initial dollar investment. ROE is return on equity, which is the ratio of a company's net income over its shareholder's equity. ROA is return on assets, which is the ratio of a company's net income over its total assets.

1.3 Risk in Finance

As stated above, the return in finance is usually the predicted return, typically an expected absolute value. As a future possible value, the predicted or expected return is surrounded by many possible returns and they are often distributed as a normal distribution. This implies the predicted or expected return contains uncertainty.

In finance, risk is defined as the uncertainty of return, or as the chance a return outcome differs from the expected value. Therefore, risk can be represented by the dispersity or concentration of those possible returns. The more dispersity or the less concentration, the high risk, and vice versa. In statistics, standard deviation or variance is used to measure the dispersity or concentration of data distribution. The standard deviation or variance then is accepted as basic indicator of risk and is widely used in finance and the related subjects.

The standard deviation is usually calculated based on relative returns rather than the absolute returns in practice, so that the size effect can be eliminated for the purpose of comparison. Such a standard deviation is referred to as volatility of an asset. The risk measured by the volatility is the total risk of the relevant asset, or the sum of the systematic risk and the nonsystematic risk.

The volatility of a company measures the volatility of return on its total asset or total capital; which is also referred to as business risk or company risk. The business risk comes from the fluctuations of markets related to the company's outputs (products and services) and inputs (land, office, raw and processed materials, interest rate of debt capital, labor and talent) as well as various events that influence the earnings and values of the company, such as work safety accidents, fire or conflagration, flood or inundation, corruption, the changes on management style and business strategy, etc. Obviously, the market may fluctuate in two dimensions of price and quantity of demand and supply.

Unfavorable fluctuations sourced from markets and some events may damage the return as well as the value of the asset. When the unfavorable conditions continue or even worsen, the solvency of the company will be damaged as well. Eventually, when the insolvency is not avoidable, the company will go (into) bankruptcy. This means the company is unable to pay the on due interest or principal of its debt obligations.

Therefore, bankruptcy risk is the possibility that the return of a company or value of its asset is down to a level at which the insolvency is inevitable. Bankruptcy risk is also named as insolvency risk, credit risk, default risk, or financial risk as in corresponding to business risk. Such risk is a particular concern to commercial banks, bond investors, and various kind of lenders or debt capital providers. However, theoretically, the equity holders lose even more. When the bankruptcy occurs, there is usually not much left to equity holders, because their claim to the asset is second to the debt holders.

In reality, the business risks vary across companies; similarly, the credit or default risks vary across bonds or loans or debts as well. Bonds issued by companies, i.e., corporate bonds, may be high rated or low rated by Standard and Poor's, Moody's and Fitch, etc. The ratings given by those rating agencies, such as AAA, AA, etc., represent the grade or degree of the credit or default risk of the bonds or debts. Comparing to corporate bonds, government bonds, especially those issued by the central government of a country, have the lowest default risk, or even no default risk at all, because the central government (rather than the local governments) have right to issue more currency, which implies that the central government can always pay its due interest and principal and avoid default.

Government bonds are thus known as risk free or riskless bonds. It is worth to make clear that risk here is by no means all risks; risk here is only referred to default risk. Investors of government bonds are still exposed to other risks, such as the fluctuations of bond price which may be led by the change of interest rate in market. No investment is fully free of all possible risks; if you really want to be risk free, the only way is avoiding any investment, just hide all your money under your mattress. Such kind of risk free means no any expected return. This is obviously not an ideal result.

Anyway, credit or default risk is the main risk for bond or debt investment. Investors have different risk attitudes that determine their willingness and ability to take risk. But in general, as risk rises, investors require higher return to compensate for the higher risk. Therefore, as a universal law or rule of thumb in capital market, the greater the default risk, the higher the expected return. Among various bonds, junk corporate bonds tend to have the highest default risk, hence have the highest return or yield; central government bonds with almost zero default risk have lowest return or yield. The annual return (yield) of government bonds is also named as risk free rate. So the risk free rate represents the annual return with zero default risk (rather than all kinds of risk).

As a convention, people do invest and meanwhile try to reduce unnecessary risks by some risk hedging or diversifying strategies. The total risk can be divided into systematic risk and non-systematic risk. They are also referred to as non-dispersible risk and dispersible risk respectively. Consider some stocks with the same level of risk, such as 33% in terms of volatility, if you invest only in one of them, your risk is 33%; if you invest in more of them, your total or average volatility is probably less than 33%, because the risk or uncertainty of some stocks can be partly canceled out by other stocks in your portfolio. Even so, investors normally are exposed to both systematic risk and non-systematic risk because diversification in practice can only cancel out part of the non-systematic risk.

Non-systematic risk is dispersible. When more and more securities are added into the portfolio, more and more non-systematic risk is dispersed. Theoretically, the non-systematic risks among securities are totally cancelled out when all securities in the market are added into the portfolio; and the risk remained is only systematic risk or non-dispersible risk. Obviously, this is just an ideal outcome in theory and cannot realize in practice. Investors, traders, and business managers in practice indeed can setup their risk free portfolios by using some derivative positions to hedge all the risks away. However, for risk reduction purpose, unlike diversification strategies, hedge strategies are not risk free, and even cost a lot sometime.

Systematic risks and non-systematic risks may come from different sources. Systematic risks, also known as market risks, are risks that can affect an entire economy or the overall market. Systematic risks may come from some common factors, such as political and macroeconomic changes, that affect the overall market. Other common sources of systematic risk include those changes of interest rate, inflation, foreign exchange rate, etc. Non-systematic risks, also known as specific risks or idiosyncratic risks, are those company or industry-specific risks that only affects an industry or a particular company. Examples of non-systematic risks include a change in management, a product recall, an office strike, an equipment failure, and some big customers walk away, etc.

Measuring risks is one thing, considering them is another. Finance usually considers risks via discount rate, which is in line with the discounting method. This implies the discount rate should be related with the risk of the asset. As derived above, there should be a relative safe return even on (default) risk free investment. Thus, discount rate for most (if not all) investments should be higher than the risk free rate.

Put it another way, risk free rate often forms a baseline for discount rate or fair rate of return on certain asset, because investors wouldn't accept additional risk unless the potential rate of return is greater than the risk-free rate. Anyway, as a result of the risk/return tradeoff for investment and financing decisions, a higher standard deviation or higher volatility means a higher risk—as well as a higher expected return. We will discuss the estimation of fair rate of return or discount rate in more detail in the following parts.

2 Alternatives to Determine a Discount Rate

According to the basic concept, to determine a discount rate, we had better to know the quantitative relation between volatility (risk) and fair return. However, it is not easy to quantify the fair return based on the volatility. Multiple alternatives have been used in financial community over a long history. We now examine them one by one.

2.1 The Industrial Average Rate of Return

Some financial and investment textbooks take the industrial average rate of return as a benchmark to determine the discount rate. This may have some advantage in data availability, but may be far from reaching a correct or fair result. Although high returns are often associated with high risks; it is by no means that the returns in a specific industry over a past specific period represents a fair return over that period, do not mention a fair return in future.

In fact, in any given geographic area over a given time period, the risks in some industries may be over compensated by the average returns, hence the risk premium is too high; whereas the risks in other industries may be under compensated by the average returns, hence the risk premium is too low. This is what the risk means. Risk is by no means definitely high return; rather, risk means more possibility of loss. This is the very reason for industries to be different in attractiveness, and also the primary motivation for capitals to transfer across industries.

For instance, a firm needs to choose one between two exclusive projects. Project A is in the traditional industry that the firm is currently operating in; project B is in a high growth industry. The currently average returns of the traditional industry and high growth industry are 5% and 30% respectively. The estimated returns on project A and B are 6% and 28% respectively. If the firm uses 5% and 30% as the discount rate respectively for project A and B, the decision is accepting A and rejecting B. Is this a right choice?

Generally speaking, accepting A and rejecting B implies that the firm chooses a 6% expected return at the cost of a 28% expected return. This is obviously a bad choice. If the risk in the high growth industry is not significantly higher than the traditional industry, accepting A and rejecting B is obviously a bad decision. Therefore, to make a right choice, the firm needs compare the returns of 6% and 28% with the risk-matched (or risk-adjusted) discount rates rather than the average rate of returns respectively in the two industries.

2.2 The Opportunity Cost of Capital

The opportunity cost of capital is one of the most used methods to estimate the discount rate without caring about the uncertainty of the opportunity cost itself.

As a concept from economics, opportunity cost is created for decision-making. Economics tells us that we need to choose because of the scarcity of resources (such as assets). When we make a choice (i.e. make a decision), we forgo other alternatives to use the asset. The opportunity cost is the return of the best one among all the forgone alternatives. Therefore, if the opportunity cost is lower than the return of the chosen alternative, the decision (choice) is right; otherwise, the decision (choice) is wrong.

Believe it or not, the opportunity cost illustrates perfectly the important features of economics (as indicated in previous chapters): decision-oriented but not necessarily feasible. For most resources, such as capitals, there are numerous or even unlimited alternatives to use (invest). As discussed previously, the decisions or choices in practice are usually made within a tight deadline, or even made with a greater randomness. This implies the best one among all the forgone alternatives is too difficult to find, because there are numerous forgone alternatives to invest by the capitals in terms of the nations, areas, industries, sectors, projects, sizes, technologies, models, assets, partners as well as the mixes of these factors.

In fact, opportunity cost may be more useful for explanation after the decision rather than for decision—making in advance. As for determining the opportunity cost of capital, different analysts or decision makers have different views about the potential opportunities for the capital; hence will inevitably work out different opportunity costs of a certain capital. In such a way, which opportunity cost of capital is right to be the discount rate? A question is: why do we determine the discount rate based on the returns of other forgone projects which are nothing to do with the risk of the invested project?

For further understanding the potential problem in taking opportunity cost as discount rate, consider a typical scenario in practical investment decision making. Suppose a company intends to choose one or two best projects from the ten potential projects to implement investments. Those projects are different from each other in terms of size or the amount of initial investment and the risk of future returns.

For comparing those projects, the initial amount of capital needed for investment and the future annual returns (earnings) are estimated for every project. The IRR (internal rate of return) then can be worked out based on the capital amount and annual returns for each project. Further, the projects can be evaluated in turn according to the NPV (net present value) rule, which is the best way in project evaluation or capital budgeting. The NPV of a project is the difference of the sum of all its future returns discounted at a discount rate subtracts its initial investment. As the discount rate is the opportunity cost of the capital invested, which should be the IRR of the best project among the rejected projects, obviously, only one project finally has a positive NPV, which is the project with highest IRR.

A small project may have very high IRR, such as 80% or something like that. This implies that the annual returns of other projects should all been discounted at a discount rate of 80%, while as the annual returns of this small projects are discounted at a discount rate lower than 80%. As 80% is higher than all the IRRs of other projects, the NPVs of other projects cannot be positive except the small project. Based on the NPV rule, undoubtedly, the small project will be chosen finally because it is the only feasible project.

A key question here is: is this a right selection or right decision?

Obviously, this is not a right decision because other projects may have much larger NPVs if the discount rate is appropriate rather than this 80%. In another word, the selection of this small project may result in obtaining a small NPV at a cost of losing or rejecting a larger NPV.

Anyway, this may not be a necessary situation. In practice, the potential project group may consist of more similar projects. Suppose the potential project group consists of only projects of same size, which means the ten projects need same amount of capital. Then as the previous process, the annual returns and IRRs of every projects are projected or calculated. Then, based on the IRR of the best project among the other nine projects, the NPV of each project is worked out in turn. Needless to say, still only one project has a positive NPV, and other projects should be rejected. The selected project is the one with highest IRR.

A key question here still is: is this a right selection or right decision?

As a common knowledge, IRR represents the expected annual rate of return of a project; and normally, high (rate of) return is corresponding to high risk. So this selection or decision implies that a project with highest risk is chosen. The higher the risk the better? Of course not! Consequently, this is not a right selection or right decision. But why the project with highest risk is chosen? Obviously, the reason is that the opportunity cost of the capital is used as discount rate for the calculation of NPVs!

In sum, we have tested the opportunity cost of the capital under two scenarios: the potential projects are different in size and risk under the first scenario, and the potential projects different in risk but same in size under the second scenario. Under both scenarios, based on the opportunity cost of the capital as a discount rate, the final selection or decision is not right. In another word, the opportunity cost of the capital is not qualified as a discount rate under both scenarios.

Let's test it further under the third scenario: no difference in both size and risk among all the potential projects. Please note that this is the only scenario left.

Under this scenario, similarly, the project with highest IRR will be selected. But the difference is that this is the right selection or right decision this time. In another word, the opportunity cost of the capital seems qualified as a discount rate to support the NPV calculation and capital budgeting under this scenario. However, this seems not a good news for certain. To choose the best project among ten or some alternative projects under this scenario, it is not necessary to calculate NPVs. The best project can be selected out just based directly on the known IRRs. Because now, all the alternative projects have same size and risk, the higher IRR means definitely the larger absolute return! So, it does not need to bother the NPV calculation; the discount rate is not necessary either. Further, the opportunity cost as a concept or as a variable is not necessary either!

Therefore, under all scenarios, the opportunity cost of the capital as a discount rate is either wrong or unnecessary! So, it is neither feasible nor reasonable to determine the discount rate based on the opportunity cost of the capital. Put it another way, the opportunity cost of the capital is not qualified as a discount rate.

Such a conclusion may go beyond the original expectations of most (if not all) people. Why do we reach such a conclusion? It is actually an inevitable conclusion.

The emerging of finance as a subject and its independence from economics are neglected by most nowadays scholars. However, this is a mile stone in the history of economics. The emerging and independence of finance is vital for understanding of finance and economics.

As a matter of fact, modern finance was emerging and independent from economics during last 50s, marked by Markowitz' portfolio theory [1]. Why the portfolio theory? Because this theory provides an exhibition of quantitatively trading-off between risk and return for the first time. Quantitatively trading-off between risk and return is a unique characteristic of finance. This is in line with the object of research in finance.

Finance focuses on the exchange or transaction of assets, while economics focuses on the exchange or transaction of commodities. As a consequence, it is vital in finance to value assets, while it is vital in economics to value commodities. The value of assets is determined by its risk and (expected) return; naturally, finance tries to quantitatively trading-off between risk and return. However, risk and (expected) return are not salient features for commodities. Commodities cannot be valued based on their features in risk and (expected) return. It seems commodities valuation is much more difficult than assets valuation.

It seems that the utility determines the commodity value. But utility as a determinant factor is too hard to measure or quantify, so do the supply and demand based on utility; needless to say their quantitative relation with the value of the commodity. Therefore, valuation feasibility is a fundamental problem in economics. But asset valuation based on the two determinants of risk and return is more feasible than commodity valuation. On one hand, the risk and return of an asset are easy to measure than the utility of a commodity; on the other hand, finance has found the effective method to incorporate the risk and return into value of an asset, which is the discounting method or DCF method (discounting cash flows).

Actually, asset is also a kind of commodity. The emerging and independence of finance as a subject thus implies the study (valuation) on a part of commodity (asset) can step further ahead of other parts. Since such a relation between finance and economics, it cannot be expected the difficult problem in finance to be solved by traditional methods in economics. The estimation of discount rate is just a difficult problem in finance, opportunity cost is a traditional method in economics. So the above conclusion, the opportunity cost of the capital is not qualified as discount rate, although contradictory to prevailing orthodoxies and practices, is actually naturally right and for sure. Further, other methods in economics are not possible to solve the problems concerning discount rate either.

2.3 The Cost or Average Cost of Capital

Most (if not all) financial, investment and valuation books as well as related research papers confuse the cost of capital with the discount rate and often adopt the weighted average cost of capital (WACC) as the discount rate for valuing assets and project. Actually, capital cost and discount rate are different from each other. Capital cost is the result of financing (decision), and may not be directly related to the asset risk; but asset risk is the main reason of the discounting calculation in investment decision and hence the discount rate.

Firms raise their debt and equity capitals by issuing securities (such as bonds and stocks). Generally speaking, the transaction of securities in market is a zero-sum game, which means that the gain of one party is just the loss of its counterparty. From a point of view, neglecting market intermediaries such as lawyers and investment bankers, the issuer and investor of securities are counterparties to each other. Thus the gain of investor is just the loss of the issuer, i.e. the return on investment is just the capital cost for financing.

Both the capital gain of the investor and the capital cost of the security issuer (fund raiser) are determined by the issuing price of the security. As a common knowledge, the issuing price is possible to be right or not right (in another word, the security is possible to be undervalued or overvalued in the issuing market), which implies that the cost of capital is possible to be either fair or unfair.

Therefore, the cost of capital first of all is a concept referring to the actual capital cost by issuing securities at the market price. More often than not, the capital cost cannot reflect its risk correctly for sure because of the uncertainty of the market price. Such a capital cost is fundamentally different from discount rate. The discount rate used in discounting is at least aimed at matching with the risk of the asset or reflecting the relevant risk correctly.

Capital cost is qualified as discount rate only under the assumption that the relevant asset or security is fairly valued in market. On the other hand, valuation or discounting aims at working out a true or fair value of the asset under the assumption that the asset is normally misvalued in market. The contradiction of the two assumptions implies that using costs of capital as discount rates to value securities or assets is a cyclic calculation and makes no sense. Obviously, only valuation based on true discount rate can find the true value of a security hence can find whether or not it is fair-valued in the market.

Some people may argue that the capital cost here means the correct or fair capital cost. This is not a good excuse. Generally speaking, every word or phrase has its original and basic meaning. As the original and basic meaning, capital cost is the actual rather than fair or correct capital cost, whereas fair or correct capital cost is a synonym of discount rate. Anyway, the capital cost derived from the market price of a security is an actual capital cost, and is more often unfair or incorrect, and is not qualified to be a discount rate.

Fundamentally speaking, discount rate is a concept for decision-making, with a characteristic of forward looking. Such kind of concept should pursue a goal or standard of rightness or reasonableness. For example, in order to fairly value a company, the discount rate used must be correct (not necessarily equal to the actual cost of capital). In contrast, the cost of capital is a concept for description, with a characteristic of backward looking. Such kind of concept should pursue a goal or standard of accordance to the actual situation. For example, if a company issued a bond with mains terms as: face value 1000 dollars, maturity 5 years, coupon rate 10%, frequency of interest payment once a year. Suppose the issuing price of the bond is 1500 dollars (for unknown reasons). What is the capital cost of the capital raised from the bond issuance? Obviously, it is 0%, which is the same as the yield to maturity of the bond. However, when the company invest those capital into a project, should it use

0% as the discount rate? Of course not. Such extreme or obvious case may not be so common in reality, but anyway, the fundamental difference of capital cost and discount rate in concept determines that they are not likely to be equal in quantity.

The reason to discount the future cash flows is that the future returns have two drawbacks: one is that they are later in time than the initial capital expenditure; the other is that they are uncertain in size or risky. The drawbacks implies that the investors need to be compensated for their investment: one is the compensation for the deferment of returns, which can be referred to as time premium; the other is the compensation for the risk of returns, which can be referred to as risk premium. Therefore, the discount rate is the sum of the time premium and the risk premium. As future cash flows are definitely later in time and uncertain in size, both the time premium and the risk premium are positive. Hence the discount rate is definitely positive, or cannot be zero or negative in theory.

As to the capital cost, no matter equity capital or debt capital, can be negative in theory, although it is not common in reality. Consider the above corporate bond, how much is the capital cost of this bond if its issuing price is 1800 dollars? Obviously, the capital cost of this bond can only be negative in such a case. Understandingly, it is more possible for an equity or stock (than for a bond or a debt capital) to issue at an over high price, which implies the capital cost of equity may be more likely to be negative.

Most finance-related books and research papers state that the cost of equity is higher than the cost of debt. This is a similar mistake resulted from the concept chaos about capital cost and discount rate. As the result of financing, equity cost is not necessarily higher than debt cost. There are multiple factors influencing the issuing prices of both equity and debt. When the equity is issuing at a very high price, its cost is obviously possible to be lower than the debt cost or even can be negative. This is the primary reason for listed firms in China or other emerging market to prefer equity financing rather than debt financing. Anyway, capital cost depends on the related financing decisions, or the securities' issuing prices.

Therefore, the cost of capital can be positive or negative, whereas the discount rate can only be positive; the discount rate for equity investment should be higher than that for debt investment; whereas the equity cost can be higher or lower than the debt cost depending on the issuing price of the relevant equity and debt. Thus, after revising the confused concepts about the capital cost and discount rate, it is clear that capital cost has no certain relationship with the asset risk; it is not theoretically sound to determine the discount rate based on capital cost; no matter it is opportunity cost, equity cost or weighted average cost.

Although finance is still on its initial stage, it is beneficial for both investment decision and financing decision by distinguishing capital cost and discount rate. On the one hand, financing decision is also an important and independent decision in most firms. Other things being equal, good financing decisions will lower firms' capital costs. The capital cost hence is an important indicator for firms to evaluate their financing decisions. Otherwise, if a financing decision is evaluated based on the discount rate, the evaluation can make no sense because the discount rate has nothing to do with the efforts in financing decision.

On the other hand, misusing capital cost as discount rate may lead to mistakes in investment decision. For instance, firm A and firm B are seeking good projects. They find project X and have the same judgement on the risk and return of it. The annual estimated returns of project X is 30% for both A and B. However, the capital cost is 10% in firm A, and is 30% in firm B as a result of some financing mistakes. Now, how can A and B make their decisions on project X? If they misuse their capital cost as the discount rate (guided by the prevailing textbooks), A will accept project X; and B will reject it.

There is a logic dilemma here, because either A or B is wrong. Just think what will happen after the above decisions. After that, A will invest in X and operate; B will go on to seek better projects. There are obviously two possibilities. One is that B succeeds to find a better project. If it is true, the decision of A is wrong. The other possibility is that B fails to find a better project. In this case, the decision of B is wrong; because during the period in which A runs X and gets annual returns of 30%, B can only get a much lower returns on its unused capital (for instance just deposit the capital in bank and gets some poor interest) although the cost of this capital is 30%. Obviously, the logic dilemma is coming from the wrong choice of the discount rate.

Therefore, the actual cost of capital cannot be used as a discount rate. Moreover, at any decision-making point, the company has only one cost of equity capital, debt capital and WACC, and there are usually multiple investment projects waiting to be considered; even just one project can be selected finally, it needs to be compared and selected among multiple projects. This single cost of capital can't match the average risk of these projects, let alone the risk of each project. How can it be used as the discount rate for evaluating each project? Actually, capital cost as the result of financing is irrelevant to investment decision. It is wrong to incorporate irrelevant cost into consideration for decision-makings.

Furthermore, financing decisions determine the cost of capital, but not the discount rate. For example, a company made serious mistakes in the capital raising process, resulting in the cost of capital as high as 50%. Now the best project among all the potential projects under consideration in their capital budgeting has the expected rate of return of 30%. If the company use the capital cost 50% as discount rate, there will be no project worthy of investment in their capital budgeting and their capital will have to wait idle. How long will the capital remain waiting? If they insist on using 50% as the discount rate, they will have no investment opportunity until a project with the expected rate of return exceeding 50% appears. Just imagine, in the process of waiting, other companies continue to have new projects launched, with profits as high as 30%, 40%, 50%, etc., while this company has been waiting because of the high cost of capital. Is this a right decision?

Therefore, using WACC (weighted average cost of capital) and costs of other capitals (equity and debt) to represent discount rate is not correct in concept although it has been prevailing widely for quite a long time. It is the time to correct the wrong usage of both the cost of capital and discount rate as financial terms in research paper and textbooks.

In sum, none of the above alternatives in determining discount rate is right in theory, though they are widely used in practice and widely spread in textbooks. Those alternatives are even wrong in concepts. As a matter of fact, those alternatives are non-professional methods, but they played an important role in a long history because modern finance as a subject did not come up until last 50s. The professional methods came up following the born of modern finance. This is the main topic of the following part.

3 The Capital Asset Pricing Model

Anyway, discount rate is a hot topic in finance. There are a lot of literature on this topic. The most well-known is Sharpe CAPM, capital asset pricing model, which is originated from Markowitz's Portfolio theory. We take a look at both of them in this section.

3.1 Portfolio Theory

As mentioned earlier, the foundation of finance is an axiom: risk and return determine (asset) value; the symbol of modern finance is Markowitz's Portfolio theory. Why? Because finance is devoted to determine (asset) value based on the trade-off between risk and return, while Markowitz's Portfolio theory introduces firstly the theory and technique about how to trade-off quantitatively between risk and return. Portfolio theory thus represents the emerging of finance as a subject of science or theory; trade-off quantitatively between risk and return thus becomes a unique and salient feature of finance.

3.1.1 Basics of Trade-Off Between Risk and Return

In the plane coordinate system of Fig. 1, the horizontal axis shows the risk in standard deviation of annual rate of return or volatility; the vertical axis shows the return in annual rate of return. Each point in the space of the first quadrant represents an investment opportunity or a stock in the market. Now, how will you make a decision or selection?

Obviously, the points represented by capital letters A, B, C, D, E and F are relative batter than those points represented by small letter and numbers. For instance, point C is better than point c1, c2, etc. because they have same expected return but point C has lowest risk. Point b5 has relative higher return and lower risk than point a6, hence is better than a6; similarly, c4 is better than b5, d3 is better than c4, e2 is better than d3, f1 is better than e2, F is better than f1. We cannot find a point better than F

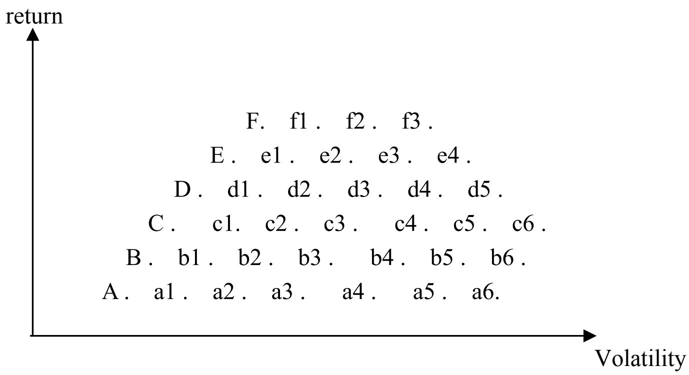


Fig. 1 Security selection based on the trade-off between risk and return

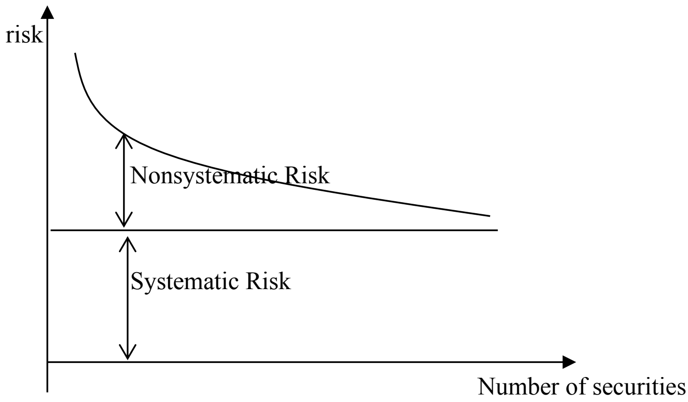


Fig. 2 The systematic risk and non-systematic risk

in the space because there is no point northwest to point F. For the same reason, we cannot find points better than A, B, C, D and E.

Normally, in an investment space, for every level of return, there is one security that offers the lowest possible risk, and for every level of risk, there is a security that offers the highest return. These best securities combined as a line, which is referred to as the efficient frontier. As the line of ABCDEF shown in Fig. 1. Understandingly, analysts try their best to find a portfolio on the efficient frontier; investors try their best to reach this efficient frontier.

Now, a question is: because the securities A, B, C, D, E and F are relative batter or best in the market for their risk and return features; investors should buy them rather than other securities represented by small letters and numbers in Fig. 1. Will such a selection rule out those securities southeast to the investment frontier?

Do not worry, this will not happen. The risk and return features of those securities are in the dynamic process of changing. During the trading time, in which investors can make investment or selection decisions, the position of every security is changing, or is moving towards different direction at different speed. For instance, the price of security a6 may go down rapidly resulted from the asymmetric force of buying and selling in the market. Its expected return thus goes up under an assumption that its expected future price keeps unchanged and hence its position will move towards northwest. Similarly, the securities now on the frontier may move southeast because their return will move down along with their prices are pushed higher and higher by the buyers.

3.1.2 The Risk and Return of a Portfolio

Portfolio refers to the combination of two or more assets / securities. A broad concept of portfolio also includes a single security or asset. Companies usually invest in a variety of assets and businesses, so they can also be regarded as portfolios.

How about the return and risk of a portfolio, or what is the relationship among the risk of a portfolio and the risks of the component securities? Let us start from a relatively simple case, a portfolio with just two component securities.

(A) **The portfolio with two component securities**

It is relatively easy to consider the return side. Obviously, the return of a portfolio is the weighted average of the yield of each component security.

$$\mathrm{E(rp)} = \sum \mathrm{wi} \cdot \mathrm{ri} \quad (\mathrm{i} = 1, 2, \ldots, \mathrm{n})$$

where, E(rp) = expected return of the portfolio; wi = the weight of security i in total investment; ri = expected return rate of securities i. n = the number of component securities consisted in the portfolio.

The risk of a portfolio is not so easy to work out as measured by the standard deviation of return because the deviations of two or more securities can partly canceled out each other if they are put into a portfolio. Take a portfolio with two securities as an example, the variance (square of volatility or standard deviation) of such a portfolio is not weighted average of the variances of the two securities, but is derived as following:

$$\begin{aligned}
&\sigma_P^2 = \mathrm{E}[\mathrm{R_P} - \mathrm{E}(\mathrm{R_P})]^2 \\
&= \mathrm{E}\{\mathrm{w_1R_1} + \mathrm{w_2R_2} - [\mathrm{w_1E(R_1)} + \mathrm{w_2E(R_2)}]\}^2 \\
&= \mathrm{E}\{\mathrm{w_1}[\mathrm{R_1} - \mathrm{E(R_1)}] + \mathrm{w_2}[\mathrm{R_2} - \mathrm{E(R_2)}]\}^2 \\
&= \mathrm{w_1^2E}[\mathrm{R_1} - \mathrm{E(R_1)}]^2 + \mathrm{w_2^2E}[\mathrm{R_2} - \mathrm{E(R_2)}]^2 \\
&+ 2\mathrm{w_1w_2E}[\mathrm{R_1} - \mathrm{E(R_1)}][\mathrm{R_2} - \mathrm{E(R_2)}] \\
&\text{i.e.,} \\
&\sigma_P^2 = \mathrm{w_1^2}\sigma_1^2 + \mathrm{w_2^2}\sigma_2^2 + 2\mathrm{w_1w_2}\sigma_{1,2}
\end{aligned}$$

where,

σ_P^2 = the variance of a portfolio; R_P = the return of a portfolio; $E(R_P)$ = the expected return of a portfolio; w_1 = the weight of security 1; R_1 = the return of security 1; w_2 = the weight of security 2; R_2 = the return of security 2; $E(R_1)$ = the expected return of security 1; $E(R_2)$ = the expected return of security 2. Further, $\sigma_1^2 = E[R_1\text{-}E(R_1)]^2$ = the variance of security 1; $\sigma_2^2 = E[R_2\text{-}E(R_2)]^2$ = the variance of security 2; $\sigma_{1,2} = E[R_1\text{-}E(R_1)][R_2\text{-}E(R_2)]$ = the covariance of security 1 and security 2.

The above equation can also be expressed by correlation coefficient, ρ.

$$\rho_{1,2} = \sigma_{1,2}/(\sigma_1\sigma_2)$$

ρ ranges from − 1 to + 1. where, $0 < \rho \leq 1$: positive correlation; $\rho = 1$: perfect positive correlation; $-1 \leq \rho < 0$: negative correlation; $\rho = -1$: perfect negative correlation; $\rho = 0$: no correlation. i.e.,

$$\begin{aligned}\sigma_P^2 &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{1,2} \\ &= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2(\sigma_1\sigma_2)\rho_{1,2}\end{aligned}$$

Risk of a portfolio of two securities in terms of the standard deviation is:

$$\begin{aligned}\sigma_P &= \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{1,2}} \\ &= \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2(\sigma_1\sigma_2)\rho_{1,2}}\end{aligned}$$

For some special values of ρ, the formula can be reduced.

(1) when $\rho_{1,2} = 1$, perfect positive correlation:

$$\sigma_P = w_1\sigma_1 + w_2\sigma_2$$

(2) when $\rho_{1,2} = 0$, no correlation:

$$\sigma_P = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2}$$

(3) when $\rho_{1,2} = -1$, perfect negative correlation:

$$\sigma_P = |w_1\sigma_1 - w_2\sigma_2|$$

Obviously, because of the asset correlations, the total portfolio risk, or standard deviation, most of the case, except the case where the correlation coefficient $\rho = +1$, is lower than what would be calculated by a weighted sum of the (two) component securities in the portfolio.

For example, assume that the expected rate of return of securities A and B, Era = 22%, Erb = 16%, the risk of securities A and B, σa = 0.21, σb = 0.15, and the returns of the two securities are uncorrelated, i.e., the correlation coefficient of A and B is zero. If the investor is willing to accept a 15% risk (the standard deviation of portfolio is 15%), what would you suggest? How would you change if the investor wants to minimize the risk?

Since the two securities are not correlated in return, they can diversify risk of each other if they are put into a portfolio. Investing only in security B has the risk of 15% in terms of volatility or standard deviation. But this is not the best choice. Alternatively, the investor can invest a portfolio with security A account for 68.57% and security B account for 32.43%, then the standard deviation of the portfolio is 15%, but the expected return of the portfolio is 20.1%, which is higher than the expected return of the security B, 16%.

If the investor wants to minimize the risk, our suggestion is: invest a portfolio with security A account for 33.74% and security B account for 66.26%, then the standard deviation of the portfolio is 12.21%, which is lower than the risk of security B; but the expected return of the portfolio is 18.0%, higher than the expected return of the security B, 16%.

(B) **The portfolio with n component securities**

For a portfolio with 2 securities,

Since $\sigma_{1,1}^2 = \sigma_1^2, \sigma_{2,2}^2 = \sigma_2^2$, then,

$$\sigma_P^2 = \mathrm{w}_1^2\sigma_1^2 + \mathrm{w}_2^2\sigma_2^2 + 2\mathrm{w}_1\mathrm{w}_2\sigma_{1,2} = \sum_{i=1}^{2}\sum_{j=1}^{2} w_i w_j \sigma_{i,j}$$

Then, for a portfolio with n securities,

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{i,j} = \sum_{i=1}^{n} w_i^2\sigma_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\(j\neq i)}}^{n} w_i w_j \sigma_{i,j}$$
$$= \sum_{i=1}^{n} w_i^2\sigma_i^2 + 2\sum_{i=1}^{n}\sum_{\substack{j=1\\(j>i)}}^{n} w_i w_j \sigma_{i,j}$$

In the case of the component securities uncorrelated to each other, the variance of a portfolio of two securities is:

$$\sigma_P^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2.$$

For a portfolio with n securities, suppose they are uncorrelated to each other and each having variance σs, the variance of the portfolio is:

$$\sigma_P^2 = \left(\frac{1}{n}\right)^2 \sigma_s^2 + \left(\frac{1}{n}\right)^2 \sigma_s^2 + \cdots + \left(\frac{1}{n}\right)^2 \sigma_s^2$$
$$= \frac{n}{n^2}\sigma_s^2 = \frac{\sigma_s^2}{n}$$

The standard deviation of the portfolio is:

$$\sigma_{\mathrm{p}} = \frac{\sigma_s}{\sqrt{n}}$$

Obviously, the risk of the portfolio in terms of the variance or the standard deviation is decreasing along with the increasing of the component securities. It seems the risk of the portfolio in terms of the variance or the standard deviation will go towards zero. But this is not the case when the component securities correlated to each other.

3.1.3 The Systematic Risk and Non-systematic Risk

For the common case that the component securities correlated to each other, suppose the variance of each component security equals to their average variance, i.e., σ_i^2 = var; the covariance between each two of the component securities equals to their average covariance, i.e., $\sigma_{i,j}$ = cov > 0; assume all weights are (1/n).

Risk of the portfolio in term of variance when n goes to infinity is,

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{i,j} = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1\\(j\neq i)}}^{n} w_i w_j \sigma_{i,j}$$
$$= \mathrm{n} \times (1/\mathrm{n})^2 \times \mathrm{Var} + \mathrm{n} \times (\mathrm{n}-1) \times (1/\mathrm{n})^2 \times \mathrm{Cov}$$
$$= (1/\mathrm{n}) \times \mathrm{Var} + \left(\mathrm{n}^2 - \mathrm{n}\right) \times \left(1/\mathrm{n}^2\right) \times \mathrm{Cov}$$
$$= (1/\mathrm{n}) \times \mathrm{Var} + (1 - 1/\mathrm{n}) \times \mathrm{Cov}$$
$$= \mathrm{Cov}\ (\text{when n goes to infinity})$$

Surprisingly, for a well-diversified portfolio, the variance of each security disappeared; only the average covariance, cov, remained, which does not equal to zero. Therefore, when the component securities (positively) correlated to each other, you cannot diversify out all the risks by add more component securities in the portfolio. Put it another way, there is still some risk remained in the portfolio even you put all the securities in the market into your portfolio.

The risk remained after adding all the securities into the portfolio or that cannot be eliminated through diversification is called systematic risk or non reducible risk. It is closely related to the national and global political and economic situation and comes from the uncertainties faced by most companies in a country or all over the world, such as inflation, economic depression, adjustment of interest rate and government regulation policies, etc.

The risk that can be eliminated by diversification is called nonsystematic risk or reducible risk. It comes from the specific situation of the company, including its unique strategic choice, product marketing, operation plan and internal control, etc. Companies with different businesses and factors will have good or bad deviation from the general situation of the overall economy, which is usually the risk that can be eliminated by diversified investment.

The securities here only refer to those traditional or ordinary securities, such as stocks, bonds, foreign exchange, etc., excluding the derivatives, such as futures and options. There are normally positive correlations among the returns of the ordinary securities. If derivatives like futures and options are allowed to put into the portfolio, with derivative position perfectly negative correlated with the underlying securities, theoretically there is no risk that cannot be eliminated, hence no systematic risk remained.

3.2 Sharpe's Capital Asset Pricing Model

Markowitz's Portfolio theory reveals that adding more securities in a portfolio can reduce the risk of the portfolio while increase the return of it. This appeal to William Sharpe, who is a Ph.D. candidate at the very first. William Sharpe was so lucky that he got a chance soon to work together with Markowitz at Rand Corporation and could consulting directly with Markowitz about the Portfolio theory. Sharpe thus immersed himself in his intensive study on portfolio theory and the related issues, and found his capital asset pricing model (CAPM) finally.

3.2.1 The Tangency Portfolio

According to the portfolio theory, if you bring all the risky securities into your portfolio (with the weights as their proportion in the market), you can eliminate all the nonsystematic risks, and push the efficient frontier towards northwest at utmost.

As shown in Fig. 3. Without diversification, the original efficient frontier is ABCDEF; Portfolio Theory states that adding different assets to a portfolio can diversify risk without sacrificing corresponding return. Therefore, the efficient frontier can be pushed northwestern when more securities were added in the portfolio. The dash line represents the most northwest position of the efficient frontier to be pushed. Theoretically, the point on this best efficient frontier consists all the risky securities in the market.

Amazingly, most of the points on this best efficient frontier can be further improved by adding a risk free security into the portfolio.

This risk free security is usually the central government bond. As mentioned before, this is a default free security. If the yield to maturity of the government bond

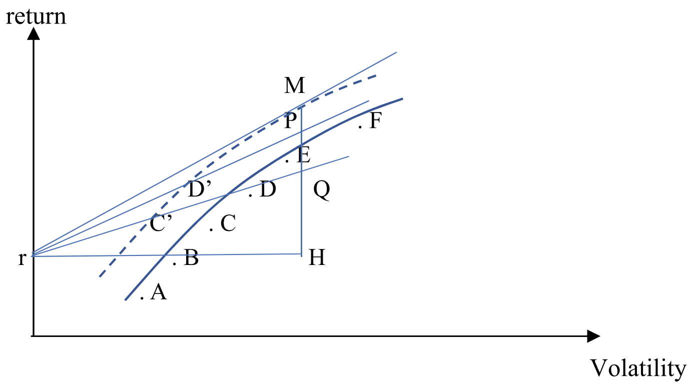


Fig. 3 Security selection based on the trade-off between risk and return

is r, and you buy it today and hold it until its maturity, you can get an annual return of r for sure, because the central government can definitely pay the interest and principal of the bond.

This means the risk of the risk free bond can be regarded as zero, i.e., its volatility (standard deviation) or variance is zero. In the plane (Cartesian) coordinate system consists of return and volatility like Fig. 3, based on the risk and return characteristics, the point of the risk free bond is obviously on the vertical axis at a height of r.

Further, the return of this bond is uncorrelated to the return of the risky securities in the market as well as their portfolio. Put it another way, the correlation coefficient between the government bond and any point on the dash line in Fig. 3 is zero.

Imagining a portfolio consisting of the risk free bond and any point on the dash line with weights wr and wt respectively. Let σ be the volatility of any point or portfolio on the dash line, and rp, σp be the expected return and volatility of the portfolio, then,

$$\mathrm{rp} = \mathrm{r} \times \mathrm{wr} + \mathrm{rt} \times \mathrm{wt}$$

$$\sigma \mathrm{p} = \sqrt{0^2 \times wr^2 + \sigma^2 \times wt^2}) = 0 \times \mathrm{wr} + \sigma \times \mathrm{wt}$$

Therefore, the return and volatility of the portfolio are the weighted average return and volatility of the two component securities (the government bond and the risky portfolio) respectively. This implies that the line formed with the portfolio points of the government bond and the risky portfolio is a straight line as rC', rD', rM in Fig. 3.

Those lines are referred to as capital allocation line (CAL). Since there are countless portfolio points on the efficient frontier, there are countless CALs. It is easy to judge that among those CALs, rM is the best one. The points on rM is better than that on rC' and rD' because the return is higher at the same risk level. The points on rM are indifferent to each other. Point M is the tangent point at which the straight line rM is tangent with the efficient frontier. The portfolio represented by M is referred to as tangency portfolio.

3.2.2 The Sharpe Ratio

In Fig. 3, the CALs intersect with a downwards straight (vertical) line MH from the tangent point M;

the M is the highest intersections among those intersections. Suppose M_1, M_2, …, M_n represent those intersections, then, $MH > M_1H > M_2H > \ldots, > M_nH$.

The Sharpe ratio describes how much excess return will be expected for each unit of risk (measured by the volatility or standard deviation) borne by a portfolio. That is,

$$\text{Sharpe ratio} = \left(r_p - r\right)/\sigma_p$$

where, r_p is the (expected) return of portfolio; r is the risk-free rate; (r_p-r) then is the portfolio's excess return; σ_p is the standard deviation of portfolio's return.

Apparently, the Sharpe ratio is the slope of the CALs. i.e., MH/rH, M_1H/rH, M_2H/rH, …, M_nH/rH, etc. The higher the Sharpe ratio, the higher the CALs, the more attractive the portfolio. One way to enlarge the Sharpe ratio in practice is to improving the diversification of the portfolio because diversification can push the efficient frontier move towards northwest. In Fig. 3, the best CAL is the straight line tangent to the most efficient frontier at a point M, which is also referred as capital market line (CML).

The Sharpe ratio was developed by Nobel laureate William F. Sharpe. It can be used as an ex-ante tool in investment decision making where the expected returns are put into the calculation to work out the estimated Sharpe ratios and the best portfolio then can be found based on those ratios. Alternatively, it can also be used in the performance evaluation of portfolio management as an ex-post tool where actual returns are put in the formula.

For example, an investor is considering adding some slightly negative related stocks to their existing portfolio which has a return of 20% last year. The current risk-free rate is 3.3%, and the volatility of the portfolio's returns was 18%, the Sharpe ratio is:

$$(20\% - 3.3\%)/18\% = 92.8\%$$

Based on a careful calculation, adding these stocks to the portfolio will lower the expected return to 15% in the coming year; the portfolio's volatility is also expected to drop to 10%. Assume the risk-free rate will not change over the coming year.

Then, the expected Sharpe ratio changes to:

$$(15\% - 3.3\%)/10\% = 117\%$$

Therefore, based on the Sharpe ratio, adding those stocks can improve the portfolio.

The CML represent the best or fair relationship between risk and return under the rational choice of investors. This fair relationship may not be always held for various reasons in various scenarios. As in the cartesian coordinate system in Fig. 3, investors possess greater return relative to risk when their portfolios fall above the CML; whereas they possess lower return relative to risk when their portfolios fall below the CML. Therefore, smart investors should seek portfolios above the CML and avoid portfolios below the CML.

3.2.3 Sharpe CAPM

As previous analyses, the tangency portfolio is the best portfolio as well as the most diversified portfolio, which also referred as market portfolio. It consists of all the securities in the market, including all the risky securities and the risk free securities. Hence all the nonsystematic risks are eliminated and the portfolio only bears systematic risk.

About 60 years ago, based on such a portfolio, William Sharpe finds a model to describe the theoretical or fair relationship between return and risk.[1] Such a model, conceptually or theoretically, can be used to determine the investors' required rate of return or risk adjusted discount rate based on the asset risk. Sharpe's model takes the form:

$$E(R_i) = r + \beta_i[E(R_m) - r] \tag{1}$$

where:

$E(R_i)$	is the expected return or required rate of return on the ith capital asset;
r	is the risk free rate of interest such as interest from government bonds;
β_i	(the beta coefficient) is the sensitivity of the asset returns to market returns;
$E(R_m)$	is the expected return of the market;
$E(R_m)–r$	is known as the average market risk premium.

Sharpe's model is referred to as capital asset pricing model (CAPM). The CAPM states that the appropriate rate of return on an asset is the sum of a risk free rate and a risk premium. The risk free rate represents the compensation to the deferment of the return; and the risk premium represents the compensation to the risk borne by the investment (asset). The structure of the CAPM as a discount rate model seems sound enough.

Further, the risk premium in CAPM is the product of two factors. One is the average risk premium of the market, i.e. $E(R_m)–R_f$; the second is the Beta coefficient of the asset, i.e. β_i. As the market risk premium is given, the Beta coefficient (or

[1] Sharpe [2]. Treynor [3], Lintner [4], Mossin [5] also found the model independently; Sharpe found the CAPM around the end of 1962 and published it in 1964.

simply referred to as Beta) measures the asset risk. Hence Beta has been a new risk indicator since Sharpe CAPM.

Depending on Sharpe CAPM, people realize the ideal to determine the discount rate based on the risk of the relevant asset or investment rather than those irrelevant factors like capital cost, opportunity cost (the return of other investment opportunities), etc.

The model is proved as following.

Consider a portfolio consists of a risky security i and the market portfolio. Suppose the weight of the security i are a (<1); then the weight of the market portfolio is (1-a).

Using E(Rp), E(Ri), E(Rm) to represent the expected return of the portfolio, the security i and the market portfolio. The expected return of the portfolio then is:

$$\mathrm{E(Rp)} = \mathrm{a\ E(Ri)} + (1 - \mathrm{a})\mathrm{E(Rm)}$$

Calculate the partial derivative of a on the two sides:

$$\frac{\partial E(Rp)}{\partial a} = \mathrm{E(Ri)} - \mathrm{E(Rm)}$$

The standard deviation of the portfolio is:

$$\sigma_p = \sqrt{\sigma_m^2(1-a)^2 + \sigma_i^2 a^2 + 2a(1-a)\sigma_{m,i}}$$

Calculate the partial derivative of a on the two sides:

$$\frac{\partial \sigma_p}{\partial a} = \frac{-\sigma_m^2 + a\sigma_m^2 + a\sigma_i^2 + \sigma_{m,i} - 2a\sigma_{m,i}}{\sqrt{\sigma_m^2(1-a)^2 + \sigma_i^2 a^2 + 2a(1-a)\sigma_{m,i}}}$$

Then,

$$\frac{\partial E(Rp)/\partial a}{\partial \sigma_p/\partial a} = \frac{[E(Ri) - E(Rm)]\sqrt{\sigma_m^2(1-a)^2 + \sigma_i^2 a^2 + 2a(1-a)\sigma_{m,i}}}{-\sigma_m^2 + a\sigma_m^2 + a\sigma_i^2 + \sigma_{m,i} - 2a\sigma_{m,i}}$$

When a $\to$ 0, (1 $-$ a) $\to$ 1, the above portfolio tends to be the market portfolio,

Therefore, E(Rp) $\to$ E(Rm), $\sigma_p \to \sigma_m$, the above formula becomes:

$$\frac{\partial E(Rm)}{\partial \sigma_m} = \frac{[E(Ri) - E(Rm)]\sigma_m}{\sigma_{m,i} - \sigma_m^2}$$

$\frac{\partial E(Rm)}{\partial \sigma_m}$ is the slope of the capital market line, i.e.,

$$\frac{\partial E(Rm)}{\partial \sigma_m} = \frac{E(Rm) - r}{\sigma_m}$$

Then,

$$\frac{[E(Ri) - E(Rm)]\sigma_m}{\sigma_{m,i} - \sigma_m^2} = \frac{E(Rm) - r}{\sigma_m}$$

Then,

$$E(R_i) - E(R_m) = \frac{[E(Rm) - r]\big[\sigma_{m,i} - \sigma_m^2\big]}{\sigma_m^2}$$

Then,

$$\begin{aligned} E(R_i) &= E(R_m) + \frac{[E(Rm) - r]\big[\sigma_{m,i} - \sigma_m^2\big]}{\sigma_m^2} \\ &= E(R_m) + \frac{\sigma_{m,i}}{\sigma_m^2}[E(R_m) - r] + r - E(R_m) \end{aligned}$$

That is,

$$E(R_i) = r + \frac{\sigma_{m,i}}{\sigma_m^2}[E(R_m) - r]$$

Let

$$\beta_i = \frac{\sigma_{m,i}}{\sigma_m^2},$$

Then,

$$E(R_i) = r + \beta_i[E(R_m) - r]$$

Sharpe CAPM is the first strict logic based model to describe the relationship between risk and return. It is actually the only theoretical model in this respect so far.[2] Apart from Sharpe CAPM, no model describes clearly and certainly about the risk-return relationship. Hence this model is widely used in determining the appropriate discount rate. William Sharpe wins Nobel Prize in Economics in 1990 for this remarkable contribution in finance.

[2] Arbitrage pricing theory (APT) initiated by Stephen Ross [6] also describes the return and risk relation. APT holds that the expected return of an asset can be modeled as a linear function of various macro-economic factors or theoretical market indices, where sensitivity to changes in each factor is represented by a factor-specific beta coefficient. However, as a model based on empirical data, the APT has no certain variables and model form, thus it is not theoretical sound and it is less popular in academic and practical research.

3.3 Defects of Sharpe CAPM

Sharpe's CAPM is an epoch-making breakthrough, which has greatly promoted people's understanding of the relationship between risk and return and hence the discount rate. It is the first close formed model in history to link the discount rate with the risk of investment or corresponding assets. Compared with the previous determination of discount rate by means of capital opportunity cost or actual cost, it represents a great progress in both theory and practice.

However, Sharpe's CAPM is not perfect and does not really solve the problem of determining the discount rate. Most importantly, it derives the model based on the whole market portfolio, simply on the assumption of the nonsystematic risk being eliminated completely by diversification. This is neither in line with the facts, nor with the principle of prudence in decision-making, nor people's expectation for the determination of discount rate for a long time. This is a fly in the ointment and a serious defect in theory as well as in practice.

Put it another way, Sharpe CAPM takes into account only the systematic risk or market risk rather than the total risk or the firm (or project) risk. Based on Sharpe CAPM, the discount rate equal to the risk free rate plus only systematic risk premium rather than the total risk premium. This is equivalent to state, the discount rate $=$ risk free rate $+$ part risk premium. This is obviously not allowed in actual decision making. In most situations when a firm makes financial decisions (such as project or business investment decision), it should take into account both the systematic risk and the non–systematic risk, or the total risk.

Therefore, the discount rate worked out based on the Sharpe CAPM is smaller than what it should be. This does not agree with the rules of thumb in decision making, or the prudential principle, and cannot be acceptable in theory and practice. The Sharpe CAPM hence is not a real or final solution to discount rate estimation, and cannot shoulder the responsibility of considering risks in financial analyses and the relevant decisions making.

The research of capital asset pricing originally aimed at the reasonable or fair rate of return of securities, but financial research after Sharpe CAPM deviated gradually from this original purpose, and turned to describing or explaining the actual rate of return, including the various market anomalies. While most research efforts are competing to add new factors and betas to enhance the interpretation power of the model, CAPM is misunderstood as a formula to predict or explain stock returns or a tool to select stocks.

In fact, the Sharpe CAPM represents the progress of financial theory, enhances the scientificity of discount rate estimation and helps to obtain the correct discount rate. Whether the model meets the actual investment return or the cost of capital is of no great importance, because the actual investment income or financing cost is not necessarily at the fair level since the existence of the stock market bubble and various contingency factors.

Increasing the number of pricing factors and beta by imagination will inevitably lead to the subjective and arbitrary choice of the form of the model and the independent variables factored in. An obvious question is: if one beta is not enough, how many are enough? In particular, how many betas does it need to solve the problem? Unsurprisingly, although the follow-up studies provide tremendous literature and saturation coverage, they make almost no significant improvement in CAPM. There is no discount rate model that can replace Sharpe CAPM so far. Therefore, we will not review the literature in detail.

Anyway, no method so far for determining discount rate is sound in theory, including the Sharpe CAPM, though they are widely used in practice and widely spread in textbooks. As such, it is urgent to find a convincing method or model to determine the discount rate, because valuation is the core function of finance and the discount rate is an inevitable input for the prevailing valuation as well as the various financial analyses and decisions making. Obviously, a theoretically sound method or model should determine the discount rate based on the total risk of the project or the asset under consideration. This is the main topic in the remaining part of this chapter.

4 Certainty Equivalent and Risk Equivalent

The basic indicator of risk is the standard deviation or volatility of asset returns. For considering total risk via discount rate, an effective method or model should be able to answer a question such as: when the standard deviation or volatility of the asset's expected return increases by, say 1%, how much should the discount rate change?

It is just one way to consider risk via discount rate in valuation and finance; another way is to consider risk via the returns (cash flows or earnings) being discounted. Specifically speaking, if we can transform the predicted future return into its certainty equivalent, which is smaller than the original predicted return because of the risk, we then only need to discount the certainty equivalent at the risk free rate. In other words, the (predicted) future certainty equivalent only needs time compensation.

As incorporating total risk into discount rate seems too difficult, we now try firstly to consider risk in the return side, or try to model the certainty equivalent.

4.1 On Certainty Equivalent

The "certainty equivalent" is a positive idea to incorporate risk in valuation and financial decisions. Traditionally, the certainty equivalent is defined as the certain returns as equally attractive to investors as the corresponding predicted uncertain returns.

However, neither academic nor practical efforts so far can provide a convenient and reliable method to work out this certainty equivalent. In practice as well as in many textbooks, the suggested method to obtain the certainty equivalent is the experience–based subjective estimation. Specifically, in order to transfer the predicted value of return into its "certainty equivalent", one should determine first a "certainty equivalent coefficient" (referred to as certainty coefficient hereafter) based on "experience"; then multiply the predicted return by the "certainty coefficient" to get the certainty equivalent of the return.

Understandingly, the certainty coefficient is supposed to vary between 0 and 1, because the certainty equivalent should be smaller than its corresponding predicted or expected value, and it increases as the estimated risk decreases. These are all we know so far about the certainty equivalent and the certainty coefficient. Beyond such basic knowledge, like what are the common influential factors, what are the relations between the certainty coefficient and its influential variables, and so on, are all unknown. Many people even believe that the influential factors of the certainty coefficient vary too much across assets and cases, it is impossible to abstract common key variables and build a general model.

Although the concept of the certainty equivalent is constructive, it cannot play a proper role in finance and valuation without an effective model. In addition, as none of the prevailing methods to determine the risk premium is right in theory, most of the financial calculations are lack of theoretical foundation without the reliable certainty equivalents. Take NPV (net present value) as an example. Being widely regarded as the most perfect method in capital budgeting, if only the systematic risk rather than the total risk can be incorporated in the investment decision, the NPV method is actually just something for show.

There are indeed some (but not many) scholars try to model the certainty equivalent, such as Hennessy and Lapan (2006), Kimball [7], Gollier and Pratt (1996), Becker and Sarin (1987), etc. However, these studies stem from an identical idea: the certainty equivalent of an uncertain future value is the certain value that can bring in the same expected utility without risk. Therefore, the problem of modeling the equivalent is translated into the problem of modeling the utility. These studies thus rely completely on the utility function. Unfortunately, the utility, as a fundamental economic concept, is neither objective nor measurable; there is no possibility to build an objective or reliable model of utility.

Viewing of this, we will not follow their suit. We need to find a new and creative way for modelling the certainty equivalent as well as other alternatives to incorporate the total risk into the discount rates as well as the valuations and financial decisions.

4.2 Risk Equivalent Model

The certainty equivalent is smaller than its corresponding expected value. We define the difference between the expected value and the certainty equivalent as the risk equivalent. Use X to represent the predicted or expected value, d to represent the certainty coefficient, CE to represent the certainty equivalent and RE to represent the risk equivalent, then.

$$\mathrm{RE} = \mathrm{X} - \mathrm{CE} \tag{2}$$

$$\mathrm{CE} = \mathrm{X} - \mathrm{RE} = \mathrm{Xd} \tag{3}$$

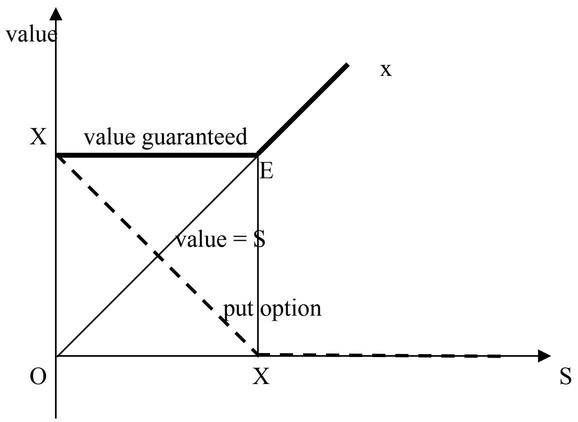
Equation 3 transfers the problem of modeling the certainty equivalent into the problem of modeling the risk equivalent, i.e. the certainty equivalent is its corresponding expected value minus the value reduction caused by the risk. Therefore, the next step is naturally to scrutinize and value the risk. As a general definition, risk is uncertainty. According to the common sense, risk here is mainly refers to the situation that the actual value of a return is lower than its predicted value.

As shown in Fig. 4, S represents the possible values of the predicted return; X represents the special predicted value of the return. Obviously, every point on a straight line x embarking from the zero point and upright rising at 45 degrees always has a scale at the ordinate (vertical) axis equaling its scale at the abscissa (horizontal) axis. Suppose point E on this line is the predicted point with scales at both axes being X. Therefore, the risk is represented by the line segment OE, which is part of the line x below the point E.

In order to eliminate such risk, i.e. to keep the actual point not dropping down along the line x below E when S is smaller than X, we need a guarantee. The guarantee functions (provides value) as the dashed line shown in Fig. 4, which is obviously a put option. Comparing with the risk free situation, the predicted value needs a guarantee or a put. This implies that the risk equivalent equals the value of the put.

Therefore, we can find the risk equivalent by valuing the put. According to Fig. 4, the guarantee is to guarantee the value not lower than the X at the due time. It thus is a standard European put option. The most convenient way to value such a put is using the Black–Scholes option pricing model, as introduced in last chapter:

Fig. 4 Risk equivalent = guarantee = put option



$$p = Xe^{-rT}N(-d2) - SN(-d1) \tag{4}$$

where, S is the current value of the underlying asset; X is the strike price of the put; T is the option maturity time; r is the risk-free discount rate, which can be determined based on the short–term government bonds. N(–d2) and N(–d1) represent the cumulative probability under standard normal distribution when the variable equals −d2 and −d1 respectively, and d1 and d2 can be derived with the following equations:

$$d1 = \frac{\ln(S/X) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \tag{5}$$

$$d2 = \frac{\ln(S/X) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d1 - \sigma\sqrt{T} \tag{6}$$

where σ is the annual standard deviation of the return on the underlying asset, often called as the (underlying) asset volatility.[3] Note that the numerator of Eq. 6, $\ln(S/X) + (r + \sigma^2/2)T = \ln(S/X) + rT + \sigma^2T/2 = \ln(S/X) + \ln(e^{rT}) + \sigma^2T/2 = \ln[S/(Xe^{rT})] + \sigma^2T/2$. Similarly, the numerator of Eq. 5, $\ln(S/X) + (r\text{-}\sigma^2/2)T = \ln[S/(Xe^{rT})] - \sigma^2T/2$. For the convenience of application, I would like to transform the Eqs. 5 and 6 as:

$$d1 = \frac{\ln[S/(Xe^{-rT})]}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \tag{7}$$

$$d2 = \frac{\ln[S/(Xe^{-rT})]}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = d1 - \sigma\sqrt{T} \tag{8}$$

As for our analysis of risk equivalent, in Eqs. 4, 7 and 8, S is the current value of the predicted value; X is the predicted value at the maturity date; T is the maturity time at which the return X will occur; σ is the annual standard deviation of the relative change of the predicted value; r is the risk free rate. In a risk neutral environment as assumed in option pricing model, various values of outlays and returns grow or discount at the risk free interest rate in the way of continuous compounding.

The put or the guarantee is to guarantee the predicted value not less than X at time T. Therefore, the X and T here are identical as that in the Black–Scholes option pricing model. In a risk neutral environment, S is the present value of the relevant variable which is expected to be X in time T. Then the $S = Xe^{-rT}$. Therefore, $\ln[S/(Xe^{rT})] = \ln[(Xe^{-rT})/(Xe^{-rT})] = 0$. According to Eqs. 4 and 5, for valuing the risk equivalent:

[3] According to prior empirical research, the volatilities of stock returns in traditional industries usually range from 20 to 60%, such as Turner and Weigel (1992), etc. Thus, if the sector or the firm is relative riskier, then the volatility is closer to 60%; if the sector or the firm is relative safer, then the volatility is closer to 20%.

$$d1 = \frac{\ln[S/(Xe^{-rT})]}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} = \frac{\sigma\sqrt{T}}{2} = \sigma\sqrt{T/4} \tag{9}$$

$$d2 = \frac{\ln[S/(Xe^{-rT})]}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = -\frac{\sigma\sqrt{T}}{2} = -\sigma\sqrt{T/4} \tag{10}$$

Hence,

$$\begin{aligned} P &= Xe^{-rT}N(-d2)-SN(-d1) \\ &= Xe^{-rT}[N(-d2) - N(-d1)] \\ &= Xe^{-rT}\left[N\left(\sigma\sqrt{T/4}\right)-N\left(\sigma\sqrt{T/4}\right)\right] \\ &= Xe^{-rT}\left\{N\left(\sigma\sqrt{T/4}\right)-N\left(\sigma\sqrt{T/4}\right)\right\} \end{aligned}$$

Or,

$$P = Xe^{-rT}\left[2N(\left(\sigma\sqrt{T/4}\right) - 1\right] \tag{11}$$

Note that the certainty equivalent and the risk equivalent as well as the corresponding predicted or expected value are all occurring at the due time T, rather than occurring at present. However, the value of the put derived from the Black–Scholes model is a present value. Therefore, we should bring the above put value back to its "future value" for the purpose of valuing the risk equivalent, i.e.,

$$\begin{aligned} RE &= Pe^{rT} \\ &= Xe^{-rT}\left[2N\left(\sigma\sqrt{T/4}\right) - 1\right]e^{rT} \\ &= X\left[2N\left(\sigma\sqrt{T/4}\right) - 1\right] \end{aligned} \tag{12}$$

Equation 12 is the model of risk equivalent.

4.3 Certainty Equivalent Model

Based on Eqs. 3 and 12,

$$\begin{aligned} CE &= X-RE \\ &= X-X\left[2N(\sigma\sqrt{T/4}) - 1\right] \\ &= 2X\left[1-N\left(\sigma\sqrt{T/4}\right)\right] \end{aligned} \tag{13}$$

Equation 13 is the model of certainty equivalent, which was first published by Zhang [8, 9, 10].

Further, based on Eqs. 3 and 13,

$$\mathrm{d} = \mathrm{CE/X} = 2\left[1 - \mathrm{N}\left(\sigma\sqrt{T/4}\right)\right] \tag{14}$$

Equation 14 is the model of certainty coefficient, which was first published byZhang [8, 9, 10]. Corresponding to the certainty coefficient, we define and model the risk equivalent coefficient, or risk coefficient, v:

$$\mathrm{v} = 1 - \mathrm{d} = 2\mathrm{N}\left(\sigma\sqrt{T/4}\right) - 1 \tag{15}$$

Based on Eqs. 14 and 15, the certainty coefficient and the risk coefficient depend on two common influential factors: the volatility σ and the time horizon T. So we finally abstract the common influential factors of certainty equivalent, risk equivalent and their coefficients. As a common sense, increase in the volatility σ and the time horizon T will decrease the certainty coefficient and increase the risk coefficient. At one extreme, when one of the volatility σ and the time horizon T approaches infinite value, the certainty coefficient will be zero and the risk coefficient will be 1; At another extreme, when one of the volatility σ and the time horizon T approaches zero, the certainty coefficient will be 1 and the risk coefficient will be zero.

Let us examine the outcomes of the Eqs. 14 and 15. When the σ or T equals 0, $\sigma\sqrt{T/4} = 0$, $\mathrm{N}(\sigma\sqrt{T/4}) = 0.5$, d = 1 and v = 0; when the σ or T approaches ∞, $\sigma\sqrt{T/4} = \infty$, $\mathrm{N}(\sigma\sqrt{T/4}) = 1$, d = 0 and v = 1. Thus, these models illustrate that: the certainty coefficient reaches its maximum of 1 and the risk coefficient reaches its minimum of 0 for today's or certain "predicted" value; the certainty coefficient reaches its minimum of 0 and the risk coefficient reaches its maximum of 1 for the infinite future or completely uncertain "predicted" value; the larger of the volatility or/and the further of the time distance, the bigger of the risk equivalent and the smaller of the certainty equivalent.

These features obviously make sense and imply that the certainty coefficient model and the risk coefficient model (Eqs. 14 and 15) as well as the risk equivalent model and the certainty equivalent model (Eqs. 12 and 10) are correct in theory. To distinguish conveniently from other certainty equivalent related models (such as those based on utility functions), I refer to the Eqs. 12–15 as ZZ risk equivalent model (ZZ RE model), ZZ certainty equivalent model (ZZ CE model), ZZ certainty equivalent coefficient model (ZZ REC model), ZZ risk equivalent coefficient model (ZZ CEC model) respectively.

Now we find the convincing models to incorporate the total risk into consideration for valuation and financial decisions. Finance now again gets out from the theoretical crisis concerning the fundamental trading off between risk and return.

4.4 A Numeric Example

As a fundamental innovation in finance, the ZZ certainty equivalent related models can be used in various financial analyses and decisions, especially when we have no effective model to incorporate total risk into valuation and financial decisions. We illustrate in this section some basic applications of this series of models in capital budgeting and risk pricing based on a simple numerical example.

Consider a project G, the initial capital outlay and the predicted follow-up cash flows each year during the 10 years' project life are shown in Table 1. In addition, the annual risk free rate is 5% which assumed to be unchanged in 10 years.

4.4.1 NPV Based on the ZZ Certainty Equivalent Related Models

If we do not care the risk at all, we can use the risk free rate 5% to discount the predicted cash flows. Under such circumstances, the NPV of this project is:

$$\begin{aligned} \text{NPV} &= \sum_{t=1}^{10} \frac{CF_t}{(1+5\%)^t} - \text{CF}_0 \\ &= 384.26{-}200 = 184.26 \end{aligned}$$

Now, assume the firm will consider risk via the certainty equivalent. If the appropriate volatility is estimated as 25%, based on the ZZ CE model and the ZZ CEC model, the certainty coefficient and the certainty equivalent are shown in Table 2.

Theoretically, the certainty equivalent should be discounted at risk free rate. Discounting the certainty equivalents in Table 2 at 5%, the NPV of the project G is:

Table 1 Cash flows forecasted for project G

Year	0	1	2	3	4	5	6	7	8	9	10
Cash flow	– 200	10	30	50	70	90	90	70	50	30	10

Table 2 The certainty equivalents of the predicted cash flows (volatility is 25%)

Year	0	1	2	3	4	5	6	7	8	9	10
Cash flow	– 200	10	30	50	70	90	90	70	50	30	10
CEC	1.00	0.90	0.86	0.83	0.80	0.78	0.76	0.74	0.72	0.71	0.69
CE	– 200	9.01	25.79	41.43	56.18	70.19	68.35	51.86	36.18	21.23	6.93

CEC = Certainty coefficient
CE = Certainty equivalent

Table 3 The certainty equivalents of the predicted cash flows (volatility is 15%)

Year	0	1	2	3	4	5	6	7	8	9	10
Cash flow	– 200	10	30	50	70	90	90	70	50	30	10
CEC	1.00	0.94	0.92	0.90	0.88	0.87	0.85	0.84	0.83	0.82	0.81
CE	– 200	9.40	27.47	44.83	61.65	78.01	76.88	58.99	41.60	24.66	8.13

CEC = Certainty coefficient
CE = Certainty equivalent

$$\begin{aligned}\text{NPV} &= \sum_{t=1}^{10} \frac{CE_t}{(1+5\%)^t} - \text{CE}_0 \\ &= 299.26 - 200 = 99.26\end{aligned}$$

If the appropriate volatility is estimated as 15% rather than 25%, then the certainty coefficient and the certainty equivalent are shown in Table 3.

Discounting the certainty equivalents in Table 3 at 5%, the NPV of the project G is,

$$\begin{aligned}\text{NPV} &= \sum_{t=1}^{10} \frac{CE_t}{(1+5\%)^t} - \text{CE}_0 \\ &= 332.78 - 200 = 132.78\end{aligned}$$

In the above calculations and in Tables 2 and 3, the certainty coefficient decreases as the "maturity", T, and the volatility, σ, increase. These are the insights we get from the ZZ certainty coefficient model and ZZ risk coefficient model beyond the traditional common knowledge that certainty coefficient ranges from 0 to 1.

4.4.2 Risk Pricing Based on the ZZ Risk Equivalent Related Models

In addition to incorporate the risk in investment decision, the certainty coefficient and risk coefficient models, of course, can be used to support other related decisions. For instance, in reality, there are indeed guarantees or insurances for promising corresponding cash flows or values, these kinds of guarantees or insurances may be priced based on the certainty coefficient and the risk coefficient. Tables 4 and 5 calculate the risk coefficient and the risk equivalent of the corresponding cash flows in Tables 2 and 3.

For consistency, we use the same discount rate to discount the predicted or expected value, the certainty equivalent and the risk equivalent. Based on the volatility 25%, the sum of the present value of each year's risk equivalent is 85.00, which is just the difference between the total value of the project without consideration of risk and the total present value of its certainty equivalent, i.e. 384.26–299.26 = 85.00. Obviously, ignoring the transaction cost (such as the operating cost and profit of the guarantee or the insurance company), the fair price of the guarantee or the insurance for the predicted cash flows should be 85.00.

Table 4 The risk equivalent of the predicted cash flow (volatility is 25%)

Year	0	1	2	3	4	5	6	7	8	9	10
Cash flow	– 200	10	30	50	70	90	90	70	50	30	10
REC	0.00	0.10	0.14	0.17	0.20	0.22	0.24	0.26	0.28	0.29	0.31
RE	0.00	0.99	4.21	8.57	13.82	19.81	21.65	18.14	13.82	8.77	3.07

REC = Risk equivalent coefficient
RE = Risk equivalent

Table 5 The risk equivalent of the predicted cash flow (volatility is 15%)

Year	0	1	2	3	4	5	6	7	8	9	10
Cash flow	– 200	10	30	50	70	90	90	70	50	30	10
REC	0.00	0.06	0.08	0.10	0.12	0.13	0.15	0.16	0.17	0.18	0.19
RE	0.00	0.60	2.53	5.17	8.35	11.99	13.12	11.01	8.40	5.34	1.87

REC = Risk equivalent coefficient
RE = Risk equivalent

Based on the volatility 15%, the sum of the present value of each year's risk equivalent is 51.48, which is also just the difference between the total value of the project without consideration of risk and the total value of its certainty equivalent, i.e. 384.26–332.78 = 51.48. Similarly, neglecting the transaction cost, the fair price of the guarantee or the insurance for the predicted cash flows should be 51.48. This fair price is lower than that when the volatility is 25% (85.00), which obviously makes sense.

In risk business like insurance and guarantee, a terminology named "risk cost" is quite similar to the risk equivalent here, which refers to the expenses and reduced expected economic benefits that need to pay due to the existence of risk and risk accidents. Apparently, the concepts, logic and models related to the risk equivalent are helpful for the insurance companies and guarantee or bonding companies to estimate the risk cost and to price their product/service, especially when they launch new and innovative products or services.

For the insurance and guarantee companies, it is very important to value or price the risk as precisely as possible. On one hand, they need to take risk to attract customers; on the other hand, they need to make their risk cost precisely to compete with their peer companies. Under such a circumstance, the right model and the precise data are the two necessary factors to ensure their success. When it is too difficult to estimate some variables, a wise choice is following the conservative or prudential principle. For instance, the risk-taking companies can somewhat overestimate the risk (such as σ), hence derive a relative high risk cost.

The discovery of new and effective pricing method may be possible to play roles beyond what illustrated in the above example, i.e., set a standard or guidance for pricing. Pricing is an essential step for financial innovation. For the guarantee or insurance firms or other financial firms, new and more effective pricing method may be a must for their business expansion. For instance, the celebrated Black–Scholes model and its applications have been a key ingredient to the booming of various financial derivatives in past decades.

5 Risk Premium and a New CAPM

Based on the solution to quantification of risk equivalent and certainty equivalent as well as the certainty equivalent coefficient, the problem of considering total risk in financial decisions has been solved. These solutions cast light on the problem of determining the discount rate, or incorporating the total risk into the discount rate.

5.1 Modeling Risk Premium and a New CAPM

As mentioned previously, the reasons to discount include time compensation (time premium) and risk compensation (risk premium). The discount rate is the sum of the time premium and the risk premium, just as what demonstrated in Sharpe CAPM.

The common financial calculation "discounting" hence can be divided into two steps: (1) risk discounting—discount the risky or uncertain cash flow at risk premium to get its certainty equivalent at due time; (2) time discounting—discount the certainty equivalent at risk free rate (time premium) to get its value at present time.

This implies the certainty equivalent in concept is the result of discounting the expected value at the risk premium based on the year the expected value occurs. Therefore, the certainty coefficient can be regarded as the "risk discount factor". Discounting in the way of "continuous-compounding, the "time discount factor" is e^{-rT}. Use c to denote the annual risk premium, similarly, the "risk discount factor" is e^{-cT}, then,

$$\mathrm{d} = 2\left[1 - \mathrm{N}\left(\sigma\sqrt{T/4}\right)\right] = \mathrm{e}^{-\mathrm{c}\mathrm{T}} \tag{16}$$

Hence,

$$\ln\left\{2\left[1 - \mathrm{N}\left(\sigma\sqrt{T/4}\right)\right]\right\} = -\mathrm{cT} \tag{17}$$

Or,

$$\mathrm{c} = -\ln\left[2 - 2\mathrm{N}\left(\sigma\sqrt{T/4}\right)\right]/\mathrm{T} \tag{18}$$

Because $2[1–N(\sigma\sqrt{T/4})] = d < 1$, then, $\ln[2–2\,N(\sigma\sqrt{T/4})] < 0$, hence $c = –\ln[2–2\,N(\sigma\sqrt{T/4})]/T > 0$, i.e. c is a positive number. Obviously, Eq. 18 is a risk premium model incorporating total risk rather than only systematic risk.

Let k to represent risk adjusted discount rate, and r to represent risk free rate; note that the risk adjusted discount rate is the sum of the time premium and the risk premium, then,

$$k = r + c = r - \ln\left[2 - 2N\left(\sigma\sqrt{T/4}\right)\right]/T \qquad (19)$$

Of course, as a discount rate, k in Eq. 19 incorporates the total risk. Equation 18 and 19 were first published by Zhang [6, 7, 8]. For the convenience to distinguish with the Sharpe CAPM, we refer to Eqs. 18 and 19 as ZZ risk premium model and ZZ capital asset pricing model or ZZ CAPM respectively.

5.2 Comparison Between Sharpe CAPM and ZZ CAPM

We now have two models concerning the relationship between risk and return. One is Sharpe CAPM, which takes only the systematic risk into account; the other is the ZZ CAPM, which takes total risk (the systematic risk and the non–systematic risk) into account.

Then, a question is: which is better for determining a discount rate? This implies that we need to make a choice between Sharpe CAPM and ZZ CAPM. It is thus necessary to make a systematic comparison between the two models.

Firstly, as revealed in Chapter "Finance and Its Fundamental Problems", finance is a decision-oriented science rather than a descriptive science. As a decision benchmark, what we need in discounting is a required (rate of) return. The required return is neither an ex-post actual one nor a predicted one. Both the ex-post actual return and the predicted return are possible to mismatch with the asset risk; but the required return should theoretically match with the asset risk. Since Sharpe CAPM and ZZ CAPM can derive a required return respectively, both are right in this sense.

Secondly, the required return derived by Sharpe CAPM matches with only part of the asset risk, i.e. the systematic risk; whereas the required return derived by ZZ CAPM matches with the asset's total risk. Normally, for various financial and economic decision-making, what we need to consider is total risk rather than only part risk (regardless it is systematic risk or not). Matching with part risk is equivalent to mismatching with total risk. The Sharpe CAPM in such sense is not right for determining discount rate.

Thirdly, the discount rate derived by the Sharpe CAPM can be lower than the risk free rate. Based on the Sharpe CAPM, the (systematic) risk premium is the product of the beta and the average risk premium of the whole market, $[E(R_m)–R_f]$. The two factors are possible to be negative; and any one of them being negative will lead to the risk premium negative and the discount rate then is less than the risk free rate.

This makes no sense. On the other hand, the discount rate derived by the ZZ CAPM is always higher than the risk free rate, because the risk premium derived by the ZZ risk premium model is always positive. This implies that the ZZ CAPM is sounder in theory and more reliable in practice.

Fourthly, the Sharpe CAPM reflects the diversification effect; the more securities brought into the portfolio, the more diversification effect, and the lower of the discount rate. While this may make some sense, especially for investment in security market; there is another "diversification effect" captured by the ZZ CAPM, which is the effect of diversification among periods; the longer the investment lasting, the more risks are cancelled out among periods and then the lower the discount rate can be. This reveals an important principle: the longer of the investment lasting, the lower of its (annual) risk.[4] Hence the successful experience from long–term investment (such as Warren E. Buffett) can be explained by the ZZ CAPM.

Fifthly, it seems that the Sharpe CAPM and the ZZ CAPM are representing the two extremes. One is full-diversification; the other is zero-diversification in terms of the variety of securities. In fact, the ZZ CAPM can also consider easily the diversification effect across securities. Replacing the volatility of an asset by the volatility of the portfolio under consideration, the discount rate matches with the portfolio risk can be easily derived based on the same ZZ CAPM, hence consider the diversification effect properly.

Sixthly, there may be some difficulties in estimating the σ of an asset or the σi,j of two assets, but this should not be a big hurdle in application of the ZZ CAPM. One reason is that the σ of an asset or the σi,j of two assets is already the most basic risk measures in finance. Most risk measures, such as beta, value at risk, etc., are based on the basic variance of assets. The second reason is that the Sharpe CAPM has been widely used for decades and it (and the beta) relies on the σ of an asset and the σi,j of any two assets heavily. In other words, the application of the Sharpe CAPM proves the feasibility of the application of the ZZ CAPM.

5.3 A Numeric Example

Again for the example in last section, when the volatility level is 15% and 25% respectively, the corresponding c and k are calculated based on the ZZ risk premium model and the ZZ CAPM as shown in Tables 6 and 7 respectively.

In Tables 6 and 7, the c, or the annual risk premium rate as well as the risk–adjusted discount rate, decreases as the due time goes on. As explained above, this is because the investment is better time-diversified as the number of period increases; more and more risks are canceled out among periods. Nevertheless, the certainty coefficient still keeps decreasing from year to year; similarly, the total discount factor, e^{-kT},

[4] Some scholars find that the annual risk premium and the risk-adjusted discount rate should decrease along with the time extending into further future, such as Martin [11], Gollier and Weitzman [12], Gollier et al. [13], etc.

Table 6 The estimation of risk adjusted discount rate k (volatility is 15%)

Year	1	2	3	4	5	6	7	8	9	10
c	0.062	0.044	0.036	0.032	0.029	0.026	0.024	0.023	0.022	0.021
d	0.940	0.916	0.897	0.881	0.867	0.854	0.843	0.832	0.822	0.813
k	0.112	0.094	0.086	0.082	0.079	0.076	0.074	0.073	0.072	0.071
e^{-kT}	0.894	0.828	0.772	0.721	0.675	0.633	0.594	0.558	0.524	0.493

Note e^{-kT} is the discount factor incorporating both systematic and non–systematic risk. The discounting can be divided into risk discounting and time discounting, i.e., $e^{-kT} = e^{-(r+c)T} = e^{-rT}e^{-cT}$; e^{-cT} and e^{-rT} are "risk discount factor" and "time discount factor" respectively

Table 7 The estimation of risk adjusted discount rate k (volatility is 25%)

Year	1	2	3	4	5	6	7	8	9	10
c	0.105	0.076	0.063	0.055	0.050	0.046	0.043	0.040	0.038	0.037
d	0.901	0.860	0.829	0.803	0.780	0.759	0.741	0.724	0.708	0.693
k	0.155	0.126	0.113	0.105	0.100	0.096	0.093	0.090	0.088	0.087
e^{-kT}	0.857	0.778	0.713	0.657	0.607	0.563	0.522	0.485	0.451	0.420

calculated based on the total discount rate k, also remains decreasing from year to year. This implies that the further in future of a variable, the larger part of its value is discounted out or the smaller part of its value remains as present value. This obviously makes sense.

Please note that the ZZ models of certainty equivalent and its coefficient, the ZZ models of risk equivalent and its coefficient, the ZZ risk premium model and the ZZ CAPM are mutually consistent in logic. In other words, we'll get the same result either via discounting the certainty equivalent at the risk free rate or via discounting the predicted cash flow at the risk adjusted discount rate. For example, for the volatility level of 25%, the discounting results of the cash flows from project G via the two methods are shown in Table 8.

Please note in Table 8, the present value I and the present value II are identical. Present value I (the second row) results from discounting the predicted cash flow at the risk adjusted discount rate, i.e. predicted cash flow multiplied by e^{-kT}, while present value II (the sixth row) results from discounting the certainty equivalent at the risk free rate, i.e. certainty equivalent multiplied by e^{-rT}. Obviously, the certainty equivalent model or certainty coefficient model and the risk premium model or CAPM are mutually consistent.

6 Summary

We now solve the fundamental problem of incorporating total risk into required rate of return or discount rate as well as into asset valuation or financial decisions with

Table 8 The consistency of the certainty equivalent and the ZZ CAPM

Year	0	1	2	3	4	5	6	7	8	9	10
Cash flow	– 200	10	30	50	70	90	90	70	50	30	10
k		0.155	0.126	0.113	0.105	0.100	0.096	0.093	0.090	0.088	0.087
e^{-kT}	1	0.857	0.778	0.713	0.657	0.607	0.563	0.522	0.485	0.451	0.420
Present value I	– 200	8.57	23.34	35.66	46.00	54.66	50.64	36.55	24.25	13.54	4.20
CE	– 200	9.01	25.79	41.43	56.18	70.19	68.35	51.86	36.18	21.23	6.93
e^{-rT}	1	0.951	0.905	0.861	0.819	0.779	0.741	0.705	0.670	0.638	0.607
Present value II	– 200	8.57	23.34	35.66	46.00	54.66	50.64	36.55	24.25	13.54	4.20

a series of models: the ZZ models of certainty equivalent and its coefficient, the ZZ risk equivalent and its coefficient, the ZZ risk premium model and the ZZ CAPM. Neither the forms nor the variables of these models are chosen subjectively, just like the model series of stock valuation and theoretical ratios in Chapter "Stock and Equity Valuation: Where Discounting Does Not Work". In other words, both the forms and the variables of these models are derived via strict logic processes. Also similar to the innovative models in last chapter, the models in this chapter are simple in form and feasible in practice, hence will benefit the related financial calculations, such as the estimation of the certainty equivalent as well as the risk–adjusted discount rate, etc. As the important and fundamental innovations, these models will support more effectively the related financial decisions, such as investment decision, stock valuation and asset pricing as well as risk management.

References

1. Markowitz H. Portfolio selection. J Financ. 1952;7(1):77–91.
2. Sharpe WF. Capital asset prices—a theory of market equilibrium under conditions of risk. J Financ. 1964;19(3):425–42.
3. Treyno JL. Toward a Theory of Market Value of Risky Assets, John Wiley & Sons, Inc., 1999.
4. Lintner J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. Review of Economics and Statistics, 1965:47(1):13–37.
5. Mossin J. Equilibrium in a Capital Asset Market, Econometrica. 1966;34(4):768–783.
6. Ross SA. The arbitrage theory of capital asset pricing. J Econ Theor. 1976;13(3):341–60.
7. Fama EF, French KR. Common Risk factors in the returns on stocks and bonds. J Financ Econ. 1993;33:3–56.
8. Zhang Z. A theoretical model of certainty equivalent. Manag World. 2009:168–9.
9. Zhiqiang Z. The capital asset pricing model accounting for total risk. Manage World. 2010;4:177–8.
10. Zhang Z. Certainty equivalent, risk premium and asset pricing. Front Bus Res China. 2010;4(2):325–39.
11. Weitzman ML. Why the far-distant future should be discounted at its lowest possible rate. Journal of environmental management. 1998;36(3):201–8.

12. Gollier C, Weitzman ML. How should the distant future be discounted when discount rates are uncertain?. Economics Letters. 2010;107(3):350–3.
13. Gollier C, Hammitt JK. The long-run discount rate controversy. Annual Review of Resource Economics. 2014;6(1):273-95.

Further Readings

14. Zhang Z, Yu M. Discount rate and capital cost: how to determine the discount rate. Res Financ Issues, 2014;(12): 11–7.
15. Cai Y, Judd K, Steinbuks J. A nonlinear certainty equivalent approximation method for dynamic stochastic problems. Quantitative Economics. 2017 Mar;8(1):117–47.
16. Wilcox NT. Random expected utility and certainty equivalents: Mimicry of probability weighting functions. Journal of the Economic Science Association. 2017 Dec;3(2):161–73.
17. Geissel S, Sass J, Seifried FT. Optimal expected utility risk measures. Statistics & Risk Modeling. 2018 Jan 1;35(1-2):73–87.
18. Chudziak J. Certainty equivalent under cumulative prospect theory. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 2019 Jun;27(3):415–28.
19. Fama EF, French KR. A five-factor asset pricing model. J Financ Econ. 2015;116(1):1–22.

Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing

Contents

Last chapter derives a new CAPM, which factors in total risk rather than only systematic risk and solves the problem of discount rate for total asset or total capital. This chapter aims at finding a proper way to determine the fair rate of return on corporate debt, or the discount rate for debt capital. Next chapter aims at solving the problem of discount rate for equity capital, after that, the problem of discount rate or asset pricing will be completely solved.

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_8

Debt in this chapter can be referred to as either corporate debt or bank loan. Despite of the transaction cost, cost of debt of the levered firm is just the return obtained by a bank from the relevant business loan. The method or model derived in this chapter is applicable as well for banks to determine interest rate on business loans.

1 Debt/Loan Risk

There are different terms to express debt or loan risk. In theory and practice, phrases like credit risk, default risk, financial risk, bankruptcy risk as well as insolvency risk, all have the close or same meaning to debt or loan risk.

1.1 The Description of Debt/Loan Risk

Here are two common understandings about debt or loan risk.

1. Debt or loan risk as a baseline of fluctuation of firm value

As shown in Fig. 1, along with the fluctuation of the firm EBIT and thus the firm value (the wave dash line), the firm keep going forward safely unless a downward move goes so deep that the firm value is pushed lower than its debt value (the black horizontal line closer to the horizontal axis). Once the firm value is lower than its debt value, the firm cannot repay its due debt (interest and principal), the debt holders will suffer losses. This is the situation when debt risk leads to the actual losses.

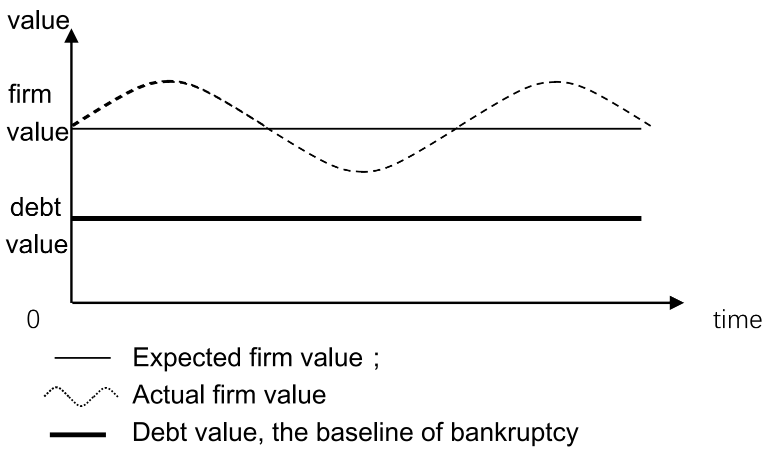


Fig. 1 The risk of debt, equity and the firm

Obviously, most of the time, the equity holders bear the whole risk from the EBIT and the value fluctuation of the firm, so that the debt holders bear no risks. Only when the bankruptcy takes place, the debtholders bear some of the risks. Theoretically, the equity gets nothing or loses everything in the firm when such disaster happens.

In sum, the debt is not influenced much by the fluctuation of the earnings and value of the firm in normal situation; only in case of bankruptcy, the equity has nothing left whereas the debt also suffers more or less depending on how much value the firm left. In such a case, the equity value more often moves down to zero and the debt value is equal to the firm value. In other cases, the downward going earnings may erode the annual return of equity (earnings per share or dividend) without negative effect on the return (interest) of debt. However, even in such a case, the downward going earnings may lead to a slight decrease of the debt value in market because the increase of the bankruptcy probability in future.

2. Debt or loan risk as a portfolio with options

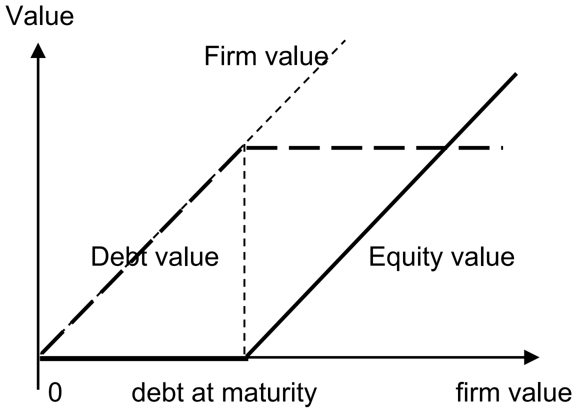
As we have known, a firm value is equal to the sum of its debt value and equity value, as shown in Fig. 2. In Fig. 2, the black dash kinked line represents the value pattern of debt at maturity. At the maturity date, if the firm value larger than the debt value, the claim of the debt holders is fully satisfied and the debt is valued as a risk free debt. If the firm value less than the fully paid debt value, the debt suffers from a bankruptcy, and the debt value is as large as the firm value; in another word, the less the firm value, the lower the debt value.

The debt vale or the kinked line in Fig. 2 can be composed in either of two ways, as shown in Figs. 3 and 4 respectively.

The debt holders as a whole can be viewed as that they own the firm but give the owner a free call option on the firm with exercise price at the maturity value of the debt, as shown in Fig. 3.

In such a case, at the maturity date, if the firm value is lower than the exercise price or the maturity value of the debt, the equity holder will not exercise the call; the debt holder continues to own the firm and has debt value equal to the firm (slop

Fig. 2 Firm value = debt value + equity value



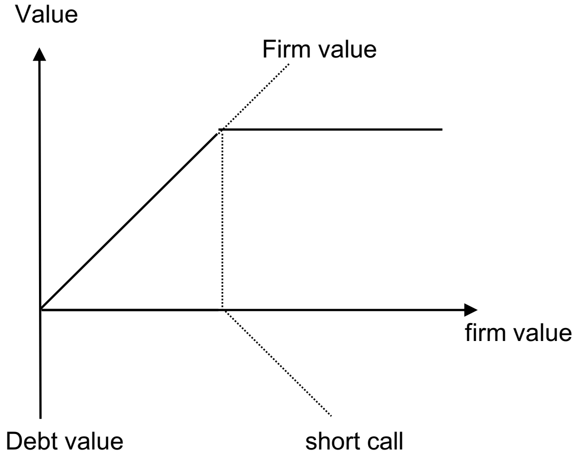


Fig. 3 Debt value = firm value + short call

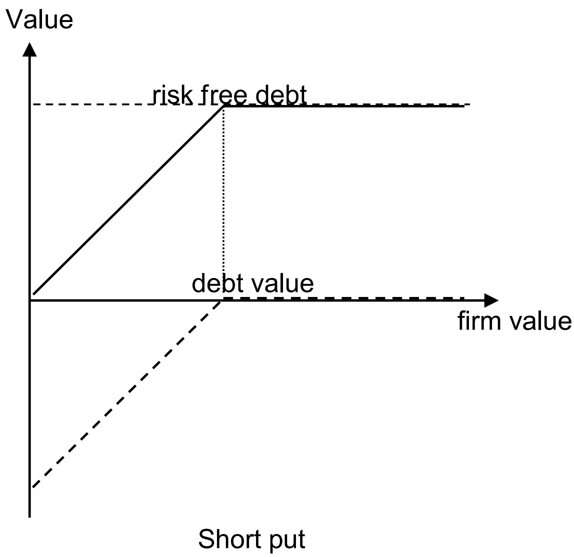


Fig. 4 Debt value = risk free debt + short put

part of the kinked line). Otherwise, if the firm value is higher than the exercise price, the equity holder will certainly exercise the call; the debt holder has to sell the firm to the equity holder at the exercise price and has debt value equal to the risk free debt (horizontal part of the kinked line).

For instance, a company has debt with book value 100 and maturity 3 years, suppose the present value of the debt is the same as its book value. The volatility of the whole company is estimated as 0.25, the risk free rate over 3 years in the market is 4%. How much is the fair value of the debt when the company value are 80, 100, 150, 200, 300 respectively? Then, we have the data of the relevant inputs:

$S = 80, 100, 150, 200, 300, Xe^{-rT} = 100,$[1] $T = 3, r = 4\%, \sigma = 0.25.$
Based on Black–Scholes model,
Value of the equity as a call option,

$$c = SN(d1) - Xe^{-rT}N(d2) = 7.39, 17.14, 54.91, 101.40, 200.13 \text{ respectively.}$$

Correspondingly, the fair value of the debt are,

$$80 - 7.39,\ 100 - 17.14,\ 150 - 54,\ 91,\ 200 - 101.40,\ 300 - 200.13$$
$$= 72.61,\ 82.86,\ 95.09,\ 98.60,\ 99.87 \text{ respectively.}$$

Alternatively, the debt holders as a whole can be viewed as that they own a risk free debt but give the owner a free put option on the firm with exercise price at the maturity value of the debt, as shown in Fig. 4.

In such a case, at the maturity date, if the firm value is lower than the exercise price or the maturity value of the debt, the equity holder will exercise the put, sell the firm to the debt holder at a price of the risk free debt; the debt holder receives the firm and has the debt value equal to the firm (slop part of the kinked line). Otherwise, if the firm value higher than the exercise price, the equity holder will certainly not exercise the put; the debt holder get its claim full paid and the debt value equal to the risk free debt (horizontal part of the kinked line).

For instance, go on with the previous case, the company has debt with book value 100 and maturity 3 years. Then, the data of the relevant inputs are the same:

$$S = 80, 100, 150, 200, 300, Xe^{-rT} = 100, T = 3, r = 4\%, \sigma = 0.25$$

Based on Black–Scholes model,
Value of the relevant put option,

$$p = Xe^{-rT}N(-d2) - SN(-d1) = 27.39, 17.14, 4.91, 1.40, 0.13 \text{ respectively.}$$

Correspondingly, the fair value of the debt are,

$$100 - 27.39, 100 - 17.14, 100 - 4, 91, 100 - 1.40, 100 - 0.13$$

$$= 72.61, 82.86, 95.09, 98.60, 99.87 \text{ respectively.}$$

Obviously, those results based on the put option values are exactly the same as the previous results based on the call option values.

[1] Note that $Xe^{-rT} = 100$, rather than $X = 100$.

1.2 The Composition of Interest Rate

When a bank provides a loan to a firm, it involves some risks. As mentioned in some financial literature, these risks include inflation risk, liquidity risk, duration or maturity risk and default risk, etc. Understandingly, the interest rate of a loan may be structured as:

$$\begin{aligned} \mathrm{y} &= \text{ risk free rate } + \text{ debt risk premium} \\ &= r + \text{ debt risk premium} \end{aligned} \tag{1}$$

Premium here means additional payment to compensate the relevant risk. Generally, the interest rate or debt cost should be determined (loan/debt pricing) in such a way that the lender is compensated according to the corresponding risk, i.e. the debt risk.

As a common perception, the debt risk premium equals to the sum of all risk premiums for inflation risk, liquidity risk, duration or maturity risk and default risk. A common interest rate equation then is written as following:

$$\mathrm{k} = \mathrm{rr} + \mathrm{ri} + \mathrm{rl} + \mathrm{rm} + \mathrm{rd} \tag{2}$$

where,

k: the interest rate, or the fair return or cost on debt;
rr: real risk free rate;
ri: the Inflation risk premium;
rl: the Liquidity risk premium;
rm: the Maturity risk premium;
rd: the Default risk premium, or bankruptcy risk premium.

An interesting question is: what do you think about Eq. 2? Does it reveal all the risks bore by the debt holder? Does it express the right relations between the dependent and independent variables?

Normally, risk free rate is defined as the yield on (central) government bond. The bond yield is the expected rate of return implied in the bond price. Simply speaking, “risk free rate” is the rate of return on riskless investment or asset. But every investment or asset has risk. The risk free here is actually just referred to “no default risk”. In another word, risk free rate should provide premiums for the risks other than default risk, such as inflation risk, etc.

In bond market, investors will naturally incorporate the expected inflation into a deal. So the bond yield (including the yield on government bond) or the risk free rate in reality includes the inflation premium originally. It is not necessary to invite trouble to isolate the inflation premium from the risk free rate, and then add back the inflation premium to the real risk free rate, i.e., rr + ri. In another word, it is better to use directly the nominal risk free interest rate or simply refer to as “the risk free rate”.

Next, what is the liquidity risk premium? Liquidity risk here means a possible lose of the debt holder because of holding debt rather than cash. The reason is that the process of changing a debt into cash may be delayed when the debt holder needs cash. On the other hand, the risk free rate is literally used to compensate the time delay of the cash inflow. In another word, the risk free rate has already compensated the liquidity risk. Therefore, the liquidity risk premium seems to be a repeated consideration for compensating the time delay in addition to the risk free rate. So, the "rr + ri + rl" in Eq. 2 needs to rewrite as "r", i.e. the nominal risk free rate.

Similarly, there may be other risk premiums not incorporated in Eq. 2 either, such as interest rate risk, etc. It is not a problem so long as we use the nominal risk free rate in the model rather than the real risk free rate, because literally, risk free rate is equivalent to default risk free rate, which means, the nominal risk free rate includes the risks premiums for risks other than default risk as well.

Further, the default risk increases along with the increase of the debt maturity. The maturity risk hence is just an influential factor of the default risk, rather than a stand-alone risk. That is, the maturity risk premium is a repeated consideration for compensating the default risk in addition to the default risk premium. So, the "rm + rd" in Eq. 2 needs to rewrite as "rd", i.e. the default risk premium.

In sum, the inflation risk, the liquidity risk and the maturity risk are all unnecessarily mentioned after the risk free rate and default risk; their risk premiums in the model are unnecessary either. In another word, the Eq. 2 is wrong, because it takes into account some risk premiums repeatedly. The correct equation should be:

$$\mathrm{k} = \mathrm{r} + \mathrm{rd} \tag{3}$$

Equation 3 states that the fair rate of return or cost of debt is the sum of risk free rate and its default risk premium. The debt risk is thus the default risk.

As well known in financial community, the problem to determine risk free rate is solved. The key problem now is how to determine the (default) risk premium.

1.3 Debt Cost and Bankruptcy Cost

Previous discussion reveals that the fair and reasonable interest rate of loan or the cost of debt is the sum of the risk free rate and the (default) risk premium. However, it is a tough problem to determine the (default) risk premium.

Finance is a discipline that studies the value of assets, and draws decision-making conclusions based on the value of assets. The common calculations in finance are to derive the value of assets based on its future risk and return. The return (such as profits, earnings, cash flows, etc.) increases value, while the risk (uncertainty of return) reduces value. Understandingly, among the three variables, risk, return and value, anyone can be derived based on the other two.

The debt risk premium, or the default risk premium, is a kind of return. It thus can be derived based on the relevant risk and value. It is imaginable that the debt risk is determined by the company risk and its leverage (debt) ratio. But, what is the value of the debt risk? As a matter of fact, it is named as bankruptcy cost in finance, and has been hotly researched for solving the problem of optimal capital structure. Unfortunately, the quantification of bankruptcy cost remains unsolved so far.

The quantification of bankruptcy cost is a theoretical center for many financial problems; and is the key for solving the problem of optimal capital structure, and the debt or loan pricing as well. We will discuss and solve it later in this chapter. Here, we just make clear the relation between the debt risk premium and the bankruptcy cost.

Let c be the (annual) risk premium of a risky debt or loan; T be the maturity of the debt; X be the value of the corresponding risk free debt; and BC be the relevant bankruptcy cost. Then the value of the risky debt is Xe^{-cT}. Conceptually, the value of a risky debt can also be obtained by subtraction of the bankruptcy cost (BC) from a corresponding risk-free debt (X). Then,

$$Xe^{-cT} = X - BC \tag{4}$$

Then,

$$e^{-cT} = 1 - BC/X \tag{5}$$

Based on elementary mathematics,

$$-cT = \ln(1 - BC/X) \tag{6}$$

Thus,

$$c = -\ln(1 - BC/X)/T \tag{7}$$

Obviously, once the quantification of bankruptcy cost solved, the debt risk premium can easily be worked out based on Eq. 7.

2 Valuation of Debt Guarantee

A debt guarantee is an agreement promising that, should one party default on a debt or loan, the other party would be responsible to pay it.

An additional party promising to pay back the debt when it is needed reduces the risk for overall default, and enhance the credit of the original borrower. A debt guarantee thus is often used by small and median firms as a way of securing a loan or reducing the interest rate of a loan or the cost of debt.

Default literally is the same as bankruptcy and are used interchangeably in finance; we do not differentiate them either in this book. The valuation of a debt guarantee is obviously helpful for the quantification of bankruptcy cost, because debt guarantee literally undertakes all the bankruptcy or default risk. The analyses and calculations in this section use the similar method as in Zhang and Yu [1].

2.1 Valuing Debt Guarantee: Simple Cases

1. Guaranteed defined in terms of option

Suppose that company L lends company B the debt capital with the due principal and interest amount of X, and company G guarantees this debt.

The guarantee is to ensure that company B can repay the debt, or, the company L can receive the principal and interest of the debt regardless of the situation and value of company B, including the case that the company B is bankrupt.

Assume that company B now only has the debt borrowed from company L. When the value of company B is greater than or equal to the book value of debt X, the intrinsic value of guarantee G is zero; When the value of company B is less than the book value of debt, the intrinsic value of guarantee G is greater than zero. As the value of company b decreases to zero, the intrinsic value of guarantee G increases to the book value of debt X. The guaranteed value is shown in Fig. 5.

It can be seen that in terms of option, the debt guarantee is equivalent to a put option on the company. The exercise price of this put option is the book value of the debt and the maturity time is the maturity time of the debt.

2. Valuing debt guarantee: example

Since the guarantee is equivalent to a put option on the company, we can certainly get the value of the debt guarantee by calculating the value of such a put option. We might as well estimate the value of guarantee with the help of Black–Scholes model.

According to the Black–Scholes model, to estimate the value of the guarantee, we need to know:

(1) The current price of the underlying asset, i.e. the current value of the company S;
(2) Exercise price X, which is equal to the book value of the debt;
(3) Maturity time t, equal to the maturity time of the debt;
(4) Annual volatility σ of the return of the company asset;
(5) Risk free interest rate r.

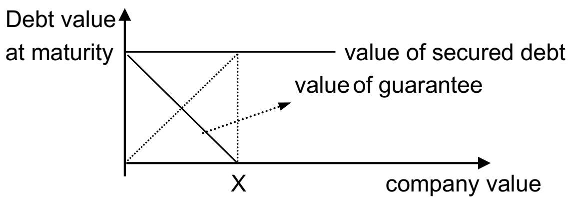


Fig. 5 The debt guarantee defined in terms of option

Where, the data in (2), (3) and (5) are not difficult to obtain. The estimations of data in (1) and (4) are analyzed as following.

The value of a company is the sum of its debt value and equity value. The debt value is often simply assumed to be equal to its book value in normal situation. The equity value needs to estimate by a proper method, which is several times of its book value in normal situation. For a public company, the market value or market cap (capitalization) of its stock is often used as the simplest estimation of its equity value. For the prudent principle, the company value may be estimated excluding the current debt.

Since a company is consisted of debt and equity; debt and equity are traded separately in the market; and equity usually has a higher volatility because it bears greater risks. Therefore, it can be judged that the volatility of the company's value should be between the volatility of its equity and that of its debt.

Therefore, for public companies, the first step is to estimate the volatilities of its stocks and bonds respectively, and then, as the second step, the volatility of the company can be obtained by weighted averaging the two volatilities in a proper way. For companies without listed stocks and bonds, the volatility can be estimated based on the comparable companies, or simply calculating the volatility of its profits or sales.

For example, company E is valued to be 50 million dollars; its volatility is 25%; The only debt of the company is a 5-year loan, with the principal of 20 million dollars, the interest rate of 8%. The sum of principal and interest at maturity is: 20 million dollars $\times$ $(1 + 8\%)^5 = 29.3866$ million dollars. The risk-free interest rate is 5%.

That is: $S = 50$; $X = 29.3866$; $r = 5\%$; $T = 5$; $\sigma = 25\%$。

Based on the Black–Scholes model,

$$\mathrm{d1} = 1.6775,\ \mathrm{N}(-\mathrm{d1}) = 0.0467$$
$$\mathrm{d2} = 1.1185,\ \mathrm{N}(-\mathrm{d2}) = 0.1317$$
$$\mathrm{P} = \mathrm{X}e^{-rT}\mathrm{N}(-\mathrm{d2}) - \mathrm{SN}(-\mathrm{d1}) = 0.6776 \text{ million dollars.}$$

Therefore, the value of the debt guarantee is 0.6776 million dollars, accounting for 0.6776/20 = 3.3878% of the guaranteed debt. Despite the time value of money, the annual cost is: 3.3878%/5 = 0.6776%. Apportion it at risk free rate, 5%, and a discount rate of 8%, the annual cost is 0.7453% and 0.7857% respectively.[2]

Assume other things being equal, but the company applies for another loan of 10 million dollars on the day the above loan transferred to its account. After that, the debt increases to 30 million dollars and the company value increases to 60 million dollars. Then: $S = 60$; $X = 30 \times (1 + 8\%)^5 = 44.0798$ million dollars; $r = 5\%$; $T = 5$; $\sigma = 25\%$.

Based on the Black–Scholes model,

[2] Please note that the annual cost is an annuity due or annuity in advance rather than an annuity in arrears or an ordinary annuity; the annual cost in amount is 0.1491 million dollars and 0.1571 million dollars respectively.

$$d1 = 1.2783,\ N(-d1) = 0.1006$$
$$d2 = 0.7193,\ N(-d2) = 0.2360$$
$$P = Xe^{-rT}N(-d2) - SN(-d1) = 2.0669 \text{ million dollars}$$

Therefore, the value of the above guarantee is 2.0669 million dollars, accounting for 2.0669/30 = 6.8897% of the guaranteed debt. Despite the time value of money, the annual cost is: 6.8897%/5 = 1.3779%; Apportion it at risk free rate, 5%, and a discount rate of 8%, the annual cost is 1.5156% and 1.5977% respectively.[3]

It is interesting to consider the additional 10 million dollars alone. The value of the guarantee for the additional 10 million debt is: 2.0669–0.6776 = 1.3893 million dollars, account for 1.3893/10 = 13.8930% of the guaranteed debt. Despite the time value of money, the annual cost is: 13.8930%/5 = 2.7786%; Apportion it at risk free rate, 5%, and a discount rate of 8%, the annual cost is 3.0561% and 3.2218% respectively.[4] Comparing with the annual costs of the previous 20 million debt, i.e., 0.6776%, 0.7453%, 0.7857%, the annual costs increase by 310.0649%, 310.0496%, 310.0547% respectively.[5]

2.2 Valuing Debt Guarantee: Complex Cases

The debt guarantee in previous discussion is just the simplest case. Debt guarantees in reality are often more complicated in some aspects.

1. Incomplete debt guarantee

The debt guarantee in previous discussion can be called as complete debt guarantee. In such a case, the customers reviewing and screening become the sole responsibility of the guarantor; the lender (such as a bank) is no longer concern of the credit grade and default risk of the customers. This is not good at the risk control and the utilization of the relevant expertise of the lender. Therefore, the incomplete debt guarantee may be better for avoiding those drawbacks.

There are two main forms of the incomplete debt guarantees. One is that the lender is responsible for a fixed or constant amount of default, and the guarantor is responsible to pay the loss beyond that amount. This can be called as excess debt guarantee. The other is that the lender and the guarantor are responsible for the certain proportion of the default loss respectively. This can be called as proportion debt guarantee.

Understandingly, in terms of option, the excess debt guarantee is equivalent to a smaller put option than the corresponding complete guarantee, shown as Fig. 6. The

[3] The annual cost in amount is 0.4547 million dollars and 0.4793 million dollars respectively.

[4] The annual cost in amount is 0.3056million dollars and 0.3222 million dollars respectively.

[5] 2.7786%/0.6776% − 1 = 310.0649%; 3.0561%/0.7453% − 1 = 310.0496%; 3.2218%/0.7857% − 1 = 310.0547%. The slightly differences of the annual cost increases are caused by the normal rounding.

proportion debt guarantee is equivalent to a lower put option than the corresponding complete guarantee, shown as Fig. 7.

As what can be seen in Fig. 6, the excess guarantee is still a complete put, but compared with the complete debt guarantee, it is a put with a smaller exercise price. For example, assuming that other things being equal, but the previous guarantee contract stipulates that once the company defaults, the guarantor needs not to pay within 6 million dollars in loss, or only needs to pay the part beyond 6 million dollars in loss.

When the loan principal is 20 million dollars, $S = 50$; $X = 29.3866 - 6 = 23.3866$; $r = 5\%$; $T = 5$; $\sigma = 25\%$.

Based on the Black–Scholes model,

$$\mathrm{d1} = 2.0860,\ \mathrm{N}(-\mathrm{d1}) = 0.0185$$
$$\mathrm{d2} = 1.5270,\ \mathrm{N}(-\mathrm{d2}) = 0.0634$$
$$\mathrm{P} = \mathrm{X}e^{-\mathrm{rT}}\mathrm{N}(-\mathrm{d2}) - \mathrm{SN}(-\mathrm{d1}) = 0.2300 \text{ million dollars}$$

Therefore, the value of the above excess guarantee is 0.23 million dollars, which is equivalent to 0.23/20 = 1.15% of the guaranteed amount; obviously, both the amount and the percentage are smaller than that of the corresponding complete debt guarantee.

An interesting question here is that, how much is the risk value left to the bank in this case? The total risk value is covered by the complete guarantee, this implies that the total risk value is 0.6776 million dollars as calculated previously. Then, the risk value left to the bank is 0.6776 − 0.2300 = 0.4476 million dollars, which is much larger than the risk value undertaken by the guarantor (0.2300 million dollars). This may seem inconceivable at the first sight, since the bank is only responsible for the first 6 million dollars loss, while the guarantor is responsible for the rest 23.3866 million dollars loss. But this is of course right, because the bank is the first in line to suffer the loss.

Fig. 6 The excess debt guarantee

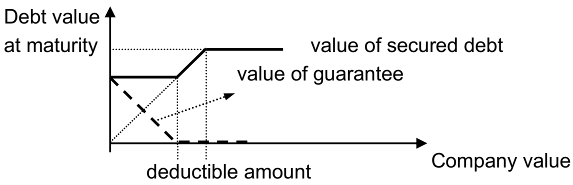
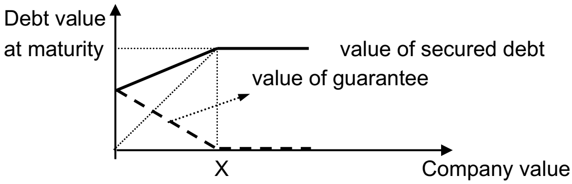


Fig. 7 The proportion debt guarantee



The relevant calculation for the proportion guarantee is relatively easy. For example, assuming that other things being equal, but the previous complete guarantee contract stipulates that once the company defaults, the guarantor and the bank undertake 85% and 15% respectively. In such a case, the guarantor and the bank proportionally undertake the complete default risk, or the put value.

Based on the previous calculation, when the loan principal is 20 million dollars and 30 million dollars, the value of debt guarantee is 0.6776 million dollars and 2.0669 million dollars respectively. In both cases, the value of 85% proportion guarantee is:

$$0.6776 \times 85\% = 0.5760 \text{ million dollars}$$
$$2.0669 \times 85\% = 1.7569 \text{ million dollars}$$

When the guaranteed debt increases by 50%, the guarantee value increases by 175.6865/57.5922–1 = 205.05%.

2. Multiple debt guarantees

A company often has multiple debts and multiple guarantees. In such a case, it needs to value those guarantees respectively. The claims of those debts may be different and indifferent in priority; the calculations are also different for those two situations.

For the case of indifference in priority, it is easy to value those guarantees. For example, the current value of a company is valued as 100 million dollars. The company has three debts with the same maturity of 4 years and same priority in repayment. Debt A, debt B and debt C is 10 million dollars, 15 million dollars and 20 million dollars in amount respectively, and their interest rates are 6%, 7% and 8% respectively. The three debts are guaranteed by three guarantors, namely GU, GV and GW. Suppose the risk free rate is 5%, and the volatility of the company is 30%.

Let us try to value the three guarantees.

First, calculate the debt value at maturity:

$$FV_A = 10 \times (1 + 6\%)^4 = 12.6248 \text{ million dollars}$$
$$FV_B = 15 \times (1 + 7\%)^4 = 19.6619 \text{ million dollars}$$
$$FV_c = 20 \times (1 + 8\%)^4 = 27.2098 \text{ million dollars}$$

Note that,

$$FV_A + FV_B = 12.6248 + 19.6619 = 32.2867 \text{ million dollars}$$
$$FV_A + FV_B + FV_C = 32.2867 + 27.2098 = 59.4965 \text{ million dollars}$$

Then, for the total guarantee,

$$S = 100;\ X = 59.4965;\ r = 4\%;\ T = 4;\ \sigma = 30\%$$

Based on the Black–Scholes model,

$$d1 = 1.4321,\ N(-d1) = 0.0761$$
$$d2 = 0.8321,\ N(-d2) = 0.2027$$
$$P = Xe^{-rT}N(-d2) - SN(-d1) = 2.6698 \text{ million dollars}$$

Therefore, the value of the above excess guarantee is 2.6698 million dollars, which represents the total value of the three guarantees.

calculate the proportion of these debts in total debt:

Based on the value at maturity,

$$P_A = 12.6248/59.4965 = 21.22\%$$
$$P_B = 19.6619/59.4965 = 33.05\%$$
$$P_C = 27.2098/59.4965 = 45.73\%$$

Then, the values of the three guarantees of GU, GV and GW are:

$$VU = 2.6698 \times 21.22\% = 0.5665 \text{ million dollars}$$

$$VV = 2.6698 \times 33.05\% = 0.8823 \text{ million dollars}$$

$$VW = 2.6698 \times 45.73\% = 1.2210 \text{ million dollars}$$

$$\text{Obviously, } VU + VV + VW = 2.6698.$$

For the case of difference in priority, the calculation is more complicated. For the previous case, suppose other things being equal, but the priority of three debts is different. Specifically, debt A is first, debt B is second and debt C is third in the repayment line. Table 1 lists the intrinsic values of guarantees for debt A, B and C under some values of the company at the debt maturity date.

Table 1 Company value and the intrinsic values of guarantee U, V and W

Company value	0	5	10	15	20	25	30
VU	12.62	7.62	2.62	0	0	0	0
VV	19.66	19.66	19.66	17.28	12.28	7.28	2.28
VW	27.21	27.21	27.21	27.21	27.21	27.21	27.21
Company value	35	40	45	50	55	60	65
VU	0	0	0	0	0	0	0
VV	0	0	0	0	0	0	0
VW	24.49	19.49	14.49	9.49	4.49	0	0

* 12.62, 19.66 and 27.21 are the approximate values rounded from 12.6248, 19.6619, 27.2098 respectively

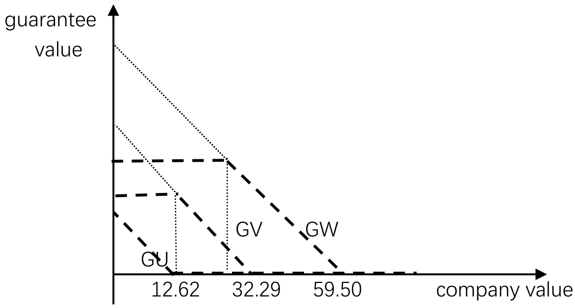


Fig. 8 Multiple debt guarantees and put options

The intrinsic values of the three guarantees can be described as what shown in Fig. 8. The curves of the intrinsic values show three put options with exercise price 12.62, 32.29 and 59.50 respectively; and defined by the options: guarantee U is the put option with exercise price 12.62, the guarantee V is the difference of the put options with exercise price 32.29 and 12.62, the guarantee W is the difference of the put options with exercise price 59.50 and 32.29. Therefore, we need to value the three options first.

The known variables are: S = 100; X = 12.62, 32.29 and 59.50 respectively; r = 4%; T = 4; σ = 30%. As calculated previously, When X = 59.50, the option put value is 2.6698 million dollars. Then, the values of the other two smaller put options can also be calculated based on Black–Scholes model.

When X = 12.62,

$$\begin{aligned}&\mathrm{d1} = 4.0165,\ \mathrm{N}(-\mathrm{d1}) = 0.000003\\&\mathrm{d2} = 3.4165,\ \mathrm{N}(-\mathrm{d2}) = 0.000317\\&\mathrm{P} = 0.00046 \text{ million dollars}\end{aligned}$$

When X = 32.29,

$$\begin{aligned}&\mathrm{d1} = 2.4507,\ \mathrm{N}(-\mathrm{d1}) = 0.0071\\&\mathrm{d2} = 1.8507,\ \mathrm{N}(-\mathrm{d2}) = 0.0321\\&\mathrm{P} = 0.1705 \text{ million dollars}\end{aligned}$$

Therefore, the values of the three guarantees are:

$$\begin{aligned}&\mathrm{VU} = 0.00046 \text{ million dollars}\\&\mathrm{VV} = 0.1705 - 0.00046 = 0.17 \text{ million dollars}\\&\mathrm{VW} = 2.6698 - 0.1705 = 2.4993 \text{ million dollars}\end{aligned}$$

Note that, VU + VV + VW = 0.00046 + 0.17 + 2.4993 = 2.6698, which is the same as the case of indifference in priority of debt claim.

No matter for the guarantor or the guaranteed, the value of the guarantee calculated by the option principle is undoubtedly an important decision-making basis.

3 Bankruptcy Cost: Concept and Measurement

The first section reveals that the quantification of bankruptcy cost is the key to derive the risk premium of the debt or loan. Now, let's consider it.

3.1 Discussion: Decisional Concept Versus Statistical Concept

The cost of bankruptcy is a hot topic in finance. In 1958, Modigliani and Miller established the MM model I[6] on capital structure, which ignited the enthusiasm of financial scholars to solve the problem of optimal capital structure.

The problem of optimal capital structure is well known for its toughness in finance. Modigliani and Miller proved with impeccable reasoning and rigorous logic that the optimal capital structure (debt ratio) can be found by trading off between tax shield and the bankruptcy cost related to the debt capital.

The quantification of the tax shield and the bankruptcy cost thus become the keys to open the door of optimal capital structure. In 1963, Modigliani and Miller published their MM model II, which provides a method to calculate tax shield. As the bankruptcy cost is the last fortress in the way to solve the problem of optimal capital structure, over the past 60 years, more and more related literature has carried out in this area.

However, for various reasons, most of those literature reveals some facts or data about bankruptcy cost, but none of them provides a convincing method or model to calculate the bankruptcy cost. Depending on those data-based findings, even numerous in quantity, it is no way to trade off between the tax shield and the bankruptcy cost, do not mention to determine the optimal capital structure.[7]

Based on Zhang [4], there has been a misunderstanding in bankruptcy cost in the relevant research, which hinders or blocks the problem to be solved.

For the purpose of trading-off between the tax shield and the bankruptcy cost, or decision making on capital structure, the bankruptcy cost should be a decisional concept, and needs to be considered in advance or beforehand, rather than being described or estimated afterwards. Put it another way, the bankruptcy cost

[6] We will discuss it in detail in Chapter "Tax Shield, Bankruptcy Cost and Optimal Capital Structure".

[7] Until 2009, Zhang Zhiqiang corrected the misunderstanding of the concept, established the bankruptcy cost model by using option pricing method, solved the problem of bankruptcy cost measurement, and further derived the optimal capital structure model.

is a forward-looking and decisional concept, rather than a backward-looking and statistical concept.

Just like finance as a science is different from that as a business, bankruptcy cost as a concept in theory is different from that in practice. In practice, bankruptcy cost is more an afterwards statistical concept, which refers to the actual expenses occur when a company goes through bankruptcy process. In theory, the bankruptcy cost is a decisional or beforehand concept for debt financing or capital structural decision, which is referred to the cost of bankruptcy risk or uncertainty that a company may not be able to repay its debt principal or interest one day in future and hence cannot go on its business.

Unfortunately, affected by the wide-spread perception, the academic concept of bankruptcy cost has also been added in too much statistical ingredients. Previous literature on the bankruptcy cost adopted the mothed of backward-looking statistical description with almost no exception, i.e., describe or calculate what and how much costs occur when the sample companies go to bankrupt. The well known findings by those statistical research is that bankruptcy cost covers direct and indirect costs.

The direct cost includes the costs occur when a company inters into the bankruptcy process, such as legal, accounting, and other professional fees, debt reorganization costs, etc. The indirect cost[8] includes those costs and expenses related to the potential bankruptcy (although maybe faraway), such as lost sales, rising costs, declining margins, additional inputs of management time and effort, etc.

For instance, Warner (1977) investigates 11 bankrupt railroad companies, and estimates that direct bankruptcy cost accounts for 5.3% of the company value.[9] Based on 31 highly levered companys, Andrade and Kaplan (1998) estimate that the total bankruptcy cost accounts for about 10% to 20% of the company value. Davydenko, Strebulaev, Zhao (2012) estimate the cost of default for an average defaulting company to be 21.7% of the market value of total assets.[10]

Undoubtedly, the statistics of bankruptcy costs vary a lot as percentages to company values across industries, stages (distance from the bankruptcy), markets, environments, etc. Actually, even a stable statistical bankruptcy cost can be found, it is not helpful to the decision making on capital structure or debt financing. For supporting the relevant decision, we need a forward looking or ex-ante bankruptcy cost model rather than a backward looking or ex-post statistical results.

Obviously, we need to revise the prevailing concept about bankruptcy cost before we can impartially measure it. Since bankruptcy risk reduces the company value, it should be all right to define it as the company value reduction resulted from the debt financing or bankruptcy risk. Please note that the decisional concept of bankruptcy

[8] A similar concept is financial distress cost, which is said to include also agency cost, etc. They are actually all aroused from the debt financing and covered by the bankruptcy cost. To avoid unnecessarily conceptual confusions, we will not use those new concepts.

[9] J. and Warner J. B. (1977), Bankruptcy Costs: Some Evidence, The Journal of Finance, 32: 337–347.

[10] Sergei A. Davydenko, Ilya A. Strebulaev, Xiaofei Zhao, A Market-Based Study of the Cost of Default, The Review of Financial Studies, v 25, n 10, 2012.

cost is not a predicted future cost either, but rather, it is a value today or it measures the present value or cost of the corresponding bankruptcy risk.

3.2 Valuing Bankruptcy Cost: An Option Pricing Method

Suppose now a company has (total) debt with book and market value X and maturity T (years) over which the debt value grows annually at risk free rate, r. As shown in Fig. 9, on the maturity date, if the company remains healthy, its value is larger than Xe^{rT}, and the debt value is Xe^{rT} (the right part of the dash line); if the company cannot repay the debt principal and interest, and the debt value is smaller than Xe^{rT}, the company has to go bankrupt; In such a case, the company value falls below its debt book value (Xe^{rT}) and the achievable debt value (fair or market value) equals to the company value (the left part of the dash line). The left part of the dash line in Fig. 9 depicts exactly such feature of debt.

While the left part of the dash line in Fig. 9 depicts the default or bankruptcy risk, how much is or how to value this risk? Generally, the cost needed to eliminate the risk is the (negative) value of the risk, i.e. the bankruptcy cost. From the view of the lender point, a debt guarantee can eliminate the bankruptcy risk, and it is the same from the point of the borrower. This implies that the value or cost of the bankruptcy is the guarantee of the debt. Fortunately, the discussions in previous section provide us enough insights for valuing the debt guarantee.

Specifically, the debt guarantee is equal to a put option in value. For the current case, a put option with exercise price of Xe^{rT} and maturity T (the debt maturity) removes exactly the bankruptcy risk. Hence, the value of the put is the bankruptcy cost. Put it another way, the bankruptcy cost can be derived by valuing the relevant put option. Note that the put to hedge the bankruptcy risk is a standard European option. Thanks to Black and Scholes (1973), the value of such a put is:

$$\mathrm{Put} = Xe^{-rT}N(-d2) - SN(-d1) \tag{8}$$

where, N(–d2) and N(–d1) are the cumulative probabilities under standard normal distribution when the Z values equal to –d2 and –d1 respectively, and,

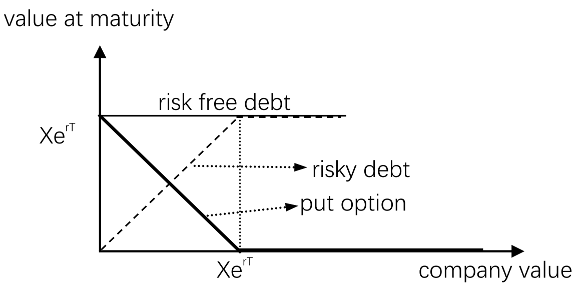


Fig. 9 Bankruptcy cost = put option

$$\mathrm{d1} = \frac{\ln(S/X) + \left(r + \sigma^2/2\right)T}{\sigma\sqrt{T}} \tag{9}$$

$$\mathrm{d2} = \frac{\ln(S/X) + \left(r - \sigma^2/2\right)T}{\sigma\sqrt{T}} = \mathrm{d1} - \sigma\sqrt{T} \tag{10}$$

Equations 8–10 are the standard Black–Scholes option pricing model. Where, σ is the volatility (standard deviation of annual return) of the underlying asset; X is the exercise price of the option which is a future value at the maturity of the option. Note that we use X as the present value of the debt in this chapter as well as in the following chapters, hence our X is equivalent to $\mathrm{Xe^{-rT}}$ in above standard option pricing model. As $\ln(\mathrm{S/X}) + \mathrm{rT} = \ln\left[\mathrm{S}/\left(\mathrm{Xe^{-rT}}\right)\right]$ and $\left(\sigma^2/2\right)\mathrm{T}/(\sigma\sqrt{T}) = \sigma\sqrt{T}/2$. Thus, replacing the $\mathrm{Xe^{-rT}}$ in Eqs. 8–10 with X, the bankruptcy cost (BC) is:

$$\mathrm{BC} = \mathrm{XN}(-\mathrm{d2}) - \mathrm{SN}(-\mathrm{d1}) \tag{11}$$

where, S and X are the current values of the company and its debt respectively. Note that X/S = L, which is the debt or leverage ratio of the company, then,

$$\mathrm{d1} = \frac{\ln(S/X)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} = -\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \tag{12}$$

$$\mathrm{d2} = \frac{\ln(S/X)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = -\frac{\ln(L)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = \mathrm{d1} - \sigma\sqrt{T} \tag{13}$$

where, σ is the volatility of the company return, or more precisely, is the standard deviation of annual return on the total asset of the company; T is the maturity (years) of the debt, defined as the average life of all debts in the company.

Different from other descriptive bankruptcy cost model, Eq. 11 is a decision-oriented bankruptcy cost model. Assume a typical company as a base case: the value of the company is 100, the average debt maturity T = 5, the volatility of the company return σ = 25%.[11] Based on Eq. 11, when the debt ratio L = 30%, 40%, 50%, 60%, 70%, 80%, the bankruptcy cost is 0.1653, 0.7275, 1.9885, 4.1393, 7.2553, 11.3230 respectively, which accounts roughly for 0.1653%, 0.7275%, 1.9885%, 4.1393%, 7.2553%, 11.3230% of the company value respectively. Figure 10 depicts the tendency of the bankruptcy cost changes along with the increasing of the debt ratio.

The above bankruptcy cost model was first published by Zhang [2],[3] , for convenience, can be referred to as ZZ bankruptcy cost model. The ZZ bankruptcy cost model is derived based on correct concepts and strict logic reasoning rather than chosen subjectively the form and the variables incorporated. In this sense, these objective models (Eqs. 11–13) are found by logic reasoning rather than built or

[11] The empirical range of σ; for equity is around 20–40%; so the reasonable range of σ; for the whole firm is about 15–35%. See Cumby et al. [1].

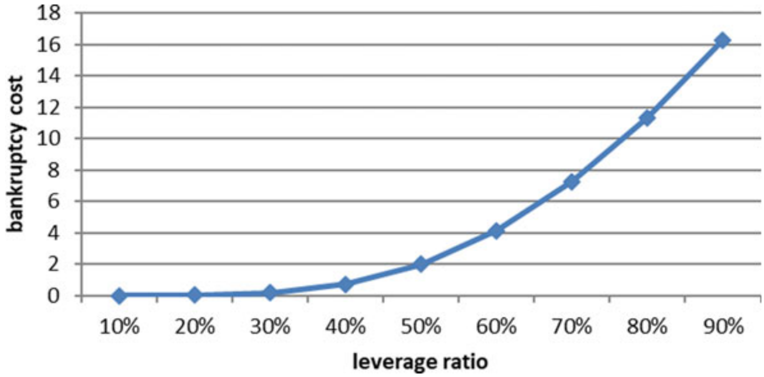


Fig. 10 The bankruptcy cost and the debt ratio in the base case

designed by imagination. This is actually the sign of the real solution to a problem, which also ensures the theoretical soundness and practical validness of the models. These are actually the same features possessed by the Black–Scholes option pricing model as well as all the previous and following ZZ models in this book.

4 Debt/Loan Pricing

Financial products are much different from others in pricing. Financial asset pricing aims at finding the reasonable required rate of return of the (capital) asset. The interest rate is the price of company debt or bank loan. Neglecting the transaction cost, the cost of debt is equal to the interest rate or price of loan (as a required rate of return by the bank or the investor). The reasonably required rate of return for providing debt capital should be matched with the risk of the debt. The risk is hard to measure and a rate of return to match it is even harder to find. This is the challenge and also the attractiveness of finance.

4.1 The Solution to Debt/Loan Pricing

As previously reveals, the key problem to determine the interest rate of a debt/loan is to find the default risk premium or simply referred to as risk premium.

The discussions in the first section and previous section reveal the relation between debt risk premium and bankruptcy cost as well as the model of the bankruptcy cost, as shown in Eqs. 7 and 11 respectively. Now, put Eq. 11 into 7,

$$c = -\ln\{1 - [\mathrm{XN}(-\mathrm{d2}) - \mathrm{SN}(-\mathrm{d1})]/\mathrm{X}\}/\mathrm{T}$$

$$= -\ln\{1 - [N(-d2) - (S/X)N(-d1)]\}/T$$
$$= -\ln\{[N(d2) + (S/X)N(-d1)]\}/T \quad (14)$$

Note that, debt ratio L = X/S, then S/X = 1/L, thus,

$$c = -\ln[N(d2) + N(-d1)/L]/T \quad (15)$$

As c is the risk premium of a risky loan, then, add it to the risk free rate is just the fair or appropriate interest rate, y, of a loan or debt, which is,

$$y = r - \ln[N(d2) + N(-d1)/L]/T \quad (16)$$

where, d1 and d2 are defined as Eqs. 12 and 13 respectively.

Now we have solved the problem of debt or loan pricing with elementary mathematics and simple process. Based on Eq. 16, the fair interest rate is an function of risk free rate (r), the volatility of return on the firm (σ) and the leverage or debt ratio of the firm (L) as well as the debt maturity (T).

Please note that $c = -\ln[N(d2) + N(-d1)/L]/T > 0$, hence the interest rate is larger than the risk free rate with the same maturity. This can be proved as following. Based on ZZ bankruptcy cost model and its derivation, bankruptcy cost (BC) is the value of a European put, and: BC = XN(–d2) – SN(–d1). This implies $XN(-d2) - SN(-d1) > 0$. Then, $Xe^{-cT} = X - [XN(-d2) - SN(-d1)] < X$ Then, $e^{-cT} < 1$. Because T > 0, hence c > 0.

4.2 Mutual Corroboration with Merton Model

The above derivation process was first published by Zhang [5], however, the model had been published by Robert C. Merton [4] through different derivation process. Robert C. Merton [4] worked out a loan risk premium model based exactly on the default risk.[12] Merton firstly worked out a parabolic partial differential equation for the debt F of a firm with value of V, which is written as:

$$\frac{1}{2}\sigma^2 V^2 F_{vv} + rVF_v - rF - F_t = 0 \quad (17)$$

where subscripts denote partial derivatives; t is the length of time to maturity; σ^2 is the variance of the return on the firm per unit time; r is risk free rate. An initial condition and two boundary conditions of Eq. 17 can be written as:

$$F(V, 0) = \min[V, B] \quad (18)$$

[12] This reflects that how correctness and preciseness Merton's concept and reasoning is.

$$F(V, t)/V \leq 1 \tag{19}$$

$$F(0, t) = f(0, t) = 0 \tag{20}$$

where B is the full value of the debt at its maturity; f is the current value of firm's equity. Merton found that Eq. 17 with conditions of 6.18–6.20 can be solved by using of the Black–Scholes model for a European call option and the basic relationship of $F = V - f$. The derived model of the debt value is:

$$F[V, t] = Be^{-rt}\left\{\phi\left[h_2\left(d, \sigma^2 t\right)\right] + \phi\left[h_1\left(d, \sigma^2 t\right)\right]/d\right\} \tag{21}$$

where

$$d \equiv Be^{-rt}/V \tag{22}$$

$$h_1\left(d, \sigma^2 t\right) \equiv -\left[\sigma^2 t/2 - \ln(d)\right]/(\sigma\sqrt{t}) \tag{23}$$

$$h_2\left(d, \sigma^2 t\right) \equiv -\left[\sigma^2 t/2 + \ln(d)\right]/(\sigma\sqrt{t}) \tag{24}$$

Transfer the debt value of Eq. 21 in terms of risk yields as:

$$R(t) - r = -\ln\left\{\phi\left[h_2\left(d, \sigma^2 t\right)\right] + \phi\left[h_1\left(d, \sigma^2 t\right)\right]/d\right\} \tag{25}$$

Then,

$$R(t) = r - \ln\left\{\phi\left[h_2\left(d, \sigma^2 t\right)\right] + \phi\left[h_1\left(d, \sigma^2 t\right)\right]/d\right\} \tag{26}$$

where d is the debt ratio or leverage, R(t) is the yield to maturity on the risky debt provided that the firm does not default, and $\exp[-R(t)t] = F(V, t)/B$. R(t) - r then is the (default) risk premium of the debt. Based on Eq. 26, Merton furthers his research on risk structure of interest rates in the paper.

Comparing Eqs. 16 and 26, it is easy to find that they are exactly the same equation. For convenience, they can be referred to as ZZ interest rate (debt/loan pricing) model and Merton interest rate (debt/loan pricing) model respectively. So, ZZ debt pricing model and Merton debt pricing model (Eqs. 16 and 26) can be mutual corroborated or mutual confirmed; they reach the exactly same result with different reasoning process. As a matter of fact, this is the best proof in scientific research, and better than any other mutual corroboration or mutual confirmation, such as the proof from sample data or big data, because any sample or big data is inevitably biased and incomplete, do not mention they are back ward looking rather than forward looking.

However, the mathematics used by Merton for the derivation of the model is much more complex than the common mathematics mastered by financial scholars and practitioners, while our reasoning process based on the ZZ bankruptcy model is just an application of elementary mathematics, and much easy for most financial scholars and practitioners to learn and to understand. This is an obvious advantage of the ZZ debt pricing model over the Merton debt pricing model.

4.3 Initial Numerical Test

The problem of debt pricing now has been solved. The correctness of our loan pricing model is further confirmed by the consistence with Merton model. Let us now test the effect of the model with numerical example.

Based on the ZZ debt pricing model, the fair interest rate is determined by risk free rate (r), the volatility of return on the firm (σ) and the leverage or debt ratio of the firm (L) as well as the debt maturity (T). Assume a base case firm, the maturity of the debt is 5 years; the volatility of the firm return is 25%. Further assume the risk free rate is 4%. Now we calculate the fair interest rates every time based on the vary of the debt ratio and one of the another three variables, as shown in Tables 2, 3 and 4.

The results in Tables 2, 3 and 4 show clearly, the fair interest rate is an increasing function of risk free rate (r), the volatility of return on the firm (σ) and the leverage or debt ratio of the firm (L). However, the effect of the debt maturity (T) does not keep consistent across different leverages, which needs further discussion.

Corporate debt is equivalent to bank loan, vice versa. So the Merton or ZZ debt pricing model can also be used to determine the interest rate of bank loan, and can also be referred to as Merton or ZZ loan pricing or interest rate model. However, shortly after Merton published his debt/loan pricing model, scholars found its two "defects", just as the interest rates shown in Tables 2, 3 and 4: (1) they seem unrealistically

Table 2 Interest rates when risk free rate varies from 1 to 10%

		Debt ratio								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
Risk free rate	1%	1.00%	1.01%	1.11%	1.37%	1.81%	2.43%	3.19%	4.05%	4.99%
	2%	2.00%	2.01%	2.11%	2.37%	2.81%	3.43%	4.19%	5.05%	5.99%
	3%	3.00%	3.01%	3.11%	3.37%	3.81%	4.43%	5.19%	6.05%	6.99%
	4%	4.00%	4.01%	4.11%	4.37%	4.81%	5.43%	6.19%	7.05%	7.99%
	5%	5.00%	5.01%	5.11%	5.37%	5.81%	6.43%	7.19%	8.05%	8.99%
	6%	6.00%	6.01%	6.11%	6.37%	6.81%	7.43%	8.19%	9.05%	9.99%
	7%	7.00%	7.01%	7.11%	7.37%	7.81%	8.43%	9.19%	10.05%	10.99%
	8%	8.00%	8.01%	8.11%	8.37%	8.81%	9.43%	10.19%	11.05%	11.99%
	9%	9.00%	9.01%	9.11%	9.37%	9.81%	10.43%	11.19%	12.05%	12.99%
	10%	10.00%	10.01%	10.11%	10.37%	10.81%	11.43%	12.19%	13.05%	13.99%

Note: T = 5; σ = 25%

Table 3 Interest rates when time to maturity varies from 1 to 10 years

		Debt ratio								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
Time to maturity	1	4.00%	4.00%	4.00%	4.00%	4.03%	4.24%	5.03%	6.87%	10.04%
	2	4.00%	4.00%	4.00%	4.04%	4.23%	4.75%	5.74%	7.24%	9.21%
	3	4.00%	4.00%	4.02%	4.14%	4.47%	5.09%	6.02%	7.23%	8.67%
	4	4.00%	4.00%	4.06%	4.25%	4.66%	5.30%	6.14%	7.15%	8.28%
	5	4.00%	4.01%	4.11%	4.37%	4.81%	5.43%	6.19%	7.05%	7.99%
	6	4.00%	4.03%	4.17%	4.47%	4.93%	5.52%	6.20%	6.96%	7.76%
	7	4.00%	4.05%	4.23%	4.56%	5.02%	5.57%	6.20%	6.88%	7.57%
	8	4.00%	4.07%	4.28%	4.63%	5.09%	5.61%	6.19%	6.80%	7.42%
	9	4.01%	4.10%	4.34%	4.70%	5.14%	5.64%	6.18%	6.73%	7.28%
	10	4.01%	4.13%	4.39%	4.75%	5.18%	5.66%	6.16%	6.66%	7.17%

Note: $r = 4\%$; $\sigma = 25\%$

Table 4 Interest rates when the volatility of firm return varies from 5 to 50%

		Debt ratio								
		10%	20%	30%	40%	50%	60%	70%	80%	90%
Volatility of firm return	5%	4.00%	4.00%	4.00%	4.00%	4.00%	4.00%	4.00%	4.02%	4.22%
	10%	4.00%	4.00%	4.00%	4.00%	4.00%	4.02%	4.13%	4.42%	5.00%
	15%	4.00%	4.00%	4.00%	4.01%	4.07%	4.24%	4.59%	5.16%	5.93%
	20%	4.00%	4.00%	4.02%	4.10%	4.33%	4.73%	5.31%	6.06%	6.93%
	25%	4.00%	4.01%	4.11%	4.37%	4.81%	5.43%	6.19%	7.05%	7.99%
	30%	4.00%	4.08%	4.34%	4.82%	5.48%	6.28%	7.17%	8.12%	9.10%
	35%	4.02%	4.24%	4.74%	5.45%	6.31%	7.25%	8.23%	9.24%	10.25%
	40%	4.08%	4.54%	5.30%	6.23%	7.25%	8.30%	9.37%	10.42%	11.46%
	45%	4.22%	4.98%	6.00%	7.13%	8.29%	9.44%	10.56%	11.65%	12.71%
	50%	4.46%	5.56%	6.84%	8.15%	9.42%	10.64%	11.82%	12.94%	14.01%

Note: $r = 4\%$; $T = 5$

too low; and (2) they are negatively related with the loan maturity when the debt ratio gets high (Table 3) which contradicts with common intuition. So the model is neither believed to be an effective solution to loan pricing, nor widely used in banking practice. Despite its pioneering innovation, it is even absent from most commercial banking textbooks. Consequently, commercial bankers have to use some too simple methods to make loan pricing decisions, such as markup on cost,[13] etc.

[13] The markup method in loan pricing can hardly be justified in theory, because the price or interest rate of a loan should be related to its (default) risk rather than its cost.

5 Application Issues of the Model

The previous research and above numerical test reveal two "obvious defects" of the Merton or ZZ loan pricing or interest rate model, although they are mutual corroborated with each other. On the other hand, the process to derive the model is simple and shows clearly that no special assumptions needed for the reasoning or derivation process. However, as a common sense, based on the regular preconditions and strict logic and quantitative process, the model should be correct and useful. Now, let's try to solve the mystery. The analyses and calculations in the following two sections use the similar method as in Zhang [6].

5.1 Incremental Debt Consideration

As a matter of fact, the calculation in Tables 2, 3 and 4 implies that the firm has zero debt before the loan under consideration. Such assumption is incorrect for most (if not all) firms. When a firm is provided a loan, the loan more often is just an incremental debt rather than total debt for the firm. For the convenience of demonstration, assume the considered loan increases the firm's debt ratio by 10%, corresponding to Tables 2, 3 and 4, we can recalculate the interest rates of the loan as shown in Tables 5, 6 and 7 respectively.

The interest rates in Tables 5, 6 and 7 become higher than their corresponding values in Tables 2, 3 and 4. For instance, when the risk free rate is around 3–4%, if the loan increases the debt ratio from 50 to 60%, the fair interest rate varies from 7.52 to 8.52%, or around 8%. The unrealistic low interest rate disappears; the interest rates in Tables 5, 6 and 7 seem regular and acceptable. Obviously, the wrong results in Tables 2, 3 and 4 come from the misuse of the loan pricing model rather than the model itself.

5.2 Transaction Cost Consideration

Conceptually, the interest rates in Tables 5, 6 and 7 are just break-even rates for banks to compensate their risks from loans. There are also transaction costs in reality, i.e., banks have some operating costs, and banks need some profits as well, otherwise they are not able to continue their businesses over long run.

These costs and profits can push the fair interest rates further higher than the rates in Tables 5, 6 and 7. To demonstrate conveniently, simply assuming the necessarily annual operating costs and profits equal to 1.5% and 0.5% of total loans respectively, i.e., for covering operating costs and expenses and acquiring moderate profit, the interest rates will increase by a further 2% in total. So we have the final fair interest rates as shown in Tables 8, 9 and 10 (corresponded to Tables 2, 3 and 4 respectively).

Table 5 Interest rates of 10% incremental debt when risk free rate varies from 1 to 10%

		Debt ratio								
		0–10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	80–90%
Risk free rate	1%	1.00%	1.03%	1.30%	2.14%	3.59%	5.52%	7.74%	10.10%	12.48%
	2%	2.00%	2.03%	2.30%	3.14%	4.59%	6.52%	8.74%	11.10%	13.48%
	3%	3.00%	3.03%	3.30%	4.14%	5.59%	7.52%	9.74%	12.10%	14.48%
	4%	4.00%	4.03%	4.30%	5.14%	6.59%	8.52%	10.74%	13.10%	15.48%
	5%	5.00%	5.03%	5.30%	6.14%	7.59%	9.52%	11.74%	14.10%	16.48%
	6%	6.00%	6.03%	6.30%	7.14%	8.59%	10.52%	12.74%	15.10%	17.48%
	7%	7.00%	7.03%	7.30%	8.14%	9.59%	11.52%	13.74%	16.10%	18.48%
	8%	8.00%	8.03%	8.30%	9.14%	10.59%	12.52%	14.74%	17.10%	19.48%
	9%	9.00%	9.03%	9.30%	10.14%	11.59%	13.52%	15.74%	18.10%	20.48%
	10%	10.00%	10.03%	10.30%	11.14%	12.59%	14.52%	16.74%	19.10%	21.48%

Note: $r = 4\%$; $\sigma = 25\%$

Table 6 Interest rate of 10% incremental debt when time to maturity varies from 1 to 10 years

		Debt ratio								
		0–10%	10–20%	20–30%	30-40%	40–50%	50–60%	60–70%	70–80%	80–90%
Time to maturity	1	4.00%	4.00%	4.00%	4.00%	4.14%	5.31%	9.77%	19.76%	35.34%
	2	4.00%	4.00%	4.01%	4.16%	5.00%	7.36%	11.68%	17.75%	24.96%
	3	4.00%	4.00%	4.06%	4.49%	5.79%	8.20%	11.61%	15.70%	20.15%
	4	4.00%	4.01%	4.17%	4.84%	6.29%	8.47%	11.19%	14.20%	17.34%
	5	4.00%	4.03%	4.30%	5.14%	6.59%	8.52%	10.74%	13.10%	15.48%
	6	4.00%	4.06%	4.44%	5.37%	6.76%	8.47%	10.33%	12.25%	14.16%
	7	4.00%	4.10%	4.58%	5.54%	6.86%	8.37%	9.97%	11.58%	13.16%
	8	4.00%	4.14%	4.70%	5.67%	6.90%	8.26%	9.65%	11.04%	12.38%
	9	4.01%	4.19%	4.81%	5.77%	6.92%	8.15%	9.38%	10.59%	11.75%
	10	4.01%	4.25%	4.90%	5.85%	6.92%	8.03%	9.14%	10.20%	11.23%

Note: $r = 4\%$; $\sigma = 25\%$

Table 7 Interest rate of 10% incremental debt when the volatility of firm return varies from 5 to 50%

		Debt ratio								
		0–10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	80–90%
Volatility of firm return	5%	4.00%	4.00%	4.00%	4.00%	4.00%	4.00%	4.00%	4.17%	5.81%
	10%	4.00%	4.00%	4.00%	4.00%	4.01%	4.12%	4.75%	6.49%	9.60%
	15%	4.00%	4.00%	4.00%	4.04%	4.29%	5.11%	6.72%	9.12%	12.06%
	20%	4.00%	4.00%	4.05%	4.36%	5.22%	6.74%	8.82%	11.27%	13.90%
	25%	4.00%	4.03%	4.30%	5.14%	6.59%	8.52%	10.74%	13.10%	15.48%
	30%	4.00%	4.15%	4.87%	6.26%	8.13%	10.26%	12.50%	14.75%	16.95%
	35%	4.02%	4.46%	5.73%	7.59%	9.72%	11.95%	14.16%	16.30%	18.36%
	40%	4.08%	4.99%	6.82%	9.03%	11.32%	13.58%	15.75%	17.80%	19.75%
	45%	4.22%	5.74%	8.06%	10.52%	12.92%	15.18%	17.30%	19.29%	21.15%
	50%	4.46%	6.67%	9.40%	12.05%	14.51%	16.77%	18.85%	20.78%	22.56%

Note: r = 4%; T = 5

Table 8.8 Interest rates (including the costs and profits of banks) of 10% incremental debt when risk free rate varies from 1 to 10%

		Debt ratio								
		0–10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	80–90%
Risk free rate	1%	3.00%	3.03%	3.30%	4.14%	5.59%	**7.52%**	9.74%	12.10%	14.48%
	2%	4.00%	4.03%	4.30%	5.14%	6.59%	**8.52%**	10.74%	13.10%	15.48%
	3%	5.00%	5.03%	5.30%	6.14%	7.59%	**9.52%**	11.74%	14.10%	16.48%
	4%	6.00%	6.03%	6.30%	7.14%	8.59%	**10.52%**	12.74%	15.10%	17.48%
	5%	7.00%	7.03%	7.30%	8.14%	9.59%	**11.52%**	13.74%	16.10%	18.48%
	6%	8.00%	8.03%	8.30%	9.14%	10.59%	**12.52%**	14.74%	17.10%	19.48%
	7%	9.00%	9.03%	9.30%	10.14%	11.59%	**13.52%**	15.74%	18.10%	20.48%
	8%	10.00%	10.03%	10.30%	11.14%	12.59%	**14.52%**	16.74%	19.10%	21.48%
	9%	11.00%	11.03%	11.30%	12.14%	13.59%	**15.52%**	17.74%	20.10%	22.48%
	10%	12.00%	12.03%	12.30%	13.14%	14.59%	**16.52%**	18.74%	21.10%	23.48%

Note: $r = 4\%$; $\sigma = 25\%$; The bold figures are used as the dividing line. The left part represents the feasible area of loan, and the right (including the bold figures) part represents the infeasible area of loan because of the higher default risk

Table 9 Interest rates (including the costs and profits of banks) of 10% incremental debt when time to maturity varies from 1 to 10 years

		Debt ratio								
		0–10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	80–90%
Time to maturity	1	6.00%	6.00%	6.00%	6.00%	6.14%	7.31%	**11.77%**	21.76%	37.34%
	2	6.00%	6.00%	6.01%	6.16%	7.00%	**9.36%**	13.68%	19.75%	26.96%
	3	6.00%	6.00%	6.06%	6.49%	7.79%	**10.20%**	13.61%	17.70%	22.15%
	4	6.00%	6.01%	6.17%	6.84%	**8.29%**	10.47%	13.19%	16.20%	19.34%
	5	6.00%	6.03%	6.30%	7.14%	**8.59%**	10.52%	12.74%	15.10%	17.48%
	6	6.00%	6.06%	6.44%	7.37%	**8.76%**	10.47%	12.33%	14.25%	16.16%
	7	6.00%	6.10%	6.58%	7.54%	**8.86%**	10.37%	11.97%	13.58%	15.16%
	8	6.00%	6.14%	6.70%	7.67%	**8.90%**	10.26%	11.65%	13.04%	14.38%
	9	6.01%	6.19%	6.81%	7.77%	**8.92%**	10.15%	11.38%	12.59%	13.75%
	10	6.01%	6.25%	6.90%	7.85%	**8.92%**	10.03%	11.14%	12.20%	13.23%

Note: $r = 4\%$; $\sigma = 25\%$; The bold figures are used as the dividing line. The left part represents the feasible area of loan, and the right (including the bold figures) part represents the infeasible area of loan because of the higher default risk

Table 10 Interest rates (including the costs and profits of banks) of 10% incremental debt when the volatility of firm return varies from 5 to 50%

		Debt ratio								
		0–10%	10–20%	20–30%	30–40%	40–50%	50–60%	60–70%	70–80%	80–90%
Volatility of firm return	5%	6.00%	6.00%	6.00%	6.00%	6.00%	6.00%	6.00%	6.17%	**7.81%**
	10%	6.00%	6.00%	6.00%	6.00%	6.01%	6.12%	6.75%	**8.49%**	11.60%
	15%	6.00%	6.00%	6.00%	6.04%	6.29%	7.11%	**8.72%**	11.12%	14.06%
	20%	6.00%	6.00%	6.05%	6.36%	7.22%	**8.74%**	10.82%	13.27%	15.90%
	25%	6.00%	6.03%	6.30%	7.14%	**8.59%**	10.52%	12.74%	15.10%	17.48%
	30%	6.00%	6.15%	6.87%	**8.26%**	10.13%	12.26%	14.50%	16.75%	18.95%
	35%	6.02%	6.46%	7.73%	**9.59%**	11.72%	13.95%	16.16%	18.30%	20.36%
	40%	6.08%	6.99%	**8.82%**	11.03%	13.32%	15.58%	17.75%	19.80%	21.75%
	45%	6.22%	7.74%	**10.06%**	12.52%	14.92%	17.18%	19.30%	21.29%	23.15%
	50%	6.46%	**8.67%**	11.40%	14.05%	16.51%	18.77%	20.85%	22.78%	24.56%

Note: r = 4%; T = 5;The bold figures are used as the dividing line. The left part represents the feasible area of loan, and the right (including the bold figures) part represents the infeasible area of loan because of the higher default risk

Obviously, Tables 8, 9 and 10 represent the correct usage of the Merton or ZZ loan pricing or interest rates model. The interest rates in Tables 8, 9 and 10 seem not only higher than their corresponding values in Tables 2, 3 and 4 but also higher (rather than lower) than the common or regular loan interest rates in reality.

This may be caused by the effective risk management and the over competition in banking industry, or simply lack of a clear benchmark or an effective loan pricing model in actual loan decisions. An important feature of loans is that as a product, they are strictly homogenous or indifferent for customers. This implies an inevitable fiercer competition among loan providers (banks etc.), which is a force to push the interest rate down. On the other hand, the uses of credit enhancement measures are very common in loan market, such as guarantees, etc., which is a force to pull the interest rate down.

5.3 The Safe-Guard Line

Anyway, the interest rates in Tables 8, 9 and 10 are the theoretical prices of credit loan without guarantees. They are a little higher than the actual interest rates. This proves that the Merton or ZZ interest rates or loan pricing model has no such a "flaw" or disadvantage that the interest rates derived from the model are too low.

However, in Table 9, in the columns with high debt ratio, the interest rates are still negatively related with the debt maturity. That means, after the adjusts based on the incremental consideration and transaction cost consideration, the second "flaw" or disadvantage of the model still exists. Let us move on to this second "flaw" now.

In Tables 8, 9 and 10, we can see some unusual "high interest rates", as high as 15%, 20% or even close to 30%. There are actually many types of players in financial market; each kind of players has its own territory. Some players domain an area featured as high risk and high return, such as venture capitals, whereas some players domain an area featured as low risk and low return, such as commercial banks.

So the returns or interest rates on loans provided by banks should not be too high. As the capital provider of the capital market, banks are characterized by providing "low-risk and low-cost" debt capital. As a matter of fact, the unusual "high interest rate" means that this is not the area that suitable to banks. Put it another way, the unusual "high interest rate" means that banks should not take the risk to provide the loans to the relevant customers.

This further means that the loan interest rate model or the results derived from it, like Tables 8, 9 and 10, can provide not only the fair interest rates, but also the boundaries or limits for banks to provide loans. Put it another way, this is not the disadvantage of the model, but rather, it is the advantage of the model. As an interest rate or loan pricing model, it works well and can play a role beyond its duty.

The normal rate of return of stock investment[14] is usually about 10%. The normal return of loan, i.e. interest rate, should be slower than that of stock investment, because a loan (investment) is safer than a stock (investment). How much should it specifically be lower? Banks can control the tightness scale according to internal and external conditions during specific application, such as the supply and demand of capital in the market, the credit risk level in average in the market. Limited to the theme and space, it will not be discussed further here.

As an example, assuming that the loan interest rate should be two percentage points lower in average, then the interest rate is 8% at most, which is 4 percentage points higher than the risk-free interest rate. According to the standard of "risk-free interest rate + 4%", a line of whether to lend or not can be set up, as shown in bold figures in Tables 8 and 10. With bold numbers as the boundaries, loans can be provided in the left area; bold numbers and the right area are where the loan should be refused.

Thus, Merton or ZZ interest rates model is a real solution to the loan pricing and loan decision-making related problems, besides the interest rate determination, it can also set up a safe-guard line for doing loan business. In the area where loans are feasible, there are no unusually high interest rates; More importantly, the interest rate is no longer negatively related with loan maturity. The second flaw is gone too!

Further, since it can be used to decide the feasibility of the loan, it can naturally be used to determine the loan size, that is, the loan amount being accepted or provided at most. For example, in Table 9, suppose the debt ratio of a company is limited to 50%. If the company applies for a loan of 80 million dollars with a maturity of two years when the debt ratio is 40%, the loan will increase the debt ratio of the company to 60%. Then the bank can approve at most 32 million dollars[15] at the interest rate of 7.00%; otherwise, the debt ratio will exceed 50%, and the interest rate will exceed the safe-guard line.

Therefore, Merton or ZZ interest rates model is really a versatile problem-solving tool, and can of course be used in practice. It can solve all the three core decision-making problems in banking or loan business, that is, to determine the acceptance or approval, the amount and the interest rate of the loan. Moreover, the model is flexible enough for banks to adjust the tightness of risk control (rather than an arbitrary boundary of risk free rate + 4%) according to the market environment and needs of the bank.

For example, when the market situation is bad and the risk of default increases, it can be strictly controlled, and the standard of loan approval can be more tight, such as that the interest rate is less than "risk-free interest rate plus 3%"; In the case of fiercer competition, in order to expand or maintain the customer base or market share, it can be approved according to a looser standard, such as "risk-free interest

---

[14] See Aswath Damodaran, Equity Risk Premiums (ERP): Determinants, Estimation and Implications—The 2008–2015 Edition at Damodaran Online (http://pages.stern.nyu.edu/~adamodar/).

[15] Note that the answer is 32 million dollars, rather than: 80 million dollars/2 = 40 million dollars. The calculation is a little more complicated here. Readers interested in solving the problem can try to work out by yourselves. Such an exercise will help you to understand the following part better.

rate plus 5%" etc. Obviously, the Merton or ZZ interest rates model can support banks by making more precise calculation and decisions, and so that maximizing the business potential.

6 Determine Incremental Leverage

The problem of loan pricing has already been solved by the Merton or ZZ loan pricing model; the widely perceived serious flaws are proved to be resulted from misunderstood and misused of the model. What left in loan or debt pricing now is to thoroughly understand and properly use the model (for evaluating the feasibility and deciding the size as well as the interest rate of a loan).

6.1 Logic of Incremental Leverage

In the calculation of Tables 8, 9 and 10, the company's incremental leverage or debt ratio is simply assumed without specific discussion. It actually need to calculate and a couple of specific problems need to solve in the actual application.

First of all, it should be clear that the debt ratio here should be the debt ratio including the loan applied, rather than the original debt ratio before applying for the loan. It is easy to understand by an extreme case, i.e., the company applies for a loan when it has no debt. Obviously, we cannot put the original debt ratio of 0% into the model.

Further, if the debt includes the potential loan, when calculating the debt ratio, the company value must also include the value of the loan, rather than the original value of the company when applying for loans. As a consequence, the company's value will also increase synchronously, which will increase the workload of debt ratio calculation.

For example, if the company increases equity financing at the same time, so that the debt ratio may remain unchanged or even decline, the calculation based on the increasing the debt ratio caused by the new loan cannot be applied. Therefore, the calculation of the incremental leverage is a problem in the application of the Merton or ZZ loan pricing model, and need to be considered and solved correctly.

Specifically, the measurement of debt ratio should ensure that the debt ratio increases after the loan provided, rather than remains unchanged or decreases. In order to specifically measure the changes in the company's debt ratio caused by loans, the company value used in calculating the company's debt ratio should remain unchanged. According to the above analysis, the total value of the company should include the value of the current debt financing and equity financing.

Then, on the basis of the above company value, the company's debt ratio before and after the loan is calculated by excluding the loan and including the loan as the debt value respectively. In this way, the impact of changes in the company's total value is excluded, the impact of loans on the company's debt ratio is measured separately, and it is also ensured that the company's debt ratio increases due to loans.

Moreover, calculating in such a way not only maintains the consistency of the company's value, but also ensures the consistency of the absolute value and relative ratio of incremental debt.

6.2 Numerical Illustration of Incremental Leverage

It might be helpful to consider a common application scenario for discussion and verification. For example, company M is considering to invest in project F, it needs an additional 8 million dollars for the investment. Besides the debt ratio L, the other three variables that determine the loan interest rate are: the risk-free interest rate r = 4%, the company value volatility σ = 20%, and the loan maturity T = 3 years.

The original value of the company was 12 million dollars, of which the debt capital accounted for 50%, that is, the equity and debt capital were 6 million dollars respectively. The additional investment of 8 million dollars can maintain the original debt ratio, that is, the new loan is 4 million dollars (50%), and the new equity financing is 4 million dollars. Any other ratio may also be used. For example, loans of 1 million dollars, 2 million dollars, 3 million dollars, 4 million dollars, 5 million dollars, 6 million dollars, 7 million dollars and 8 million dollars can be used, and the gap be made up by equity financing, and then the equity and debt capital can jointly meet the needs of the investment.

Suppose the reasonable operating cost and profit are 2% and 1% of the loan amount respectively. Calculate the corresponding debt ratio and loan interest rate under these schemes, as shown in Table 11.

In Table 11, project F is priced at historical cost; that is, the value of the project is equal to the capital invested, or the NPV of the project is zero. The value of the company is fixed after joining the project F, that is, 20 million dollars. Therefore, as long as there are loans, the debt ratio is higher than the original debt ratio, and the higher ratio is consistent with the ratio of new liabilities or loans to the value of the company. The incremental interest rate in the table is calculated based on 4.00% when the debt ratio is 30% (this is the value after rounding to two decimal places, not exactly 4.00%; more decimal places are taken into account in the actual subsequent calculations).

For example, If the loan is 5 million dollars, the company's debt will reach 11 million dollars (= 600 + 500) and the debt ratio will reach 55% (= 1100/2000). When the debt ratio is 55%, the total loan interest rate is 4.27%. Therefore, the incremental debt or loan interest rate of 5 million is: (4.27%*55% - 4.00%*30%) / (55% - 30%) = 4.59%. Plus 2% of normal operating costs and 1% of normal operating profit, the loan interest rate is 7.59%.

To sum up, we should calculate the debt ratio before and after the loan according to the company value including the loan, and use the model to calculate the loan interest rate under the two debt ratios, and then get the incremental loan interest rate

Table 11 A calculation example of interest rate of a new loan

Loan amount	0	100	200	300	400	500	600	700	800
Original debt ratio	30%	30%	30%	30%	30%	30%	30%	30%	30%
New debt ratio	30%	35%	40%	45%	50%	55%	60%	65%	70%
Loan interest rate	4.00%	4.01%	4.02%	4.06%	4.14%	4.27%	4.46%	4.73%	5.09%
Incremental interest rate	No loan	4.04%	4.09%	4.18%	4.34%	4.59%	4.92%	5.36%	5.91%
Plus 2.0% cost	No loan	6.04%	6.09%	6.18%	6.34%	6.59%	6.92%	7.36%	7.91%
Plus 1.0% profit	No loan	7.04%	7.09%	7.18%	7.34%	7.59%	7.92%	8.36%	8.91%

Note
Loan amount is in million dollars
Original debt ratio = Original debt/new company value = 6/20
New debt ratio = (6 + loan amount)/20
Loan interest rate, worked out by the loan pricing model in this chapter
Incremental interest rate, based on 4.00% when the debt ratio is 30%
Plus 2.0% cost, break even interest rate
Plus 1.0% profit, interest rate based on target profit

according to the two loan interest rates. This is the risk cost break even interest rate; the final interest rate can be obtained by adding operating cost and target profit to this interest rate.

7 Summary

The discussion of this chapter is typical in the solution of financial problems. Obviously, the reason why this chapter can solve the problem of debt or loan pricing is largely due to the ingenious use of option pricing method. In other words, option pricing method is indeed a powerful tool to solve financial problems. Since the option pricing problem has been basically solved (by the Black–Scholes model), to solve the relevant financial problems, it depends on whether we can correctly understand the relevant problems and effectively use the option pricing model.

Based on the discussion of this chapter as well as the previous chapters, there are roughly three steps to solve financial problems.

First of all, we should correctly understand the problem. If the problems are not understood correctly, it is impossible to solve the problem. Practice in reality represents some kind of understanding, but it may not represent correct understanding. Therefore, being able to imitate the calculation or decision-making in current practice in reality does not mean the solution of the problem. For example, if we consider that the loan interest rate or debt capital cost is equal to 'the actual risk-free interest rate + inflation risk compensation + liquidity risk compensation + default risk compensation + term risk compensation", i.e., we believe that Eq. 2 is correct, then the problem of debt cost or loan interest rate can never be solved correctly.

Secondly, the solution of financial problems relies on theoretical solution. Since the timeliness is often emphasized in practice, practical methods may not represent the solution of the problem, but often the temporary or convenient method. Of course, academic or theoretical models without rigorous logic support do not represent the real solution either. The basic standard of qualified theoretical solution is that based on the basic premises, and is derived through strict logic, such as the Morton or ZZ loan pricing or interest rate model in this chapter. Here, the basic premises refer to the common important premises existing at different times and places. For example, the levered company in this chapter will have bankruptcy risk, transaction cost, etc.

The reason why it is marked by theoretical solution is that the theoretical solution based on the basic premises often represents the most difficult step. With such basic solution, if some other special problems, special factors or special requirements need to be considered in a specific application, it may not be difficult to take them into account. Such as tighter or looser in the loan control by applying Morton or ZZ loan interest rate model. At the same time, these specific application problems often have a wide variety, and with the application of theory or model, new problems continue to appear or to find, and it is impossible to determine when to thoroughly discover and solve all application problems. Therefore, the solution of financial problems can only be judged by the discovery of theoretical model based on rigorous logic or reasoning.

Thirdly, three conditions must be met in order to obtain correct or ideal results in applying the theoretical model. First, the problem must be understood correctly. Second, the model must be understood and applied correctly. Third, the variable values input into the model must be correct. It is easy to get the incorrect result without a correct understanding of the relevant problem and the relevant model. The application of Morton or ZZ loan interest rate model is a good example. The loan interest rates derived is incorrect if theoretical model is used in a wrong way. However, after correcting the errors in the application, it is found that the model is completely correct and very effective. In this regard, appropriate application research is obviously needed to effectively use the theoretical model. For example, for the application of Morton or ZZ loan interest rate model, we should firstly study how to correctly calculate the incremental change of debt ratio, just as what we do in this section.

Anyway, we find the theoretical solution of debt or loan pricing; we prove that this theoretical model is powerful enough to solve all the three problems for commercial banks to decide the acceptance, the size and the interest rate of a loan. Therefore, with the basic model of loan interest rate, based on the volatility, debt ratio, loan life of the borrowing company as well as risk-free interest rate in market, the loan decision-making related problems can be solved comprehensively. Finally, the fair interest rate is just the fair discount rate for the relevant debt capital; the model and method here thus can also be used to solve the problem of determine the discount rate for debt capital.

References

1. Zhang Z, Yu M. The value of loan guarantees based on option pricing model. China Valuation. 2013 June;6–10.
2. Zhang Z. ZZ financial discovery, University of International Business and Economics Press. 2008 November.
3. Zhang Z. Shufang Xiao, Tax Shield, Bankruptcy Cost and the Optimal Capital Structure, Accounting Research. 2009;4:47–55.
4. Robert C. Merton, on the pricing of corporate debt: the risk structure of interest rates. J Financ. 1974;29:449–70.
5. Zhiqiang Z, Basic model of loan pricing. China Asset Appraisal. 2017;(4):26–29.
6. Zhang Z. Research on the application of loan pricing model. China Appraisal J. 2021;9:14–20.

Further Readings

7. Zhang Z, Finance—fundamental problems and solutions. Springer; 2013.
8. Zhang Z. Option Theory and Corporate Finance. Huaxia Publishing House:Beijing. 1999 June.

Capital Asset Pricing: An Easy and Unified Solution

Contents

In previous chapter, we derive a discount rate model for debt based on the ZZ bankruptcy cost model. In this chapter, we prove that a discount rate model for total asset can be derived based on the ZZ debt pricing model, which is exactly the new asset pricing model derived in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing". Based on the discount rate models for debt and that for total asset, we further work out a discount rate model for equity, or equity pricing model.

Therefore, we can derive the three discount rates in two ways: one is calculating the discount rates based on the same basic ZZ CAPM by input debt, equity and company volatility respectively; the other is inputting the company volatility respectively into debt, equity and total asset pricing models. The former uses one model but three volatilities; the later uses three models but one volatility. The discount rates are calculated based on the structure of "risk free rate + risk premium" for sure in both ways. Further discussion reveals that the later way is sounder in theory, it is more consistent in logic and variables with the solutions to other problems in the subsequent chapters.

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_9

1 Systematic Risk and Total Risk

Previous chapter solves problem of how to determine discount rate for debt capital, and the model (Merton or ZZ debt interest rate or debt pricing model) is also helpful for the loan decisions in commercial banks. Now it is time to solve the problem to determine discount rate for equity capital, or the problem of equity pricing.

1.1 Equity Pricing with Sharpe CAPM

The prevailing method for equity pricing nowadays is Sharp CAPM. As introduced in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing", the following is the reduced form of the model,

$$r_i = r + \beta_i(r_m - r) \tag{1}$$

where,

r_i is the expected return or required rate of return on the ith capital asset;
r is the risk free rate of interest, or the yield to maturity on government bonds;
β_i (the beta coefficient) is the sensitivity of the asset returns to market returns;
r_m is the expected return of the market;
r_m-r is known as the average market risk premium.

The Sharp CAPM is a big progress for determining discount rate, because it is the first model that determines the discount rate based on the structure of "risk free rate + risk premium", and no other model or method can take the relevant risk into account before it. Unfortunately, as revealed in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing", the "risk premium" in the model, i.e., $\beta_i(r_m-r)$, takes only the systematic risk into account, because the risk indicator beta (β) only represents the times of the security to the whole market in terms of systematic risk. This does not meet with the rule of thumb in decision making, i.e., the the prudence principle.

As a matter of fact, it is quite strange to calculate the discount rate or fair return of equity capital by using Sharp CAPM. According to Sharp CAPM, the fair return of any stock can only be worked out based on its relationship with all other stock risks in the market. Among them, whether the stocks to be priced or other stocks, the risks that can be offset by other stocks in the market should be excluded, and the risk premium should be calculated according to the remaining risk, i.e., the systematic risk.

In theory, if all the risks are offset as nonsystematic risk by each other and no risk remained, the risk and risk premium will be calculated as zero. In other words, the risk premium of a stock may be zero, so the discount rate is equal to the risk-free interest rate. This means that the risk of the stock is equal to that of government bonds. This seems incredible. But what is more incredible is that when applying Sharp's CAPM, after deducting the nonsystematic risk, the systematic risk may be negative; if so, the risk premium then will also be negative. The discount rate of the stock then will be lower than the yield of treasury bonds, and can even be negative in theory.

However, Sharpe CAPM is the first and only model to calculate the discount rate in connection with risk over the long term after its publication. Comparing with capital cost and other methods, it is sounder in theory. After all, based on Sharpe CAPM, a specific discount rate can be obtained based on the risk of each investment project, rather than a unified discount rate for the whole industry or the whole company (WACC). Therefore, the obvious and serious theoretical and quantitative defects did not hinder the wide dissemination and application of Sharpe CAPM.

Although neither sound in theory nor reliable in practice, there is no better model to replace the Sharpe CAPM over decades. The nowadays discount rates determination is heavily relied on the Sharpe CAPM. The prevailing convention in academy and practice now is to determine the discount rate for equity by the Sharpe CAPM, and to determine the discount rate for debt according to the actual debt capital cost, and to determine the discount rate for the total asset by weighted averaging the discount rate for equity and that for debt.

1.2 Systematic, Nonsystematic and Total Risk

Under the reasonable structure of "risk-free interest rate + risk premium", Sharpe CAPM actually expresses "risk-free interest rate + partial risk premium". Since Sharpe CAPM was published, it has been questioned a lot. In the face of the tide of doubts, there have been various "authoritative and classic explanations".

For example, one explanation is that although the company doing a single business or a few businesses will take the total risk, the shareholders of the company can eliminate the nonsystematic risk by diversifying their investment in the stock market; Shareholders then will not and should not require the risk premium according to the total risk. Another explanation is that the market does not recognize nonsystematic risks and will not give corresponding risk premium; What can be realized in the market is only the systematic risk premium.

Such explanations are obviously far-fetched and incredible. In the face of any decisions, investment, financing and operating, of course, the decision-maker should consider the total risk. It does not matter whether the investor does not require it or the market does not give premium for the non systematic part of the total risk, the total risk should be considered based on the principle of prudence.

Anyway, failure to consider the total risk has become a worrying stone in the financial field. There has been a temptation to make up the lost non systematic risk premium. But how much is this missing part of risk premium, or how much it accounts for in the total risk? This is an interesting and tough question.

The other side of the question is: how much the systematic risk or Sharpe CAPM accounts for in the total risk? Unfortunately, this is not well explored in financial research. Hence it is hard to find a convinced theoretical or empirical conclusion, although it is helpful for understanding and using the Sharpe CAPM.

We now try to do some exploration.[1]

As revealed in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing", the variance of the portfolio consisting of n stocks is:

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{i,j} = \sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ (j \neq i)}}^{n} w_i w_j \sigma_{i,j}$$

There are n items before the plus sign and n(n−1) after the plus sign. Assume that the stand alone (total) risks of all stocks in the market are the same and equal to the average risk of all stocks in the market, that is, their standard deviations of return are: $\sigma_i = \sigma_j = \mathrm{Var} = \sigma$; assume further that all the covariances of any two stocks in the market are equal to the average of the covariances of any two stocks in the market, that is, their covariances of return are: $\sigma_{i,j} = \mathrm{Cov}$. Now, suppose a portfolio consisted of n component stocks with equal weights, 1/n. Then, the variance of the portfolio is,

$$\begin{aligned} & \mathrm{n} \times (1/\mathrm{n})^2 \times \mathrm{Var} + \mathrm{n} \times (\mathrm{n} - 1) \times (1/\mathrm{n})^2 \times \mathrm{Cov} \\ & = (1/\mathrm{n}) \times \mathrm{Var} + \left(\mathrm{n}^2 - \mathrm{n}\right) \times \left(1/\mathrm{n}^2\right) \times \mathrm{Cov} \\ & = (1/\mathrm{n}) \times \mathrm{Var} + (1 - 1/\mathrm{n}) \times \mathrm{Cov} \end{aligned}$$

When the number of the component stocks, n, tends to infinity (∞), i.e., the portfolio consists of all the stocks in the market, the variance of the portfolio is Cov. Understandingly, the variance of the portfolio now, Cov, represents the systematic risk, because the nonsystematic risks are all eliminated by the fully diversification.

Suppose the return correlation coefficient (ρ) between any two stocks is 0.5 (the correlation coefficient, ρ, generally is between 0–1, the middle value is then 0.5).

Then their covariance $\mathrm{Cov} = \sigma_{1,2} = \sigma_1\sigma_2\rho_{1,2} = 0.5\sigma^2$.

The standard deviation or volatility of the portfolio of the whole market is:

$(\mathrm{Cov})^{0.5} = (0.5\sigma^2)^{0.5} = 70.7\%\sigma$

The stand alone risk (σ) of a single stock represents the total risk (average total risk as assumed previously); The risk of stock portfolio (standard deviation or volatility) of

[1] It is often believed that conclusion needs support from data; the more data, the more convincing. As a matter of fact, data is also a resource and perhaps cost as well. For the same conclusion, the less data, the better. The reason is obvious: less data usually means less work load of research or the high efficiency of research; while less data also means the saving of reading time as well as cost or higher reading efficiency of readers. On the other hand, whether a conclusion is more convincing is usually determined by the logic behind. The reason is again obvious, the strict or correct logic represents total (correct) data, which is bigger than any sample data as well as any big data. Therefore, here as well as in other places in this book, we try to derive conclusion based on strict reasoning and necessary data rather than only the past incomplete data with unknown mistakes.

the whole market represents systemic risk. Therefore, based on the above calculation, on average, systematic risk accounts for about 70% of the total risk. As a result, the nonsystematic risk accounts for about 30% of the total risk on average.

This implies that to determine the discount rate based on the Sharpe CAPM, the risk premium is about 30% lower than what it should be on average. For example, if the risk free rate is 3%; the risk premium based on the Sharpe CAPM is 7%, the discount rate then is 10%. But for an average case, the right discount rate should be: $3\% + 7\%/(1-30\%) = 13\%$, rather than the 10% based on the Sharpe CAPM. This will inevitably result to a significant error either for firm valuation or for capital budgeting (investment decision).

1.3 Debt, Equity and Company Volatility

The main problem of Sharp CAPM is that it calculates risk premium based on systematic risk rather than total risk, or beta (β) rather than sigma (σ).

Traditionally, risk is defined as uncertainty, and measured by the standard deviation (or variance) of return on asset. The standard deviation of return is also called as volatility of the asset, usually symbolized by sigma (σ). Therefore, volatility or sigma (σ) is the original and authentic risk indicator; more importantly, it represents total risk rather than systematic risk. This implies that determining the risk premium based on volatility can make up the main or fundamental defect of the Sharp CAPM.

Risk is indeed one aspect for people to care when making the future oriented decision, but it is somehow infeasible to measure. Even, the future possible returns are hard to estimate. Just imagine that, how many possible returns every year are there in the future, and how much are the probabilities of those possible returns? It is difficult to predict the future expected return (one for every year); it is much more difficult to predict the future possible returns (two or more for every year).

To make things more complicated, the total risk of a company is undertaken by the equity holders and the debt holders respectively. Literally, to solve the problem of discount rate, we should find ways to determine the discount rates for equity, debt and total asset respectively; and to determine those discount rates, we need to know the volatilities of equity, debt and total assets as well as the relationships among them.

For instance, if the volatilities of equity and debt of a company are 33% and 7% respectively; the leverage or debt ratio of the company is 40%, that is, equity ratio is 60%.[2]

It is imaginable that the correlation between the debt and the equity of a same company is between complete positive correlation (correlation coefficient 100%) and no correlation (correlation coefficient 0%). When the debt and the equity is 100% correlated, the volatility of total asset or total capital of the company, or the volatility of the company, is: $33\% \times 60\% + 7\% \times 40\% = 22.6\%$.

[2] Which is equivalent to 100% or complete positive corelated.

When no correlation between the debt and the equity, the volatility of the company is:

$(33\%^2 \times 60\%^2 + 7\%^2 \times 40\%^2)^{1/2} = 20.0\%.$

When the correlation coefficient (ρ) between the debt and the equity varies from 0 to 100%, take 50% as an example, the volatility of the company is:

$(33\%^2 \times 60\%^2 + 7\%^2 \times 40\%^2 + 33\% \times 60\% \times 7\% \times 40\% \times 50\%)^{1/2} = 20.68\%.$

Anyway, the relationships among the volatilities of equity, debt and total asset is interesting and also important. Unfortunately, this is not well explored in financial research. Hence it is hard to find a convinced theoretical or empirical conclusion in this respect, although it is helpful for determining discount rate based on total risk.

We now try to do some exploration.

Assume a base case: the firm value is 100, the book and fair present value of the debt is 50, the fair starting value of the equity is then also 50. This is the situation when the volatility is zero. Now, consider a couple of volatilities of the company, such as 8%, 13%, 19%, 26%. That is, what are the volatilities of equity and debt under those cases?

Note that the 4 volatilities belong to the common volatilities, or represent common situations in risk. Those volatilities can be represented by the possible annual returns (−12%, −9%, −6%, −3%, 0%, 3%, 6%, 9%, 12%), (−20%, −15%, −10%, −5%, 0%, 5%, 10%, 15%, 20%), (−28%, −21%, −14%, −7%, 0, 7%, 14%, 21%, 28%), (−40%, −30%, −20%, −10%, 0, 10%, 20%, 30%, 40%) respectively. The returns will affect the value of the company (starting value is 100), and further affect the value of equity and debt. But the volatility of the company affects the equity more and the debt less. This determines the volatility of equity is larger than the company, and the volatility of debt is smaller than the company.

The possible values of the company can be worked out based on the starting value and the annual possible returns. Based on the option-like features of the equity and debt as revealed in Chapters "Option Pricing and Valuation of Contingent Cash Flow" and "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing", the possible values of the equity and debt can be worked out based on the possible values of the company. Then, the possible returns of the equity and debt can be worked out based on their possible values. Further, the possible leverage ratios can be worked out respectively based on the book value and the estimated fair value of the debt. The results of those calculation are shown in Table 1.

Table 1 is the low risk situation. The volatilities of the equity, debt and company can be worked out based on their possible returns, which are 16.71%, 0.00% and 8.22% respectively. That is, when the risk of the company is low and the leverage is about the average level (50%), the volatility of the debt is close to zero; while the volatility of the equity is around 16–17% (or more generally 15–20%). Based on the returns on the debt and equity, their covariance is 0.00%, close to but not equal to zero. Further, the correlation coefficient can be derived as 76.36% (or more generally 70–80%).

Table 1 Debt, equity and company volatility (8%)

returnf (%)	−12	−9	−6	−3	0	3	6	9	12
firm	88	91	94	97	100	103	106	109	112
equity	38	41	44	47	50	53	56	59	62
debt	50	50	50	50	50	50	50	50	50
returne (%)	−27.44	−19.85	−12.78	−6.19	0.00	5.83	11.33	16.55	21.51
returnd (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
leverb (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64
leverm (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64

returnf = return of the company; firm = value of the company; equity = value of the equity; debt = value of the debt; returne = return of the equity; returnd = return of the debt; leverb = leverage ratio based on the book value of the debt; leverm = leverage ratio based on the fair value of the debt

The calculations in Table 1 are based on a volatility of the company, 8%. Changing the volatility to 13%, doing the same calculation, the results are shown in Table 2. Similarly, based on the company volatility of 19% and 26%, the results can be obtained respectively as shown in Tables 3 and 4.

Table 2 Debt, equity and company volatility (13%)

returnf (%)	−12	−9	−6	−3	0	3	6	9	12
firm	88	91	94	97	100	103	106	109	112
equity	38	41	44	47	50	53	56	59	62
debt	50	50	50	50	50	50	50	50	50
returne (%)	−27.44	−19.85	−12.78	−6.19	0.00	5.83	11.33	16.55	21.51
returnd (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
leverb (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64
leverm (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64

The same as Table 1

Table 3 Debt, equity and company volatility (19%)

returnf (%)	−12	−9	−6	−3	0	3	6	9	12
firm	88	91	94	97	100	103	106	109	112
equity	38	41	44	47	50	53	56	59	62
debt	50	50	50	50	50	50	50	50	50
returne (%)	−27.44	−19.85	−12.78	−6.19	0.00	5.83	11.33	16.55	21.51
returnd (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
leverb (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64
leverm (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64

The same as Table 1

Table 4 Debt, equity and company volatility (26%)

returnf (%)	−12	−9	−6	−3	0	3	6	9	12
firm	88	91	94	97	100	103	106	109	112
equity	38	41	44	47	50	53	56	59	62
debt	50	50	50	50	50	50	50	50	50
returne (%)	−27.44	−19.85	−12.78	−6.19	0.00	5.83	11.33	16.55	21.51
returnd (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
leverb (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64
leverm (%)	56.82	54.95	53.19	51.55	50.00	48.54	47.17	45.87	44.64

The same as Table 1

Table 2 is the relative low risk situation. The volatilities of the equity, debt and company can be worked out based on their possible returns, which are 28.68%, 0.06% and 13.69% respectively. That is, when the risk of the company is relatively low and the leverage is about the average level (50%), the volatility of the debt is close to zero; while the volatility of the equity is around 28–29% (or more generally 25–35%). Based on the returns on the debt and equity, their covariance is 0.016%, close to but not equal to zero. Further, the correlation coefficient can be derived as 86.85% (or more generally 80–90%).

Table 3 is the relative high risk situation. The volatilities of the equity, debt and company can be worked out based on their possible returns, which are 40.87%, 0.89% and 19.17% respectively. That is, when the risk of the company is relatively high and the leverage is about the average level (50%), the volatility of the debt is close to zero; while the volatility of the equity is around 40–41% (or more generally 35–45%). Based on the returns on the debt and equity, their covariance is 0.341%, close to but not equal to zero. Further, the correlation coefficient can be derived as 93.40% (or more generally 90–95%).

Table 4 is the high risk situation. The volatilities of the equity, debt and company can be worked out based on their possible returns, which are 59.05%, 3.89% and 27.39% respectively. That is, when the risk of the company is relatively high and the leverage is about the average level (50%), the volatility of the debt is around 1–5%; while the volatility of the equity is around 58–60% (or more generally 55–65%). Based on the returns on the debt and equity, their covariance is 2.208%, around 0–5%. Further, the correlation coefficient can be derived as 96.18% (or more generally 95–99%).

Summarize the findings in Tables 1, 2, 3 and 4, along with the risks or the volatilities of company increases, the risks or the volatilities of the equity and debt increase; the volatility of the debt remains bellow 1% most of the time, except when the risk of the company is high.[3] The covariances of the equity and debt remains also bellow 1% most of the time, except the high risk situation; their correlation coefficient

[3] The volatility of the debt can also be high when the leverage level is high. We shall not do numerical illustration further here since it is easy to understand.

Table 5 The volatility and correlation between debt and equity

firm (%)	8.22	13.69	19.17	27.39
equity (%)	16.71	28.68	40.87	59.05
debt (%)	0.00	0.06	0.89	3.89
covariance (%)	0.00	0.016	0.34	2.208
cor coe (%)	76.36	86.85	93.40	96.18

firm = company volatility; equity = equity volatility; debt = debt volatility; covariance = covariance between debt and equity; cor coe = correlation coefficient between debt and equity

increases along with the increase of the company volatility, from around 76–97%. Those features are shown in Table 5, which can provide good intuition for relevant understanding.

The possible returns in any year in the future are too uncertain to predict. In practice, the standard deviation of future returns is usually calculated based on time series data of the historical returns of the same asset. The standard deviation based on historical returns then need to be adjusted according to the expected changes over the future period. The adjustment is often omitted or ignored in practice for convenience which implies that the risk in future is the same as past.

For instance, the volatility of a stock or equity is often calculated as following:

Based on the daily closing price of a stock over a continuous period of past time, the daily rate of return based on the daily price change of the stock is calculated first; and then the standard deviation of the daily rate of return is calculated; and further, this standard deviation is annualized to derive the standard deviation of the annual rate of return, which is the volatility of the stock or equity.

The volatility of bond or debt capital can be derived in a similar way. If a company has not issued regular credit bond[4] or no enough bond trading prices available, the debt volatility can be obtained based on the bond trading data of a similar company or company in the same sector or industry. With both the volatilities of equity and debt of a company, the volatility of total asset or total capital of the company can be calculated based on the leverage or debt ratio of the company.

Anyway, volatility as a basic risk indicator is much easier to estimate than beta. To estimate the beta in the Sharpe CAPM, we also have to depend past data, and as revealed in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing", it is the covariance of the return of security i and the market ($\sigma_{m,i}$) divided by the variance of the return of the market ($\sigma_m{}^2$), i.e., $\beta = \sigma_{m,i}/\sigma_m{}^2$.

[4] Rather than bonds with guarantees, or bonds embedded options like convertible bonds, callable bonds, etc.

2 Equity Pricing and Total Asset Pricing

The Sharpe CAPM does not solve the problem of the discount rate, because it is not capable to take the total risk into consideration. Theoretically, the right equity pricing or asset pricing model should derive the discount rate based on total risk or the volatility. Chapter "Certainty Equivalent, Risk Premium and Asset Pricing" provides such a model, i.e., the ZZ CAPM. It is obviously worthy to try to derive the discount rates for three capitals based on the ZZ CAPM.

2.1 Based on the ZZ CAPM

The ZZ CAPM takes the form as,

$$\mathrm{k} = \mathrm{r} - \ln\left[2 - 2\mathrm{N}\left(\sigma\sqrt{T/4}\right)\right]/\mathrm{T} \tag{2}$$

where,

k is equivalent to the ri in Eq. 1;

r is the risk free rate of interest;

σ is the volatility of the asset, or the standard deviation of the return on the asset;

T is the time in year, which can be integer or noninteger, but usually a serial integer number with value 1, 2, 3, … in sequence.

By incorporating total risk into the discount rate, the ZZ CAPM avoid the defect of Sharpe CAPM. Specifically, we can obtain the discount rate for equity by putting the volatility of equity into the model; we can also obtain the discount rate for debt by putting the volatility of debt into the model. Of course, if we know the volatilities of equity and debt, as well as the leverage and the correlation between the equity and debt, we can also work out the volatility of the company. Then, we can also derive the discount rate for total asset or total capital by putting the volatility of company into the model.

For instance, based on the intuition from Tables 1, 2, 3, 4 and 5, consider a common case, the risk free rate is 4%, the volatilities of equity and debt are 35% and 1.5% respectively, the debt and equity ratios of the company is 40% and 60% respectively. Further assume the correlation coefficient between the debt and the equityis 85%.

Then the volatility of the company is:

$(35\%^2 \times 60\%^2 + 1.5\%^2 \times 40\%^2 + 35\% \times 60\% \times 1.5\% \times 40\% \times 85\%)^{1/2} =$ 21.26%.

Literally, put the volatility of equity, debt and the company into the ZZ CAPM, we can work out the discount rate for equity, debt and the company respectively,

The discount rate for debt:

$$
\begin{aligned}
k_d &= r - \ln\left[2 - 2N\left(\sigma d\sqrt{T/4}\right)\right]/T \\
k_d &= 4\% - \ln\left[2 - 2N\left(1.5\%\sqrt{T/4}\right)\right]/T
\end{aligned} \tag{3}
$$

The discount rate for equity:

$$
\begin{aligned}
k_e &= r - \ln\left[2 - 2N\left(\sigma e\sqrt{T/4}\right)\right]/T \\
k_e &= 4\% - \ln\left[2 - 2N\left(35\%\sqrt{T/4}\right)\right]/T
\end{aligned} \tag{4}
$$

The discount rate for the company:

$$
k = r - \ln\left[2 - 2N\left(\sigma d\sqrt{T/4}\right)\right]/T \tag{5}
$$

where, σ represents the volatility of the company.

$$
k = 4\% - \ln\left[2 - 2N\left(21.26\%\sqrt{T/4}\right)\right]/T
$$

The results are shown in Tables 6, 7 and 8.

Just as revealed in Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing". The discount rates in Tables 6, 7 and 8 are the break-even discount rates

Table 6 The discount rates for equity, debt and total asset (year 1–10)

Year	1	2	3	4	5	6	7	8	9	10
k_d (%)	4.60	4.42	4.35	4.30	4.27	4.25	4.23	4.21	4.20	4.19
k_e (%)	18.96	14.87	13.07	11.99	11.26	10.72	10.30	9.96	9.68	9.45
k (%)	12.85	10.36	9.26	8.61	8.16	7.83	7.58	7.37	7.20	7.05

Table 7 The discount rates for equity, debt and total asset (year 10–100)

Year	10	20	30	40	50	60	70	80	90	100
k_d (%)	4.19	4.14	4.11	4.10	4.09	4.08	4.07	4.07	4.06	4.06
k_e (%)	9.45	8.18	7.62	7.29	7.07	6.90	6.78	6.68	6.59	6.52
k (%)	7.05	6.27	5.93	5.73	5.59	5.48	5.41	5.34	5.29	5.25

Table 8 The discount rates for equity, debt and total asset (year 100–1000)

Year	100	200	300	400	500	600	700	800	900	1000
k_d (%)	4.06	4.04	4.04	4.03	4.03	4.03	4.02	4.02	4.02	4.02
k_e (%)	6.52	6.16	6.01	5.92	5.86	5.82	5.79	5.76	5.74	5.73
k (%)	5.25	5.01	4.91	4.85	4.81	4.78	4.76	4.74	4.73	4.72

view from the point of risk cost. There are also operating cost and fair profit margin need to be considered. If the operating cost and fair profit margin in total is 2% of the amount of the capital asset, then the final discount rates should be the discount rates in the tables plus 2%. This is quite easy, we will not illustrate in detail.

The total risk is factored into the three discount rates with the form of "risk free rate + risk premium". Obviously, the calculation to determine the three discount rates are sounder and more reliable than the method based on the Sharpe CAPM.

However, we cannot stop here. This is not a final solution, because we have derived a discount rate model specialized for debt, i.e., the ZZ debt or loan pricing model in previous chapter, which is different from the model based on the ZZ CAPM, i.e., Eq. 3. But, a scientific problem normally has only one right answer. So we need to make clear that which is the right answer, or at least, which is better.

2.2 *Based on the ZZ Debt Pricing Model*

The ZZ debt or loan pricing model derived in Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing" is:

$$\mathrm{k_d = r - ln[N(d_2) + N(-d_1)/L]/T} \tag{6}$$

where,

k_d is the discount rate for debt;

r is the risk free rate;

L is the leverage or debt ratio;

T is the maturity of the debt, which is equivalent to the T in ZZ CAPM.

d_1 and d_2 are calculated as Eqs. 7 and 8.

$$\mathrm{d_1} = \frac{\ln(S/X)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} = -\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \tag{7}$$

$$\mathrm{d_2} = \frac{\ln(S/X)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = -\frac{\ln(L)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = \mathrm{d1} - \sigma\sqrt{T} \tag{8}$$

where,

σ is the volatility of the company.

Comparing Eq. 6 with the kd based simply on the ZZ CAPM, Eq. 3, the Eq. 6 or ZZ debt pricing model measures the risk of the debt by two variables: the company volatility and the company leverage; while the Eq. 3 depends only on the volatility of the debt. Of course, the volatility or risk of the debt depends also on the company volatility and the company leverage. But anyway, we now have two methods to take them into account.

The problem is: which is better? Or which is closer to the right answer?

It is not easy to make a clear judgement. Literally, the company risk or company volatility as a unified pricing base is more meaningful, which may represent or reflect the risk of the business or the industry. On the other hand, the debt risk or the debt volatility may vary across industries as well as companies. Therefore, from the application point of view, if the industry or business volatility is available and can be used when estimating the company risk, the Eq. 6 or ZZ debt pricing model is more convenient than the Eq. 3. But the further insight needs deeper analyses.

As a common knowledge, equity (investment) is riskier than debt (investment). Please note that this is a conclusion on one unit capital or marginal capital basis. Specifically, other things being equal, for the same unit of incremental capital, equity (investment) is riskier than debt (investment). However, this is not true on the total capital (investment) basis. For instance, a company doing certain businesses with certain capital mix. The certain company risk is partaken by the debt holders and the equity holders. Consider what if all the debt capital is transferred into equity capital. Obviously, the equity holders take all the company risk; similarly, if all the equity capital is transferred into debt capital, the debt holders take all the company risk. Such a leverage change does not cause the change of the total company risk. So when the company uses just one capital, whether it is debt or equity, the risk burdened by the debt or equity holders is the same, i.e., the total risk of the company.

Put it another way, when the L increases to 100%, the debt capital is equivalent to the total capital; the "interest rate" now derived from ZZ debt pricing model becomes the appropriate rate of return on the total capital. As the total risk of the company remains as the same as before, obviously, such an "interest rate" represents the weighted average fair return on equity and debt or the appropriate discount rate for a firm when it uses both equity capital and debt capital.

Based on Eqs. 7 and 8, when L = 100%,

$$\mathrm{d}_1 = -\ln(L)/\left(\sigma\sqrt{\mathrm{T}}\right) + \left(\sigma\sqrt{\mathrm{T}}\right)/2 = \left(\sigma\sqrt{\mathrm{T}}\right)/2 \tag{9}$$

$$\mathrm{d}_2 = -\ln(\mathrm{L})/\left(\sigma\sqrt{\mathrm{T}}\right) - \left(\sigma\sqrt{\mathrm{T}}\right)/2 = -\left(\sigma\sqrt{\mathrm{T}}\right)/2 \tag{10}$$

Based on Eqs. 9 and 10, $\mathrm{d}_2 = -\mathrm{d}_1$.

$$\text{Then } \mathrm{N}(\mathrm{d}_2) = \mathrm{N}(-\mathrm{d}_1) = 1 - \mathrm{N}(\mathrm{d}_1) = 1 - \mathrm{N}\left(\sigma\sqrt{T/4}\right) \tag{11}$$

Put Eq. 11 into Eq. 6, note that L = 100%,

$$\begin{aligned} \mathrm{k_d} &= \mathrm{r} - \ln[\mathrm{N}(\mathrm{d}_2) + \mathrm{N}(-\mathrm{d}_1)/\mathrm{L}]/\mathrm{T} \\ &= \mathrm{r} - \ln\{[1 - \mathrm{N}(\mathrm{d}_1)] + [1 - \mathrm{N}(\mathrm{d}_1)]\}/\mathrm{T} \\ &= \mathrm{r} - \ln\left[2 - 2\mathrm{N}\left(\sigma\sqrt{T/4}\right)\right]/\mathrm{T} = \mathrm{k} \end{aligned} \tag{12}$$

Equation 12 is the same as the ZZ CAPM. This implies that the ZZ CAPM is the model of discount rate for total asset or total capital, and it can be derived from the ZZ debt pricing model when the leverage increases to 100%. Obviously, the ZZ CAPM and the ZZ debt pricing model share the same bases in variables (company volatility) and same logics. Put it another way, if the volatility of business or company is available, the discount rate for debt can be derived based on the ZZ debt pricing model, and the calculation is not further depended on the additional volatility of debt.

This is in line with a rule of our research, that is, finding the solution or solve the problem as simply as possible. Simpleness here means easy to understand, easy to learn, easy to use as well as easy to prove, hance saving readers time and brains as well as knowledge in understanding and application, as an example, the easier to estimate or obtain the value of the independent variables, the better of the model. This may be not emphasized by other literature or research, but this is an important rule or principle in our research. That is why no sophisticated math and newfangled concepts (such as neural network methods, etc.) in this book and in our solution as well as in the relevant proving process. Briefly speaking, the purpose of research is solving the problem or answering the questions for readers and users, rather than showing the author's knowledge to readers, not to mention to make troubles or obstacles for understanding and solving problems as well as using models.

2.3 The Unified Solution to Asset Pricing

We find in previous section the logic bridge between the ZZ CAPM and the ZZ debt pricing model; or it is better to say, we find no logic gap between the ZZ CAPM and the ZZ debt pricing model; they share the exact same concept and logic. The previous analysis provides also two surprised gains. One is the new derivation of ZZ CAPM via ZZ interest rate model, which is much easier and simpler than the derivation in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing"; the other is the further clarification that the ZZ CAPM derived in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing" is the model of discount rate for the total asset.

Literally, debt financing or the leverage will affect the partition of the total risk of a company, rather than changing or altering the size or magnitude of the total risk. This implies that the changing of leverage will change the discount rate of debt and equity, but will not change the discount rate of the company, that is, the weighted average of the discount rates of debt and equity will remain constant.

Thus, the discount rate model for equity is hopeful to be worked out based on the weighted average relationship among discount rates for debt, equity and total asset. This can be confirmed or reinforced by the discussion of optimal capital structure and the proof of ZZ optimal leverage model in later chapters of this book, i.e. the influence of the leverage or debt ratio on the company value is small and neglectable.

Now, let us derive the discount rate for equity capital. Use k, ke and kd to represents the discount rate for total capital, equity capital and debt capital respectively. The weight of debt capital in total capital is L as before, then the weight of equity capital in total capital is (1−L). The discount rate for total capital is the weighted average of the discount rate for debt capital and the discount rate for equity capital. Then,

$$\mathrm{k} = (1 - \mathrm{L})\mathrm{k_e} + \mathrm{Lk_d} \tag{13}$$

then,

$$\mathrm{k_e} = \frac{k - Lk_d}{1 - L} \tag{14}$$

put Eqs. 2 and 6 into Eq. 14,

$$\begin{aligned} \mathrm{k_e} &= \frac{\left\{\mathrm{r} - \ln\left[2 - 2\mathrm{N}\left(\sigma\sqrt{T/4}\right)\right]/\mathrm{T}\right\} - L\{\mathrm{r} - \ln[\mathrm{N}(\mathrm{d_2}) + \mathrm{N}(-\mathrm{d_1})/\mathrm{L}]/\mathrm{T}\}}{1 - L} \\ &= \mathrm{r} + \frac{L\{\ln[\mathrm{N}(\mathrm{d_2}) + \mathrm{N}(-\mathrm{d_1})/\mathrm{L}]\} - \ln\left[2 - 2\mathrm{N}\left(\sigma\sqrt{T/4}\right)\right]}{(1 - L)T} \end{aligned} \tag{15}$$

Equation 15 is an equity pricing model, which can be used to find the fair discount rate for equity capital. For consistence and convenience, Eq. 15 can be referred to as ZZ equity pricing model or ZZ discount rate model for equity. Where, $\mathrm{d_1}$ and $\mathrm{d_2}$ can be derived by using Eqs. 7 and 8. Fortunately, although a little complex, the ZZ equity pricing model again is a closed form solution with a wonderful structure of "risk free interest rate + risk premium".

Fortunately, we find the fundamental and unified solution to asset pricing. The ZZ CAPM, ZZ debt pricing model and ZZ equity pricing model (Eqs. 2, 6 and 15), or the ZZ CAPM series as an abbreviation, can be used to determine the discount rates for total asset, debt and equity respectively. Those models have some unprecedented and important features, such as: (1) they all take the wonderful or perfect structure of "risk free rate + risk premium"; (2) they are all based on the same total risk, i.e. the volatility of the company, rather than the systematic risk; (3) they all share the strict and unified logic, i.e., the logic from the bankruptcy cost.

In addition, the solutions or the models are sound in theory and convenient in practice; even, they are helpful to reduce the difficulties and burdens in asset pricing learning, since the ZZ CAPM series are close-formed equations and their derivations need only elementary mathematics. In some sense, they are objectively discovered rather than subjective "designed" or "hypothesized". Specifically, the simplest path to understand the solution is: starting from the ZZ bankruptcy cost model to ZZ interest rate model, and then to ZZ CAPM, and finally ZZ equity pricing model.

A numerical example illustration is helpful to get more intuition from those series of models. Consider again the case in Sect. 2.1, i.e., the risk free rate is 4%, the volatilities of equity and debt are 35% and 1.5% respectively, the debt and equity ratios of the company is 40% and 60% respectively. Further assume the correlation coefficient between the debt and the equityis 85%.

Just as before, the volatility of the company is:

$(35\%^2 \times 60\%^2 + 1.5\%^2 \times 40\%^2 + 35\% \times 60\% \times 1.5\% \times 40\% \times 85\%)^{1/2} = 21.26\%$.

Then, we have, $r = 4\%$, $L = 40\%$, $\sigma = 21.26\%$, input those data into the ZZ CAPM series, i.e., the ZZ debt pricing model, the ZZ equity pricing model and the ZZ CAPM, we can work out the discount rates for debt, equity and total asset respectively over the following T years, shown as Tables 9, 10 and 11.

Comparing the results in Tables 9, 10 and 11 with the corresponding results in Tables 6, 7 and 8, obviously, there is no big difference. The discount rates for total asset or the company are the same, the discount rates for debt and equity are slightly different between the two calculations. The differences are shown in Tables 12, 13 and 14.

Apparently, all the previous Tables 6, 7, 8, 9, 10 and 11 show that the three discount rates decrease along with time extending into the future. While Tables 6, 7 and 8 show all the three discount rates keep monotonic decreasing, Tables 9, 10 and 11

Table 9 The discount rates for equity, debt and total asset (T = 1–10)

Year	1	2	3	4	5	6	7	8	9	10
k_d (%)	4.00	4.01	4.04	4.09	4.15	4.21	4.27	4.32	4.37	4.41
k_e (%)	18.74	14.60	12.75	11.62	10.84	10.25	9.78	9.40	9.08	8.82
k (%)	12.85	10.36	9.26	8.61	8.16	7.83	7.58	7.37	7.20	7.05

Table 10 The discount rates for equity, debt and total asset (T = 10–100)

Year	10	20	30	40	50	60	70	80	90	100
k_d (%)	4.41	4.67	4.77	4.81	4.83	4.84	4.84	4.84	4.84	4.84
k_e (%)	8.82	7.35	6.71	6.34	6.09	5.92	5.78	5.68	5.59	5.52
k (%)	7.05	6.27	5.93	5.73	5.59	5.48	5.41	5.34	5.29	5.25

Table 11 The discount rates for equity, debt and total asset (T = 100–1000)

Year	100	200	300	400	500	600	700	800	900	1000
k_d (%)	4.84	4.80	4.76	4.74	4.72	4.71	4.70	4.69	4.68	4.67
k_e (%)	5.52	5.15	5.00	4.92	4.87	4.83	4.80	4.78	4.76	4.75
k (%)	5.25	5.01	4.91	4.85	4.81	4.78	4.76	4.74	4.73	4.72

Table 12 The discount rate differences between the two methods (T = 1–10)

Year	1	2	3	4	5	6	7	8	9	10
k_d (%)	−0.60	−0.41	−0.31	−0.21	−0.12	−0.04	0.04	0.11	0.17	0.22
k_e (%)	−0.22	−0.27	−0.32	−0.37	−0.42	−0.47	−0.52	−0.56	−0.60	−0.63
k (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 13 The discount rate differences between the two methods (T = 10–100)

Year	10	20	30	40	50	60	70	80	90	100
k_d (%)	0.22	0.53	0.66	0.71	0.74	0.76	0.77	0.77	0.78	0.78
k_e (%)	−0.63	−0.83	−0.91	−0.95	−0.98	−0.98	−1.00	−1.00	−1.00	−1.00
k (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 14 The discount rate differences between the two methods (T = 100–1000)

Year	100	200	300	400	500	600	700	800	900	1000
k_d (%)	0.78	0.76	0.72	0.71	0.69	0.68	0.68	0.67	0.66	0.65
k_e (%)	−1.00	−1.01	−1.01	−1.00	−0.99	−0.99	−0.99	−0.98	−0.98	−0.98
k (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

illustrate a little different feature of discount rate for debt in detail. That is, under the assumed conditions, such as the volatility and leverage, the discount rates for debt increase within about 100 years, and then decrease together with the other two discount rates.

3 The Solutions to Some Related Problems

The discount rate or the capital asset pricing is the core of finance. The findings or improvement of the fundamental solution to such problem means the important progress of financial theory, and the new solution will definitely cast light on many related issues. As a preliminary attempt, we will discuss on two controversial issues based on the ZZ CAPM series in this section. One is about the relationship among the three discount rates; the other is concerning the long run tendency of the discount rates.

3.1 The Comprehensive Application of ZZ Growth Model and ZZ CAPM

In Chapter "Stock and Equity Valuation: Where Discounting Does Not Work", a brand new stock valuation model, namely ZZ growth model, is derived. One feedback from practical application is that it is difficult to find a basis for determining the variable of required payback period in the model, n. Of course, simply speaking, the required payback period, n, is the reciprocal of the required rate of return or the discount rate. Now that we have a better answer to the discount rate, we can naturally better solve the problem of determining the required payback period.

Imagine a benchmark case. The company's earnings per share this year is 2 dollars. Assuming that there are no special factors affecting the company's earnings this year, this earnings per share then can be regarded as normalized earnings per share. Based on a careful prediction, the company's earnings are expected to grow at the average annual rate of 15% over the foreseeable future, that is, about 15 years. According to the company's internal and external conditions in the future, it is estimated that the company's stock or equity risk in terms of volatility is 30%.

Assume that the risk-free interest rate applicable to the market for a long time in future is 4%. Let us consider the stock valuation of the benchmark case company.

First, given the initial earnings per share and the average annual growth rate in the foreseeable period, it is necessary to estimate the required payback period in order to value the stock using the ZZ growth model.

In Chapter "Stock and Equity Valuation: Where Discounting Does Not Work", the required payback period was calculated according to the reciprocal of the required rate of return. But, the discount rates calculated based on ZZ CAPM are not equal to each other, that is, there are multiple discount rates. Now, which discount rate should be used to calculate the required payback period?

Of course, a simple idea is to calculate based on the average of these discount rates. However, in this way, we will encounter the problem of how many years the average is taken, because different length of the period (years) will lead to different mean values. However, there seems to be no absolutely correct method for this number of years.

In fact, the reason behind the calculation of the required payback period according to the reciprocal of the discount rate is that within the required payback period, the discount rate of each year, that is, the sum of the required rate of return, is exactly 100%. In the case of the same rate of return in each year, it is natural to use the reciprocal of the rate of return to obtain the corresponding payback period; However, if the annual rates of return are different, the required payback period can be the years when these rates of return are cumulated to 100%. For example, the cumulative sum of the required return rates over the first three years is: $16.70\%\times1 + (13.20\%\times 2\text{-}16.70\%\times1) + (11.65\%\times3\text{-}13.20\%\times2) = 11.65\%\times3 = 34.95\%$.

Based on the ZZ CAPM series model, the certainty equivalent coefficient, risk premium (compensation rate) and fair rate of return, i.e. discount rate, in the foreseeable future can be calculated, as shown in Table 15.

Based on Table 15, the accumulative total of the discount rates in the year of 11, 12, 13 are 91.99%, 98.53%, 105.00% respectively. Therefore, the required payback period is between 12 years and 13 years. Note that the incremental discount rate for the year 13 is: $105.00\% - 98.53\% = 6.47\%$. At the end of year 12, the percentage remaining unrecovered is: $100\% - 98.53\% = 1.47\%$.

Note that

$1.47\%/6.47\% = 0.23$.

Then, the required payback period is:

$12 + 0.23 = 12.23$ (years)

Table 15 The calculation of the risk premium and discount rate

Year	0	1	2	3	4	5	6	7
CEC	1.00	0.88	0.83	0.80	0.76	0.74	0.71	0.69
RP (%)	–	12.70	9.20	7.65	6.72	6.09	5.63	5.27
DR (%)	–	16.70	13.20	11.65	10.72	10.09	9.63	9.27
Year	8	9	10	11	12	13	14	15
CEC	0.67	0.65	0.64	0.62	0.60	0.59	0.57	0.56
RP (%)	4.98	4.74	4.54	4.36	4.21	4.08	3.96	3.85
DR (%)	8.98	8.74	8.54	8.36	8.21	8.08	7.96	7.85

CEC certainty equivalent coefficient; *RP* risk premium; *DR* discount rate

Then, the stock can be valued based on ZZ growth model,

$P = [(1 + g)^n - 1](1 + g)E/g = [(1 + 15\%)^{12.23} - 1](1 + 15\%) \times 2/15\% = 69.238$ (dollars).

We thus derive the value of the base case stock which is 69.38 dollars based on an precise required payback period 12.23 years. Each chapter of this book discusses and solves various tough problems in the current financial field. These solutions can be integrated or used together as needed for solving theoretical and application problems. They are certainly not limited to the ZZ growth model and ZZ CAPM series. The comprehensive application here is just an example for a little inspiration and guidance.

3.2 The Relationships Among the Three Discount Rates

In 1958, Modigliani and Miller made their pioneering research on capital structure. Their findings are well known as the irrelevance of capital structure, which is equivalent to the constant weighted average capital cost.[5] That is, along with the debt ratio increases, the total cost of capital (the required rate of return or the discount rate) will remain unchanged. Why? As they explained, when the debt ratio increases, the discount rate of debt remains unchanged, while the discount rate of the equity increases in such a way that the weighted average discount rate remains unchanged. Their constant WACC or constant discount rate was illustrated as Fig. 1.

However, Fig. 1 is not easy to understand for a lot of financial students. One puzzle is: when the discount rate for equity increases and the discount rate for debt remains constant, why their weighted average remains constant? In addition, as a common knowledge in finance, the fair discount rate for debt cannot remain unchanged as the debt ratio increases. Rather, it should increase as well. Then, it becomes even

[5] The capital cost in their paper is actually the reasonable or fair capital cost, which is equivalent to the discount rate.

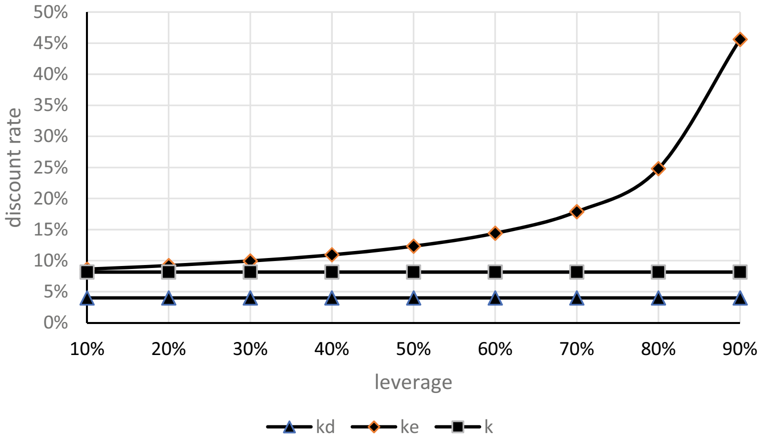


Fig. 1 Fair capital costs of debt, equity and total capital in MM model I

doubtful: when both the discount rates of debt and equity increase, why does their weighted average remain constant?

Such a question may be too tough to understand before. But now, based on the ZZ debt pricing model, equity pricing model and capital asset pricing model, when the company volatility is given, it is easy to depict the changing process of the three discount rates along with the increase of the leverage.

Let us test or illustrate the leverage effect on discount rates for the debt, equity and total asset by the previous numeric example. For instance, assume we are interested in the situation in year 5 (T = 5), the three discount rates based on the ZZ CAPM series (Eqs. 2, 6 and 15) are shown as Table 16.

The discount rates in Table 16 show that along with the increase of the leverage ratio, the discount rates for debt and equity increase, while the discount rates for total asset remain constant. Figure 2 depicts this feature.

Apparently, the ZZ CAPM series confirm that both the discount rates for debt and equity increase along with the increase of the leverage, while in MM model, only the discount rates for equity increases, the discount rate for debt remains constant. Now, the question again: when both the discount rates of debt and equity increase, why does their weighted average remains constant?

This may be difficult to understand before, but now, as depicted in Fig. 2 or revealed by the ZZ CAPM series, the reasons for the discount rate of total asset

Table 16 The changing of the discount rates along with the leverage (T = 5)

Leverage (%)	10	20	30	40	50	60	70	80	90
k_d (%)	4.00	4.00	4.03	4.15	4.43	4.89	5.52	6.30	7.19
k_e (%)	8.62	9.20	9.93	10.84	11.90	13.07	14.32	15.61	16.89
k (%)	8.16	8.16	8.16	8.16	8.16	8.16	8.16	8.16	8.16

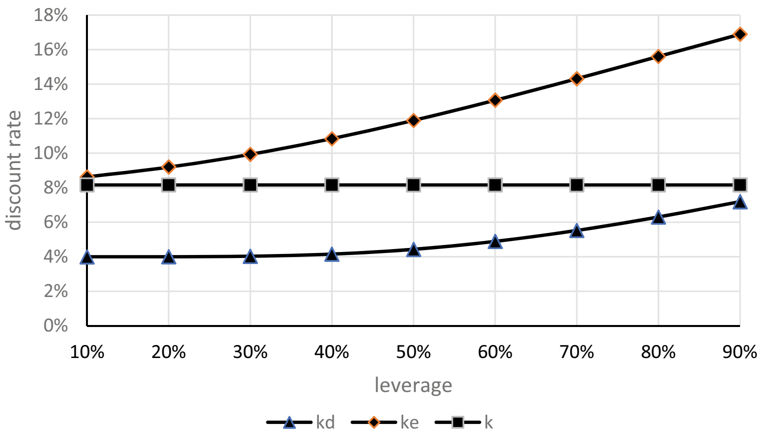


Fig. 2 Discount rates for debt, equity and total capital

remains constant is very clear. The leverage represents the weight of debt. At the start point, the leverage ratio is 0% and the weight of the equity is 100%, i.e., the total asset consists only equity, the discount rate for total asset thus is the same as that for equity; At the final point, the leverage ratio is 100% and the weight of the debt is 100%, i.e., the total asset consists only debt, the discount rate for total asset thus is the same as that for debt.

Both the discount rate for equity and debt are increasing with the increase of the leverage, but the lowest point of discount rate for equity (at the start point) is just at the same level with the highest point of discount rate for debt (at the final point); when the leverage ratio increases, the discount rates for total asset, as the weighted average discount rate, remains constant. In some sense, with the increase of the leverage, the process is that the discount rates for total asset deviates gradually from the discount rate for equity and gets closer to the discount rate for debt.

Therefore, the increase of both the discount rates of debt and equity and the constant of their weighted average are reinforced with each other. Put it another way, along with the increase of the leverage, the discount rates for debt and total asset cannot both remain constant, because in the final point, the total asset consisting only debt, their discount rates are by no means to be different. In this sense, the ZZ CAPM series or the description of Fig. 2 is sound or more plausible; while the equity cost in the MM model I or in Fig. 1 must increase much faster to ensure the WACC can remain constant.

This is not surprising, because the discount rate is not the focus of the MM model. MM model focuses on the optimal capital structure; while the ZZ CAPM series focuses on discount rate. For the understandings or conclusions about discount rates as well as the relationships among the three discount rates, the ZZ CAPM series is supposed to have advantages than the MM model.

Please note that the calculation of Table 15 is based on the real solution of discount rates and some typical data inputs, the results in the table thus can be used to increase

Table 17 The final discount rates and the leverage (T = 5)

Leverage (%)	10	20	30	40	50	60	70	80	90
k_d (%)	6.00	6.00	6.03	6.15	6.43	6.89	7.52	8.30	9.19
k_e (%)	10.62	11.20	11.93	12.84	13.90	15.07	16.32	17.61	18.89
k (%)	10.16	10.16	10.16	10.16	10.16	10.16	10.16	10.16	10.16

our intuition for determining discount rates. For practice or application purpose, adding 2% as transaction cost, the Table 15 can be transferred into Table 17.

Table 17 shows that as the debt ratio increases from very low to very high, the discount rate for debt increases from about 6% to about 9%; the discount rate for equity increases from about 10% to about 19%; meanwhile, the weight of equity in total capital decrease from very high to very low; so that the discount rate for total capital keeps constant, at a level about 10%.

The discount rate for the equity and debt at their start point is around 10% and 6% respectively. This is resulted from the ZZ equity pricing model based on the typical conditions, where risk-free interest rate is 4%, debt maturity is 5 years, the volatility of return on the firm's asset is 21.26%, the transaction cost is 2%.

When the debt ratio increases, the discount rates for equity and debt increase respectively from above and below the discount rate for the total capital; meanwhile, the discount rate for total capital, as the weighted average of the discounts for equity and debt, remains unchanged. When the debt ratio is very high, the discount rate for equity reaches as high as 19%, the discount rate for total capital remains about 10% which is close to the discount rates for debt at the end point.

The above interpretation about the constancy of the discount rate for total capital is obviously more reasonable and thorough than the explanation in the MM model I.

3.3 The Long Run Tendency of the Discount Rates

As a convention in finance, a constant discount rate is used in asset valuation and capital budgeting (calculation of net present value or NPV). Meanwhile, the long run tendency of the discount rates is being wondered for a long time: should it be increasing, decreasing or constant? And why?

When a project or asset has a long life expectancy, it is often expected to be evaluated at a decreasing discount rate, such as most government dominant investments, which are usually supposed to work over a very long period. But it seems hard to find an adequate reason for the decreasing discount rate.

A large quantity of literature has been cumulated in financial research on the tendency of the discount rates. Various factors and reasons are put forwards to explain the decreasing discount rates, such as, consumption intention, global warming, climate change, environmental protection, risk attitude, etc. Even, a couple of international symposia on this topic have convened. However, no convincing explanation has been found so far, do not mention the convincing method or model to determine the specific decreasing discount rates.

Surprisingly, it is so easy for the ZZ CAPM series to explain and to work out the decreasing discount rate. The discount rates decrease over time based on the ZZ CAPM series simply because the risk can be diversified away among the future years to some extent. Based on the calculation of Tables 6, 7 and 8, the decreasing discount rates for debt, equity and total asset over the future 10, 100 and 1000 years can be depicted as Figs. 3, 4 and 5 respectively. Based on the calculation of Tables 9, 10 and 11, the decreasing discount rates for debt, equity and total asset over the future 10, 100 and 1000 years can be depicted as Figs. 6, 7 and 8 respectively.

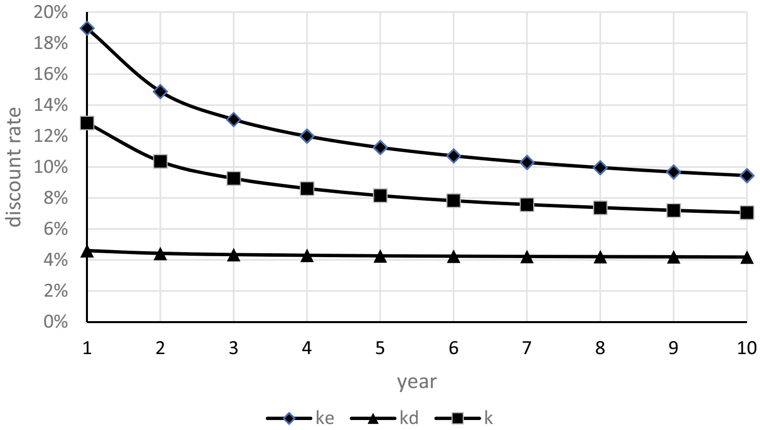


Fig. 3 The discount rates over time based on Table 6

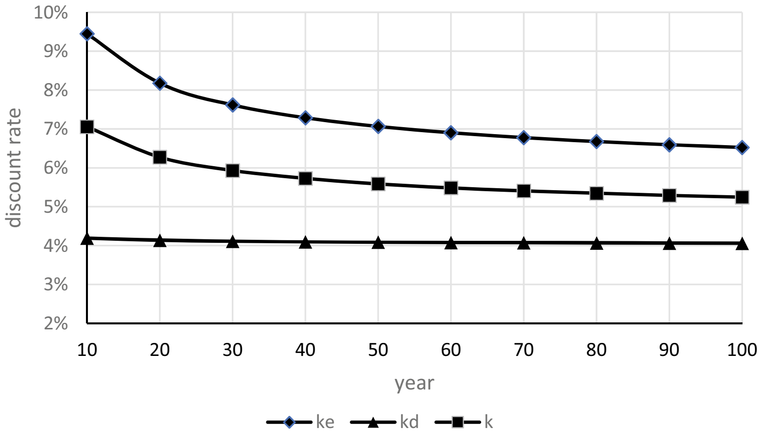


Fig. 4 The discount rates over time based on Table 7

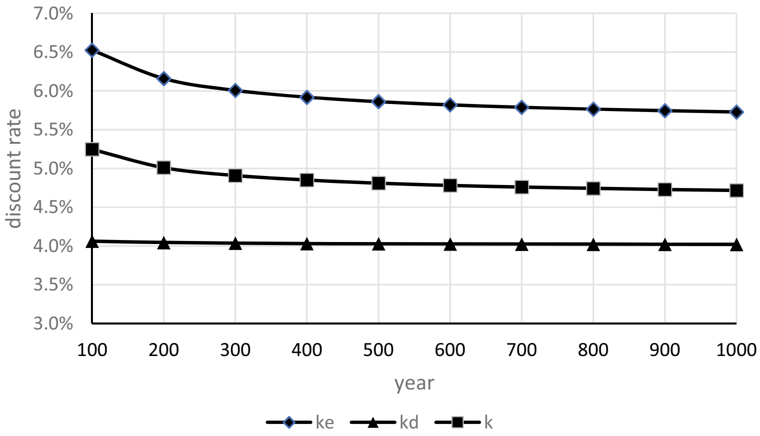


Fig. 5 The discount rates over time based on Table 8

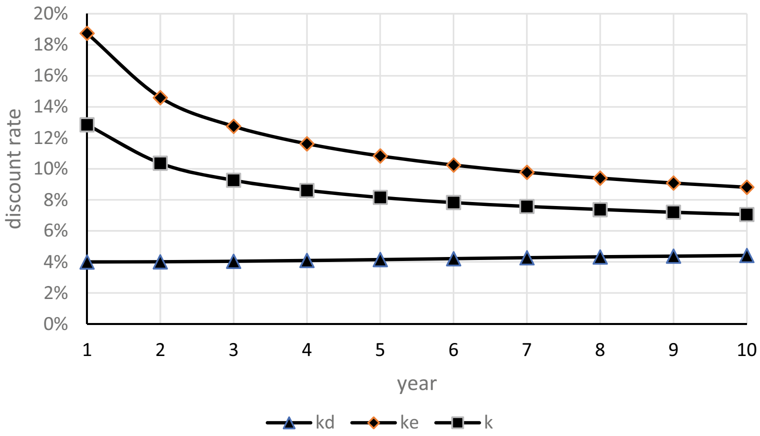


Fig. 6 The discount rates over time based on Table 9

The inconsistencies between Figs. 3 and 6, Figs. 4 and 7, Figs. 5 and 8, reflect the inconsistencies between Tables 6 and 9, Tables 7 and 10, Tables 8 and 11. It seems that the three discount rates converge faster in Figs. 6, 7 and 8 or the Tables 9, 10 and 11. It is interesting to calculate the differences between the discount rate of debt and equity over time. Based on Tables 6, 7 and 8, the differences are shown as Table 18. Based on Tables 9, 10 and 11, the differences are shown as Table 19.

As the discount rate for total asset is in between of the discount rates for debt and equity, the difference between the discount rate for total asset and any one of the other two discount rates are of course smaller than the differences shown in Tables 18 and 19. Obviously, the differences in Table 19 decrease faster than Table 18. This may reflect several effects. Tables 6, 7 and 8 just capture the operating risk; but Tables 9,

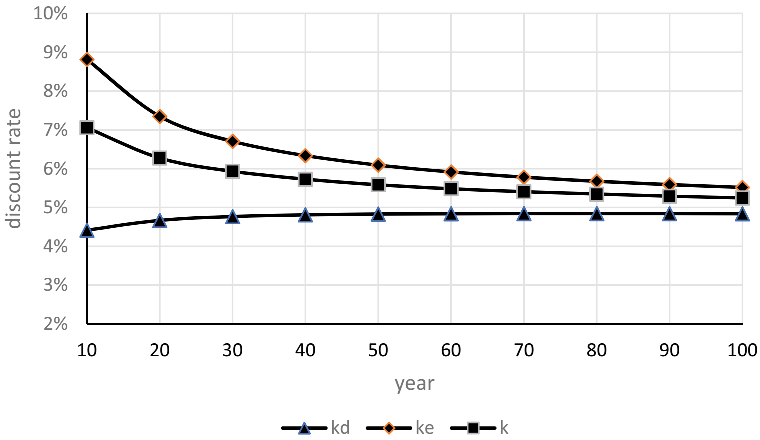


Fig. 7 The discount rates over time based on Table 10

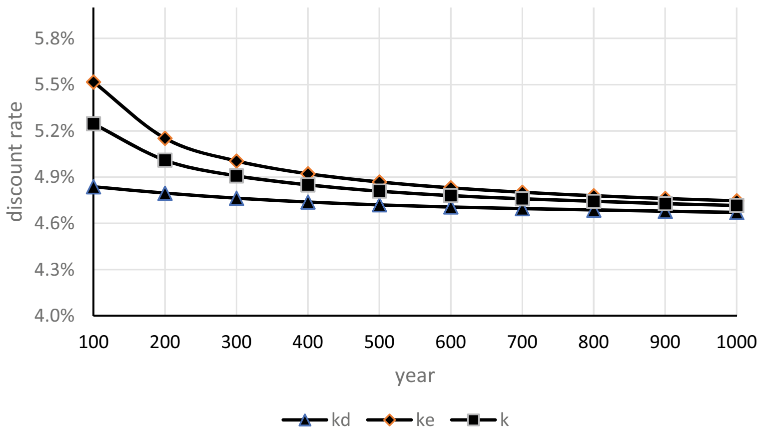


Fig. 8 The discount rates over time based on Table 11

Table 18 The differences between k_e and k_d based on Tables 6, 7 and 8

Year	1	2	3	4	5	6	7	8	9	10
Difference (%)	14.36	10.45	8.72	7.69	6.99	6.47	6.07	5.75	5.48	5.26
Year	10	20	30	40	50	60	70	80	90	100
Difference (%)	5.26	4.04	3.51	3.19	2.98	2.82	2.70	2.61	2.53	2.46
Year	100	200	300	400	500	600	700	800	900	1000
Difference (%)	2.46	2.11	1.97	1.89	1.83	1.79	1.76	1.74	1.72	1.71

Table 19 The differences between k_e and k_d based on Tables 9, 10 and 11

Year	1	2	3	4	5	6	7	8	9	10
Difference (%)	14.74	10.59	8.71	7.53	6.68	6.03	5.51	5.08	4.71	4.40
Year	10	20	30	40	50	60	70	80	90	100
Difference (%)	4.40	2.68	1.94	1.53	1.26	1.08	0.94	0.83	0.75	0.68
Year	100	200	300	400	500	600	700	800	900	1000
Difference (%)	0.68	0.36	0.24	0.18	0.15	0.12	0.11	0.09	0.08	0.08

10 and 11 capture the operating risk and the bankruptcy risk as well. Therefore, the discount rates for debt based on the ZZ CAPM series decrease more slowly over time.

Put it another way, the bankruptcy cost increases over time, the ZZ debt pricing model capture this feature, the discount rates for debt then increase over time within around the first one hundred years. As the total risk of the total asset keeps unchanged, or the discount rates for total asset decrease at the same rate, the feature of the discount rate for debt makes the discount rates for equity decrease faster and the discount rates for debt decrease more slowly, then the differences between the two discount rates decrease soon and get smaller.

In sum, the ZZ CAPM series confirm the decreasing of the discount rates over time. In addition, the ZZ CAPM series also reveal some detailed features of the decreasing discount rate, which in addition to the transaction cost, including at least the following:

(1) All the three discount rates decrease over time from a very long perspective; but the discount rates for debt increase first within decades and then decrease together with the other two discount rates.
(2) The lower bound of the decreasing discount rate is not 0, but the risk free rate; that is, the three discount rates in infinite future are close to the risk free rate.
(3) The decreasing discount rate does not mean the longer the time period the lower the risk; rather, the longer the time period the lower the risk in unit time; but the longer the time period the larger the total risk; this is reflected by the decreasing present value factor $(1/(1+k)^t$ or $e^{-kt})$ over time.

The present value factors (e^{-kt}) based on Tables 9, 10 and 11 are shown as Tables 20, 21 and 22 respectively. Obviously, all the present value factors for debt, equity and total asset are decreasing over time, which implies that the longer the time, the more in future returns is discounted off. This is a new insight revealed by the ZZ CAPM series.

Asset pricing, debt pricing and equity pricing, etc., have been the toughest problems in finance. Based on strict reasoning and unified logic as well as the findings in previous two chapters, this chapter provides the total or thorough solution to asset pricing, i.e., the discount rate models for three kinds of capitals.

Table 20 The present value factors (e^{-kT}) based on Table 9 (T = 1~10)

Year	1	2	3	4	5	6	7	8	9	10
Debt	0.9608	0.9229	0.8859	0.8491	0.8126	0.7768	0.7416	0.7078	0.6748	0.6434
Equity	0.8291	0.7468	0.6822	0.6283	0.5816	0.5406	0.5043	0.4714	0.4417	0.4140
Asset	0.8794	0.8129	0.7574	0.7086	0.6650	0.6251	0.5883	0.5545	0.5231	0.4941

Table 21 The present value factors (e^{-kT}) based on Table 10 (T = 10~100)

Year	10	20	30	40	50	60	70	80	90	100
Debt	0.6434	0.3930	0.2391	0.1460	0.0894	0.0548	0.0338	0.0208	0.0128	0.0079
Equity	0.4140	0.2299	0.1336	0.0792	0.0476	0.0287	0.0175	0.0106	0.0065	0.0040
Asset	0.4941	0.2854	0.1688	0.1011	0.0611	0.0373	0.0227	0.0140	0.0086	0.0052

Table 22 The present value factors (e^{-kT}) based on Table 11 (T = 100~1000)

Year	100	200	300	400	500	600	700	800	900	1000
Debt	0.0079	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Equity	0.0040	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Asset	0.0052	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Asset = total asset; 0.0000 in the table is a result after rounding off, rather than the exact zero

Further Readings

1. Zhang Z. Debt pricing, equity pricing and total asset pricing—a unified solution model to the discount rate problem. China Soft Sci. 2018;4:146–55.
2. Fama EF, French KR. Common risk factors in the returns on stocks and bonds. J Financ Econ. 1993;33(1):3–56.
3. Ml W. Why the far-distant future should be discounted at its lowest possible rate. J Environ Manage. 1998;36(3):201–8.
4. Cochrane JH. Presidential address: discount rates. J Financ. 2011;66(4):1047–108.
5. Gollier C, Hammitt JK. The long-run discount rate controversy. Ann Rev Resour Econ. 2014;6(1):273–95.
6. Levi Y, Welch I. Best practice for cost-of-capital estimates. J Financ Quant Anal. 2017;52(2):427–63.
7. Fama EF, French KR. Choosing factors. J Financ Econ. 2018;128(2):233–53.
8. Serhiy K, Shrihari S. Why do discount rates vary? J Financ Econ. 2020;137(3):740–51.
9. Zhang Z. ZZ financial discovery. University of International Business and Economics Press; 2008.
10. Zhang Z. Advanced finance: theoretical innovation and decision application. Peking University Press; 2012.
11. Zhang Z. Research on real option valuation. China Financial and Economic Publishing House; 2019.

Part III
Leverage and Risks

Tax Shield, Bankruptcy Cost and Optimal Capital Structure

Contents

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_10

Capital is a must for every firm. The capitals backing of businesses include debt and equity. Firms raise their capitals in capital market. When firms make their financing decision, they have to consider the problem of capital structure, i.e. the mix of equity and debt, which is often represented by the debt ratio, or referred to as leverage ratio.

As revealed in previous chapters, the company risk is partaken by the equity holders and the debt holders. The fluctuation of the future return first affects the equity holders; once the company value is reduced below the debt value, the debt holders will suffer a loss, and the company will go bankrupt. The bankruptcy risk will imperil the repayment of the debt and the subsistence of the borrowing company.

The bankruptcy risk hence needs to be quantified for various reasons. Fortunately, Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing" has solved the problem by providing the ZZ bankruptcy cost model. Based on such a theoretical model, we try to find solutions to the optimal capital structure, which perhaps is the toughest problem in finance, and has been hot researched over recent 70 years without a real solution so far in the mainstream of finance.[1]

The optimal capital structure is the debt ratio that can maximize a firm's value. Franco Modigliani and Merton Miller reveal very important insights about optimal capital structure, but they failed to solve the problem. This chapter derives an optimal capital structure model based on the trade off between the main cost and benefit of debt, i.e. the tax shield and bankruptcy cost, thus solve the problem of optimal capital structure.

1 Firm's Goal and Its Capital Structure

The fundamental goal of a firm is to maximize its value through various decisions, which mainly include investment decision, financing decision and operating decision. As for financing decision, a practical and meaningful question is: can we increase a firm's value by adjusting its capital structure? How? This is involving the issue of optimal capital structure.

The problem of optimal capital structure has been intensively studied since MM model (Modigliani and Miller, 1958, 1963),[2] but has not been effectively solved. That is, scholars have not developed a theoretically sound model to determine the optimal capital structure. Firms in reality thus have to make their capital structure decisions based on intuition or experience.

A common way to make a decision is to choose the best (optimal) one among all available alternatives via trading-off between the potential benefits and costs. As for capital structure decision, similarly, an optimal debt ratio can be determined by

---

[1] Most of the problems discussed in this book are on the top in importance in finance; however, most of the solutions in this book are not old enough to be accepted as the mainstream of financial theory.

[2] Modigliani and Miller (1958). Modigliani and Miller (1963).

trading-off between the potential benefits and costs related to the debt financing.[3] From the MM model I to the trade-off theory, the conventional research of capital structure goes along this way.

Thus, the problem of capital structure decision is equivalent to: what is the optimal debt ratio? To solve such a problem, we should know: what are the potential benefits and costs of debt financing and how do they change when the debt ratio increases or decreases? Understandingly, the optimal capital structure is the debt ratio where the net benefit of the debt financing is maximized, or the firm value is maximized.

Note that solving any problem need to assume other things being equal. To solve the problem of optimal capital structure, we should assume that other aspects of the firm, such as the business and investment as well as the size of total asset, are unchanged. While there are numerous comprehensive researches in this area, the problem of primary importance still is: other things being equal, what is the optimal debt ratio of the firm?

2 The Potential Benefits and Costs of Debt Financing

There is a wide variety of potential benefits and costs related to the debt financing. However, as a common sense, for the convenience and efficiency in actual financial decision, we should focus on the benefits and costs with direct and great importance.

2.1 General Analyses

When a firm uses debt capital, it is obliged to repay the due interests and principle at the maturity of the debt, hence has the risk of default or bankruptcy. Because of this, from investors' point of view, debt (such as corporate bond) is safer than equity. This leads to the investors' required rate of return on debt is lower than that on equity. The investors' required rate of return determines the cost of capital born by the borrowing firm. Therefore, from the firms' point of view, debt financing has a benefit of lower cost and a cost of bankruptcy risk. The cost arising from bankruptcy risk is often referred to as bankruptcy cost; and the usage of debt capital is often referred to as usage of financial leverage.

There is another important difference between debt and equity. As a world-wide rule, the cost of debt, including the interests and price discount, is paid as a cost. Such a cost is paid before corporate (income) tax, hence deducted from the firm's income. Oppositely, the cost of equity, similarly including the dividends and price discount, is paid after corporate tax hence deducted from the firm's earnings. In other

[3] You can also choose an optimal equity size or ratio by trading-off the potential benefits and costs related to the equity financing. This is equivalent to the consideration from the debt side which is the academic convention.

words, the cost of debt is treated as a real cost, but the cost of equity is treated as a part of earnings. The firms then can get corporate tax savings from debt capital. This is another benefit of debt in addition to the lower capital cost, which is referred to as tax shield in the arena of capital structure.

In sum, debt financing has two important benefits including the lower cost and the tax shield and one important cost referred to as the bankruptcy cost.

2.2 MM Model I

In 1958, Modigliani and Miller published their breakthrough papers concerning firms' capital structure decision. They revealed that, in an assumed environment without corporate tax and bankruptcy risk, a firm's value is irrelevant to its capital structure. Meanwhile, they revealed that the firm's weighted average cost of capital (WACC)[4] is also irrelevant to its capital structure. These relationships are shown in Eq. (1):

$$V_L = \text{EBIT}/\text{WACC} = \text{EBIT}/K_{SU} = V_U \tag{1}$$

where V_L is the levered firm value; EBIT is the annual earnings before interest and tax; WACC is the weighted average cost of capital in the levered firm; K_{SU} is the cost of equity in unlevered firm; V_U is the unlevered firm value. Obviously, firm L and firm U are only different in capital mix, i.e., one is levered with debt, the other is unlevered without debt.

The reasoning behind the irrelevance of capital structure or Eq. (1) is: as the debt ratio increases, the earnings remained (after paying interests or cost to debt investors) to equity investors is down-sizing and more volatile; hence the required rate of return on equity or the fair cost of equity capital is increasing. The increase in equity cost will just cancel out the cost reduction effect from the debt financing. Therefore, it is useless if a firm tries to reduce its WACC or to raise its value by adjusting the debt ratio, because the increase in low-cost debt is always followed by the increase in the equity cost, as shown in Fig. 1.[5]

---

[4] We reveal in Chapter "Certainty Equivalent, Risk Premium and Asset Pricing" that the discount rate is different from the cost of capital (including the weighted average cost of capital, WACC), because the capital cost is the result of financing (decision), and may have nothing to do with the asset risk. The discount rate is the investors' benchmark to value an asset (such as a project or a firm, etc.), and has much to do with the asset risk. However, financial scholars are used to refer to the discount rate as capital cost in the discussion of capital structure since Modigliani and Miller. We continue to use such an appellation in this chapter to avoid unnecessary chaos or difficulties for understanding. Nevertheless, readers should aware that the WACC as well as the equity and debt cost in this chapter are the appropriate or fair ones, which incorporate the relevant total risk and can be used as the discount rate to derive the value of the firm and its equity, etc.

[5] The same as Fig. 1 in Chapter "Capital Asset Pricing: An Easy and Unified Solution".

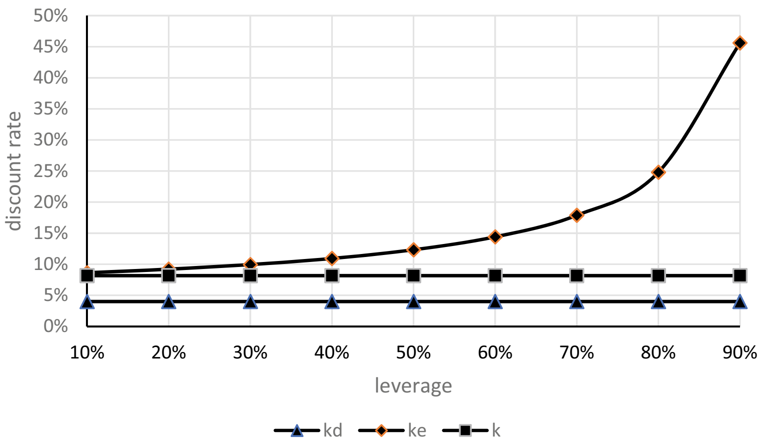


Fig. 1 Fair capital costs in MM model I

Modigliani and Miller's finding in 1958 is afterwards referred to as MM model I or the irrelevance of capital structure. Please note that MM model I has not eliminated the possibility of increasing firm value via optimizing its capital structure, because their conclusion is derived in an assumed environment without corporate tax and bankruptcy risk, which is significantly different from the reality. Nevertheless, MM model I is very important because it implies that the benefit of low-cost from debt is not an achievable benefit, thus the debt financing actually has only one important benefit, i.e. the tax shield. In other words, the cheapness of debt is no longer a benefit worthy to consider in capital structure decisions.

Thanks to Modigliani and Miller (1958), because of their contribution, to solve the problem of optimal capital structure, we need not to care about the cheapness of debt. All we should care about are the tax shield and the bankruptcy cost, or on how to maximize the difference between the tax shield and the bankruptcy cost.

3 Efforts to Value Tax Shield and Bankruptcy Cost

MM model I casts light on the problem of optimal capital structure. Since then, research on capital structure focuses on the valuation of the tax shield and the bankruptcy cost.

3.1 MM Model II

In 1963, Modigliani and Miller relaxed the condition of no corporate tax in MM model I and published a new model incorporating the tax shield. They denote the corporate tax rate by T, the size of debt by D, and the cost of debt by i, then the annual tax savings is DTi. Assume the annual tax savings are perpetual cash flows and the appropriate discount rate is i, its value then is DTi/i = DT. Equation (1) is then rewritten as:

$$V_L = V_U + DT \tag{2}$$

Equation (2) is referred to as MM model II. Modigliani and Miller fail to find an effective way to value the bankruptcy cost hence still cannot incorporate the bankruptcy cost into their new model. Unsurprisingly, with only the benefit (tax shield) taken into account, the optimal debt ratio is 100% based on the new model. This is again an unpractical conclusion.

None of MM model I and MM model II is the final solution to optimal capital structure. However, MM Model I and II open a new era for the research on optimal capital structure. Modigliani and Miller won the Nobel Prize in 1985 and 1990 respectively for their contributions to finance and economics. After that, more and more scholars try to quantify the bankruptcy cost but get no satisfied solution. Therefore, the optimal capital structure remains unsolved in theory, and firms have no theoretical tools to make their capital structure decisions.

3.2 The Trade-Off Model

After MM model II, some scholars emphasize that the optimal leverage ratio should be derived by trading-off between the benefits and costs of debt, such as Robichek and Myers (1966), Kraus and Litzenberger (1933), DeAngelo and Masulis (1980), etc. As the followers get more and more, they are referred to as the school of trade-off theory in finance. Unfortunately, they failed to model the bankruptcy cost either, and even failed to value the tax shield correctly.

Figure 2 is a typical demonstration of scholars' expectation based on the conventional trade-off theory. In the absent of bankruptcy cost, the firm value will increase proportionally with the debt ratio, just as what depicted in the MM model II. In reality, as the debt ratio increases, the actual firm value increases because of the increase of the tax shield; and then increases slowly because of the faster increase of the bankruptcy cost; and then decreases because the bankruptcy cost increases over the tax shield. The firm value goes firstly upwards and then downwards; the top point is the optimum of the capital structure.

In most (if not all) prevailing finance books, the trade-off model is usually written as:

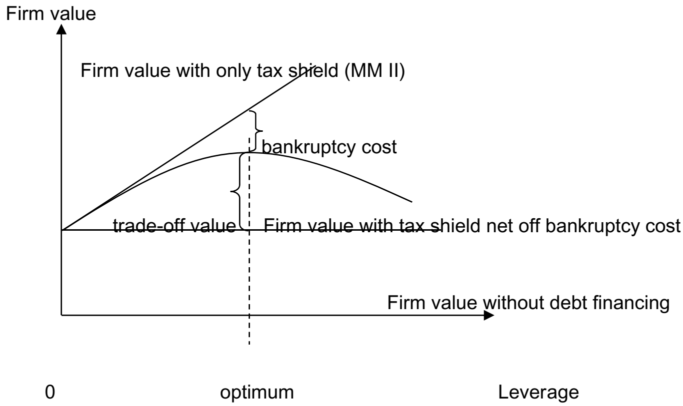


Fig. 2 Traditional expectation of the value-addition from capital structure

$$V_L = V_U + DT - \text{bankruptcy cost} \tag{3}$$

The "DT" in Eq. (3) is copied from the MM model II. As mentioned earlier, MM model II derives the tax shield as "DT" under the assumption of no bankruptcy risk; hence the firm and its annual tax savings of "DTi" can last forever. The trade-off model attempts to remedy the MM model II, i.e. to consider additionally the bankruptcy cost. But obviously, with the existence of potential bankruptcy, the annual tax savings can no longer last forever, and the value of the tax shield should be much lower than "DT".

Therefore, the trade-off model contradicts itself. On one hand, it tries to incorporate the bankruptcy cost; on the other hand, it assumes the tax savings of the firm can last forever. Such a conceptual error (self-contradictory) has not been recognized; most prevailing financial textbooks are copying and propagating such a trade-off model.

The trade-off model or trade-off school develops into "dynamic trade-off" in 1980s, which stress on determining capital structure from a long run perspective. However, the basic problems, such as the quantification of tax shield and bankruptcy cost are still beyond their solutions. Decorating with some advance labels or terminologies, such as long run, dynamic, etc., this "school" or the "model" cannot provide clear answer or even valuable insight to the problem of optimal capital structure.

3.3 The Pecking Order Theory

After MM model II, the research on capital structure has quietly changed. In terms of research methods, statistical description and regression replace professional logical reasoning; The purpose of the research has also shifted from solving problems to explaining phenomena, that is, from providing the solution or method for finding optimal or reasonable capital structure to describing and explaining the capital structure decisions in practice, including what factors affect the actual capital structure decisions. Scholars make statistical description, regression and interpretation of the capital structure in practice from various angles. Due to the different theories (which should be regarded as hypotheses) used to explain the actual capital structure, a variety of capital structure theories have been formed.

For example, Jensen and Meckling (1976), Barnea et al. (1980), Harris and Raviv (1990), diamond (1989), Stulz (1990) and Hart and Moore (1995) elaborated on the theory of agency cost. Ross (1977), Leland and Pyle (1977), Heinkel (1982) and others elaborated on signal theory. After the 1980s, some scholars speculate whether no optimal capital structure exists at all. Following the trend of the times, Myers and Majluf (1984) set up pecking order theory. The pecking order is followed by many scholars, such as Narayanan (1988), Shyam sunder and Myers (1999), Robert and Anuja (2000), Haan and Hinloopen (2003), etc., and has a great impact on the capital structure research afterwards.

The pecking order theory set a successful example for the following empirical research in capital structure. According to it, for reducing agency costs or information asymmetry, executives will follow the order to use various capitals. Specifically, in terms of information asymmetry, external financing is larger than internal financing; equity financing is larger than debt financing. Thus, companies should follow the order of internal financing, debt financing and equity financing when they raise money or funds, so there is no objective optimal capital structure.

The pecking order theory seems plausible and has been widely accepted and followed. However, it cannot be justified at least in the following aspects.

(1) As what made clear in the first chapter of this book, finance is a decisional subject; the purpose of financial theory is to set up the benchmark or provide decisional method for decision problems, like here the problem of capital structure decisions. In this sense, the pecking order theory seems not a theory. It may be a guess or a description of some phenomena, or at least not a qualified financial theory.
(2) As what made clear in the first chapter of most finance books, the decision goal of most (if not all) companies is to maximize its total value, rather than minimize the information asymmetry. In this sense, the pecking order theory misunderstands the company goal. As the goal is wrong, the conclusion is unlikely to be correct. If a company sets up its goal as minimize the information asymmetry, it should sell out all the assets, and put all the proceeds into its bank account, or lock the cash in its safe, rather than raise fund to do business.
(3) Even if a company considers the information asymmetry in its financing decision, will it put the debt before equity? Of course not! In the situation of information asymmetry, the part with information advantages has no reason to be afraid of the counterpart. In the situation of information asymmetry between the internal executives and external investors, the internal executives own information advantages over the external investors. Why should the company put debt in priority to benefit the external investors? As a matter of fact, in most markets,

public companies prefer seasoned equity offering rather than bond offering to raise additional funds.

(4) Even if a company uses various capital in order, and put the debt before the equity, the problem of optimal capital structure is still there, and need to be solved, because the company must know when it is best to turn from the debt capital to the equity capital. This is equivalent to "what is the optimal debt ratio". Therefore, the order of the capitals cannot replace or eliminate the problem of optimal capital structure.

Therefore, the pecking order theory does not stand up under close scrutiny.

3.4 Other Efforts

After MM model II and traditional trade-off theory, many research efforts are put on valuing the tax shield, such as Kane, Marcus and McDonald (1984), Miles and Ezzell (1985), Graham (2000), Arzac and Glosten (2005), Cooper and Nyborg (2006), etc. Graham (2000) estimates the capitalized tax benefits of debt to be as high as 5% of firm value. Recently, Korteweg (2010) derives that the median firm captures net benefits of up to 7.5% of firm value at its optimal leverage ratio. Van Binsbergen, Graham and Jie (2010) estimate that the gross and net benefits of debt are 10.4% and 3.5% of asset value respectively.

Unfortunately, most (if not all) of the studies focus on empirical evidence from past data and reveal little decision- or future- oriented insight. Actually, a statistical result of the tax shield (rather than a theoretical model) does not work for deriving an optimal debt ratio.

Besides the tax shield, there are even more research efforts put on the bankruptcy cost. Unfortunately, most (if not all) efforts try to find the bankruptcy cost by statistics or data processing. According to the prevailing definition, bankruptcy cost includes direct and indirect costs, which provides little insight or inspiration for better understanding the bankruptcy cost. The more detailed introduction about the direct and indirect costs and the research on them see Sect. 3 in Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing". We do not intend to repeat them here. Anyway, a statistical result of the bankruptcy cost does not work for deriving an optimal debt ratio.

In addition, the direct bankruptcy cost is actually the transaction cost of the bankruptcy, whereas the indirect cost is the real bankruptcy cost. The indirect cost occur far earlier and more frequently and is likely larger than the direct cost. In most circumstances, the firm remains healthy until and beyond the debt maturity and the direct cost will not eventually occur, but the indirect cost will surely occur, more or less. The inability to measure the indirect cost means unable to measure the real bankruptcy cost.

Even worse, the conventional division of bankruptcy cost—direct and indirect ones are the ex post concept. However, decision is always forward looking; capital structure decision is not an exception. Therefore, capital structure and other financial decisions need an ex-ante model rather than a backward looking or ex-post descriptive model based on sample data. For healthy firms to make their capital structure decision, what they need is a model relating the bankruptcy cost to the debt size or ratio, rather than a simple estimated number of bankruptcy cost based on some past data of the sample firms. It seems impossible to quantify the bankruptcy cost correctly based on the prevailing concept and statistical method.

As an ex-ante concept, the bankruptcy cost should be a present value of the expected possible future costs rather than actual costs in statistics. Since debt financing increases the bankruptcy risk and consequently reduce the firm value, we define bankruptcy cost as the firm value reduction now resulted from the future potential bankruptcy risk.[6] Such a definition reflects the important features of this cost, i.e. the contingency and the uncertainty.

3.5 Some Comments

Reviewing the capital structure research over decades, a doubt come up inevitably: why do so many theories and countless research innovations fail to solve the problem? When will the problem be finally solved? Judging by the present situation, the hope is very slim, because the relevant research seems to focus on explaining phenomena rather than solving problems; most (if not all) research abandons the search for the optimal capital structure, and do not directly answer the question how the company should make capital structure decisions.

Similar to the research on problem of discount rate, there is a big gap between the academic research and the demand of practice: practical decision needs a benchmark or method or model to make a better decision, while academic research try to answer the question like "how did practitioners make decisions and why". That is, companies need a method or model to find optimal capital structure so they can make their financing and capital structure decisions, while scholars tell them how did they make capital structure decisions in the past and why. So the answers are cumulated more and more, but they are not needed in practice. That is, the practitioners of course know by themselves how they made the financing and capital structure decisions and why; they do not need the scholars come to tell them what or how they did in the past.

Apparently, the mainstream research in finance makes a mistake in the basic feature of finance as a science. As indicated in the first chapter of this book, finance is a decisional subject, not a descriptive subject. The mission or function is to provide benchmark or method for decision-making, rather than to describe the results of actual decision-making and explain. It is an overwhelming method in nowadays financial

[6] This may be close to the prevailing concept of distress cost, which may include also agency cost, etc., but I will not distinguish those costs in detail, and just regard them as all aroused from the debt financing.

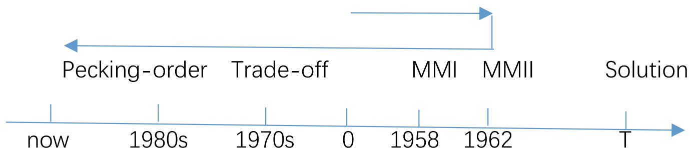


Fig. 3 The progress and retrogress of the capital structure research

research to derive conclusion based on the sample data; but the sample data are just the results of decisions in the past!

In this sense, the pecking order theory go further than the trade-off theory on the wrong path. Comparatively, the trade-off theory does not forget the purpose of the research is to support the decision, rather than to describe or to predict the result of the decision; it does not forget either the goal of decision is to maximize the company value, rather than others. Anyway, comparing with the MM model, the trade-off theory is by no means a progress but a retrogress, because it is self-contradictory, or inconsistent in logic.

Hence, if we mark the time when we know nothing about how to determine optimal capital structure as 0, and the time when we find the fundamental solution to this problem as T. The MMI and MMII push forward the relevant research towards the point T; but a lot of others pull the research back, to the opposite direction against the final solution, as shown in Fig. 3.

After pecking order theory, the confusion of the nature of finance as a discipline is becoming more and more serious. Describing and explaining phenomena, taking novel interpretation as research innovation and contribution, and naming it as a new theory, has become the absolute mainstream paradigm of financial research. In such an academic atmosphere, it is impossible for someone to identify the errors of trade-off theory and pecking order theory; do not mention to solve the problem of valuing the bankruptcy cost and determining optimal capital structure. On the contrary, more and more people have joined the ranks of followers of the two theories in pursuit of publishing papers in advanced academic journals. As a result, trade-off theory and pecking order theory have become two major factions in capital structure research during recent decades, and they are even far better known than the MM models.

4 Decision-Oriented Valuation of Tax Shield and Bankruptcy Cost

In fact, the research since MM model fully shows that avoiding the quantitative problem of bankruptcy cost, or quantifying it in the way of subjective hypothesis or post statistics, or explaining the actual capital structure afterwards, will not help to solve the problem of optimal capital structure. The optimal capital structure should be based on the correct quantification of tax shield and bankruptcy cost and the quantitative trade-off between them beforehand. After revising the conceptual errors on tax shield and bankruptcy cost, we focus on the valuation of them in this section in case of potential bankruptcy.

4.1 The Time Horizon

MM models (I and II) follow an assumption of no bankruptcy risk, so they value the relevant benefits and costs over an infinite period. In reality, however, a firm will surely go bankrupt over an infinite or long enough period. As a common case, when a firm goes bankrupt, its value falls below its debt book value and its equity value falls to zero.

Therefore, from the owner's (equity-holder's) point, considering over an infinite or long enough time horizon, avoiding bankruptcy definitely outweighs obtaining the tax shield. In such a sense, the optimal capital structure is definitely 0% debt or 100% equity, rather than 100% debt as predicted by MM model II. Of course, neither 100% equity nor 100% debt is the "right" solution for capital structure decision. The two extreme "optimal" debt ratios just reflect that the infinite period is a wrong choice of time horizon for the capital structure decision and we cannot get an effective solution this way.

The time horizon has received little attention in previous research on capital structure; but the fundamental difference between practical decision and academic assumption in the time horizon may be one of the key reasons that the optimal capital structure remains unsolved. A convincing evidence is that no executives base their (capital structure) decisions over infinite future even they do not care their tenures. So we should choose a more practical time horizon as the first step for solving the problem of optimal capital structure.

Along with the growth of a firm, it will finance round by round; capital structure decision is among the considerations of every round of financing. When a firm considers its capital structure, what it should care about is the benefits and costs of the debt determined by the current round financing, rather than those benefits and costs determined by last round financing or next round financing or financing in the distant future.

Thus, it is natural and correct to trade-off between tax shield and bankruptcy cost on the basis of the current financing round. The time before next round financing is usually determined by a firm's growth opportunity and/or its debt maturity. Most people caring a firm, whether they are insiders or outsiders, are aware easily of the firm's debt maturity; but they seldom know the time of the next financing round of the firm arising from a growth opportunity. In addition, growth opportunity is not directly related to bankruptcy, while repaying of debt principal is usually a direct cause of bankruptcy.[7] Therefore, to make things more simple and certain, the best

[7] Some scholars study the capital structure decision with possible bankruptcy before the debt maturity, such as Ju et al. (2005). For the general validity and decisional efficiency of the solution, I will

choice of time horizon for the purpose of capital structure decision should be the debt maturity determined by the current round financing.

We thus redefine the problem of optimal capital structure as: the debt ratio maximizing the difference between the tax shield and the bankruptcy cost during the debt life.

4.2 Value the Tax Shield

Now we consider how to value the tax shield during the debt life determined by the current round financing. We use S and X to denote the current market value of a firm and its debt respectively. The firm's debt or leverage ratio then is X/S. For a healthy firm, the debt market value is close to its book value. So we assume the initial market value of the debt is also its book value and is the base for calculating the interest payment.

A firm usually has various debts. Debts with maturity less than one year are current debts or short term debts among which the most common part is payables; debts with maturity longer than one year are long term debts. X in this chapter denotes all debts a firm owes, including short term debt and long term debt. Some research only considers long term debt with reasoning that most short term debts (such as payables) bear no interest cost hence will not contribute tax shield. However, the short term debts still "contribute" bankruptcy risk or bankruptcy cost just as the long term debt does. We thus have to take all short term and long term debts into account in valuing bankruptcy cost. For this reason, we should not neglect the short term debts in valuing tax shield, because we should definitely consider the identical debt when we trade-off between its tax shield and its bankruptcy cost.

Besides the debt size, the debt tax shield depends also on the interest rate, the time to maturity and the corporate tax rate. To make things simple, use b and T to represent the average interest rate and the average maturity of all short term and long term debts, and f to represent the corporate (income) tax rate. Define the perpetual tax shield as the tax shield of the debts over an infinite time horizon in absent of bankruptcy, just as the DT in MM model II. Thanks to MM model II, the perpetual tax shield of the debts starting from now is:

$$\frac{Xbf}{b} = \mathrm{Xf} \tag{4}$$

where Xf is equivalent to DT in MM model II.

Following a convention in financial research, let r to denote the annual risk free rate and assume all asset values are compounding continuously at r. Then the present value of the perpetual tax shield starting from the debt (future) maturity is:

---

focus on the most simple and common situation, where the bankruptcy can only occur at the debt maturity.

$$\frac{Xbf}{b}\mathrm{e}^{-rT} = \mathrm{Xfe}^{-\mathrm{rT}} \tag{5}$$

Therefore, value of the tax shield (TS) during the debt life is Eqs. (4) and (5):

$$\mathrm{TS} = \mathrm{Xf}-\mathrm{Xfe}^{-\mathrm{rT}} = \mathrm{Xf}\left(1-\mathrm{e}^{-\mathrm{rT}}\right) \tag{6}$$

where, X is the book value and current market value of the firm's debt; f is the corporate tax rate; r is the risk free rate; T is the maturity of the firm's debt. Since Eq. (6) is different from other theoretical and empirical tax shield model, for the convenience to be referred to, I would like to name it as ZZ tax shield model.

Let's test the ZZ tax shield model via a numerical example. Assume a typical base case, where the value of the firm and its debt is 100 and 50 respectively, the corporate tax rate f = 25%, the risk free rate r = 3.0%, the average debt maturity T = 4. Based on Eq. (6), the ZZ tax shield (present value) over the debt life is,

$$\mathrm{TS} = 50 \times 25\% \times (1-\mathrm{e}^{-3.0\% \times 4}) = 1.4135$$

For the same case, the MM tax shield over infinite tome horizon is,

$$\mathrm{TS_{MM}} = 50 \times 25\% = 12.5$$

For the same case, the MM tax shield is 12.5% of the firm value, whereas the ZZ tax shield is only 1.4135% of the firm value. This demonstrates the significant difference between the MM tax shield and the ZZ tax shield. But which is right or more convincing? They are already abundantly clear and self-evident and no additional analysis needed.

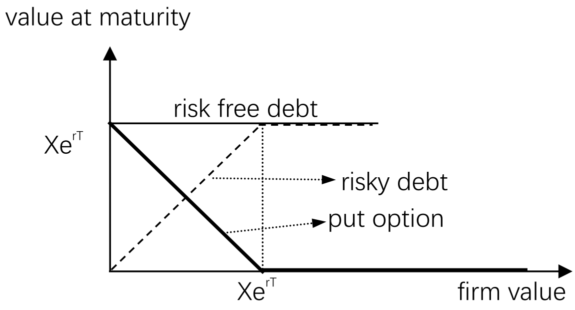
As other ZZ models in previous chapters, the ZZ tax shield model is derived based on correct concepts and strict logic reasoning rather than chosen subjectively in terms of its form and the variables incorporated. In addition, the ZZ tax shield model also makes sense because the relationships among the relevant variables revealed in the model are completely in line with common intuitions. For instance, the tax shield should be positively related with the debt size (X) and its maturity (T), corporate tax rate (f) and the risk free rate (r).[8] These relationships are all reflected in the ZZ tax shield model or Eq. (6).

4.3 *Value the Bankruptcy Cost*

It is well known in mainstream that bankruptcy cost is a tough valuation problem in finance. Fortunately, it has been solved in Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing", we now just review it briefly.

[8] Note that the risk free rate is positively related with the interest rate of the debt.

Fig. 4 Bankruptcy cost = put option



Since the put option is just enough to save the debt from bankruptcy risk, the value or cost of the put option is just the bankruptcy cost, as shown in Fig. 4.

As we use X to represent the present value of the debt, hence X in our model is equivalent to Xe^{-rT} in the standard Black–Scholes option pricing model. Thus, replacing the Xe^{-rT} in Black–Scholes model by X, the ZZ bankruptcy cost model is:

$$\mathrm{BC} = \mathrm{XN}(-\mathrm{d}2) - \mathrm{SN}(-\mathrm{d}1) \tag{7}$$

where, S and X are the current market value of the firm and its debt respectively, and,

$$\mathrm{d}1 = \frac{\ln(S/X)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \tag{8}$$

$$\mathrm{d}2 = \frac{\ln(S/X)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = \mathrm{d}1 - \sigma\sqrt{T} \tag{9}$$

Consider the previous base case again, where the value of the firm and its debt is 100 and 50 respectively, the corporate tax rate f = 25%, the risk free rate r = 3.0%, the average debt maturity T = 4. Now, based on the relevant analyses in Sect. 1 in Chapter "Capital Asset Pricing: An Easy and Unified Solution", assume further the volatility of the firm σ = 20%. Based on Eqs. (7)–(9),

$$\mathrm{d}1 = \frac{\ln(100/50)}{20\%\sqrt{4}} + \frac{20\%\sqrt{4}}{2} = 1.9329$$

$$\mathrm{d}2 = \frac{\ln(100/50)}{20\%\sqrt{4}} - \frac{20\%\sqrt{4}}{2} = 1.5329$$

$$\mathrm{N}(-\mathrm{d}1) = 0.0266$$

$$\mathrm{N}(-\mathrm{d}2) = 0.0627$$

Then, the bankruptcy cost is,

$$\mathrm{BC} = \mathrm{XN}(-\mathrm{d}2) - \mathrm{SN}(-\mathrm{d}1) = 0.4701$$

Now the ZZ bankruptcy cost is 0.4701, or the bankruptcy cost is about 0.4701% of the firm value. As calculated earlier, the corresponding tax shield is 1.4135% of the firm value. This implies the trade-off value (net benefit) of the debt financing is only 0.9434, or 0.9434% of the firm value. As the firm now has 50% debt in its total capital, an interesting question is: does the firm over levered or under levered? This may be not easy to answer as it seems to be now; but will be very easy to answer after you read the next section.

As other previous ZZ models, the ZZ bankruptcy cost model is derived based on correct concepts and strict logic rather than chosen subjectively in terms of its form and the variables incorporated. In addition, the ZZ bankruptcy cost model also makes sense because the relationships among the relevant variables revealed in the model are completely in line with intuitions. For instance, the bankruptcy probability hence bankruptcy cost should be positively related with the debt size (X) and its maturity (T) as well as the firm value volatility (σ), and negatively related with the firm value (S).[9] These relationships are all reflected in the ZZ bankruptcy cost model or Eqs. (7)–(9).

5 Decision-Oriented Optimal Capital Structure Model

We have solved the problem of the quantification of tax shield and bankruptcy cost. This implies that it is ready to derive the optimal capital structure model based on the quantitatively trade off between the tax shield and the bankruptcy cost.

5.1 Derivation of the Model

The quantitatively trade off implies to find the largest difference between the benefit and cost. Combining Eqs. (6) and (7), or subtracting the bankruptcy cost from the tax shield, we can get the ZZ trade–off value or the net benefit of the debt financing:

$$\begin{aligned}
&\text{ZZ Trade}-\text{off value}\\
&\quad = \text{Net benefit of debt financing}\\
&\quad = \text{Tax shield} - \text{Bankruptcy cost}\\
&\quad = \mathrm{Xf}\left(1 - e^{-rT}\right) - [\mathrm{XN}(-d2) - \mathrm{SN}(-d1)] \qquad (10)
\end{aligned}$$

Mathematically, when the derivative of Eq. (10) with respect to X equals to zero, the ZZ trade-off value reaches its maximization, or the capital structure reaches its optimal level. We now try to derive the condition of the optimal capital structure.

[9] These relationships will be more easily to be understood when they related to the payoff or intrinsic value and time value of the relevant put option.

The condition of the optimal capital structure is equivalent to that the derivative of Eq. (10) with respect to X equals zero. To avoid the confusion of derivative operator "d" and the "d" in option pricing model and the ZZ leverage related models, I'd like to use an apostrophe (') to denote the derivative with respect to X in the derivation. Note that,

$$\begin{aligned}&\text{ZZ Trade}-\text{off value}\\&\mathrm{F(X)} = \mathrm{fX}\left(1-\mathrm{e}^{-\mathrm{rT}}\right)-[\mathrm{XN}(-\mathrm{d2})-\mathrm{SN}(-\mathrm{d1})]\end{aligned}$$

where 0 < X < S

$$\begin{aligned}&\text{Let } \mathrm{F}'(\mathrm{X}) = 0, \text{ i.e.,}\\&\left\{\mathrm{fX}\left(1-\mathrm{e}^{-\mathrm{rT}}\right)-[\mathrm{XN}(-\mathrm{d2})-\mathrm{SN}(-\mathrm{d1})]\right\}' = 0\end{aligned} \tag{11}$$

$$\begin{aligned}&\left\{\mathrm{fX}\left(1-\mathrm{e}^{-\mathrm{rT}}\right)-[\mathrm{XN}(-\mathrm{d2})-\mathrm{SN}(-\mathrm{d1})]\right\}'\\&= \left[\mathrm{fX}\left(1-\mathrm{e}^{-\mathrm{rT}}\right)\right]' - [\mathrm{XN}(-\mathrm{d2})-\mathrm{SN}(-\mathrm{d1})]'\\&= \mathrm{f}\left(1-\mathrm{e}^{-\mathrm{rT}}\right) - [\mathrm{XN}(-\mathrm{d2})]' + [\mathrm{SN}(-\mathrm{d1})]'\\&= \mathrm{f}\left(1-\mathrm{e}^{-\mathrm{rT}}\right)-\mathrm{N}(-\mathrm{d2}) - \mathrm{X}[\mathrm{N}(-\mathrm{d2})]' + \mathrm{S}[\mathrm{N}(-\mathrm{d1})]'\end{aligned} \tag{12}$$

Since $[\mathrm{N(x)}]' = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}$

$$[\mathrm{N}(-\mathrm{d2})]' = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(-d2)^2}{2}}(-\mathrm{d2})' \tag{13}$$

$$\begin{aligned}(-\mathrm{d2})' &= \left[-\frac{\ln(S/X)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}\right]'\\&= -\frac{\ln'(S/X)}{\sigma\sqrt{T}} = -\frac{(X/S)(S/X)'}{\sigma\sqrt{T}}\\&= \frac{-(1/X)}{\sigma\sqrt{T}} = \frac{1}{X\sigma\sqrt{T}}\end{aligned} \tag{14}$$

Thus,

$$[\mathrm{N}(-\mathrm{d2})]' = \frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2}{2}} \tag{15}$$

And,

$$\mathrm{X}[\mathrm{N}(-\mathrm{d2})]' = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2}{2}}\frac{1}{\sigma\sqrt{T}} \tag{16}$$

Similarly,

$$
\begin{aligned}
[\mathrm{N}(-\mathrm{d}1)]' &= \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d1)^2}{2}}(-\mathrm{d}1)' \\
&= [-\frac{\ln(S/X)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2}]'\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d1)^2}{2}} \\
&= \frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d1)^2}{2}} \qquad (17)
\end{aligned}
$$

As $\mathrm{d}2 = \mathrm{d}1 - \sigma\sqrt{T}$, or $\mathrm{d}1 = \mathrm{d}2 + \sigma\sqrt{T}$,

$$
\begin{aligned}
&\text{Then } (\mathrm{d}1)^2 = (\mathrm{d}2 + \sigma\sqrt{T})^2 \\
&= (\mathrm{d}2)^2 + 2(\mathrm{d}2)\sigma\sqrt{T} + \sigma^2\mathrm{T} \\
&= (\mathrm{d}2)^2 + 2[\frac{\ln(S/X)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2}]\sigma\sqrt{T} + \sigma^2\mathrm{T} \\
&= (\mathrm{d}2)^2 + 2\ln(\mathrm{S}/\mathrm{X}) - \sigma^2\mathrm{T} + \sigma^2\mathrm{T} \\
&= (\mathrm{d}2)^2 + 2\ln(\mathrm{S}/\mathrm{X}) \qquad (18)
\end{aligned}
$$

According to (17) and (18),

$$
\begin{aligned}
[\mathrm{N}(-\mathrm{d}1)]' &= \frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d1)^2}{2}} \\
&= \frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2+2\ln(S/X)}{2}} \\
&= \frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2}{2}}\mathrm{e}^{-\ln(S/X)} \\
&= \frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2}{2}}\mathrm{e}^{\ln(X/S)} \\
&= \frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2}{2}}\frac{X}{S} \\
&= \frac{1}{S\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2}{2}} \qquad (19)
\end{aligned}
$$

Thus,

$$
\mathrm{S}[\mathrm{N}(-\mathrm{d}1)]' = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{(d2)^2}{2}}\frac{1}{\sigma\sqrt{T}} \qquad (20)
$$

Based on (19) and (20),

$$
\mathrm{X}[\mathrm{N}(-\mathrm{d}2)]' = \mathrm{S}[\mathrm{N}(-\mathrm{d}1)]' \qquad (21)
$$

Based on (21) and (10),

$$\begin{aligned} F'(X) &= \left\{fX\left(1-e^{-rT}\right)-[XN(-d2)-SN(-d1)]\right\}' \\ &= f\left(1-e^{-rT}\right)-N(-d2)-X[N(-d2)]' + S[N(-d1)]' \\ &= f\left(1-e^{-rT}\right)-N(-d2) \end{aligned} \tag{22}$$

Thus, the condition of optimal capital structure is then simplified as:

$$\begin{aligned} &f\left(1-e^{-rT}\right)-N(-d2) = 0 \quad \text{or} \\ &f\left(1-e^{-rT}\right) = N(-d2) \end{aligned} \tag{23}$$

Further, to make sure when $f(1–e^{-rT}) = N(–d2)$ the ZZ trade-off value or net benefit of the debt financing is being maximized rather than being minimized, we need to prove the second derivative of F(X) is negative, i.e. [F'(X)]' < 0.

According to (22),

$$\begin{aligned} \left[F'(X)\right]' &= \left[f\left(1-e^{-rT}\right) -N(-d2)\right]' \\ &= [-N(-d2)]' \\ &= -\frac{1}{X\sigma\sqrt{T}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(d2)^2}{2}} \end{aligned} \tag{24}$$

Because $\frac{1}{X\sigma\sqrt{T}} > 0$, $\frac{1}{\sqrt{2\pi}} > 0$, $e^{-\frac{(d2)^2}{2}} > 0$,

Therefore, [F'(X)]' < 0, i.e., when the $F'(X) = f(1–e^{-rT}) –N(–d2) = 0$, the ZZ trade-off value F(X) gets maximized. Thus, Eq. (23) is the condition of optimal capital structure, which can be referred to as the model of optimal capital structure, or the condition (model) of optimal capital structure.

Define the debt or leverage ratio (capital structure) as L = X/S, then ln(S/X) = ln(1/L) = -ln(L). Based on Eq. (9),

$$-d2 = -\frac{\ln(S/X)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} = \frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \tag{25}$$

Thus, Eq. (23) can be rewritten as:

$$N[\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}] = f\left(1 - e^{-rT}\right) \tag{26}$$

The leverage or debt ratio L satisfying Eq. (26) is then the optimal leverage ratio. It is easy to find the optimal leverage, L, based on Eq. (26) by using of the "goal seek" function in Excel. So the problem of optimal capital structure is solved.

We can also resort to the inverse cumulative distribution function or "probit function" to "solve out" the "L" in Eq. (26). Conceptually, probit(p) = N^{-1}(p), probit[N(p)] = p, thus, take “probit” of the two sides of Eq. (26),

$$\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} = \mathrm{probit}\left[\mathrm{f}\left(1 - \mathrm{e}^{-\mathrm{rT}}\right)\right] \tag{27}$$

$$\begin{aligned}\ln(\mathrm{L}) &= \{\mathrm{probit}\left[\mathrm{f}\left(1 - \mathrm{e}^{-\mathrm{rT}}\right)\right] - \frac{\sigma\sqrt{T}}{2}\}\sigma\sqrt{T} \\ &= \left[\mathrm{probit}\left(\mathrm{f} - \mathrm{fe}^{-\mathrm{rT}}\right)\right]\sigma\sqrt{T} - \sigma^2\mathrm{T}/2\end{aligned} \tag{28}$$

Then,

$$\begin{aligned}\mathrm{L} &= \mathrm{e}^{[probit(f - f\mathrm{e}^{-rT})]\sigma\sqrt{T} - \sigma^2 T/2} \\ &= \exp\{\left[\mathrm{probit}\left(\mathrm{f} - \mathrm{fe}^{-\mathrm{rT}}\right)\right]\sigma\sqrt{T} - \sigma^2\mathrm{T}/2\}\end{aligned} \tag{29}$$

Equations (26) or (29) provide a clear and simple way to derive optimal leverage or optimal capital structure based on the ZZ tax shield and ZZ bankruptcy cost, and can be referred to as ZZ optimal leverage or optimal capital structure condition and model respectively. For convenience, we can refer to the ZZ tax shield model, ZZ bankruptcy cost model, the ZZ optimal leverage condition and the ZZ optimal leverage model as ZZ leverage model series. The ZZ leverage model series were first published by Zhiqiang Zhang (2008, 2009). By the way, it is convenient to use the “NORMSINV” function in Excel to derive the result of probit (). So it is easy to find the optimal leverage ratio by using the ZZ leverage model series.

Now, for any given firm, based on the reliable estimates of the conditional variables, including the corporate tax rate f, the debt maturity T, the risk free interest rate r and the volatility of the firm value σ, we can determine its optimal debt or leverage ratio. One thing worthy to mention is that the optimal debt ratio derived through the ZZ leverage model is a ratio based on market values of the debt and equity rather than their book values, which is in line with the tradition now in capital structure research.

Consider again the previous base case firm, where the corporate tax rate f = 30%; the risk free rate r = 3.0%; the debt maturity T = 4; the return volatility σ = 20%. The optimal debt ratio of the firm can be derived easily by applying the ZZ optimal leverage model, i.e.,

$$\begin{aligned}&\sigma\sqrt{T} = 20\% \times \sqrt{4} = 0.4; \\ &\sigma\sqrt{T} - \sigma^2\mathrm{T}/2 = 0.4 - 0.4^2/2 = 0.32; \\ &\mathrm{f} - \mathrm{fe}^{-\mathrm{rT}} = 25\% - 25\% \times \mathrm{e}^{-3.0\% \times 4} = 2.826989\%; \\ &\mathrm{probit}\left(\mathrm{f} - \mathrm{fe}^{-\mathrm{rT}}\right) = \mathrm{probit}(2.826989\%) = -1.90685; \\ &\left[\mathrm{probit}\left(\mathrm{f} - \mathrm{fe}^{-\mathrm{rT}}\right)\right]\sigma\sqrt{T} - \sigma^2\mathrm{T}/2 \\ &\quad = -1.90685 \times 0.4 - 0.4^2/2 = -0.8427;\end{aligned}$$

Then, the optimal leverage is:

$$L = e^{-0.8427} = 43.05\%.$$

The optimal leverage is 43.05% rounding to two decimals. Hence the firm now (actual debt ratio is 50%) is over levered. Imaginably, even the firm can adjust its leverage to the optimal level (43.05%), no much potential in value addition can be expected.

5.2 Basic Features of the Model

The ZZ leverage model takes four independent variables into account, i.e. the corporate tax rate (f), the risk free rate (r), the debt maturity (T) as well as the firm value volatility (σ). Again, as all previous ZZ models, the ZZ leverage model is derived based on correct concepts and strict logic rather than chosen subjectively in terms of its form and the variables incorporated. Comparing with other capital structure models so far, the ZZ leverage model has several advantages, which include but do not limit to:

(1) It solves the problem of optimal capital structure clearly and definitely via trading-off between the core benefit and cost of debt capital (tax shield and bankruptcy cost), rather than just describe the results or specific sample data of actual capital structure decisions which is equivocal in properness or not necessarily correct or optimal.
(2) The derivation goes along a common and simple idea: trading off between the main benefit and cost of debt capital via rigorous logic process. The main benefit and cost of debt financing are the tax shield and bankruptcy cost respectively. So the final optimal model is based on some fundamental breakthroughs, such as the properly modeling of the tax shield and bankruptcy cost.
(3) Correct concepts are essential for solve financial problems. The concepts of the tax shield and bankruptcy cost in prevailing capital structure research is not correct in terms of time horizon. The ZZ leverage model is derived based on a correct time horizon, i.e. the debt maturity determined by the financing under consideration.
(4) The ZZ leverage model is based on the most essential and simple assumptions, such as the existence of both corporate tax and bankruptcy risk; other details are assumed as simple as possible so long as they are roughly realistic, such as that bankruptcy is assumed only possible at debt maturity. This is necessary for a fundamental solution which should be flexible enough for adjustments in various special applications.
(5) All the variables incorporated in the ZZ leverage model are the direct and important determinants of the optimal capital structure. While numerous factors have

some influences on a firm's capital structure decision, some of them are indirectly related to the decision, such as the industry that the firm operating in, the firm's business strategy, market competition, macroeconomic factors, etc.; some of them are unimportant factors, such as the personality as well as the education and other background of the executives, etc.

The academic research on capital structure over past decades contributes countless "new theories" and "innovations" but fail to solve the problem. One reason is that those studies do not differentiate direct and indirect factors, and incorporate more and more indirect and farther indirect factors into their new models. However, a correct model can only incorporate direct factors or variables. Put it another way, confusing direct factors with indirect factors implies those models cannot be correct, and cannot solve the problem. The ZZ leverage model and its derivation prove that the direct factors of optimal capital structure include: the corporate tax rate (f), the risk free rate (r), the debt maturity (T) as well as the firm value volatility (σ). Most (if not all) important indirect factors are incorporated into the four direct variables. For instance, the factors related to future market and macro economy conditions are reflected in the current firm value (S) and its volatility (σ).

Some of the prior studies include bankruptcy probability in their models in the sense that the higher the bankruptcy probability the larger the bankruptcy cost. Actually, the bankruptcy probability is implied already in the firm's volatility and the debt maturity. As a common sense, the bankruptcy probability is positively related with the firm's volatility and the debt maturity. This is already reflected by the ZZ bankruptcy cost model, i.e., the bankruptcy cost is also positively related with the firm's volatility and the debt maturity; similar features can also be found in the ZZ optimal leverage model in which the optimal debt ratio is negative related with the firm's volatility and the debt maturity. It is thus not necessary to further incorporate the bankruptcy probability into the model. In addition, as an independent variable, the bankruptcy probability is too difficult to estimate objectively for a healthy firm. No one can judge the quality of the estimation even afterwards. So incorporating a subjective probability will inevitably hurt the objectivity or quality of the model.

On the other hand, the indirect factors like personality and background of the executives may be the actual influential factors in capital structure decision-making, or the factors lead to bias of capital structure decision, but are irrelevant to the optimal capital structure or rational capital structure decision, hence are not necessary to be incorporated into the model.

(6) Unlike the prevailing stochastic or other complex models, the ZZ leverage model is an explicit and analytical model, involves less mathematics and the calculation is easy by using the common software, such as Excel, etc.; there is no immeasurable variable like utility etc.; the independent variables are easy to estimate based on commonly accessible data. While the volatility (σ) may be not as easy as other variables to estimate, it is the most basic and most common measure of risk. Hence, the ZZ leverage model is totally feasible in terms of the

understanding and application in practice. This feature is actual essential for a good financial model, since finance is a practice oriented science.

(7) Unlike the prevailing dynamic models, the ZZ leverage model is a simple and static model. Dynamic model, though fashionable in current academic research, is infeasible in practice, because it is neither possible nor necessary for a firm to adjust its debt ratio continuously or dynamically. Since capital structure adjustment involves adjustment cost, most firms only want to adjust their debt ratios occasionally; and the best choice is adjusting when they need additional capital. As a firm raises capitals from round to round, it can adjust its capital structure towards the optimum again and again. For such a purpose, an effective and efficient static optimal capital structure model is enough.

All above features are obviously essential for the model to remain its properness across times and markets and for further theoretical research and practical application.

5.3 Basic Insights from the Model

The basic relationships revealed.

The ZZ leverage model incorporates four influential variables and reveals the relationships between the four variables and the optimal capital structure: the corporate tax rate (f), the risk free rate (r), the debt maturity (T) as well as the firm value volatility (σ).

Based on the ZZ leverage model, the optimal debt ratio is positive related with the tax rate and negative related with the firm's volatility; these are obviously make sense and in line with the common intuition. The relationships of the four influential variables to the optimal debt ratio revealed by the model are summarized in Table 1.

While some relationships revealed by the model are easy to understand, such as that between the optimal leverage and the corporate tax rate and the firm value volatility, some relationships in the model are not so easy to understand, such as that between the optimal leverage and the other two influential variables (r and T).

Based on the ZZ leverage model, the optimal debt ratio is positive related with the risk free rate and negative related with the maturity of the debt. These seem somehow not consistent with the common intuition. The risk free rate determines various interest rates to a large extent. When the risk free rate and interest rate

Table 1 The relationships between the influential variables and the optimal debt ratio (" + " represents positive related, and "–" represents negative related)

Conditional variable	Denotation	Relation
Corporate income tax rate	f	+
Risk-free interest rate	r	+
Maturity of the debt	T	–
Firm volatility	σ	–

increase, firms are supposed to use less debt, so is it really possible that they lift their leverages in such a case?

There is a possibility that the increase of the debt ratio and the decrease of the debt size coexist on the condition that the equity capital decreases more than the debt does. That is true when the interest rate increases. The interest rate increase represents the increase in the debt cost. In reality, the debt cost cannot increase alone without a concurrent increase in equity cost. The capital transfer in capital market will rebalance the cost relationship between debt and equity. Therefore, when the interest rate increases, the equity cost will increase as well. The increase of the equity cost will restrain the use of equity and result in the increase of the debt ratio. Thus, when the risk free rate increases, the less uses of debt and equity can decrease and increase the optimal debt ratio respectively. But what is the net effect of the two contrary movements? It seems hard for us to decide by intuition. Fortunately, the ZZ leverage model tells us that the net effect is increase in optimal debt ratio.

As for the variable of debt maturity T, a common reasoning is that long term debt is somehow more similar to equity, or is safer (from the issuers' point of view) than short term debt. The debt maturity and the optimal leverage ratio hence should be positively related to each other. Leland and Toft (1996), Stohs and Mauer (1996), among others, stand for this viewpoint. On the contrary, Dennis, Nandy and Sharpe (2000) show that leverage is inversely related to debt maturity by their regressions. They argue that this happens because agency costs may be limited by reducing leverage and shortening debt maturity.

It is actually not necessary to bother with agency cost to explain the relationship between the debt maturity and the optimal debt ratio, since the reasoning can be quite simple based on the ZZ leverage model. Uncertainty will increase with the increase of the debt maturity, which will increase the bankruptcy cost (cost from debt financing). The optimal debt ratio thus should decrease as the debt maturity increase. A fact in reality supports this reasoning, i.e. other things being equal it is easier to borrow more short term debt than long term debt.

The value-addition potential from optimizing the leverage.

How much can financing or capital structure decision add value to a firm? Many people want to know the answer to this question but previous research fails to answer. Traditionally, scholars expect a significant value-addition from capital structure, just as something shown in Fig. 2, which is usually appearing in finance textbooks.[10]

Without a convincing optimal capital structure model, Fig. 2 is drawn based on subjective guess, which is inevitably too subjective. For instance, is the value-addition potential from optimizing the leverage really as large as that shown in Fig. 2?

Based on the ZZ tax shield model and the ZZ bankruptcy cost model, take the base case as an example, where, $f = 25\%$; $r = 3.0\%$; $T = 4$; $\sigma = 20\%$, the tax shield, bankruptcy cost and firm value can be calculated; the results shown as Table 2.

[10] Such as Stephen A. Ross, Randolph W. Westerfield, Jeffrey Jaffe, Corporate Finance (Chap. 16), The McGraw Hill Companies, Inc., 2005; Richard A. Brealey, Stewart C. Myers, Principles of Corporate Finance (Chap. 18), The McGraw Hill Companies, Inc., 2003.

Table 2 The tax shield, bankruptcy cost and firm value (unlevered firm value = 100)

Leverage (%)	10	20	30	40	50	60	70	80	90
TS	0.2827	0.5654	0.8481	1.1308	1.4135	1.6962	1.9789	2.2616	2.5443
BC	0.0000	0.0001	0.0079	0.0936	0.4701	1.4606	3.3712	6.3912	10.5713
FV	100.2827	100.5653	100.8402	101.0372	100.9434	100.2356	98.6077	95.8704	91.9730

TS tax shield, *BC* bankruptcy cost, *FV* firm value

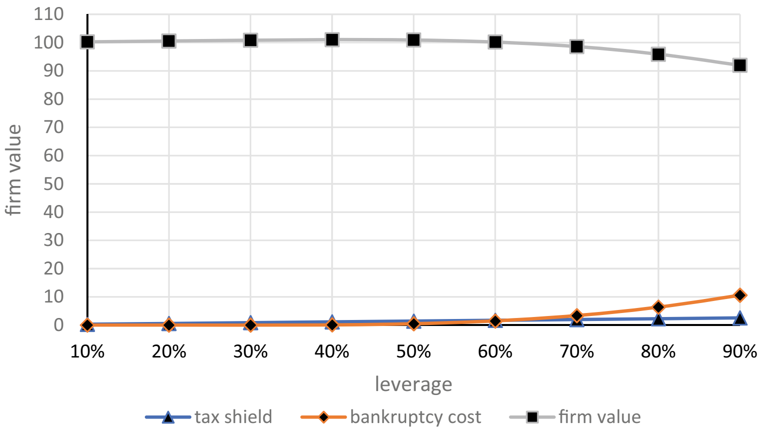


Fig. 5 The tax shield, bankruptcy cost and firm value (unlevered firm value = 100)

The curves of the tax shield, bankruptcy cost and firm value is depicted as shown in Fig. 5, which is based on Table 2.

Because the firm and its related variables are typical in value, and the ZZ tax shield model and ZZ bankruptcy cost model are all derived based on correct concepts and strict logics, the curves and their positions are objective and have practical meaning.

It is interesting and meaningful to compare Fig. 5 with Fig. 2. There are firm values with and without trade-off value in both figures. Figure 2 is depicted based on the subjective imagination of trade-off school; while Fig. 5 is depicted based on the ZZ tax shield and bankruptcy cost model as well as the typical variable values. In both figures, the curves of firm value with trade-off value (levered firm value) are rising first and falling then; and the curves of firm value without trade-off value (unlevered firm value) are level.

However, there is significant difference between Fig. 5 and 2. The curve of levered firm value in Fig. 2 is much higher over the curve of unlevered firm value, which implies the big potential of the value-addition from the adjustment of the capital structure, or that the trade-off value accounts for a large part in the levered firm; whereas the curve of levered firm value in Fig. 5 is just slightly higher over the curve of unlevered firm value only at around the optimal leverage, which implies that the trade-off value is not an significant part of the levered firm value, or the potential value-addition from capital structure decision or optimization is limited or insignificant.

In addition, according to Fig. 2, over-leverage does not hurt the firm value; rather, leverage always adds value to the firm, no matter it is under-levered or over-levered. However, it is much different in Fig. 5, comparing with optimal leverage, the cost of over-leverage is asymmetrically higher than the cost of under-leverage. When it is under-levered, the firm loses the potential trade-off value, but the potential value-addition is very limited. Take the base case as an example, where $f = 30\%$, $r = 3.0\%$, $T = 4$, $\sigma = 20\%$, the optimal leverage ratio $L = 43.05\%$. This means that when the firm has unlevered value S = 100, the optimal size (value) of the debt X = 43.05. At this optimal point,

$$\text{Tax shield} = \mathrm{Xf}\left(1 - \mathrm{e}^{-\mathrm{rT}}\right) = 43.05 \times 30\% \times (1 - \mathrm{e}^{-3.0\% \times 4}) = 1.2170$$
$$\text{Bankruptcy cost} = \mathrm{XN}(-\mathrm{d2}) - \mathrm{SN}(-\mathrm{d1}) = 0.1639$$

This firm (a typical firm) will get value-addition of 1.0531 (=1.2170–0.1639) or around 1% of the original firm value. This implies the firm can get a 1% value addition if it can keep optimal usage of debt capital until the debt maturity. In reality, no firm can keep its leverage on optimum during any year because the value of its equity fluctuates endlessly in the market, needless to say during the whole debt life. Thus the value-addition from capital structure decision or adjustment is much lower than what widely expected or guessed so far. Therefore, the appreciation potential from leverage depicted in Fig. 2 is just an unrealistic imagination without solving the problem of optimal capital structure.

On the other hand, over-leverage will damage much of the firm value. For instance, if the typical firm uses more debt and reaches a leverage ratio of 61.9368%, the tax shield (1.7509) is just cancelled out by the bankruptcy cost (1.7509); the trade off value is zero; the levered firm cannot obtain any value-addition from the usage of debt. If leverage ratio of the firm reaches 90% or 95%, based on the ZZ tax shield and bankruptcy cost models, the tax shield is 2.5443 and 2.6856 respectively; the bankruptcy cost is 10.5713 and 13.0808 respectively; the trade off value is −8.0270 and −10.3952 respectively; the levered firm gets a negative value-addition so that the firm value decreases from 100 to 91.9730 and 89.6048 respectively, or decreases by −8.0270% and −10.3952% respectively.

6 Explanations to Some Capital Structure Puzzles

Since the problem of optimal capital structure have remained unsolved for decades, more and more empirical observations are found inconsistent with academic hypotheses, which are referred to as capital structure puzzles. We explore some of them and provide alternative explanations in this section based on the ZZ leverage model.

6.1 Why Financial Conservatism

Based on overvalued tax shields and undervalued bankruptcy costs, most relevant studies suggest higher optimal leverage ratios. Consequently, the leverage ratios in reality seem too low. Graham (2000, 2001) finds that conservative debt policy (under-levered financing) is a persistent and pervasive capital structure puzzle. According to Tserlukevich (2008), the optimal market leverage ratio generally falls between 43 and 73%; while the common market leverage ratios reported by various surveys lie between 29 and 35%.

Financial scholars create a terminology of "financial conservatism" to explain the gap between the academic hypotheses and the reality, which implies that firms are too conservative in debt financing. However, based on a typical case, the optimal leverage ratio is 43.05% based on the ZZ leverage model, which is not so high as subjectively assumed by previous studies. Thus a more convincing explanation may be: the "theoretical standards" are too radical, and the actual leverage ratios are roughly appropriate.

In addition, we also find in previous section, the value-addition resulted from the optimal usage of debt capital is very limited (only around 1% of the firm value at most). Especially, the cost of being overlevered is asymmetrically higher than the cost of being underlevered. For the typical firm, the optimal leverage ratio is around 43%, which can add value 1%; but if the firm reaches a leverage ratio of 90%, its value will decrease by around 8%. Under-leverage will not hurt the firm value much; but over-leverage will led to a big loss of the firm value.

Why so many firms behave conservatively in financing? Now the reason cannot be clearer. Firms do not know how to find a precise optimal leverage, but they sense intuitively that over leverage may damage their firms' value seriously. In such a case, the lower or even zero debt ratios are all rational choices in absence of a reliable optimal leverage model, because it is not worthy to risk a big loss for a very limited value addition.

6.2 Why No Leverage Target

Prior research documents that many firms have no leverage target.[11] Now, based on the ZZ leverage model, we can explain this phenomenon as follows.

Firstly, there has been no reliable model with theoretical soundness and practical feasibility to support actual capital structure decisions, but executives sense that the potential of the value-addition via debt financing is very limited, so it is not worthy to put much times and efforts on determining and maintaining a "leverage target".

Secondly, there are various constrains for many firms to raise their necessary capitals in reality. Thus firms may just choose the most convenient way to raise their capitals. They do not care whether the capitals are equity or debt, or the effect of their choice on their capital structures, so long as their debts do not bring in too much bankruptcy risk.

Thirdly, because both debt and equity financing have transaction costs, adjusting capital structures will involve adjustment costs. Firms prefer to only adjust their capital structures by the chance of financing and let the leverage drift alone between

[11] See among others, Bradley et al. (1984), Graham and Harvey (2001).

the financing rounds, regardless how far of their capital structures deviates from the leverage targets (if they have the targets).

A firm may reach its optimal capital structure at a debt/equity issuance. Then the firm value changes inevitably with its business performance and the capital market conditions. This implies the leverages of most firms deviate from optimums most of the time. Thus, researchers actually cannot find leverage targets for most firms based on sample data.

6.3 Why Averse-Change with Profitability

Theoretically, firms can use more debt financing when they are expected to get more profitable. However, some scholars find significant exceptions, i.e., the leverage ratio decreases as the profitability increases, such as Strebulaev (2007).

When firms decide on whether or not to adjust capital structures, they should trade-off between the adjustment cost and the value-addition from the adjustment. Based on the above discussions, as the value-addition from adjusting capital structure is very limited, firms usually do not change their debt size before the next round financing.

However, if a firm gets more profitable, its value will increase immediately in the market and its debt ratio will consequently decrease. Therefore, an increase in profitability "naturally" lowers leverage by increasing future profits and the current firm value. As the same reasoning, a decrease in profitability "naturally" reduces firm value and raises leverage.

In short, because firms only adjust their capital structures periodically, their debt ratios drift naturally in a way of averse-changes with profitability most of the time.

6.4 Why Over Stable Leverage

Some scholars find that some firms do not adjust their capital structure in time even when some important conditions change significantly,[12] such as the changes of the corporate tax rate and the risk free rate. Now we can explain this phenomenon as follows.

Firstly, there has been no decision-oriented model to determine the optimal leverage; firms do not know how much debt they should add or reduce to reach the optimum after the relevant conditions change. To avoid unnecessary risk, they just copy the past experiences, and maintain their capital structures even they have adjusting opportunities.

Secondly, based on the ZZ leverage model, the optimal debt ratio is not very sensitive to some conditions, such as the corporate tax rate, the risk free rate, etc., although they are important influential variables for capital structure decisions.

[12] See among others, Lemmon et al. (2008).

For instance, other things being equal in the base case, but the corporate tax rate changes from 25 to 15% or 35%, based on the ZZ leverage model, the optimal debt ratio will change to 39.51% and 45.73% respectively, which is not far from the original 43.05%. Based on the ZZ trade off value model, when the corporate tax rate becomes 15% or 35%, the firm value at the optimum (39.51% and 45.73%) is 100.59 and 101.56 respectively. If the leverage is still at 43.05% (which is no longer the optimum), the firm value is 100.57 and 101.54 respectively; both are 0.02% less than the firm value at the new optimum. Obviously, the leverage adjustment makes no sense even in absence of adjustment cost.

The analysis above can also explain why firms seem to have optimal leverage ranges rather than precise optimal leverage targets. We thus conclude: theoretically, there is indeed a precise optimal leverage for every firm at a certain time; but practically, a firm needs to rebalance its leverage only when it deviates far away from the optimum.

6.5 Why Pecking Order

Myers and Majluf (1984) suggest that firms follow financing priorities to minimize information asymmetry between the firm's insiders (executives) and the outsiders (shareholders), i.e., they prefer internal sourced capitals like depreciation funds, retained earnings, etc. to outside sourced capitals; when they have to raise outside sourced capitals, they prefer debt to equity. This is referred to as pecking order hypothesis or pecking order theory.

Based on what we find so far, as the potential value-addition from the leverage adjustment is very limited, the leverage decision is not as important as other management decisions for a firm to pursue its value-maximization target, such as investment and operating decisions. Therefore, firms are more likely to choose a convenient way to raise their capital, so long as the debt capital is not much as a fraction of their total capital.

For most firms, the internal sourced capitals are the most convenient capitals. So, those internal sourced capitals are often the first ordered capitals. Beyond those capitals, however, there should be no certain pecking order for most firms. Firms in different environment have different favorite pecking order. For instance, while some researchers find that firms prefer debt to equity, publicly listed firms in China prefer equity to debt. There is actually no certain or unified pecking order for firms across markets and periods; firms simply choose their pecking order according to the ever-changing internal and external conditions.

There may be a pecking order for every firm over a certain period to financing its business, but the reason behind this order is not as complicated as guessed in prevailing research literature, such as information asymmetry, etc. It is just because every firm has its own advantage over various financing channels, so it may choose equity or debt first. Besides, the information asymmetry is not likely to be the consideration of top importance in firms' decision. The ultimate target for most firms is maximizing its value which should be the first ordered consideration in financing as well as capital structure decisions. If reducing information asymmetry outweighs maximizing firm value, firms should not do any business; they should just sell all the assets and deposit all the proceeds in banks.

The pecking order hypothesis actually cannot eliminate the problem of optimal capital structure. Suppose firms really make financing decision based on their "pecking order", there is still a problem of "what is the best debt ratio at which they should turn to the second ordered capital". More importantly, while most scholars regard the pecking order hypothesis as contrary to the optimal capital structure, it is actually not. Financing convenience may outweigh optimizing capital structure in some circumstances. That firms have pecking order for financing does not necessarily mean that they have no optimal capital structure.

In academic area, every issue is regarded as most important in the relevant field; but in practice, the importance varies a lot across managerial decisions. Firms make the important decisions carefully and make the less important decisions at their convenience. Firms may sense that there is an optimal leverage, but they just do not care, because optimizing capital structure is not as beneficial as regarded by scholars.

6.6 Why Market Timing

Some scholars find that the equity market timing is an influential factor on firms' capital structures, such as Graham and Harvey (2001), Baker and Wurgler (2002), Ilch (2004), etc. They suggest that firms issue new shares when their shares are overvalued and repurchase outstanding shares when their shares are undervalued. Consequently, fluctuations in stock prices affect firm's capital structures, and the current capital structure is the cumulative outcome of past attempts to time the equity market.

Many scholars regard the market timing theory or hypothesis as a contrary to the optimal capital structure (trade-off theory). This is a misunderstanding similar to the relation between the optimal capital structure and the pecking order hypothesis. The ZZ leverage model reveals that the pure capital structure consideration is not as important as other managerial or financial decisions, so it has less priority in firms' decision lists.

As analyzed above, for a typical firm, even it keeps continuously on optimal leverage during the debt life, the potential value-addition at most is 0.95%. However, the issuing price of stock may vary far beyond 10% at different time. Therefore, it is a rational choice for a firm to make more efforts on (equity) market timing rather than capital structure, although there is indeed an optimal leverage ratio in theory.

Please note this is by no means that the market timing is a factor considered in capital structure decision; rather, the market timing and the capital structure are two competing decisions with different priorities. If you cannot give enough consideration to both decisions, you have to ensure the market timing is considered first. In other words, your capital structure may deviate from its optimal level because of your market timing decision.

Obviously, while the market timing hypothesis may explain some actual capital structures (not necessarily optimal capital structure) to some extent, it cannot offer any help for solving the problem of optimal capital structure.

6.7 Dynamic Consideration?

In order to better explain the capital structures in reality, some scholars try to find insights beyond one single-period, such as Kane et al. (1984) and Brennan and Schwartz (1984), etc. Their arguments are referred to as the dynamic trade-off theory. Dynamic trade-off models also try to consider the option values embedded in deferring leverage decisions to the next period, such as Fischer et al. (1989)Goldstein et al. (2001) and Strebulaev (2007).

In the dynamic models or hypothesis, a firm's leverage responds less to short-run equity fluctuations and more to long-run value changes. The optimal financial leverage choice today depends on what is expected to be optimal in the next period. Thus, the dynamic trade-off theory explains the chaotic actual leverages by this way: although they seem not to be optimal in current period, they are expected to be optimal in next period.

Appearing brilliant with the "dynamic" label, the dynamic hypothesis is actually too academic. Firms may seem respond less to short-run fluctuations and more to long-run changes in their leverage decisions. However, this is not because they care their capital structure so much that they make a foresighted decision, but that they do not very care their capital structure so they only rebalance their leverage inactively.

Similar to the dynamic hypothesis, some scholars try to explain various capital structures puzzles by resorting to more factors far from tax shield and bankruptcy risk (cost), such as various macroeconomic factors, investment decision factors, and so on. However, the fact is that comparing with other important issues, it is not necessary for firms to care their capital structure so much; and they actually do not consider so much in their capital structure decisions. In other words, in absence of capital structure decision, they may still need to consider those macro or micro factors for the value creation of the firm.

6.8 Why Not 0% Debt in Absence of Corporate Tax

A fact seems contradictory to the trade off between tax shield and bankruptcy cost in capital structure decision. Some scholars find that firms during early years in American had some debt in their capital mix even when there is no corporate income tax.

In the environment without corporate tax, firms financing with debt cannot obtain any tax shield. If they made their capital structure decision based on trading off between tax shield and bankruptcy cost, the final decision should be 0% debt, because there is no tax shield but bankruptcy cost. This seems hard to explain the fact that debt financing was also used when corporate income tax did not exist.

The basic idea of optimal capital structure is: other things being equal, a firm should trade off between tax shield and bankruptcy cost to make its capital structure decision. The “other things” include of course the firm’s business mix. A firm often needs a certain amount of capital before a deadline for setting up a new business. The debt capital is often more convenient to obtain than the equity capital. To avoid loss the business opportunity, a firm may use debt financing to ensure the business opportunity, because the net benefit of the business opportunity is usually much larger than the bankruptcy cost of the debt financing.

In such a case, the firm should give priority to the trade off between the benefit and cost of the business opportunity rather than between the tax shield and bankruptcy cost. In another words, the investment is more important than the financing. Just as what revealed by the ZZ leverage related models, the net benefit of the new business is likely much larger than the net benefit (cost) of debt financing over equity financing. That is why firms still use debt capital even when there is no corporate income tax and debt financing has net cost rather than net benefit. Anyway, the ZZ leverage model reveals also that too much debt may damage the firm value significantly. So, if it is too urgent to raise additional equity capital and the net benefit of the new business cannot cancel out the net cost of debt financing, the firm should give up the business opportunity.

Therefore, the fact that firms had some debt in their capital mix when there is no corporate income tax is just the evidence that the value addition from leverage decision is not so important and the insights from the ZZ leverage related models are correct.

7 Summary

Optimal capital structure has been a bad headache in financial community. In front of the research difficulties and the impact of the empirical research “fashion”, the relevant studies actually deviate from how to find the optimal capital structure to how to explain the existing capital structure (disregard it is optimal capital structure or not) in recent decades.

Unfortunately, the very limited efforts left to the real research on “optimal capital structure” are mostly based on the wrong concepts of tax shield and bankruptcy cost—the two most important concepts in optimal capital structure studies. That is why the problem of optimal capital structure remains unsolved after intensive studies of half century.

In this chapter, we revise the prior erroneous concepts and redefine the tax shield and bankruptcy cost on one round financing basis. Based on the correct concepts of tax shield and bankruptcy cost as well as the MM models and Black–Scholes model, we develop the ZZ model series on leverage, which include the ZZ tax shield model and the ZZ bankruptcy cost model as well as the ZZ optimal leverage condition and the ZZ leverage model.

The ZZ leverage (capital structure) model accounts for four basic variables to determine the optimal debt ratio, namely corporate income tax rate, risk-free interest rate, debt maturity and firm volatility. Based on the typical values of the influential variables, the optimal debt ratio is 43.05% based on the ZZ leverage model. The more important insight revealed may be that the value loss for under leverage is significantly lower than the value loss from the over leverage. Therefore, it is straightforward to explain the long-lasting puzzle of "financial conservatism" as well as the other puzzles in capital structure.

The quantitation of bankruptcy cost is one of the logic bases for the ZZ leverage model series as well as the ZZ asset pricing model series. The three basic variables, risk-free interest rate, debt maturity and firm volatility, among the four basic variables in the ZZ leverage model, are also used in the ZZ asset pricing model series. This reinforces the correctness and consistence in logic of our models or solutions to various related financial problems.

References

1. Modigliani F, Miller MH. The cost of capital, corporation finance and the theory of investment. Am Econ Rev. 1958;48:261–97.
2. Modigliani F, Miller MH. Corporate Income taxes and the costs of capital: a correction. Am Econ Rev. 1963;53:433–43.
3. Kraus A, Litzenberger RH. A State-preference model of optimal financial leverage. J Finance. 1973;28:911–22.
4. DeAngelo H, Masulis R. Optimal capital structure under corporate taxation. J Financ Econ. 1980;8:7–29.
5. Myers SC, Majluf NS. Corporate financing and investment decisions when firms have information that investors do not have. J Financ Econ. 1984;13:187–222.
6. Narayanan MP. Debt versus equity under asymmetric information. J Finan Quant Anal. 1988;21(1):39–51.
7. Miles J, Ezzell JR. Reformulating the tax shield: a note. J Finance. 1985;40(5):1485–92.
8. Graham JR. How big are the tax benefits of debt? J Finance. 2000;55:1901–41.
9. Korteweg AG. The net benefits to leverage. J Finance. 2010;65:2137–70.
10. Ju N, Parrino R, Poteshman AM, Iisbach MS. Horses and rabbits? Trade-off theory and optimal capital structure. J Finan Quant Anal. 2005;40:259–82.
11. Leland H, Toft K. Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. J Finance. 1996;51:987–1019.
12. Tserlukevich Y. Can real options explain financing behavior? J Financ Econ. 2008;89:232–52.
13. Bradley M, Jarrell G, Kim EH. On the existence of an optimal capital structure: theory and evidence. J Finance. 1984;39:857–78.
14. Graham JR, Harvey C. The theory and practice of corporate finance: evidence from the field. J Financ Econ. 2001;60:187–243.
15. Strebulaev IA. Do tests of capital structure theory mean what they say? J Finance. 2007;62:1747–87.
16. Lemmon ML, Roberts MR, Zender JF. Back to the beginning: persistence and the cross-section of corporate capital structure. J Finance. 2008;63:1575–608.
17. Baker M, Wurgler J. Market timing and capital structure. J Finance. 2002;57:1–32.

18. Fischer E, Heinkel R, Zechner J. Dynamic capital structure choice: theory and tests. J Finance. 1989;44:19–40.
19. Goldstein RS, Ju N, Leland HE. An EBIT-based model of dynamic capital structure. J Bus. 2001;74:483–512.

Further Readings

20. Modigliani F, Miller MH. Some estimates of the costs of capital to the electric utility industry, 1954–57. Am Econ Rev. 1966;56:333–91.
21. Miller MH. Debt and taxes. J Finance. 1977;32:261–76.
22. Zhang Z. Determining the optimal capital structure based on revised definitions of tax shield and bankruptcy cost. Front Bus Res China. 2009;3(1):120–44.
23. Myers SC. The capital structure puzzle. J Finance. 1984;39:575–92.
24. Haugen RA, Senbet LW. Bankruptcy and agency costs: their significance to the theory of optimal capital structure. J Finan Quant Anal. 1988;23:28–48.
25. Baskin J. An empirical investigation of the pecking order hypothesis. Financ Manage. 1989;18:26–35.
26. Harris M, Raviv A. The theory of capital structure. J Finance. 1991;46(1):297–355.
27. Myers SC. Still searching for optimal capital structure. J Appl Corp Financ. 1993;6:4–14.
28. Leland HE. Corporate debt value, bond covenants, and optimal capital structure. J Finance. 1994;49:1213–52.
29. Andrade G, Kaplan SN. How costly is financial (not economic) distress? Evidence from highly levered transactions that became distressed. J Finance. 1998;53:1443–93.
30. Leland HE. Agency cost, risk management, and capital structure. J Finance. 1998;53:1213–43.
31. Graham JR. Do personal taxes affect corporate financing decisions? J Public Econ. 1999;73:147–87.
32. Morellec E. Asset liquidity, capital structure and secured debt. J Financ Econ. 2001;61:173–206.
33. Morellec E. Can managerial discretion explain observed leverage ratios? Rev Finan Stud. 2004;17:257–94.
34. Leland H. Predictions of default probabilities in structural models. J Investment Manage. 2004;2:1–16.
35. Leary MT, Roberts MR. Do firms rebalance their capital structures? J Finance. 2005;60:2577–619.
36. Sigrist F, Hirnschall C. Grabit, Gradient tree-boosted Tobit models for default prediction. J Bank Finance. 2019;102:177–92.
37. Hull JC. Risk management and financial institutions. 3rd ed. John Wiley & Sons; 2012. p. 249.
38. Zhang, Z. Volatility and company life expectancy—research based on ZZ Bankruptcy cost model, research on financial issues; 2018;91–99.
39. Zhang Z. Modeling corporate bankruptcy risk and its application. Monthly J Finance Acc. 2021;11:1–6.

Some Extensive Discussions of ZZ Leverage Model

Contents

We solve the problem of optimal capital structure in previous chapter. The ZZ leverage model series represent the fundamental solutions to optimal capital structure. For the application of the model, we need to consider some specific conditions in practice. We discuss on some extensive issues in this chapter, then explore the applications of the model under some specific conditions, finally illustrates the applications via a case.

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_11

1 Problem Solving and Research Assumptions

The purpose of scientific research is to solve problems. This is a necessary feature of science. The decisional science in social science should meet this requirement too. But the question is whether a problem is solved in decisional science is not as clear as in natural science. We try to make an intensive discussion on this issue here. Of course, the discussion is limited to the scope of decisional science or finance.

1.1 Theory and Problem Solving

Within the financial circle, it is less discussed on the issues concerning problems solving. Therefore, people lack the right and in-depth understanding about how to solve problems and how to judge whether a problem is solved. For example, as a pair of concepts, theory and practice, many people may think that practice solves problems and theory is just empty talk or talk on paper, and does not solve problems. Perhaps to the surprise of most people, problem solving actually depends not on practice, empirical research, or even applied research, but on theoretical research.

In the field of finance, the goal of practice is to make money; rather than to solve problem. For example, how should the (reasonable) loan interest rate be calculated? Such a problem can be ignored in practice. An interest rate can be determined according to intuition, experience, market and other standards. Perhaps the interest rate determined in this way is unreasonable and incorrect, but it doesn't matter as long as it does not affect financial institutions to make money or to do business. On the other hand, market opportunities are often an more important factor affecting the financial institutions to make money. The market opportunities may disappear in unexpected time. This means that timely or quickly decision-making is more important than correct decision-making. It is impossible for banks to wait until the theoretical sound solution was worked out and then make their interest rates and loan decisions.

Strictly speaking, decision-making in practice is certainly not to solve problems; But to solve problems that can be solved according to common sense or experience, that is, problems that can be solved without or with little time. Of course, it can also be said that material wealth is created by practice; The scientific research creates theory, which is the spiritual wealth of human society. Theory will contribute to the creation of material wealth, but theory itself is not material wealth. For example, with the theoretical model of interest rate, the determination of loan interest rate will be more accurate, and the bank's loan and interest rate decision-making will be more efficient and accurate; However, the direct determinant of the bank's profit is the bank's decision-making, not the theoretical model of interest rate.

In academic research, empirical research is another kind of research stands side by side with theoretical research. Related to the wrong belief that practice solves problems, there seems to be a common misunderstanding in the financial field, that is, empirical research is closer to practice, so it is more likely to solve problems. In fact, empirical research focuses on statistical description based on sample data rather than solving problems. Since data represent the relevant phenomena or the results of past decisions, empirical research is naturally the description of characteristics or law imagined from the phenomena or the results of past decisions.

Obviously, it is impossible to solve the problem based on the results of such decisions. For example, when people in banks cannot calculate the loan interest rate, they make loan decisions based on their intuition or following the crowd. Can we get to know how to calculate the loan interest rate by processing their actual loan interest rates afterwards? Similarly, when people in companies can't calculate their optimal capital structures, they just make their financing decisions intuitively or randomly. Can we calculate the optimal capital structure or debt ratio by processing their actual debt ratios afterwards?

Therefore, empirical research only describes the decision-making results after the event, and it is impossible to solve the decision-making problem, because the decision-making is considered in advance. Even, the purpose of empirical research is not to solve the problem, that is, it does not intend to solve the problem. Only theoretical research intends to and can solve problems. However, for various reasons (the reasons are worth discussing, see the following chapters for details), empirical research has occupied an absolute advantage in quantity of published papers over decades.

Theoretical research often obtains the basic solution of a decision-making problem, such as ZZ asset pricing model, ZZ optimal capital structure model and so on. These models may also need to be adjusted according to specific application scenarios and factors and the special requirements of users (such as companies, banks, etc.). How to adjust according to various specific or special factors is the task of applied research. As a bridge between theory and practice, applied research is very important. But generally speaking, compared with the establishment of the basic model, the adjustment according to various specific factors is simple and easy.

Moreover, it is conceivable that new application scenarios and specific factors will continue to appear over time; In other words, after the completion of theoretical solution, applied research is a continuous process; As long as the corresponding business still exists in practice, there are always application problems that need to be studied and solved. Therefore, it is impossible to have a clear boundary between the completion of research and noncompletion in applied research. Of course, there is no landmark result that (all) problems have been solved. In other words, as long as there are applications, there may be application problems that have not been solved. Therefore, whether a problem is solved can only be judged based on whether the theoretical solution (such as the model in finance) is found or worked out.

Theoretical solutions or models reveal decision variables and the internal relationship between decision variables and answers. The decision variables here are often the most basic or important variables, and the internal relationship is often the relationship in the simplest case. Because of this, theoretical solutions or models are often not suitable for simply or directly application; In other words, users must give play to their subjective initiative in practice, or need applied research to supplement and deal with or solve problems not involved in the theoretical model. It can also be understood that there is no such theoretical model in the world, that is, whether the user understands correctly, uses correctly, and inputs data correctly, the correct conclusion can be drawn. On the contrary, from understanding to using to the estimation of the value of decision variables, as long as it is incorrect in any one step, the conclusion is likely to be incorrect. Theoretical research to solve problems, applied research to build a bridge and correct application in practice are the necessary trilogy of solving financial problems from theory to practice.

1.2 Universal Premise Assumptions

As mentioned above, the theoretical model should incorporate in the simplest or most basic decision variables and represent the simplest and most basic relationship between decision variables and solutions or answers. In other words, the theoretical model is a universal conclusion under the universal premises.

The "universal" here should be easy to understand and accept, that is, the theoretical model needs to be widely applicable, not only applicable to specific time, place or conditions. It needs to be further understood that in order to ensure wide applicability, we should not consider more preconditions or variables, but on the contrary, we should simplify and consider preconditions and variables as much as possible. It is necessary to deduce the conclusion or model under the basic premise, that is, to abstract general premises from various specific scenes and omit special premises. In other words, the "universal" here represents simplicity and basic. When such a simple or basic conclusion or model is applied to a specific scene, it needs to be adjusted according to the specific situation.

On the other hand, there are preconditions or variables that must be considered. Once these preconditions or variables that must be considered are incorporated, conditions and variables should be considered as little as possible. Taking the optimal capital structure as an example, the premise cannot be simplified to not consider the corporate income tax and bankruptcy risk; In terms of variables, it cannot be simplified to exclude tax shield or bankruptcy cost. Of course, this is not to say that it is wrong not to consider tax shield or bankruptcy costs, but that such a conclusion cannot solve the problem. The reason why MM model does not solve the problem of optimal capital structure is just that it fails to consider bankruptcy cost. Of course, this is not to deny the theoretical contribution of MM model. MM model is a precious semi-finished product in the process of solving the problem. Although it is not the final solution, it is helpful to solve the problem finally.

As the simplest, most basic and most common premise assumption, it can be imagined that such premise assumption is easy to reach consensus, even self-evident. This is also the main reason why many models in this book do not specify their premise assumptions. Of course, before discussing the optimal capital structure, this topic is relatively inconvenient to illustrate with examples.

It can be seen that the solution of the optimal capital structure problem should be based on the following universal assumptions: corporate income tax and bankruptcy risk; The basic variables that should be considered are: tax shield and bankruptcy cost, or the basic variables that determine tax shield and bankruptcy cost, such as corporate income tax rate, corporate risk (volatility), etc. Of course, it is better to determine the specific variables not by subjective selection, but objectively based on strict logical reasoning, just like many models in this book. In fact, such universal or common premises need not be explained; there is nothing to explain for the same assumption as the reality. What needs to be explained is the assumptions that are different from the reality, such as those in the MM model.

1.3 Assumptions of the ZZ Leverage Model

As mentioned above, comparing with MM model, the most important assumptions of the ZZ leverage model are the coexistence of corporate tax and bankruptcy risk, which should be the universal assumptions for theoretical research on optimal capital structure.

The universal assumptions are relatively simple assumptions with least number of variables which can meet the most important requirements and features of the relevant decision-making, so that the problem can be solved with relative soundness and easiness.

Research assumptions are relative in properness except the universal assumptions for theoretical research; they may vary over time, since the practical conditions vary over time. Thus, if we do not further our discussion on the application aspect of the ZZ leverage model, we need not discuss on the assumptions of the model.

For the application of the ZZ leverage model, some additional conditions and influential variables need to be further incorporated into the model. In other words, the assumptions of the model are the differences in the premises of the model from the reality.

The main assumptions of the ZZ leverage model include:

(1) The optimal debt ratio calculated is on the reasonable or fair value basis, rather than the book value basis.
(2) All debts are loans or bonds on credit; and there is no mortgage, guarantee and other credit enhancement measures.
(3) The liabilities are all conventional liabilities, there is no liability with unclear boundary between equity and creditor's rights such as convertible bonds, perpetual bonds, etc.
(4) The firm value grows compounding-continuously at risk free rate, or the return matches risk, so there is no abnormal growth of equity and firm value.
(5) No short or long term inter-firm investments, in other words, no part of a firm's assets functioning concurrently as the support of other firm's debt obligation.

(6) The company does not provide or accept various forms of external debt guarantees, including various asset mortgages and guarantees.
(7) The debt is fair-priced so that the variable of the interest rate of the company debt can be cancelled out in the process of the model derivation.
(8) No transaction cost for both equity financing and debt financing; hence no adjustment cost for firms to change their capital structures.
(9) No personal income tax on personal income from debt or equity investment.
(10) As assumed in the Black–Scholes option pricing model, the firm value follows a geometric Brownian motion (GBM) with a constant volatility σ so that we can apply the option pricing model.

The assumptions (1)–(9) may be somehow different from the reality. For instance, the guaranteed or mortgage debts are common in reality; there are indeed transaction costs for both equity financing and debt financing; there may be indeed taxes on personal income from debt investment and equity investment; firms may grow faster or slower than the risk free rate or even expected to grow negatively in the future; there are indeed inter-firm investments, such as receivables and various equity investments; the debt may be more or less mispriced because of various reasons; etc. We try to relax these assumptions and find the relevant special solutions in next section.

As for assumption (10), although the firm value may move away from GBM sometime in reality due to special events, GBM is still a reasonable approximation of firm value dynamics except for rare events; and it is also among the simplest of stochastic processes. Within reasonable range, we should prefer simplicity to accuracy, because simplicity is a rare advantage for theoretical solutions and practical decisions. No models can be based on assumptions completely in line with the reality. Models on simple assumptions give more efficiencies and flexibilities for applications as long as the assumptions are roughly realistic.

There are also other assumptions implied in our model, such as that the business activities of a firm are unchanged by its capital structure, that the bankruptcy is only possible at the debt maturity when the firm value less than its debt full value (Xe^{rT}), etc. In reality, the firm activities may be changed more or less by its capital structure, but it is not commonly important. Similarly, a firm may go to bankrupt before or after its debt maturity, or the bankruptcy threshold may be somehow higher or lower than its debt value on maturity; but these cases are no longer difficult to solve with the ZZ leverage model series.

Financial issues in reality are complex; no model can take all assumptions and factors into account. Assumptions behind a financial model different from reality can make an endless list, such as constant volatility, constant risk free rate, no dividends, perfect capital market, etc. Nevertheless, for the simplicity, feasibility, efficiency and flexibility of the model, we will not consider further the less important factors or the indirect factors. As mentioned earlier, most of the indirect factors have been incorporated already into the four independent variables of the ZZ leverage model series.

As a matter of fact, the most important assumptions of the ZZ leverage model are still the coexistence of corporate tax and bankruptcy risk, which is also the most different assumptions of the ZZ leverage model from that of the MM model. Such a difference in assumptions implies that the ZZ leverage model is a real solution to optimal capital structure, and hence can be applied in practice.

2 Some Application Extensions

As theoretical models, the ZZ leverage models have clear logic bases, hence is not difficult to extend for incorporating additional assumptions and variables. We intend to explore the applications of the ZZ leverage model in various specific situations in this section by relaxing some of the assumptions mentioned in last section and account for more common specific factors in actual capital structure decisions.

2.1 Abnormal Growth

We have already accounted for firm value growth at a risk-free rate in the ZZ leverage model by applying the option pricing model. A firm may grow faster or slower during its debt life in reality. Suppose a firm is expected to grow at an abnormal rate of g in addition to r. How should we account for the additional growth rate of g?

Firm value S is the present value of its value at the debt maturity, S_T. In absence of abnormal growth, $S_T = Se^{rT}$; the firm value now is $S = Se^{rT}e^{-rT}$. If the firm's debt X reaches its optimal level now, the firm reaches its optimal leverage ratio L = X/S. Other things being equal, when the values of the firm and its debt grow normally, the optimal leverage ratio at the debt maturity is the same as now, i.e., $L = (Xe^{rT})/(Se^{rT}) = X/S$.

With the abnormal growth at g in addition to r, the firm values now and on debt maturity are Se^{gT} and $Se^{rT}e^{gT}$ respectively. Suppose the debt value is not affected by the firm value's growth, or it remains growing at a risk-free rate and to be Xe^{rT} on the maturity. The debt ratio at debt maturity is: $(Xe^{rT})/(Se^{rT}e^{gT})$. This implies that to reach the optimal leverage X/S at the debt maturity, the leverage ratio now should be:

$$\begin{aligned} L = X/S &= \left(Xe^{rT}\right)/\left(Se^{rT}e^{gT}\right) \times e^{gT} \\ &= e^{gT + probit(f - fe^{-rT})\sigma\sqrt{T} - \sigma^2 T/2} \\ &= \exp\{gT + \left[probit\left(f - fe^{-rT}\right)\right]\sigma\sqrt{T} - \sigma^2 T/2\} \end{aligned} \tag{1}$$

As an example, assume the same base case in previous chapter, i.e., the corporate tax rate f = 30%; the risk free rate r = 3.0%; the debt maturity T = 4; the return volatility σ = 20%, except that the firm has an abnormal growth rate g = 7.0% (so that the total growth rate is 10.0%). Suppose that the firm intends to reach optimal leverage at the debt maturity, i.e. $X/(Se^{gT}) = 43.05\%$; thus the optimal debt ratio now should be: $43.05\% \times e^{7\% \times 4} = 59.96\%$.

Note that the above reasoning can be applied to satisfy any growth pattern as well as any leverage policy of the firm. If the firm is expected to grow abnormally at 7% during only 2 years, or the firm intends to reach an optimal leverage in the end of year 2 for some reasons, the optimal leverage now is $43.05\% \times e^{7\% \times 2} = 49.52\%$.

It is commonly believed that firms with high ROA (return on total assets) should use more debt capital and have high leverages, because debt financing can increase share holders' return (ROE, return on equity) in such cases. The above analyses confirm this common sense. As illustrated in above analyses, the higher the growth rate, the higher the optimal leverage ratio. Since high growth rates are usually resulted from high return[1] (ROA), hence the higher the ROA, the higher the optimal leverage ratio.

2.2 Bankruptcy Expectancy

Bankruptcy expectancy implies the firm grows negatively during the debt life. We can account for such bankruptcy expectancy similarly as the "abnormal growth", except that the growth rate will be negative rather than positive.

Suppose the base case firm value is originally 100, and its debt value is originally 43.05. Now the firm value is expected to be 20 (< 43.05) at the debt maturity, hence the firm will go bankrupt then. The annual growth rate during the debt life is $\sqrt[4]{20/100} - 1 = -33.13\%$. Assume the same inputs as the base case firm except that the firm has a total growth rate -33.13% (so that the abnormal growth rate is $g = -33.13\% - 3.0\% = -36.13\%$). Hence the optimal capital structure now is: $43.05\% \times e^{-36.13\% \times 4} = 10.15\%$.

The optimal leverage ratio decreases sharply comparing to the original level of 43.05%. Actually, the bankruptcy expectancy may even drive the optimal leverage ratio down to zero. When a firm moves towards bankruptcy, it normally suffers from continuous loss without hope to recover to be profitable. Thus it has little chance to carry the loss forwards and to benefit from the tax shield of the debt capital. The debt can only contribute bankruptcy cost to the firm in such a case. This justifies an optimal leverage ratio of 0%.

Hence firms with bankruptcy expectancy can hardly borrow in reality. However, firms before bankruptcy may have high debt ratios. This just reflects that the ex-bankruptcy or distressed firms postpone their repaying of the due debts and their firm value decline sharply. The relatively high debt ratios are by no means the optimal leverages in such a case. This reminds us that an optimal leverage model cannot be derived from or to justified by sample data. In other words, theoretical research

[1] Growth rate is determined by the rate of return of the firm's investment and its dividend policy. Given the dividend policy, the growth rate is positively related to the rate of return.

(based on correct concepts and strict logic) is the only possible way to solve the problem of optimal capital structure.

2.3 Market Value Versus Book Value

Following the financial research convention, the ZZ leverage model series are on market value or fair value basis, but as some surveys revealed, firms concern more on their leverage on book values basis.[2] Hence, it makes sense to consider how to translate the fair value leverage into its corresponding book value leverage.

The fair value can also be referred to as theoretical value or bubble free value. Literally or normally, fair value should be greater than book value, since there should be some value addition coming from the effective use of the relevant asset; fair value should be smaller than market value or market price, since there should be some market bubble in the market value or market price of an asset, especially for a stock. But often, the market value or market price is used as the fair value of an asset for convenience.

The fair value of debt is usually close to its book value as long as the firm keeps healthy, so the debt fair/book ratio is usually close to 1. The fair value of equity may be several times of its book value. The equity fair value of a publicly traded firm is the product of its fair or market value per share and number of shares outstanding. The equity fair value of a privately held firm can be determined according to the similar publicly traded firms. Therefore, it is possible to estimate the equity fair/book ratio of a firm.

For the base case firm, assume the fair/book ratio of equity is 2, i.e., the book/fair ratio is 1/2; the fair/book ratio of debt is 1. As the optimal debt ratio L (=X/S) at fair value is 43.05%; the optimal equity ratio (=1 − L) at fair value then is 56.95%. Then, the optimal leverage ratio L at book value is: 43.05%/(43.05% + 56.95%/2) = 60.19%.

For the convenience in expression, hereafter, we refer to debt or leverage ratio on book value as book debt or leverage ratio, and debt or leverage ratio on fair value as fair debt or leverage ratio.

The result can be checked this way: the book equity ratio is: 1 – 60.19% = 39.81%. Thus, the fair debt ratio is: 60.19%/(60.19% + 2 × 39.81%) = 43.05%. This illustrates that the optimal book debt ratio of 60.19% is equivalent to the optimal fair debt ratio of 43.05%. To generalize the above process, denote the optimal book leverage by L′, the fair/book ratio of the equity by m, the optimal fair leverage ratio is:

$$\mathrm{L} = \frac{L'}{L' + m(1 - L')} \tag{2}$$

[2] Such as Servaes and Tufano (2006).

Thus, the model of optimal book leverage ratio is:

$$\mathrm{L}\prime = \frac{m}{1/L + m - 1} \tag{3}$$

where L is determined by the ZZ leverage model, and the fair/book ratio of debt is assumed to be 1. Thus, we can find the optimal book leverage ratio based on Eq. (3) and the ZZ leverage model. Clearly, when m = 1 in Eq. 3, L′ = L.

2.4 Guaranteed Debt

Guaranteed debts are even more common than credit debts in reality, especially for small firms. When a firm has debt guaranteed by an outside or independent guarantor, it has to pay for the guarantee. Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing" reveals that the fair guarantee fee equals to a put option in value which is used to hedge the bankruptcy risk. So, the bankruptcy cost of the guaranteed debt doubles (200%) that of the credit debt. Based on Eq. (12) in Chapter "Tax Shield, Bankruptcy Cost and Optimal Capital Structure", for a regular credit debt, the optimal leverage, L, satisfies:

$$\mathrm{N}[\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}] = \mathrm{f}\left(1 - \mathrm{e}^{-\mathrm{rT}}\right) \tag{12}$$

Then, for a guaranteed debt, the optimal leverage, L, satisfies:

$$2\mathrm{N}[\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}] = f(1 - \mathrm{e}^{-rT}) \tag{4}$$

That is,

$$\begin{aligned} \mathrm{L} &= \mathrm{e}^{probit(f/2 - f\mathrm{e}^{-rT}/2)\sigma\sqrt{T} - \sigma^2 T/2} \\ &= \exp\{\left[\mathrm{probit}\left(\mathrm{f}/2 - \mathrm{fe}^{-\mathrm{rT}}/2\right)\right]\sigma\sqrt{T} - \sigma^2\mathrm{T}/2\} \end{aligned} \tag{5}$$

According to Eq. (5), the effect of debt guarantee on the optimal capital structure is equivalent to a half reduction of the corporate tax rate.

Consider the base case again, other things being equal, i.e. f = 25%, r = 3.0%, T = 4, σ = 20%, but the firm needs a guarantee for its debt and pays a fair upfront guarantee fee. Based on Eq. (5), the optimal leverage now is 38.39%, which is lower than the original optimal leverage of 43.05%. This reflects the guarantee makes the debt financing less attractive. Why some small firms or startups with high growth rates have lower leverage ratios? Why do they prefer equity financing? Obviously, one reason is that debt financing is less attractive to them, or debt guarantees are difficult and costly to find.

2.5 Transaction Costs

There are normally transaction costs associated with financing. When firms issue bonds, there are underwriting spreads as well as registration and legal fees. When firms borrow from banks, there may be also some implicit or explicit costs in addition to the interest costs. Similarly, when firms raise funds with equity, there are also or even more substantial transaction costs as a fraction of the amount of the capital raised.

In case of the transaction costs, firms had better to adjust their capital structures by the chance of financing, so that the adjustments will not generate additional transaction costs. On the other hand, a firm's capital structure will change automatically along with the fluctuations of its market value. This implies that it is not possible for a firm to keep on optimal leverage via continuous or dynamic adjustment. That is to say, a firm has to let its capital structure suboptimal in between the adjustments or financing rounds. Leary and Roberts (2005) show that the adjustment costs having a persistent effect on leverage.

Firms can adjust their capital structures by issuing debt or equity at financing. When a firm adjusts its capital structure by the chance of financing, the transaction costs of debt or equity or both are necessary in despite of their capital structure adjustments. The adjustment cost as an additional cost is insignificant in such a case. However, when firms adjust their capital structures between two financing rounds, they need to issue additional equity and prepay some debt, or issue additional debt and repurchase some outstanding equity. In either case, the adjustment cost is the sum of the two-way transaction costs of debt and equity financing which is obviously significant.

Firms thus should adjust their capital structures by the chance of financing, so that they can avoid most of the additional adjustment costs. Under such a consideration, viewing from the debt side, the adjustment cost is the additional transaction cost of debt over equity financing, which can be either positive or negative. The additional debt transaction cost results from similar reasons as that of the guarantee fee, such as the uncertainty of the business or the insecurity of the debt repayment, and has the same features, such as both are the upfront payments. Thus, the adjustment costs can be treated in a similar way as that of the guarantee fee.

Suppose the expected adjustment cost of the debt financing sums to h times the guarantee fee, the optimal leverage, L, satisfies:

$$(1+\mathrm{h})\mathrm{N}[\frac{\ln(L)}{\sigma\sqrt{T}}+\frac{\sigma\sqrt{T}}{2}]=f(1-\mathrm{e}^{-rT}) \tag{6}$$

Or,

$$L = e^{probit[f(1-e^{-rT})/(1+h)]\sigma\sqrt{T}-\sigma^2 T/2}$$
$$= \exp\{\left[\text{probit}\left(f - fe^{-rT}\right)/(1+h)\right]\sigma\sqrt{T} - \sigma^2 T/2\} \quad (7)$$

According to Eq. (7), the effect of an adjustment cost of the debt financing on the capital structure decision is equivalent to replacing the corporate tax rate f by f/(1 + h). Exemplifying again with the base case firm, assume the same values for f, r, σ and T, except that the firm now needs to pay an additional upfront transaction cost (adjustment cost) proportional as 30% of the debt guarantee fee, since f/(1 + 30%) = f/1.3. Based on Eq. (7), the optimal capital structure is now 41.17%.

The adjustment cost makes the debt financing slightly less attractive. Of course, if the debt is cheaper than equity in terms of transaction cost, it will be more attractive. For instance, other things being equal, but comparing with the equity financing, the debt financing has a negative additional transaction cost of −30% of the guarantee fee. Based on Eq. (7), the optimal leverage ratio is 45.91%. Thus, the transaction cost advantage makes the debt financing more attractive.

2.6 Personal Income Tax

As the same reasoning of the transaction cost, for the capital structure decision, the personal income tax is only relevant when there is significant difference in personal income tax between debt income and equity income. Anyway, we can extend the ZZ leverage model to incorporate this factor whenever it is necessary.

The debt interest as personal income is usually taxed at a higher rate than the income from the stock, which results in a "personal tax penalty" to the interest income. Therefore, investors usually demand a higher pre-tax return for their debt capitals. Miller (1977) tried to improve the MM model by factoring in the effect of the personal tax, Miller's formula is,

$$V_L = V_U + [1 - \frac{(1-f)(1-f_{pe})}{(1-f_{pd})}]D \quad (8)$$

where V_L is the value of the firm with debt in its capital; V_U is the value of the firm with debt in its capital; D is the amount of debt; f is still the tax rate on corporate income, f_{pe}, f_{pd} are the tax rates on personal income from equity investment and debt investment respectively. Thanks to Miller's contribution, we can factor in the "personal tax penalty" by using Miller's formula to adjust f in the ZZ leverage model,

$$N[\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}] = [1 - \frac{(1-f)(1-f_{pe})}{(1-f_{pd})}]\left(1 - e^{-rT}\right) \quad (9)$$

or,

$$L = e^{\mathrm{probit}\left\{\left[1-\frac{(1-f)(1-f_{pe})}{(1-f_{pd})}\right](1-e^{-rT})\right\}\sigma\sqrt{T}-\sigma^2T/2}$$

$$= \exp\{\mathrm{probit}[\frac{(1-f_{pd})-(1-f)(1-f_{pe})}{(1-f_{pd})} \times (1-e^{-rT})]\sigma\sqrt{T}-\sigma^2 T/2\} \tag{10}$$

Again, take the base case as an example, other things being equal, i.e. $f = 25\%$, $r = 3.0\%$, $T = 4$, $\sigma = 20\%$, assume further $f_{pe} = 10\%$, $f_{pd} = 20\%$. Thus $[1 - \frac{(1-f)(1-f_{pe})}{(1-f_{pd})}]$ $= 15.63\%$. Based on Eq. (10), the optimal leverage ratio is $L = 39.78\%$.

Comparing with the original optimum 43.05%, the personal tax penalty makes the debt financing slightly less attractive. Understandingly, if the personal tax rate on equity income exceeds that on debt income, i.e. a special situation with negative personal tax penalty, the debt financing will be more attractive. For instance, other things being equal, but $f_{pd} = 0\%$. Then $[1 - \frac{(1-f)(1-f_{pe})}{(1-f_{pd})}] = 32.50\%$. Based on Eq. (10), the optimal leverage ratio is 45.12%, the debt financing now is more attractive.

2.7 Inter-Firm's Investments

Consider ten firms in an economy with firm value 100, 200, 300, …, 1000 million dollars respectively. The total value of these firms is 5500 million dollars. If these firms are all the "typical" firms (like the base case firm), their optimal debts should be: $43.05\% \times 100 = 43.05$, $43.05\% \times 200 = 86.10$, $43.05\% \times 300 = 129.15$, …, $43.05\% \times 1000 = 430.5$ million dollars respectively. The optimal debt ratio for the ten firms as a whole is 43.05%, or the sum of the ten firms' optimal debt are $43.05\% \times 5500 = 2367.8$ million dollars.

However, firms in reality are not isolated. They have various relations with each other. A common relation is that they invest capital in or receive capital from other firms. The investments may take various forms, such as receivables (payables) or some long-term investments as debt or equity. Because of this, part of a firm's value may be also part of another firm's value. For instance, a firm's receivables may be just another firm's payables.

Therefore, getting rid of the double counting, the total value of the above ten firms is less than 5500 million dollars. If every firm has 30% value as the investments (equity or/and debt) in other firms, the total assets of the ten firms, which can be used to support their debt repayments, is $(1 - 30\%) \times 5500 = 3850$ million dollars. Based on the asset value without double counting, the total optimal debt of the ten firms is $43.05\% \times 3850 = 1657.43$ million dollars. This implies the average optimal debt ratio is 30.13% (= (1 – 30%) × 43.05%) rather than 43.05% when all the firms are regarded as isolated to others.

The capital structure in reality is not only the choice of the firm (debtor), but rather the result of the negotiation between the firm and the lender (creditor), such as the commercial banks. Lenders want their money safe, or they want every debt has enough assets at its back. Even the above ten firms prefer a capital structure of 43.05% debt, lenders who are rational enough to know the inter-firm's investment may insist that the debt ratio of every firm is not obviously more than 30.13%.

The extensive applications above illustrated that the ZZ leverage model series is flexible enough to accommodate capital structure decisions in various real situations. Obviously, the extensive discussions in this section enrich or enlarge the content of the ZZ leverage model series.

3 A Case Study Based on ZZ Leverage Model Series

We illustrate in this section the application of the ZZ Leverage Model series by an actual case study on a Chinese public company, Haier, including the determination of Haier's basic optimal debt ratio based on the ZZ Leverage Model and the final optimal debt ratio after adjustments based on the specific scenario of Haier, as well as the analyses and evaluation on the past and current capital structure of Haier.

3.1 Haier and Its Leverage

Haier was founded in 1984. Under the leadership of its main founder Zhang Ruimin and his team, Haier started from a neighbourhood factory on the verge of bankruptcy, promoted development with quality and service, and rapidly grew into the most competitive and influential household appliance brand in China. Today, Haier has become a provider of a full range of household appliances and the first brand of white appliances in China and the world. The company has established dozens of manufacturing bases, overseas sales companies and multiple comprehensive R&D centers around the world. According to the data released by Euromonitor, the world's authoritative market research organization, Haier has ranked first in the global sales of refrigerators, washing machines, wine cabinets, freezers and other household appliances and goods for many years.

Haier's shares went public on the Shanghai Stock Exchange in October 1993. After the initial public offering (IPO), the company's business has developed by leaps and bounds, and continues to develop new products and businesses. The company's businesses, assets and profits have increased steadily and become a veritable blue chip in China's stock market. As a listed company, Haier was once referred to as Qingdao Haier, and then changed to Haier Zhijia from July 2019. For convenience, we use Haier in our analyses.

Like other companies, Haier's capital structure decision also depends on the manager's experience and intuition because the optimal capital structure problem has not been solved in mainstream of finance. Table 1 shows the assets, liabilities and debt ratio, i.e. capital structure of Haier in recent 20 years. It can be said that it is the result of the company's capital structure decision. Of course, it is also possible that among the various decisions of the company, the capital structure decision is not important, and this result is only the incidental result of various decisions.

It can be seen that in the past 20 years, Haier's debt ratio has generally shown an upward trend; It was as low as 10.67% in 2005 and as high as 71.37% in 2016; In recent years, it has been basically stable between 65 to 70%. According to preliminary judgment, in such a huge span, it is unlikely that the debt ratio of each year represents a reasonable level. Maybe it means that Haier has no target capital structure; Or top executives do not care much about capital structure decision-making. In any case, as long as there are models or methods that can calculate the optimal capital structure

Table 1 Debt and debt ratios of Haier (in million yuan except the debt ratio)

Year	Current	Noncurrent	Total debt	Equity	Total asset	L (%)
2001	1613.31	2.21	1615.52	4932.18	6942.41	23.27
2002	1903.11	–	1903.11	5089.95	7394.14	25.74
2003	1392.26	144.61	1536.87	5388.97	7372.71	20.85
2004	783.45	146.87	930.32	5718.52	7107.06	13.09
2005	721.73	1.37	723.09	5598.70	6777.50	10.67
2006	2106.94	39.80	2146.74	5785.94	8476.75	25.33
2007[a]	4043.22	89.61	4132.84	7056.13	11,188.97	36.94
2008[b]	4444.54	85.00	4529.54	7701.06	12,230.60	37.03
2009[c]	8698.04	46.88	8744.91	8752.24	17,497.15	49.98
2010[d]	18,673.35	1105.02	19,778.37	9488.79	29,267.16	67.58
2011	25,932.66	2251.88	28,184.54	11,538.95	39,723.48	70.95
2012	31,341.24	2920.94	34,262.18	15,426.14	49,688.32	68.95
2013	38,005.67	3016.05	41,021.72	19,994.14	61,015.86	67.23
2014	41,628.06	4258.43	45,886.49	29,119.97	75,006.46	61.18
2015	39,783.31	3775.10	43,558.41	32,402.26	75,960.67	57.34
2016	73,452.86	20,222.07	93,674.92	37,580.37	131,255.29	71.37
2017	76,894.45	27,818.65	104,713.10	46,750.01	151,463.11	69.13
2018	80,081.66	31,487.60	111,569.27	55,130.28	166,699.54	66.93
2019	95,609.74	26,854.64	122,464.38	64,989.86	187,454.24	65.33
2020	109,392.85	25,955.64	135,348.49	68,111.01	203,459.50	66.52

[a] Current = current debt
[b] Noncurrent = noncurrent debt
[c] L = debt ratio
Source compiled according to the data of Haier's financial statements for each year

or the reasonable range of capital structure of Haier, these doubts and guesses will come out.

3.2 The Optimal Leverage Based on ZZ Leverage Model

According to the ZZ leverage model, that is, the optimal debt ratio, depends on four variables. Among them, corporate income tax rate, f, and risk-free rate, r, are variables reflecting macroeconomic conditions; the debt maturity, T, and volatility, σ, are variables reflecting the situation of the company. The values of the four basic variables are estimated according to the situation of China and Haier Company as following.

(1) Corporate tax rate

According to the current situation in China, the corporate income tax rate is usually 25%. In the future, the income tax rate of 25% will be applied in the long term.

(2) Risk free rate

Risk free rate is the rate of return without default risk. According to the convention in financial sector, the risk-free rate can be estimated according to the yield to maturity of treasury bonds. In order to ensure that the estimated risk-free rate is applicable before the debt maturity, the treasury bond used for estimating yield should has maturity equal to or longer than the debt maturity. Since the debt maturity of domestic companies usually does not exceed 3 years, it should be reasonable to estimate the risk-free rate by taking the simple average of treasury bond yields with maturity of 1–5 years (Table 2).

The average of the treasury bond yields is 2.51%; we shall use 2.51% as the risk-free rate hereafter in the calculations.

(3) Market price and volatility

The volatility in ZZ leverage model refers to the volatility of the company's overall return, which usually needs to be estimated separately from equity capital and debt, and then the overall volatility of the company can be obtained.

According to the industry practice, the volatility of the company's equity can be calculated based on the trading price data of its listed shares for a continuous period of time, and then annualized to obtain the annual volatility. Table 3 shows the closing prices of Haier's shares in the recent three months, i.e. from July to September 2021.

Table 2 The quotation of treasury bonds (yield in percentage)

Maturity	1	2	3	5	Average
Yield (%)	2.3327	2.4912	2.5122	2.7134	2.51

Source China bond information network on September 30, 2021

Table 3 The closing prices of Haier stock (in yuan) on trading days in the three months from July to September 2021

Date	6–30	7–1	7–2	7–5	7–6	7–7	7–8	7–9
Price	25.54	26.56	25.33	24.7	24.87	25.08	24.75	25.29
Date	7–12	7–13	7–14	7–15	7–16	7–19	7–20	7–21
Price	25.43	27.21	26.9	26.88	26.39	26.88	26.56	27.21
Date	7–22	7–23	7–26	7–27	7–28	7–29	7–30	8–2
Price	26.56	25.87	25.6	24.08	24.08	23.33	24.61	25.53
Date	8–3	8–4	8–5	8–6	8–9	8–10	8–11	8–12
Price	26.68	26.43	26.66	26.39	26.91	27.43	27.49	27.21
Date	8–13	8–16	8–17	8–18	8–19	8–20	8–23	8–24
Price	27.26	27.93	26.96	26.75	26.47	25.74	26.18	26.52
Date	8–25	8–26	8–27	8–30	9–1	9–2	9–3	9–6
Price	26.65	25.75	26.2	25.65	28.96	28.41	28.44	28.8
Date	9–7	9–8	9–9	9–10	9–13	9–14	9–15	9–16
Price	28.58	28.16	27.79	28.19	27.31	26.64	26.02	25.71
Date	9–17	9–22	9–23	9–24	9–27	9–28	9–29	9–30
Price	26.8	26	25.25	24.66	24.89	25.63	26.03	26.15

Source Dongfang fortune.com

Based on the data in Table 3, the continuous compound daily returns of Haier stock are calculated in this period. The calculation results are shown in Table 4.

Based on the yield data in Table 4, the standard deviation or the daily volatility can be worked out, which is 2.9118%. The annual volatility of Haier's stock then can be derived by annualizing the daily volatility based on 240 trading days a year, which is 45.11%.

Haier has not issued regular bonds except some convertible corporate bonds to support business development. However, the volatility of convertible bonds is similar to that of stocks, and its volatility of return cannot represent the risk of its debt. Since China's listed companies usually issue convertible bonds when they intend to raise debt in the market, it is difficult to find the data of ordinary bonds issued by similar comparable companies. Therefore, it is difficult to directly calculate the volatility of its debt. For simplicity and convenience, simply assume that the bond volatility is 1/10 of its stock, that is, the volatility of debt is 4.51%.

To find the overall volatility of the company, we need to know the volatility of its equity and debt as well as the proportions of the debt and equity in the total capital. Based on the debt ratios in Table 1. The average over the past five years was 67.86%, which is greatly affected by the extreme value in 2016; The average of the past 10 years, 66.49%, is relatively more representative. Therefore, take 66.49% as Haier's long-term stable debt ratio level; the corresponding equity ratio is 33.51% (= 1 – 66.49%).

Table 4 The daily returns of Haier stock on trading days in the three months from July to September 2021

Date	6–30	7–1	7–2	7–5	7–6	7–7	7–8	7–9
Return	–	0.0392	−0.0474	−0.0252	0.0069	0.0084	−0.0132	0.0216
Date	7–12	7–13	7–14	7–15	7–16	7–19	7–20	7–21
Return	0.0055	0.0677	−0.0115	−0.0007	−0.0184	0.0184	−0.0120	0.0242
Date	7–22	7–23	7–26	7–27	7–28	7–29	7–30	8–2
Return	−0.0242	−0.0263	−0.0105	−0.0612	0.0000	−0.0316	0.0534	0.0367
Date	8–3	8–4	8–5	8–6	8–9	8–10	8–11	8–12
Return	0.0441	−0.0094	0.0087	−0.0102	0.0195	0.0191	0.0022	−0.0102
Date	8–13	8–16	8–17	8–18	8–19	8–20	8–23	8–24
Return	0.0018	0.0243	−0.0353	−0.0078	−0.0105	−0.0280	0.0169	0.0129
Date	8–25	8–26	8–27	8–30	9–1	9–2	9–3	9–6
Return	0.0049	−0.0344	0.0173	−0.0212	0.1214	−0.0192	0.0011	0.0126
Date	9–7	9–8	9–9	9–10	9–13	9–14	9–15	9–16
Return	−0.0077	−0.0148	−0.0132	0.0143	−0.0317	−0.0248	−0.0235	−0.0120
Date	9–17	9–22	9–23	9–24	9–27	9–28	9–29	9–30
Return	0.0415	−0.0303	−0.0293	−0.0236	0.0093	0.0293	0.0155	0.0046

Note that 66.49 and 33.51% represent the debt and equity ratio in book value. What is needed to calculate the company volatility is the debt and equity ratio in reasonable or fair value. According to the closing price on September 30, 2021, the price to book ratio (P/B ratio) of Haier's stock (equity) is 3.33.

However, the stock price in market inevitably consists some bubble, this 3.33 may not represent the fair price to book ratio of Haier stock. We thus had better to calculate the bubble free or fair price to book ratio of Haier stock based on the fair or theoretical price to book ratio model derived in Chapter "Stock and Equity Valuation: Where Discounting Does Not Work", which takes the form,

$$\mathrm{P/B} = \left[(1+\mathrm{g})^{\mathrm{n}} - 1\right](1+\mathrm{g})\mathrm{r_e}/\mathrm{g} \tag{11}$$

where, g is the average annual growth rate of earnings over foreseeable future; n is investor's required rate of return; r_e is the future annual return on equity.

Haier has achieved rapid expansion and growth since its listing; its average annual growth rate of earnings has reached 16% in past 20 years; and the growth is more than likely to continue in the coming years. For convenience, just simply assume an average annual growth rate of 10% over the foreseeable future (such as 10–20 years).

As a blue chip stock in China's stock market, Haier's risk is lower than the average level. Based on an average level of investor's required rate of return of 10%, the required payback period of 10 years. Then a little longer required payback period on Haier stock, such as 11 years, seems to be a fair or good estimation.

Simply taking the return on equity as 12%, fair P/B ratio of Haier stock then is:

$$\left[(1+10\%)^{11}-1\right](1+10\%)\times 12\%/10\% = 2.4461\ \text{times}$$

Then the debt ratio in fair value is:

$$66.49\%/(66.49\% + 33.51\% \times 2.4461) = 44.79\%;$$

The corresponding equity ratio is:

$$100.00\% - 44.79\% = 55.21\%$$

Given the respective proportion and volatility of the company's equity and debt, we also need to know the correlation between the return of equity and debt in order to calculate the overall volatility of the company. Considering that under normal circumstances, the fluctuation of EBITs will affect the earnings of equity, but will not affect the return of the company's debt. Therefore, it can be assumed that the return of the company's debt capital is not related to the return of its equity capital (i.e. the correlation coefficient is 0). Then the overall volatility of the company's capital is:

$$\begin{aligned}\sigma &= (\sigma e^2 we^2 + \sigma d^2 wd^2 + 2we\sigma ewd\sigma d \times 0)^{0.5}\\ &= (4.51\%^2 \times 44.79\%^2 + 45.11\%^2 \times 55.21\%^2)^{1/2} = 24.99\%\end{aligned}$$

(4) The average debt maturity

China's listed companies generally have a preference for equity financing, that is, they prefer equity to debt in financing. When equity financing is limited or not available at the moment, short-term debt or current debt will be used temporarily or circularly to satisfy the capital demand, so that the equity capital can be used to replace the short-term debt capital whenever the equity capital is available.

It can be seen from the data in Table 1 that Haier has a similar tendency. In the past 20 years, Haier's current liabilities accounted for an average of 81.38% of all debts. Of course, if the data of the first 10 years (2001–2010) are calculated, the situation is more obvious, and the current liabilities account for 96.39%. Of course, this also shows that the preference for equity financing has improved in recent years. Based on the data of recent five years, this proportion decreased to 76.69%.

Haier's financial report does not disclose the maturity of long-term debt in detail, but it has disclosed the amount of long-term or noncurrent debt due within one year in the balance sheet. According to the annual report data of 2019 and 2020, the amounts are 7317.1389 million yuan and 7522.7249 million yuan respectively. In addition, the total current debt were 95,609.7374 million yuan and 1093.9285 million yuan respectively, and the total noncurrent debt were 26,854.6386 million yuan and 25,955.6384 million yuan respectively. Based on those data, the ratio of the noncurrent debt due beyond one year to that within one year can be calculated as,

$$2019{:}26854.6386/7317.1389 = 3.67(\text{Times})$$
$$2020{:}25955.6384/7522.7249 = 3.45(\text{Times})$$

The difference between the two years is insignificant, indicating that this multiple is relatively stable and representative. The average is:

$$(3.67 + 3.45)/2 = 3.56(\text{Times})$$

Assuming that the maturity time of the company's noncurrent debt is evenly distributed, note that the minimum and maximum maturity times of noncurrent debt are 1 year and 4.56 year (=3.56 + 1) respectively. The average term of Haier's noncurrent debt then is calculated as follows:

$$(1 + 4.56)/2 = 2.78(\text{years})$$

It is understandable that the data in recent 5 years are more fair and representative for the future. Therefore, the following calculation is based on current debt accounting for 76.69%. Accordingly, noncurrent debt accounted for 23.31% (= 1 – 76.69%). Assuming that the average term of current debt is 0.5 years, the average term of Haier's debt is:

$$2.78 \times 23.31\% + 0.5 \times 76.69\% = 1.03(\text{years})$$

Note that the average term here is actually the average remaining life of the debt. For the existing debt, the remaining life is shorter than the contract life. Considering the actual debt financing scenario, the remaining life of the new debt is the longest, that is, it is equal to the contract life. Therefore, it is necessary to calculate the average life according to the longest contract life of current debt and noncurrent debt. Note that the maximum life of current debt and noncurrent debt is 1 year and 4.56 years respectively according to the previous data. Then the average contract term is:

$$4.56 \times 23.31\% + 1 \times 76.69\% = 1.83(\text{years})$$

In other words, it makes sense to take the debt maturity of 1.03 years and 1.83 years when calculating; or take a certain length of time between 1.03 and 1.83 years. The longer the debt term, the lower the optimal debt ratio, which leads to the more cautious decision-making. In practical application, the degree of prudence can be adjusted according to the changes of internal and external conditions of the company.

Considering the preference of listing companies in China market for equity financing and current debt financing, we choose a debt term 1.8 years to calculate the optimal capital structure of Haier, which is close to the up bound 1.83 years and is likely more representative in the future.

Summarize the estimation results of the basic data in the ZZ leverage model: corporate tax rate f = 25%; risk free rate r = 2.51%; Average debt maturity t =

1.8 years; company volatility = 24.99%. Then, based on the ZZ leverage model,

$$L = e^{\text{probit}\left(25\% - 25\%\, e^{-2.51\% \times 1.8}\right) \times 24.99\% \sqrt{1.8} - 24.99\%^{2} \times 1.8/2} = 43.88\%$$

That is, Haier's optimal debt ratio is 43.88% based on the basic influential factors in the ZZ leverage model.

3.3 *The Adjustments Based on the Application Extensions*

The above optimal debt ratio, 43.88%, is only the preliminary answer to the optimal capital structure of Haier, and the final answer can only be obtained after adjustment according to the relevant situation of Haier. In reality, most companies prefer to calculate and control the capital structure according to the book value; some companies expect extraordinary growth in the future; some companies need debt guarantee, while others provide external debt guarantee; Some companies have external equity or debt investment. The following is the adjustments based on the "basic solution" of 43.88% according to the situation of Haier to obtain the final solution.

For the situation of Haier, there are individual income tax differences, book value differences, transaction cost differences, external guarantee and investment, abnormal growth and other situations, which need to adjust on the basis of the basic optimal leverage. However, when using the methods and models discussed in the previous section to adjust, there is a problem of the order of adjustment. In other words, it may be more feasible and convenient to adjust various factors in a certain order. The following is the optimized arrangement of the adjustment order in case of Haier.

Adjustment 1: The personal income tax

Suppose Haier intends to consider the difference between the personal income tax on debt income (interest) and equity income (dividend and value addition) when making capital structure decision. Given that the personal income tax rate on debt income and equity income are 20% and 10% respectively, i.e., $f_{pe} = 10\%$, $f_{pd} = 20\%$. Other things being equal, i.e., corporate tax rate f = 25%; risk free rate r = 2.51%; average debt maturity t = 1.8 years; company volatility = 24.99%. Then,

$$[1 - \frac{(1-f)(1-f_{pe})}{(1-f_{pd})}] = [1 - \frac{(1-25\%)(1-10\%)}{(1-20\%)}] = 15.63\%$$

Based on Eq. (10) and the ZZ leverage model,

$$L = e^{\text{probit}\left\{\left[1 - \frac{(1-f)(1-f_{pe})}{\left(1-f_{pd}\right)}\right]\left(1 - e^{-rT}\right)\right\}\sigma\sqrt{T} - \sigma^2 T/2}$$

$$= e^{\text{probit}\left(15.63\% - 15.63\%\, e^{-2.51\% \times 1.8}\right) \times 24.99\% \sqrt{1.8} - 24.99\%^{2} \times 1.8/2} = 41.41\%$$

That is, the optimal debt ratio decreases slightly from 43.88 to 41.41% because of the influence of the difference in personal income tax.

Adjustment 2: The guarantee and transaction cost for debt financing

As revealed in previous section, the debt guarantee and transaction cost for debt financing will reduce the optimal leverage. In case of Haier, it need not guarantee for debt financing, rather, it provides guarantee for other companies to raise debt capital. We will analyze the situation of providing guarantee for others later.

Thus, no debt guarantee adjustment needed for Haier. However, based on Eqs. (5) and (7), the adjustment of guarantee and transaction cost can be considered together. That is, Eq. (7) can be rewrite as,

$$\begin{aligned} L &= e^{\mathrm{probit}\left[f\left(1-e^{-rT}\right)/(1+j+h)\right]\sigma\sqrt{T}-\sigma^2 T/2} \\ &= \exp\{\left[\mathrm{probit}\left(f - fe^{-rT}\right)/(1+j+h)\right]\sigma\sqrt{T} - s^2 T/2\} \end{aligned} \tag{12}$$

where, j and h are the adjust factors of debt guarantee and transaction cost respectively, and $j = 1$ for debt guarantee and $j = 0$ for no debt guarantee.

Similar to other listing companies, Haier also prefers equity financing, which implies an additional capital cost of debt financing. This additional capital cost is more likely in a form of implicit cost and hard to estimate.

Other thing being equal, i.e., risk free rate $r = 2.51\%$; debt maturity $t = 1.8$ years; volatility $\sigma = 24.99\%$; but based on the adjustment 1, the corporate tax rate now $f = 15.63\%$; and further assume $j = 0$, $h = 40\%$, based on Eq. (12),

$$\begin{aligned} L &= e^{\mathrm{probit}\left[15.63\%\left(1-e^{-2.51\%\times 1.8}\right)/(1+0+0.4)\right]24.99\%\sqrt{1.8}-24.99\%^2\times 1.8/2} \\ &= 39.80\% \end{aligned}$$

Compared with the previous optimal leverage of 41.41%, it is obvious that the guarantee requirements and transaction cost reduce the company's optimal debt ratio, that is, limit the company's use of debt.

The three factors considered so far in this section, the personal income tax difference, the guarantee requirements and transaction cost, affect the optimal leverage through the corporate tax rate. Note that, $15.63\%/(1 + 0 + 0.4) = 11.16\%$. That is, according to our assumptions, the total effect of the three factors considered so far is equivalent to the corporate tax rate decreases from 25 to 11.16%.

Adjustment 3: Book value standard

It is the convention in capital structure research to calculate and express the optimal capital structure on fair value basis, but it is not convenient for the company to compare the optimal result with the actual capital structure in book value, and not convenient to use in financing and capital structure decision-making either.

Fortunately, it is not difficult to transfer the optimal leverage in fair value, 39.80%, into the optimal debt ratio in book value. As, derived in last section, the fair price to book ratio of Haier is 2.4461. Based on Eq. (3), Haier's optimal leverage in book value is:

$$\mathrm{L}' = 2.4461/(1/(39.80\%) + 2.4461 - 1) = 61.79\%$$

Adjustment 4: External guarantee

Financially sound companies like Haier do not need guarantee for debt financing, rather, they may provide guarantee to other companies for debt financing. Providing external debt guarantee is equivalent to outputting their own debt financing ability, which should reduce the companies own debt financing accordingly. Otherwise, once the guaranteed company has problems, it is easy to be involved and fall into financial crisis.

The guaranteed party is relatively unstable and it is possible that the guarantor to repaid part of the debt for the guaranteed party finally. It is reasonable to assume that the guarantor and the guaranteed party share the responsibility of the guaranteed debt based on the cost or value of the guarantee. Therefore, based on the optimal debt ratio and debt size without external guarantee, we can deduct the value of the external guarantee from this debt size to obtain the company's own optimal debt size, and further derive the optimal debt ratio based on the optimal debt size after the deduction.

In the 2020 annual report, Haier disclosed the external guarantee. The total amount of external guarantees, i.e. the balance of guarantees to subsidiaries at the end of the reporting period, decreased slightly, totaling 29,431.65 million yuan, accounting for about 44.0% of the company's net assets. At the same time, the total book assets disclosed by the company were 203,459.4959 million yuan. According to the optimal debt ratio in book value of 61.79% without external guarantee, the optimal debt size is:

$$203459.4959 \times 61.79\% = 125719.7938 \text{ million yuan}$$

According to the numerical examples in 6.2 Valuation of debt guarantee, the value of a common debt guarantee is about 10% of the amount of the guaranteed debt. For the prudence principle in decision making, we would like to deduct 15% of the guaranteed debt from the original optimal debt size to derive the new optimal debt size of Haier. The optimal debt size is adjusted as follows:

$$125719.7938 - 29431.65 \times 15\% = 121305.0463 \text{ million yuan}$$

Accordingly, the optimal debt ratio in book value is adjusted as follows:

$$121305.0463/203459.4959 = 59.62\%$$

Adjustment 5: External Investment

In reality, the company may invest in other companies with growth potential, so as to form active external investment. Even if the company does not make any external investment except for its main business, it will inevitably have accounts receivable due to the credit sales of products or services, that is, it will passively become an investor in the debt of other companies. These intentional or unintentional external equity or debt investments are repeatedly listed as assets in two or more companies, which may be listed in the asset side of the balance sheet of one company and in the liability and equity side of the other company.

Therefore, in a national or local economy, if there is no double calculation, the real total asset value of all companies will be less than the sum of their respective asset values. Of course, the company's optimal capital structure or optimal debt ratio is based on the premise that all the company's assets are used to ensure the repayment of the company's own debts. If some of the company's assets need also to ensure the repayment of the debt of the invested company, the optimal debt ratio should be lower than that under the same conditions.

Table 5 shows the notes receivable, accounts receivable and long-term equity investment of Haier in recent two years, which is equivalent to a rough statistics of its external investment. This situation is similar to that of companies all over the world, which is typical, that is, external investment accounts for 20–30% of the company's total assets.

External investment as an asset is responsible to the debt repayment of two or more companies. Apparently, these repeatedly listed assets need to be deducted by a proportion. The proportion can be determined based on the possibility of the investments becoming bad debt. For simplicity, similarly to the external guarantee, assume the investment of Haier on the assets of other companies should be reduced by 15%. Haier's average ratio in the past two years was 25.56%, 15% of which was .83%.

For the above investment proportion, the optimal debt size is adjusted as,

$$121305.0463 \times (1 - 3.83\%) = 116654.2108 \text{ million yuan}$$

Accordingly, the optimal debt ratio in book value is adjusted as,

Table 5 External investment of Haier in recent two years (in million yuan)

Year	Receivable	Equity	Total	Proportion
2019	24,967.29	20,460.76	45,428.05	24.23%
2020	32,108.01	22,587.01	54,695.02	26.88%

Receivable = Notes and accounts receivable
Equity = Long term equity investment
Total = Total external investment
Proportion = Proportion of external investment in total asset

$$116654.2108/203459.4959 = 57.34\%$$

Adjustment 6: Abnormal growth

Haier has achieved remarkable high-speed growth over the long run since its listing. This unusual growth means that the optimal debt size or debt ratio determined according to the current asset value may be lower than the optimal level at the debt maturity. Therefore, the current debt size can appropriately exceed the optimal level; apparently, it is better to achieve the optimal debt ratio in the middle of the debt maturity when the company grows as expected.

Value is the fundamental goal of the company as well as its shareholders. From the perspective of shareholders, the so-called growth should refer to the value growth driven by earnings. Reasonably speaking, the effects of capital changes, such as share expansion or equity reduction, spin off and stock distribution (including cash and stock dividends) should be excluded. However, there is no more effective way to eliminate these effects, whether it is the growth of total earnings or earnings per share.

A shortcut worth using or perhaps the best choice is to represent the growth rate of earnings and net asset value by return on net assets. Because the earnings from net assets does represent the growth of earnings and net asset value, including the growth of old and new assets, the impact of changes in assets or capital scale itself is largely excluded. At the same time, the earnings in the return on net assets includes the issued stock dividends and the company's retained earnings, which is more comprehensive.

As a blue chip in China's A-share market, Haier has maintained rapid growth for a long time in the past. Table 6 shows the annual return on equity (ROE) of Haier in the 28 years since its listing. According to the above analysis, the yield data in Table 6 better reflects the growth of Haier since its listing.

Taking the return on net assets in Table 6 as the growth rate for geometric average calculation, it can be concluded that the average annual compound interest growth rate of Haier since its listing is 16.18%. In other words, if someone buys Haier shares at the initial stage of Haier shares listing and holds them all the time, he or she can theoretically obtain an average annual return of 16.18% through dividends and stock appreciation. This return exceeds the average return of stock market investment

Table 6 Return on net assets (ROE, %) of Haier since its listing

Year	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002
ROE	12.20	15.60	19.09	19.62	15.34	14.74	12.03	14.67	12.53	7.80
Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
ROE	6.85	6.46	4.07	5.43	10.20	11.34	14.89	28.98	32.36	29.38
Year	2013	2014	2015	2016	2017	2018	2019	2020	2021	
ROE	28.80	22.86	18.95	19.10	21.50	18.88	17.14	17.67	17.29	

Source Excerpted from Sina Financial Data and the average growth rate of the national economy in the same period. There is undoubtedly abnormal return and abnormal growth.

It can also be judged that Haier has the potential for abnormal growth in the future, at least in the average debt cycle of less than two years. It may be easy to judge whether there is abnormal growth, but it is not easy to measure it accurately. For simplicity, the 10% average annual abnormal growth rate within the future debt maturity will be used to discuss the adjustment accordingly on the optimal capital structure of Haier.

The abnormal growth of 10% earnings means that the equity value has an abnormal growth of 10%. Since the debt does not increase abnormally in value, to match the growth with equity, the debt should increase by 10% in scale or in size, so as to keep the debt ratio unchanged. In order to make the debt ratio close to the optimal level as much as possible during the duration of the debt, it can be arranged this way, the first half of the debt cycle is higher than the optimal level, and the second half is lower than the optimal level, which is the closest to the optimal level on the whole process.

Based on the debt cycle 1.8 years, the corresponding intermediate time of debt maturity is 0.9 years. An annual growth of 10% means an increase of 9.0% in 0.9 years. Based on the previous optimal debt size, and then adjusted according to this growth rate, we can get a new optimal debt size in 0.9 years.

The optimal debt size is adjusted as,

$$116654.2108 \times (1 + 9.0\%) = 127153.0898 \text{ million yuan}$$

Accordingly, the optimal debt ratio in book value is adjusted as,

$$127153.0898/203459.4959 = 62.50\%$$

The abnormal growth increases the level of the optimal debt ratio. In reality, the negative abnormal growth may be also possible in some companies. The negative growth will certainly decrease the optimal debt ratio level of the company.

There may be other factors worth to consider for further adjustments, but they may be relative insignificant, so we would like to stop here. This means the final optimal leverage of Haier is 62.50% in book value. The corresponding optimal leverage in fair value based on Eq. (2) is: $62.50\%/(1 + 2.4461 \times (1 - 62.50\%)) = 32.60\%$.

This is quite low comparing with what it was supposed to be in most people's mind, since Haier is a typical blue-chip share in the market, and its abnormal growth and profitability are quite impressing. Based on such a standard, the optimal leverage in fair value of many other companies should be even lower than this 32.60%.

3.4 The Trade-Off Value Analysis

Through the application of the above theoretical model and adjustment according to important scenarios, it is concluded that the optimal debt ratio of Haier is 62.50% in book value. Compared with the actual situation of Haier's capital structure in recent 20 years in Table 1, about half of the years (most of the later years) exceeds the optimal level and another half of the years (most of the former years) are lower than the optimal level.

In the past 20 years, the simple average of leverage of Haier was 48.77%, lower than the optimal level. However, this is more the reflects of the low leverage in early years of Haier; the upward trend is obvious in recent years, the average leverage is 66.49% over the past 10 years and 67.86% over the past five years, both are higher than the optimal level. It implies that Haier's finance has changed from cautious and conservative to radical. So, we should evaluate Haier's capital structure based on the situation in recent years. What impact does recent years' high leverage have on Haier?

According to the research of capital structure, when the company's debt ratio is lower than the optimal standard, it will lose value because it can not make full use of tax shield; When the company's debt ratio is higher than the optimal standard, it will also lose value because the tax shield is lower than the bankruptcy cost. It is interesting and a effective way to reveal the impact of Haier's capital structure deviation on the company by calculating the trade-off value (tax shield—bankruptcy cost) of Haier under actual leverage and comparing it with the trade-off value under the optimal debt ratio.

The following is the calculation and analysis based on Haier's situation in 2020.

In 2020, Haier's debt, equity and total capital in book value were 135,348.4887 million yuan, 68,111.0072 million yuan and 203,459.4959 million yuan respectively, and the actual debt ratio was 66.52%. According to the previous analysis, the optimal debt ratio in book value is 62.50%. This means that when the total assets remain unchanged, the scale of debt deviates (is higher than) the optimal standard by 6.43% (= 66.52%/62.50% − 1). Note that 62.50% actually corresponds to the debt ratio of 43.88% at fair value before the adjustments, and an increase of 28.23% on this basis means that the debt ratio in fair value before the adjustments reaches 6.70% (= 43.88% (1 + 6.43%).

The question is, what changes will be made to the trade-off value when the debt ratio increases from 43.88% to 46.70% in fair value. Tax shield, bankruptcy cost and trade-off value can be calculated according to the following model respectively.[3]

$$\text{ZZ tax shield} = \mathrm{Xf}\left(1 - \mathrm{e}^{-\mathrm{rT}}\right) \tag{6}$$

$$\text{ZZ bankruptcy cost} = \mathrm{XN}(-\mathrm{d2}) - \mathrm{SN}(-\mathrm{d1}) \tag{7}$$

where, S and X represent the fair value of the company and its debts respectively; d1 and d2 are calculated according to the following formula respectively.

[3] For more detail, see their derivations in Chapter "Tax Shield, Bankruptcy Cost and Optimal Capital Structure".

$$\mathrm{d1} = \frac{\ln(S/X)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \tag{8}$$

$$\mathrm{d2} = \frac{\ln(S/X)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = \mathrm{d1} - \sigma\sqrt{T} \tag{9}$$

$$\text{ZZ trade off value} = \mathrm{Xf}\left(1 - \mathrm{e}^{-\mathrm{rT}}\right) - [\mathrm{XN}(-\mathrm{d2}) - \mathrm{SN}(-\mathrm{d1})] \tag{10}$$

If the price to book ratio in fair value is still 2.4461, the debt at fair value and the total capital of the company are respectively:

$$\begin{aligned} \mathrm{X} &= 135348.4887 \text{ million yuan;} \\ \mathrm{S} &= 135348.4887 + 68111.0072 \times 2.4461 \\ &= 301954.8233 \text{ million yuan.} \end{aligned}$$

Note that the 301,954.8233 million yuan is the total capital of the company without the benifit from the debt fnancing or the trade-off value. Given the total capital of the company. Under the optimal debt ratio of 43.88%, the debt size X = 301,954.8233× 43.88% = 132,497.7765 million yuan; Under the actual debt ratio of 46.70%, the debt size X = 301,954.8233×46.70% = 141012.9025 million yuan.

Then the tax shield, bankruptcy cost, trad-off value and the value of the comapny under optimal leverage and actual leverage can be worked out based on Eqs. (6)–(10). Obviously, those calculations are too difficult or even impossible to carry out within the mainstream financial theories or tools. We will not illustrate the calculation process because of tedious details and only list the results as shown in Table 7.

As shown in Table 7, if Haier achieves the optimal capital structure, i.e. the debt ratio of 43.88%, which can creates tax shield of 1463.2585 million yuan and bankruptcy cost of 151.7447 million yuan, resulting in a net value addition (trade-off value) of 1311.5138 million yuan; given the total assets of 301,954.8233 million yuan (excluding the trade-off value) as the company owed in 2020, the company's value with the trade-off value reaches the maximum, i.e. 303,266.3372 million yuan.

According to the actual debt ratio (the book value is 66.52% or the fair value is 46.70%), there are more tax shield, i.e. 1557.2965 million yuan, and also more bankruptcy costs, i.e. 272.5165 million yuan. The trade-off value is reduced to 1284.7800 million yuan; The company's value thus reached 303239.6033 million yuan. Obviously, the company slightly overuses debt capital. Compared with the optimal situation, it loses a value of 26.7338 million yuan (= 1311.5138 – 1284.78 or 303,266.3372 – 303239.6033).

The "Invalid" line in Table 7 shows that when the debt ratio reaches 59.31%[4] of the fair value of total asset (excluding the trade-off value), the tax shield from leverage will be completely offset by the corresponding bankruptcy cost, and the net benifit of leverage or the trade-off value decrease to zero. If this is regarded as the

[4] 59.306927% more precisely.

Table 7 The impact of Haier's capital structure on the company's value (in million yuan except d1, d2 and debt ratio)

	Leverage (%)	d1	d2	Debt	Shield	Cost	Net	Firm value
Optimal	43.88	2.6245	2.2892	132,497.7765	1463.2585	151.7447	1311.5138	303,266.3372
Actual	46.70	2.4387	2.1034	141,012.9025	1557.2965	272.5165	1284.7800	303,239.6033
Invalid	59.31	1.7259	1.3906	179,080.1267	1977.6974	1977.6974	0.0000	301,954.8233

Leverage = Debt ratio
Debt = Debt size
Shield = Tax shield
Cost = Bankruptcy cost
Net = Trade-off value
Optimal = Optimal leverage
Actual = Actual leverage
Invalid = Invalid leverage, used to represent the leverage level where the benefit of tax shield from debt financing is just offset by the corresponding bankruptcy cost

maximum limit of financial leverage, Haier has exceeded the optimal standard, but has not exceeded the maximum limit. Compared with this warning line, although the actual financial leverage is excessive, it still brings an additional value of 1284.78 million yuan to Haier. Note that this is the value added brought by pure financial activity to the company in addition to investment and operation. This is the numerical result that the optimal capital structure research has long dreamed of but no way to work out over decades.

Attention should be paid to the current debt level, though it seems far from the warning line. As revealed by the ZZ leverage model, the impact of financial leverage on corporate value is asymmetric. When the debt ratio is too low, the value of loss is limited; However, when it is too high, it will significantly damage the value of the company. If the debt ratio continues to increase after exceeding the optimal level, the tax shield will maintain a slow growth, but the bankruptcy cost will accelerate growth, so that the trade-off value will quickly decrease to zero or even become negative, and the company's value will be damaged rather than benefit due to the use of financial leverage.

In fact, the above analysis is based on the premise that the company can freely choose between equity and debt. However, in practice, corporate financing may be more or less limited within a certain time; In many cases, debt may be more convenient. In order to solve the urgent need for capital, it may be necessary to go a little far in debt financing. Of course, the over leverage should be reduced whenever a chance emerges, such as the next round of financing or debt repayment.

The above discussion shows that whether use financial leverage rationally will increase or decrease the value of the company by hundreds or thousands of millions. This is obviously related to the scale of Haier. From the perspective of relative value, the increase or decrease in value accounts for only a small proportion of the company or equity value. Taking the situation of Haier in 2020 as an example, at the optimal capital structure, the trade-off value is 1311.5138 million yuan, which is 0.79% compared with the market value of the company's equity (= 1311.5138/(68,111.0072 $\times$ 2.4461)).

This does not mean that the capital structure problem is not important. On one hand, it is worth to adjust to optimal leverage and obtain the expected value addition as big as hundreds or thousands of millions yuan in case of financing for big companies like Haier since it is not difficult to find the optimal capital structure based on the ZZ optimal leverage model. On the other hand, most comanies need to maintain vigilance against overleverage because once the company's debt ratio exceeds the optimal level, the bankruptcy cost will rise rapidly, resulting in the trade-off value rapidly falling to zero and negative.

Knowing the optimal leverge and keeping the leverage below the optimal level are important for risk management in most companies, since the profitbility and value of the company inevitably fluctuate in future, and the leverage then fluctuates as a result. For example, if Haier's debt ratio reaches 70, 80 and 90% because the profitbility deteriorates beyond expectation, the trade-off value is −3849.3488, −10,935.0420, −21,941.6081 million yuan respectively, that is, excessive debt will reduce the equity value of the company by 2.31%, 6.56%, 13.17% respectively. These risk or cost are obviously not neglectable.

Overall, Haier's debt financing decision is slightly radical, but basically correct. In the absence of theory and model that can be resorted to, it is difficult to achieve such an effect by empirical and intuitive decision-making.

So far, we can fully explain a puzzle in the study of capital structure, that is, whether there is an optimal capital structure. The answer is of course yes. According to the relevant situation of Haier, its optimal capital structure can be worked out. This optimal capital structure exists objectively; unable to calculate it doesn't mean no such a optimal leverage. Whether to set up a target capital structure in practice is actually different from this problem. For example, according to Table 1, Haier's debt ratio changes constantly, so it can be said that there is no target capital structure.

Graham et al. (2001) and Brounen et al. (2006) found that most companies will set up a target leverage, it may not be meaningful to set up a target leverage without a right calculation. On the one hand, the target leverage established may not be optimal. It is of course meaningless to artificially set a non optimal target and make adjustment in case of deviation. On the other hand, based on the above analysis, it can be seen that this optimal standard is affected by many factors, and the standard of optimal leverage changes with the change of these factors. The target leverage established without a quantitative method cannot be adjusted accordingly, and of course, it is impossible to represent the optimal level over time.

4 Summary

Through the calculation of the optimal capital structure of Haier and the comparative analysis with the actual capital structure of the company, this chapter verifies the theoretical rationality and application feasibility of the ZZ optimal leverage model, including the calculation and analysis of the optimal capital structure of the company, and obtains the decision-making suggestions for the future.

The model can be adjusted and analyzed according to the specific conditions in reality, including the consideration of personal income tax, transaction cost, the book value, the abnormal growth, the external investment between companies, and the provision and need of debt guarantees. It shows that the model not only has application feasibility, but also has flexibility and adaptability, and can cope with complex and changeable specific situations.

The model can also help to find the value added potential in hundreds or thousands of millions in big companies like Haier by comparing the actual leverage with the optimal one. This means that the ZZ optimal leverage model series not only contribute to the risk management for various companies, but also have an immediate value-added effect when applied to large-scale companies.

It is worth noting that although our analyses are based on the case of Haier, the models and methods used are obviously also applicable to other companies. Relevant application scenarios and required macro and micro data are accessible in the usual financial report of other companies. This means that the application analyses of this chapter can be transplanted to most other companies. Of course, due to the limited time, the treatment of some problems has not been carried out in a more in-depth and detailed manner, such as the treatment of external guarantee and investment.

Over the past decades, the research on capital structure has mostly gained theories to explain phenomena. As an original theory to solve the problem of optimal capital structure, the ZZ optimal leverage model must have broad application prospects. The application research of this chapter preliminarily confirms the theoretical and application value of the model. This will promote the understanding of the model in the theoretical and practical circles, and enhance the confidence of relevant applied research and practical application. I believe that this is just a brick to attract jade, and there will be more applied research from various angles and levels in the follow-up.

For example, when commercial banks consider loan approval, they often set a fixed loan limit of 70%. With the help of the ZZ optimal leverage model, it is not necessary to adopt such fixed standard, 70% or so. Some companies have low risk (low volatility), a debt ratio of 80% may be acceptable; some companies have high risks and may have various adverse factors such as external guarantee and investment as well, a debt ratio of 60% may be the line to stop loans. In this way, banks can avoid risks and win customers to the greatest extent. Similarly, other financial institutions, such as guarantee and insurance, can also use similar methods to improve the scientificity and rationality of their client screening or customer selection and business decisions.

In short, with the help of the ZZ optimal leverage model, the company can timely analyze and adjust the capital structure and fully tap the value-added potential; It can also customize various optimal leverage standards for companies in various financial and non-financial industries, so as to facilitate the customer screening, risk management and relevant decision-making of industrial and commercial companies and financial institutions, as well as the application in industry management institutions.

Further Readings

1. Zhang Z, Xiao S. Tax shield, bankruptcy cost and optimal capital structure. Acc Res. 2009;(4):47–54+97
2. Solomon E. Leverage and the cost of capital. J Finance. 1963;18:273–9.
3. Robichek AA, Myers SC. Problems in the theory of optimal capital structure. J Finan Quant Anal. 1966;1:1–35.
4. Taub AJ. Determinants of the firm's capital structure. Rev Econ Stat. 1975;57:410–6.
5. Jensen MC, Meckling WH. Theory of the firm: managerial behavior, agency costs and ownership structure. J Finan Econ. 1976;3(4):305–60.
6. Ross SA. The determination of financial structure: the incentive-signaling approach. Bell J Econ. 1977;8:23–40.
7. Leland HE, Pyle DH. Informational asymmetries, financial structure, and financial intermediation. J Finance. 1977;32:371–87.
8. Brennan MJ, Schwartz ES. Corporate earnings taxes, valuation and the problem of optimal capital structure. J Bus. 1978;51:103–14.

9. Smith CW, Warner JB. On Financial contracting: an analysis of bond covenants. J Financ Econ. 1979;7:117–61.
10. Barnea A, Haugen RA, Senbet LW. A rationale for debt maturity structure and call provisions in the agency theoretic framework. J Finance. 1980;35:1223–33.
11. Heinkel R. A theory of capital structure relevance under imperfect information. J Finance. 1982;37:1141–50.
12. Marsh P. The choice between equity and debt: an empirical study. J Financ. 1982;37(1):121–44.
13. Myers SC, Majluf NS. Corporate financing and investment decisions when firms have information that investors do not have. J Finan Econ. 1984;13(2):187–221.
14. Kane A, Marcus AJ, McDonald RL. How big is the tax advantage to debt? J Finance. 1984;39:841–53.
15. Diamond D. Financial intermediation and delegated monitoring. Rev Econ Stud. 1984;51:393–414.
16. Masulis RW. Discussion: how big is the tax advantage to debt? J Financ. 1984;39(3):853–5.
17. Brennan MJ, Schwartz ES. Optimal financial policy and firm valuation. J Finance. 1984;39(3):593–607.
18. Brander JA, Lewis TR. Oligopoly and financial structure: the limited liability effect. Am Econ Rev. 1986;76(5):956–70.

Bankruptcy Probability and Firm Life Expectancy

Contents

Bankruptcy cost is a key for the solutions of many financial problems. We have solved the problem of discount rate or asset pricing and optimal capital structure based on the ZZ bankruptcy cost model. Actually, the measurement of bankruptcy probability can also be solved based on the ZZ bankruptcy cost model, and further financial problems can be solved based on the bankruptcy probability, such as the determination of firm life expectancy, the risk analysis and management of most companies.

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_12

1 Bankruptcy Risk

Theoretically, bankruptcy can be classified into current bankruptcy and overall bankruptcy. Current bankruptcy means that the cash flow of the company is insufficient to repay the due debts; Overall bankruptcy refers to insolvency, that is, the value of the company is lower than the value of the debt, and the debt cannot be paid in full through the sale of assets.

1.1 The Concept of Bankruptcy Risk

According to the conventional understanding, company bankruptcy refers to the situation that the company cannot pay off its due debts. In the case of declaring bankruptcy, the company is unable to continue its business or has to terminate its business and carry out bankruptcy liquidation of the company's property; The liquidation proceeds will be used to compensate the creditors of the company. The risk of corporate bankruptcy usually refers to the possibility of this situation.

Bankruptcy is not a good thing. Whether bankruptcy or bankruptcy risk, people dislike such kind word and often try to avoid it or even do not mention it. Therefore, there are various alternative words, for example, bankruptcy is renamed as default, bankruptcy risk is renamed as default risk, credit risk, insolvency risk, financial distress, financial failure, etc. Those phrases roughly mean the same thing. However, covering up does not help to avoid or get rid of bankruptcy. It makes sense to study how to measure in advance and how to deal with it afterwards.

As revealed in previous chapters, corporate risk can be divided into two categories: operational risk and financial risk. As mentioned above, financial risk, credit risk, default risk and bankruptcy risk all refer to the risk that the company cannot repay the principal and interest of debt in full on time. There may be different expression habits or purposes on different occasions, but no significant difference in essence among these expressions. The bankruptcy risk in this chapter and this book is identical to the phrases like credit risk and default risk. It is worth noting that the bankruptcy risk is not the whole of the company's risk, but the risk that the fluctuation of the company's income or value touches the "bottom line", that is, there is a problem or difficulty in debt repayment.

One of the most influential operations in practice is credit rating, which is actually credit risk rating, that is, bankruptcy risk rating. The internationally recognized and authoritative credit rating agencies are Standard and Poor's, Moody's and Fitch. To some extent, there have been some biased understanding to credit rating in practice. It is necessary to clarify the relevant concepts here. First of all, credit rating evaluates the credit or bankruptcy risk of companies. Secondly, the "rating" actually implies that it does not solve the quantitation of bankruptcy risk or bankruptcy probability; Therefore, the bankruptcy risk difference among companies can only be divided into levels or grades. Of course, the probability of bankruptcy of companies at different levels can be calculated on the accumulated data. However, it is not the standard for this book to discuss and solve problems, because this means derive conclusion based on ex post data; and it is impossible for companies in a risk level to repeat exactly the past situation in the future.

As a convention or common knowledge, people used to measure the risk by the possibility or probability, such as measuring bankruptcy risk by the bankruptcy possibility or probability. For many or most companies, the probability of bankruptcy every year is very small. However, this does not mean that the bankruptcy risk does not exist, let alone that the research of bankruptcy risk is not important. On the contrary, the study of bankruptcy risk is very important, because no matter how unpopular bankruptcy or similar expressions are, the company will inevitably go bankrupt or disappear in the end; Before that, the company must always guard against and consider this problem. On the other hand, all parties concerned in the capital market (such as banks, analysts, etc.) will inevitably judge companies from this perspective, including companies in danger of bankruptcy, as well as normal or healthy companies.

1.2 The Current Bankruptcy and Overall Bankruptcy

The normal repayment of debts usually requires cash rather than other assets, such as raw materials, machinery and equipment and other physical assets. From the perspective of debt repayment, companies can be divided into three categories: The first category is those are able to repay due debts with their operating cash flows; the interests of the companies and creditors can be guaranteed in this category. The second category is those are difficult to repay the due debts with operating cash flows, but the debts can be repaid without affecting the normal operation of the company through postponement or other arrangements (such as disposal of over-stocked materials or idle equipment); The interests of the company and creditors are basically unaffected. The third category is those suffered from operating losses due to the deterioration of internal and external environment, and the company's value decreases accordingly. The due debts are unable to be repaid with operating cash flows or other arrangements under the condition of maintaining the company's operation. The due debts can only be repaid through the liquidation of the company's necessary assets.

The companies in first situation are healthy companies, or normal companies. The companies in third situation are bankruptcy companies in the traditional or strict sense. The second situation is also abnormal, or in bankruptcy. But unlike the third situation, this situation can be saved through adequate efforts; The third situation cannot be saved. Therefore, bankruptcy can be divided into two types: current bankruptcy and overall bankruptcy, which correspond to the second and third cases respectively. Overall bankruptcy is bankruptcy in the real sense, that is, insolvent in the value of assets, which will eventually lead to bankruptcy liquidation. In this case, the company will have to terminate its operation or transfer its ownership, and the shareholders are likely to lose all the capital invested in the company; The interests of creditors are also likely to suffer large losses and can only recover the lent capital according to a certain proportion. If default is defined as debt delayed in repayment, the second situation can also be called default, mainly due to cash flow problems. Of course, if the company sinks into current bankruptcy and there is no effective remedy, it may deteriorate into overall bankruptcy.

The classification of bankruptcy in reality is mainly based on the classification of relevant laws in practice. Such as bankruptcy under Chap. 7 and bankruptcy under Chap. 11; roughly corresponding to the above overall bankruptcy and current bankruptcy. Under Chap. 7, the company (debtor) must stop all operations and go out of business. A trustee is appointed to sell (liquidate) the company's assets and the proceeds is used to pay off the debt. As a result, the interests of creditors are often not preserved, and there is little property left to the shareholders of the company. Under Chap. 11, the company has chance to reorganize its debt and business, and try to recover its profitability; the day-to-day business operations go on, but important business decisions must be approved by a bankruptcy court. Current bankruptcy often helps the company get out of trouble through debt and asset restructuring, while protects the interests of creditors to the greatest extent. Therefore, in the case of inevitable bankruptcy, the parties concerned will often strive for current bankruptcy as much as possible to avoid overall bankruptcy.

1.3 The Annual and Cumulative Bankruptcy Probability

The following discussion focuses on the measurement of bankruptcy risk, or quantification of the bankruptcy probability and bankruptcy loss on current bankruptcy and overall bankruptcy bases respectively.

For the preparation in concepts and quantitative tools, we define two bankruptcy probabilities: the annual bankruptcy probability, p, and the cumulative bankruptcy probability, P. The annual bankruptcy probability is the probability of bankruptcy takes place within one year; the cumulative bankruptcy probability is the probability of bankruptcy takes place in several years. The number of years for the cumulative bankruptcy probability can be an integer or a decimal, and can be greater than or less than one year.

Consider a problem, given the annual bankruptcy probability, p, how to derive the cumulative bankruptcy probability, P?

In a period of n years, the probability for a company to go bankrupt is equal to one minus the probability that the company keeps away from bankruptcy. As the annual bankruptcy probability is p; the annual probability of no bankruptcy then is $1 - p$. the probability that the company keeps away from bankruptcy over a period of n years then is $(1 - p)^n$, the cumulative bankruptcy probability hence is,

$$P = 1 - (1 - p)^n \tag{1}$$

Table 1 The cumulative bankruptcy probabilities (based on $p = 1\%$)

Years	1	2	3	4	5	6	7	8	9	10
Prob (%)	1.00	1.99	2.97	3.94	4.90	5.85	6.79	7.73	8.65	9.56
Years	15	20	30	40	50	60	70	80	90	100
Prob (%)	13.99	18.21	26.03	33.10	39.50	45.28	50.52	55.25	59.53	63.40
Years	100	200	300	400	500	600	700	800	900	1000
Prob (%)	63.40	86.60	95.10	98.20	99.34	99.76	99.91	99.97	99.99	100.00

Prob = cumulative bankruptcy probabilities

That means,

$$p = 1 - (1 - P)^{1/n} \tag{2}$$

Given the annual bankruptcy probability, we can derive the cumulative bankruptcy probability over a certain time of period based on Eq. (1). Given the cumulative bankruptcy probability over a certain time of period, we can derive the average annual bankruptcy probability based on Eq. (2).

For instance, consider a numerical example. Suppose a research reveals that the annual bankruptcy probability of company E is 1%. Based on Eq. (1), the cumulative bankruptcy probabilities over several to several hundred years are shown in Table 1. Although the bankruptcy probability in one year is not significant, the cumulative bankruptcy probabilities over decades or hundreds of years are big enough, as shown in Table 1. That may be why companies live over 100 years are rare in reality.

Similarly, we can also calculate the average annual bankruptcy probability based on the cumulative bankruptcy probability over a certain time of period. For instance, a research reveals that the cumulative bankruptcy probability of company F over 20 years is 20%. Based on Eq. (2), the average annual bankruptcy probability is:

$$p = 1 - (1 - 20\%)^{1/20} = 1.11\%$$

2 Modelling of Bankruptcy Probability

We can calculate the annual or cumulative bankruptcy probability given any one of the other. The problem now is, how can we know any one of the other?

This is an unsolved problem in mainstream finance. With the foundations prepared in previous chapters, we now have some tools to solve this problem.

2.1 The Related Research in Main Stream

In mainstream, two types of research are related to corporate bankruptcy risk. The first is the solvency ratio analysis based on financial report, which mainly comes to the conclusion that the company's solvency is high or low, so the bankruptcy risk is small or large based on the relevant ratios; The second is the research on financial early warning, which mainly uses statistical analysis to draw the important factors affecting the company's bankruptcy and the conclusion of the company's bankruptcy risk.

The solvency ratio analysis focuses on the relevant financial ratios. Specifically, it includes current ratio and quick ratio reflecting short-term solvency situation, leverage ratio and interest coverage ratio reflecting long-term solvency situation.

The specific calculation formula is as follows:

$$\text{Current ratio} = \text{current assets}/\text{current debts} \tag{3}$$

$$\text{Quick ratio} = \text{quick assets}/\text{current debts} \tag{4}$$

$$\text{Leverage ratio} = \text{total debts}/\text{total assets} \tag{5}$$

$$\text{Interest coverage ratio} = \text{EBIT}/\text{interest} \tag{6}$$

Most of the variables involved in the above formula are relatively simple and clear in concepts. Among them, quick assets refer to the net amount of current assets after deducting the component part with low liquidity. It is usually the current asset with inventory deducted. The quick ratio can be regarded as an improved version of the current ratio, but the improvement is not so ideal or successful. There is a theoretical standard in current ratio, which is 1, because for a regular company, debt need to be paid with cash; current ratio less than 1 means current asset less than current debt, which means the company does not have enough asset to transform into cash in one year for the payment of debt. However, there is not a clear standard in the quick ratio, because the part deducted from the current asset is not clear and certain.

It can be said that these ratios can usually reflect the solvency of the company to a certain extent. The stronger the solvency, the less likely it is to go bankrupt. However, the relationship between these ratios and bankruptcy risk is nebulous. Most ratios have no certain standard as a meaningful indicator of high or low risk except very rare ratios, such as the current ratio. More importantly, there is no clear quantitative relationship between the ratios and bankruptcy probability.

For example, if the interest coverage ratio of a company is three times, does this mean the bankruptcy risk is high or low, or how high is the bankruptcy risk? For another example, if a company's leverage ratio is 50% or 60%, does this mean the bankruptcy risk is high or low? Or even, if a company's current ratio is 1 or 1.2, does this mean the bankruptcy risk is high or low, and what is the probability of bankruptcy?

Further, two companies with same solvency ratio, such as the interest coverage ratio, the leverage ratio or the current ratio, may bear different bankruptcy risk or probability. This implies in addition to the factors concerned by those ratios, there are other factors or variables need to be incorporated. But what are those other factors? Obviously, the solvency ratio analysis can only show the ratios themselves, but cannot indicate the high or low of bankruptcy risk or the specific probability, and cannot reveal all the factor behind bankruptcy. In short, solvency ratio method seems too simplistic.

The research of financial early warning mainly uses statistical methods to analyze the company's financial data and try to predict whether the company will go bankrupt. It is a certain extension or improvement of the simple financial ratio analysis.

As early as 1932, Fitzpatrick tried to predict bankruptcy by using financial ratios and found that the most effective ratios are net profit/shareholders' equity and shareholders' equity/debts. From 1966 to 68, Beaver conducted a bankruptcy prediction study on 30 financial ratio, and found that three ratios were particularly effective, namely, debt protection ratio (cash flow/total debt), return on assets (net income/total assets) and leverage ratio (total debt/total assets).

A more influential study in early warning is the Z scoring model. In 1968, Professor Edward Altman of Stern Business School, New York University, compared bankrupt and normal companies and selected five most predictable ratios from the financial ratios by using multivariate statistical method, namely working capital/total assets (x1), retained earnings/total assets (x2), profit before interest and tax/total assets (x3), market value of shares/total book value of debts (x4), sales revenue/total assets (x5), and defined their weighted average value as Z value. The lower the Z value, the more likely the enterprise is to go bankrupt. According to Edward Altman's research, $Z \geq 2.68$ and $Z \leq 1.81$ represent the situation with little and great bankruptcy risk respectively. By 1977, Edward revised and improved the model and proposed Zeta model.

After the Z-score Model, the research on the early warning of bankruptcy risk put more efforts on the potential of the statistical model. One performance was the wide application of logistic regression model, which was commonly used in epidemiology and medicine field. Since the 1980s, especially since the twenty-first century, more and more newfangled methods were introduced to the study in early warning, such as the application of artificial neural network and machine learning (ML), etc.

Both traditional and new methods are statistically related in nature. The advantage of statistical methods is that conclusions can be obtained; The disadvantage is that no definite conclusion can be obtained. Because the model form, model variables and data processing methods are chosen subjectively, and because they are not constrained by professional logic or the internal mechanism of bankruptcy risk, the room for these subjective choices seems unlimited; there is almost no explicit limit on the choices of the model form, model variables and data processing methods. Therefore, it is inevitable that there are too many individual subjective and arbitrary factors in the statistical model, which are affected too much by individual subjective preferences.

Moreover, the "quality" of data in the field of social sciences is much lower than that in the field of natural sciences. In the field of Social Sciences, there has been no academic discussion and research on the contingency of drawing conclusions based on a set of past data. It should be understood that no matter how large the sample data is, it is only a drop in the ocean of the overall data base; More importantly, for decision-making needs, it should be based not on past data, but on future data. Cross professional reference may be worth advocating, but it should not be blind and hasty. In the absence of professional understanding of the problem, the application of methods from another professional field (such as statistics) will inevitably be subjective, arbitrary and even layman.

The application of statistical methods cannot get a certain model, because different data will get different models, including the variables and parameters in the model, all will be different due to different sample data; Even the same set of data will have different conclusions due to different processing (such as classification, etc.) methods and different model forms of subjective assumptions. This means that with the continuous production of new data, new models will continue to appear. Finally, the models obtained by such way will be endless and countless. In a broad sense, the purpose of scientific research is to answer questions. If there are countless answers to a question, and new answers continue to flow in, is this an answer or not an answer? Obviously, only the conclusion derived from strict logic reasoning may has uniqueness, and can be the true answer or solution to a scientific problem, especially a decisional problem.

2.2 Modelling of Bankruptcy Probability

According to the conventional understanding, the bankruptcy risk depends on the overall business risk of the company on the one hand and the financial risk on the other hand. The overall business or operation risk of the company is generally measured by the company volatility; the financial risk is generally measured by its debt or leverage ratio.

For instance, consider company A and B. If both companies have a 50% leverage ratio, but company A has higher business risk than B in terms of volatility. Then, company A has higher bankruptcy probability than B. If both companies have a business volatility of 30%, but the leverage ratio of company A is 40%, and B is 60%. Then, company B has higher bankruptcy probability than A. Questions as such seems easy to answer.

However, practical decision-making needs to know the answer to the problem like, if the business volatility of company A is 30% and B is 20%; while the leverage ratio of company A is 40%, and B is 60%. Then, which one has higher bankruptcy probability? Or what are their bankruptcy probabilities? It is much difficult to find the answer to such a question based on common sense; but the answer is easy to find as long as a model incorporating both the volatility and leverage is available.

Now, the problem is, where to find or how to setup such a model? Apparently, the mainstream research as introduced previously, such as financial ratios analysis or statistical analysis, cannot provide an ideal answer or model to such a problem. Fortunately, we can find such a model in the ZZ bankruptcy cost model, which is derived in Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing", that is,

$$\mathrm{BC} = \mathrm{XN}(-\mathrm{d}2) - \mathrm{SN}(-\mathrm{d}1) \tag{11}$$

where, BC is the bankruptcy cost; S is the current value of the company; X is the principal and the present value of the company's debt. Note that X/S = L, which is the debt or leverage ratio of the company, and,

$$\mathrm{d}1 = -\frac{\ln(L)}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2} \tag{12}$$

$$\mathrm{d}2 = -\frac{\ln(L)}{\sigma\sqrt{T}} - \frac{\sigma\sqrt{T}}{2} = \mathrm{d}1 - \sigma\sqrt{T} \tag{13}$$

As a concept, bankruptcy cost represents the expected loss resulted from the possible bankruptcy. The bankruptcy cost itself hence is also a measure concerned by relevant parties, such as the company and its creditors, investors, etc. Put it another way, the so-called bankruptcy risk needs to measure in the two dimensions, the occurrence probability and the potential loss. Consider company A and B, when they have same bankruptcy probability, they may have different bankruptcy loss. The bankruptcy loss thus is also a major concern. Other things being equal, the bigger the debt size, the bigger the bankruptcy loss or the bankruptcy cost. Therefore, bankruptcy cost and bankruptcy probability are usually the two measures drawing most attention in bankruptcy concerns.

Anyway, the quantification of bankruptcy cost has been solved in Chapter "Debt/Loan Risk, Bankruptcy Cost and Debt/Loan Pricing"; now we focus on the solution of bankruptcy probability.

Please note, in Eq. (11), N(−d2) is the probability of the (put) option being exercised at the expiration, or the value of the underlying asset being less than the exercise price; so conceptually, it is equivalent to the bankruptcy probability of the company at its debt expiration, or the probability that the value of the company is less than the principal of its debt, which represents the insolvency of the company.

Use capital P to represent this probability. Note that P represents the probability of overall bankruptcy occurring within T years, that is, the cumulative probability that the company will go bankrupt due to insolvency during the whole debt life. That is,

$$\mathrm{P} = \mathrm{N}(-\mathrm{d}2) = \mathrm{N}\left(\frac{\ln L}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}\right) \tag{7}$$

Equation (7) models the cumulative bankruptcy probability. Based on Eq. (2), the annual bankruptcy probability is,

$$\mathrm{p} = 1 - \left[1 - \mathrm{N}\left(\frac{\ln L}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}\right)\right]^{1/\mathrm{T}} \tag{8}$$

Now we solved the quantification of bankruptcy probability. Equation (7) and (8) were first published by Zhang [1, 2]. For the consistence with the ZZ bankruptcy cost model, Eqs. (7) and (8) can be referred to as ZZ bankruptcy probability model. Please note that Eqs. (7) and (8) are also consistent with the models in previous chapters in two dimensions. One is that those models are all closed form solutions; the other is that they are all logic-based rather data-based, the variables and form of the models are all derived from objective logic rather than subjective preference, so they are applicable for cases across times, places and conditions, and can represent the fundamental solutions to the relevant problems, i.e., the quantification of bankruptcy probability.

Previous discussion reveals that business and financial risk (company volatility and leverage) are the two determinants of bankruptcy probability. Now, the ZZ bankruptcy probability model takes these two variables into account based on strict logic reasoning rather than by subjective choice. In addition, the model factors in the time as well. This obviously makes sense and reflects the perfection of the ZZ bankruptcy probability model in logic.

2.3 *The Probability of Current and Overall Bankruptcy*

Previous discussion differentiates current bankruptcy and overall bankruptcy. Since the quantification of the bankruptcy probability is solved by the ZZ bankruptcy probability model, a following question is, how to use the ZZ bankruptcy probability model to measure the current bankruptcy risk and overall bankruptcy risk?

Firstly, Eqs. (7) and (8) can be used to measure the overall bankruptcy risk directly. For instance, the volatility of company A is 30% and B is 20%; while the leverage ratio of company A is 40%, and B is 60%, i.e., $\sigma_A = 30\%$; $\sigma_B = 20\%$; $L_A = 40\%$; $L_B = 60\%$. Then, for $T = 1, 2, 3, \ldots$, based on Eqs. (7) and (8), the cumulative and annual bankruptcy probabilities of company A and B are shown in Table 2.

We cannot compare the (value) bankruptcy probabilities between A and B before the derivation of the ZZ bankruptcy probability model; but now, it is quite easy to make any comparisons between A and B. For instance, based on Table 2, with the same average debt maturity, company A is always lower in bankruptcy probabilities, whether it is on cumulative basis or on annual basis. But, if the debt maturity of A is 4 years and B is 3 years, then, the cumulative overall bankruptcy probabilities of A is 10.99% and B is 9.66%, A is higher than B, though A is still lower than B on annual basis.

Table 2 The overall bankruptcy probabilities of company A and B

Year		1	2	3	4	5	6	7	8
A	d2	2.9043	1.9476	1.5036	1.2272	1.0305	0.8795	0.7576	0.6556
	N(−d2) (%)	0.18	2.57	6.63	10.99	15.14	18.96	22.44	25.60
	Annual (%)	0.18	1.29	2.26	2.87	3.23	3.44	3.56	3.63
B	d2	2.4541	1.6646	1.3014	1.0771	0.9186	0.7978	0.7008	0.6202
	N(−d2) (%)	0.71	4.80	9.66	14.07	17.91	21.25	24.17	26.76
	Annual (%)	0.71	2.43	3.33	3.72	3.87	3.90	3.88	3.82

A = company A with volatility 30% and leverage 40%; B = company B with volatility 20% and leverage 60%; Year = average debt maturity in years; N(−d2) = cumulative overall bankruptcy probabilities; Annual = annual overall bankruptcy probabilities

This implies that for overall bankruptcy or traditional bankruptcy, the ZZ bankruptcy probability model and the ZZ bankruptcy cost model are sufficient for analyses.

As revealed previously, current bankruptcy is the case that cash flow is not sufficient to pay the current due debt; overall bankruptcy is the case that money is not sufficient to pay all its debts even the company liquidates all its assets. The total debt payment depends on the total assets of the company; the current debt payment (in one year) depends on the current assets. Therefore, corresponding to the company's total leverage ratio, it makes sense to define a concept of current leverage ratio, which can be defined as the ratio of the company's current debts to current assets. It can be further understood that the current leverage ratio (current debt/current asset) is the reciprocal of the current ratio (current asset/current debt).

Current assets are assets owned by the company that are cashable within one year; Current debts are the debts that the company needs to repay within one year. Cash is needed to repay the due debt, especially the cash obtained from operation; It is abnormal to sell off non current assets (such as plant, machinery and equipment) for debt repayment. Therefore, it can be considered that in order to ensure that there is no current bankruptcy, the company should have sufficient current assets as a guarantee.

Traditionally, the minimum limit of current ratio is 1, which is equivalent to the current leverage ratio equal to 100%. It is understandable that unless the company has high value-added and best-selling inventory or convenient external financing support, if the current ratio is really equal to 1, debt repayment default or current bankruptcy is likely to occur. Of course, within mainstream finance, there has been no credible method or model to calculate the current bankruptcy probability, or the probability of default.

Now, with the ZZ bankruptcy probability model, it is easy to calculate the bankruptcy probabilities corresponding to the various common current ratios. For instance, consider again the company A and B, the volatility of company A is 30%

Table 3 The current bankruptcy probabilities of company A and B

Current ratio		0.5	0.8	0.9	1	1.1	1.2	1.5	2
CLR (%)		200	125	111	100	91	83	67	50
A	d2	−4.6960	−1.5626	−0.7774	−0.0750	0.5604	1.1405	2.6281	4.5460
	N(−d2) (%)	100.00	94.09	78.15	52.99	28.76	12.70	0.43	0.00
B	d2	−6.9815	−2.2814	−1.1036	−0.0500	0.9031	1.7732	4.0047	6.8815
	N(−d2) (%)	100.00	98.87	86.51	51.99	18.32	3.81	0.00	0.00

A = company A with volatility of current asset 15%; B = company B with volatility of current asset 10%; CLR = current leverage ratio; N(−d2) = current bankruptcy probabilities based on the maturity = 1 year

and B is 20%, suppose the volatility of the current asset is the half as that of the total asset,[1] i.e., 15% and 10% respectively for A and B. Input the current ratio or current leverage ratio into the ZZ bankruptcy probability model, their current bankruptcy probabilities are shown in Table 3.

Based on Table 3, or comparing between Tables 2 and 3, current bankruptcy probabilities are much higher than overall bankruptcy probabilities, which implies the risk of current bankruptcy are more likely to occur than overall bankruptcy. This is obviously in line with our intuition, which also makes sense. Specifically, based on the ZZ bankruptcy probability model, along with the current ratio decrease from 2 to 0.5, the current bankruptcy probabilities increase from (almost) zero to (almost) 100%. Put it another way, the current bankruptcy probabilities are very sensitive to current ratio.

Obviously, the traditional financial statement analyses are too simple and cannot reveal the basic features of the relevant financial ratios. For instance, based on Table 3, a further interesting finding is that, different from the situation in overall bankruptcy, company A is not always riskier than B in current bankruptcy, although A has a higher volatility. Rather, when the current ratio is relatively lower, such as less than 1, the current bankruptcy probabilities of A is lower than B; when the current ratio is larger than 1, the current bankruptcy probabilities of A is higher than B. This needs more and deeper discussion.

Just like our previous models, as they solve the relevant fundamental problems; there are plenty specific findings based on them as well as new questions. Put it another way, the fundamental solutions and the related questions are both the contributions of this book. As a matter of fact, new questions are often the first step to find the final solution; insightful questions are often better than stupid answers. As to the above question why the current bankruptcy probabilities of B are larger than

[1] The total asset can provide additional guarantee for the repayment of the current debt. This implies some of the repayment pressure or risk is partaken by the total asset, or the current asset is relative less risky than the total asset. But the specific measurement of the division of the risk is a little complicated. We just deal it simply by cutting the volatility of the total asset by 50% as the volatility of the current asset.

A when the current ratio is bigger than 1, one of the reasons may be that the current bankruptcy is more surely happen when the current ratio is bigger than 1, but company A is not as sure as B in current bankruptcy, just because A is more volatility than B.

ZZ bankruptcy probability and bankruptcy cost were originally the quantitative methods for analyzing bankruptcy risk. According to the above analyses, current debts depend on the sale of normal current assets, just as total debts depend on total assets. Therefore, the ZZ bankruptcy probability and bankruptcy cost model can be used to evaluate the risk of current bankruptcy and overall bankruptcy respectively.

3 Bankruptcy Risk Analysis—Case Illustration

We take Midea, Gree and Haier, China's three major household appliance giants, as examples to make a specific application and discussion of the ZZ bankruptcy probability and bankruptcy cost model in this section.

The selection of the three major household appliance giants for discussion is mainly due to the following considerations. First, these three companies are well known in China and other places over the world. The introductions in more detail thus can be omitted. Second, these three companies are blue chips or excellent companies in China's stock market, which implies that the authenticity of financial data is more guaranteed. Third, these three companies basically have no much bankruptcy risk or bankruptcy problem; Therefore, we can focus on the application of models and methods, rather than really discussion on the bankruptcy risk of the three companies, or serve as a guidance or reminder for the actual equity or debt investment or transaction. Therefore, in order to highlight the key points in model application, some issues may be simply treated or simply assumed. What we focus on here is the general model and method to analyze bankruptcy risk, including bankruptcy probability and bankruptcy cost, which hopefully can be applied to any company in the world.

3.1 The Volatilities of the Case Companies

Please note that, all the solutions, from the problems concerning capital asset pricing (part II) to that concerning the risk management (part III), are based on a same risk indicator, which is the volatility of the company (total asset or total capital).

In another word, many financial problems can be solved by putting the same company volatility into different ZZ models. Hence, the company volatility is a major work load in solving the relevant financial problems; of course, the relevant ZZ model is a must as the equipment to solve the problem. The situation is the same for using the relevant ZZ model to evaluate the bankruptcy risk of a company.

Specifically, we need to know or estimate the company value and its volatility, the value of the company's debt, the company's current assets and current debts. Under normal circumstances, except for the company's value and its volatility, these

Table 4 Assets and debts of the three companies as of December 31, 2021 (in million yuan except current ratio and leverage ratio)

Company	Cdebt	Tdebt	Casset	Tasset	CR	LR (%)
Midea	222,851.48	253,121.03	248,864.51	387,946.10	1.12	65.25
Gree	197,101.39	211,672.73	225,849.65	319,598.18	1.15	66.23
Haier	124,796.95	136,376.53	123,607.78	217,459.49	0.99	62.71

Cdebt = current debt; Tdebt = total debt; Casset = current asset; Tasset = total asset; CR = current ratio; LR = leverage ratio on book value basis

variables are equal to or close to its book value, which can be obtained directly from the company's financial report. The related financial data of Midea, Gree and Haier as of the end (December 31) of 2021 are shown in Table 4.

As most previous ZZ models, asset values in the models should be fair values rather than book values, the ZZ bankruptcy probability and bankruptcy cost model follow the same rule. As a convention in financial research, debt principal or book value is often assumed as its fair value, including current debt and noncurrent debt. We further assume that the value of the current assets of the company equals to its book value. Then, the current ratio based on book value in Table 4 can also be regarded as current ratio on fair value; but the leverage ratio in the last column of the table is just the leverage ratio in book value, because the fair value of the company does not equal to its book value.

So, first of all, we need to work out the fair value of the company's equity, which is different from its book value. Literally, the fair value of the equity is equal to its book value multiplied by a fair price to book ratio. We can derive the fair or theoretical P/B ratios for the three companies by using the ZZ P/B model derived in Chapter "Stock and Equity Valuation: Where Discounting Does Not Work" based on the expected growth rate and required payback period (risk). We have done that in Chapter "Some Extensive Discussions of ZZ Leverage Model" in the case study of Haier. For simplification, we would not repeat the process here, just assume simply that the fair P/B ratios of the three companies is 2.0 times.[2]

Then, we can work out the debt and equity ratios of the three companies on fair value basis, which is shown in Table 5.

As what we did in Chapter "Some Extensive Discussions of ZZ Leverage Model", we need to estimate now the volatility of equity and debt respectively and then derive the volatility of the company.

Let us estimate the equity volatility first. More stock price data may enhance the precision of the estimation, but in order to save space, we would estimate the equity volatility based only on the daily stock price data in one month; i.e., the closing prices of the three companies for each trading day in March 2022, as shown in Table 6.

[2] One of the possible situations is: the growth rate is 8%, required payback period is 10 years, and the expected return on equity is 12%, then, based on Eq. (29) in Chapter "Stock and Equity Valuation: Where Discounting Does Not Work", ZZ P/B $= [(1+g)^n - 1](1+g)r_e/g = 2.03$.

Table 5 Debt and equity ratios of the three companies on fair value basis

Company	Tdebt	Tequity	Tasset	LR (%)	ER (%)
Midea	253,121.03	269,650.14	522,771.17	48.42	51.58
Gree	211,672.73	215,850.90	427,523.63	49.51	50.49
Haier	136,376.53	162,165.92	298,542.45	45.68	54.32

Tdebt = total debt in fair value (in million yuan); Tequity = total equity in fair value (in million yuan); Tasset = total debt in fair value (in million yuan); LR = leverage ratio on fair value basis; ER = equity ratio on fair value basis

Table 6 Daily closing prices in March 2022 (in yuan)

Date	22.3.1	22.3.2	22.3.3	22.3.4	22.3.7	22.3.8	22.3.9	22.3.10
Midea	68.33	66.92	66.67	65.6	62.51	59.76	59.52	58.87
Gree	36.83	36.58	36.1	35.83	34.89	33.93	33.34	33.5
Haier	25.71	25.2	24.98	24.45	23.08	22.4	22.08	22.36
Date	22.3.11	22.3.14	22.3.15	22.3.16	22.3.17	22.3.18	22.3.21	22.3.22
Midea	59.38	56.69	54.4	56.99	58.7	58.9	58.45	58.58
Gree	33.8	33.03	31.2	31.75	32.07	32.26	32.16	32.45
Haier	22.48	22.02	20.77	21.56	22.23	22.2	22.24	22.38
Date	22.3.23	22.3.24	22.3.25	22.3.28	22.3.29	22.3.30	22.3.31	
Midea	58	57.58	56.09	55.8	54.65	57	57	
Gree	32.36	32.07	31.52	31.64	31.28	32.07	32.3	
Haier	22.26	22.39	21.9	21.79	21.45	22.4	23.1	

Based on the trading data, the closing prices of the three companies are 66.72, 36.77 and 25.49 yuan respectively on February 28, 2022, so the daily yield of the three companies in March 2022 can be calculated, as shown in Table 7.

The volatility of daily returns of the three companies can be derived by calculating the standard deviation of these returns. The results are 2.59%, 1.78% and 2.62% respectively. Annualizing this volatility based on 250 trading days a year, that is, the daily volatilities are multiplied by the square root of 250. Then we get the equity volatility of the three companies are 41.01%, 28.13% and 41.39% respectively.

Because the process to estimate the debt volatility is the same as that of the equity and normally no adequate debt trading data available, we would simply assume the debt volatility is 10% of the equity volatility, and assume the correlation coefficient between the debt and equity is zero (for more accurite relationship between equity and debt volatility, see Sect. 1.3 in Chapter "Capital Asset Pricing: An Easy and Unified Solution" Debt, equity and company volatility).[3] Then, based on the debt

[3] For healthy companies like those three companies, this is a reasonable assumption.

Table 7 Daily rate of return in March 2022

Date	22.3.1	22.3.2	22.3.3	22.3.4	22.3.7	22.3.8	22.3.9	22.3.10
Midea (%)	2.3844	−2.0851	−0.3743	−1.6179	−4.8249	−4.4990	−0.4024	−1.0981
Gree (%)	0.1630	−0.6811	−1.3209	−0.7507	−2.6585	−2.7901	−1.7542	0.4788
Haier (%)	0.8594	−2.0036	−0.8768	−2.1445	−5.7664	−2.9905	−1.4389	1.2601
Date	22.3.11	22.3.14	22.3.15	22.3.16	22.3.17	22.3.18	22.3.21	22.3.22
Midea (%)	0.8626	−4.6360	−4.1234	4.6512	2.9564	0.3401	−0.7669	0.2222
Gree (%)	0.8915	−2.3045	−5.6998	1.7475	1.0028	0.5907	−0.3105	0.8977
Haier (%)	0.5352	−2.0675	−5.8441	3.7330	3.0603	−0.1350	0.1800	0.6275
Date	22.3.23	22.3.24	22.3.25	22.3.28	22.3.29	22.3.30	22.3.31	
Midea (%)	−0.9950	−0.7268	−2.6218	−0.5184	−2.0825	4.2102	0.0000	
Gree (%)	−0.2777	−0.9002	−1.7299	0.3800	−1.1443	2.4942	0.7146	
Haier (%)	−0.5376	0.5823	−2.2128	−0.5035	−1.5727	4.3336	3.0772	

Table 8 The estimation of the total volatility

Company	volatilityd (%)	weightd (%)	volatilitye (%)	weighte (%)	volatilityt (%)
Midea	4.10	48.42	41.01	51.58	21.25
Gree	2.81	49.51	28.13	50.49	14.27
Haier	4.14	45.68	41.39	54.32	22.56

volatilityd = debt volatility; weightd = debt weight; volatilitye = equity volatility; weighte = equity weight; volatilityt = total volatility

and equity ratios derived previously, we can estimate the overall volatilities for the three companies, as shown in Table 8.

Therefore, the total volatility of the three companies is 21.25%, 14.27% and 22.56% respectively. As assumed previously, the current asset guarantees the repayment of the current debt, while the total asset guarantees the repayment of total debt.

Literally, both asset and debt are involved in an endless turnover process, i.e., a process that the noncurrent asset turn into current asset, and the noncurrent debt turn into current debt. This implies that the total asset somehow is also responsible or takes part in the repayment of the current debt. However, for total asset, there is no further other asset takes some of its responsibility to repay the total debt. Put it another way, there is some difference between current asset and total asset in concern of the relevant debt repayment; but the difference is complicated.

For simplicity, we just assume current asset is relative less risky than the total asset, or specifically, the volatility of current asset is half of that of the total asset. Hence the volatility of current asset is 10.63%, 7.14%, 11.28% respectively for Midea, Gree and Haier.

Besides the volatility, for the bankruptcy risk analyses, we should also estimate the debt maturity, including current debt and total debt. There are no abnormal cases

Table 9 The estimation of the debt maturity

Company	Tdebt	Cdebt	Ldebt	Cweight (%)	Lweight (%)	Tmaturity
Midea	253,121.03	222,851.48	30,269.55	88.04	11.96	0.8995
Gree	211,672.73	197,101.39	14,571.34	93.12	6.88	0.8360
Haier	136,376.53	124,796.95	11,579.58	91.51	8.49	0.8561

Tdebt = total debt (in millions); Cdebt = current debt (in millions); Ldebt = noncurrent debt (in millions); Cweight = weight of current debt; Lweight = weight of noncurrent debt; Tmaturity = maturity of total debt

in the debts of the three companies. For simplicity, just assume the maturities of their current and noncurrent debt are 0.75 and 2 years respectively. Based on the actual proportion of their current and noncurrent debt, the weighted average maturities of their total debts are calculated as shown in Table 9.

As mentioned above, the bankruptcy risks of the three companies are evaluated respectively in terms of overall bankruptcy and current bankruptcy.

3.2 *The Overall Bankruptcy Risk*

Summarize the relevant data estimated previously for the overall bankruptcy risk, as shown in Table 10.

Based on the data in Table 10, take Midea as an example,

The volatility within total debt life,

$$\sigma\sqrt{\mathrm{T}} = 21.25\%\sqrt{0.8995} = 20.15392\%;$$

$$\mathrm{d1} = -\frac{\ln 48.42\%}{20.15\%} + \frac{20.15\%}{2} = 3.69936$$

$$\mathrm{d2} = -\frac{\ln 48.42\%}{20.15\%} - \frac{20.15\%}{2} = 3.49782$$

$$\mathrm{N}(-\mathrm{d1}) = 0.0108071\%$$

Table 10 The data estimated for overall bankruptcy risk analyses

Company	Tdebt	Tasset	LR (%)	Tvolatility (%)	Tmaturity
Midea	253,121.03	522,771.17	48.42	21.25	0.8995
Gree	211,672.73	427,523.63	49.51	14.72	0.8360
Haier	136,376.53	298,542.45	45.68	22.56	0.8561

Tdebt = total debt (in millions); Tasset = total asset (in millions); LR = leverage ratio; Tvolatility = total volatility; Tmaturity = total or weighted average debt maturity

Table 11 The overall bankruptcy risk analyses

Company	d1	d2	N(−d1) (%)	N(−d2) (%)	BC
Midea	3.6993611	3.4978219	0.0108071	0.0234537	2.869642
Gree	5.2905543	5.1559648	0.0000061	0.0000126	0.000638
Haier	3.8579257	3.6491878	0.0057177	0.0131535	0.868667

BC = bankruptcy cost (in million yuan)

$$N(-d2) = 0.0234537\%$$

Hence overall bankruptcy probability $P = 0.0234537\%$.

The overall bankruptcy cost,

$$\begin{aligned} BC &= 253{,}121.03 \times 0.0234537\% - 522{,}771.17 \times 0.0108071\% \\ &= 2.87 \text{ (million yuan)} \end{aligned}$$

Similarly, the bankruptcy probability and bankruptcy cost of Gree and Haier can be calculated in the same way, and the results are summarized in Table 11.

In Table 11, column N(−d2) is the bankruptcy probability; Strictly speaking, it is the overall bankruptcy probability in a debt cycle. It can be seen that the lowest is Gree, which is almost zero. Haier and Midea are slightly higher, around 0.02%, which are basically negligible. This reflects the actual situation of the three home appliance giants. From the absolute number of bankruptcy costs, the difference between the three companies is more obvious. Gree's bankruptcy cost is negligible, while Midea and Haier are more than 2.8 and 0.8 million yuan respectively. Obviously, the figures are also insignificant compared with the overall volume of the companies.

Healthy companies often have small bankruptcy probability and bankruptcy cost, which are confirmed in the above analyses. The calculations in Table 11 also confirm the soundness and capability of the ZZ bankruptcy probability model and bankruptcy cost model in bankruptcy risk analyses. There may not be much difference between the three companies based on intuition and conventional methods. It is conceivable that if rating companies are allowed to rate, the three companies are likely to belong to the same risk level, such as AAA. The calculation results of the two models are even more precise and clear than the ratings. The models are like a microscope, which can accurately enlarge the bankruptcy risk of the company and facilitate the naked eye to identify the bankruptcy risk difference caused by subtle differences between companies. It can be said that they can be clearly observed through model calculation.

Of course, further or deeper analyses can also be conducted based on the above discussion. For instance, the bankruptcy probabilities in Table 11 are bankruptcy probabilities within a debt cycle. Because the debt cycles are different among the three companies, they can be enhanced in comparability through annualization based on Eqs. (2) or (8). That is,

Table 12 The bankruptcy probabilities of the three companies over years

Year	Annual	2	3	4	5	6	7	8	9	10
Midea (%)	0.03	0.05	0.08	0.10	0.13	0.16	0.18	0.21	0.23	0.26
Gree (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Haier (%)	0.02	0.03	0.05	0.06	0.08	0.09	0.11	0.12	0.14	0.15
Year	11	12	13	14	15	16	17	18	19	20
Midea (%)	0.29	0.31	0.34	0.36	0.39	0.42	0.44	0.47	0.49	0.52
Gree (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Haier (%)	0.17	0.18	0.20	0.21	0.23	0.25	0.26	0.28	0.29	0.31

Annual = annual bankruptcy probabilities

$$\text{Midea: } \mathrm{p} = 1 - (1 - \mathrm{P})^{1/\mathrm{n}} = 1 - (1 - 0.0234537\%)^{1/0.8995} = 0.02607436\%;$$

$$\text{Gree: } \mathrm{p} = 1 - (1 - \mathrm{P})^{1/\mathrm{n}} = 1 - (1 - 0.0000126\%)^{1/0.8360} = 0.00001509\%;$$

$$\text{Haier: } \mathrm{p} = 1 - (1 - \mathrm{P})^{1/\mathrm{n}} = 1 - (1 - 0.0131535\%)^{1/0.8561} = 0.01536367\%.$$

Because the debt cycles are not much different among the three companies, the differences in annual bankruptcy probabilities among the three companies are similar to their differences in a debt cycle. Based on the annual bankruptcy probabilities, we can further calculate the bankruptcy probabilities of the three companies over any period of years based on Eq. (1), which is interesting because it can further enlarge the differences among the three companies over a long run. Some results of the relevant calculation are shown in Table 12.

In Table 12, the cumulative bankruptcy probabilities of the three remain less than 1% over 10 or 20 years, and Gree even remain to be zero in bankruptcy probabilities rounding to the two decimal. This may reflect the current situation or situation over near future of the three companies, but may be not the situation of most companies. Apparently, the cumulative bankruptcy probabilities in the table depend on the annual bankruptcy probabilities. These annual bankruptcy probabilities are unlikely to remain constant over long time in the future, and are likely to increase because it is not easy for the companies to keep on the top in multiple dimensions.

3.3 The Current Bankruptcy Risk

Similar to previous section, summarize the relevant data estimated previously for the current bankruptcy risk, as shown in Table 13.

Based on the data in Table 13, take Midea as an example,

The volatility within the current debt life,

Table 13 The variable values estimated for current bankruptcy risk analyses

Company	Cdebt	Casset	CL (%)	Cvolatility (%)	Cmaturity
Midea	222,851.48	248,864.51	89.55	10.63	0.75
Gree	197,101.39	225,849.65	87.27	7.14	0.75
Haier	124,796.95	123,607.78	100.96	11.28	0.75

Cdebt = current debt; Casset = current asset; CL = current debt ratio; Cvolatility = volatility of current asset; Cmaturity = current debt maturity

$$\sigma\sqrt{\mathrm{T}} = 10.63\%\sqrt{0.75} = 9.20\%;$$

$$\mathrm{d1} = -\frac{\ln 89.55\%}{9.20\%} + \frac{9.20\%}{2} = 1.245843$$

$$\mathrm{d2} = -\frac{\ln 89.55\%}{9.20\%} - \frac{9.20\%}{2} = 1.153828$$

$$\mathrm{N}(-\mathrm{d1}) = 10.64111\%$$

$$\mathrm{N}(-\mathrm{d2}) = 12.42854\%$$

Hence current bankruptcy probability P = 12.42854%.
The current bankruptcy cost,

$$\begin{aligned}\mathrm{BC} &= 222{,}851.48 \times 12.42854\% - 248{,}864.51 \times 10.64111\% \\ &= 1215.26\ (\text{million yuan})\end{aligned}$$

Similarly, the bankruptcy probability and bankruptcy cost of Gree and Haier can be calculated in the same way, and the results are summarized in Table 14.

In Table 14, column N(−d2) is the current bankruptcy probability in a current debt cycle. As shown, the current bankruptcy risks (bankruptcy probabilities and costs) of the three companies are significantly higher than that of the overall bankruptcy; This is because the current leverage ratio of the three companies is significantly higher than the total leverage ratio. Haier's current leverage ratio is higher than 100%, reaching 100.96%. Midea and Gree are also as high as 89.55% and 87.27% respectively.

Table 14 The current bankruptcy risk analyses

Company	d1	d2	N(−d1) (%)	N(−d2) (%)	BC
Midea	1.245843	1.153828	10.64111	12.42854	1215.2561
Gree	2.167928	2.104189	1.50821	1.76810	78.6670
Haier	−0.049168	−0.146855	51.96072	55.83770	5456.2459

BC = current bankruptcy cost (in million yuan)

Their current bankruptcy probabilities are hence relatively high, especially Midea and Haier, which are 12.43% and 55.84% respectively. That is to say, the probability of default is about 12.43% and 55.84% respectively if the repayment of current debts depends only on the current asset. As mentioned earlier, this is not a special study on the bankruptcy risk of Haier or the three companies. The relevant calculation is based on several simplified assumptions, so the conclusion may not be reasonable or reliable.

Even this is indeed the final conclusion, nothing needs to be worried too much, because even the current bankruptcy probability exceeds 50%. When the overall bankruptcy probability is not high, default will not occur at the end, and generally there will be no more serious situation. From the perspective of Haier, it may be worth paying attention to, because abnormal arrears will damage the company's reputation; Moreover, it may increase the management workload and cause other unnecessary troubles. Therefore, Haier may need to arrange more carefully to match the cash outflow with the inflow; If necessary, short-term financing is worth to consider.

On the other hand, the current bankruptcy probabilities are much different among the three companies. The lowest is Gree, which is less than 2%. Midea and Haier are much higher, which are 12.43% and 55.84% respectively. This may reflect the actual situation of the three home appliance giants. That is, although they are almost the same as no overall bankruptcy concerns, but they may be much different in the pressure of the current debt repayment. This is confirmed by the absolute number of bankruptcy costs. The differences among the three companies are more obvious. Gree's current bankruptcy cost is 78.67 million yuan, and is negligible comparing with the size of the company, while Midea and Haier are around 1200 and 5500 million yuan respectively. Although they are also insignificant comparing with their sizes, they are dozens of times as big as Gree.

This once again reflects the microscopic effect of the ZZ bankruptcy probability model and bankruptcy cost model, which can accurately reflect and amplify the bankruptcy risk of the company and help to identify the subtle differences in bankruptcy risk between or among companies. This is particularly important, because the calculation will be too late and of little significance when a company becomes more obvious in insolvency. Therefore, for the risk assessment of customers or potential customers, banks can use ZZ bankruptcy probability model and bankruptcy cost model to calculate the bankruptcy risk of all customers, and then sort and screen them.

The differences are also confirmed by the size and the proportion of "cash and cash equivalents" in their assets. At the end of 2021, Midea, Gree and Haier hold "cash and cash equivalents" 71,875.56, 116,939.30 and 45,857.17 million yuan respectively. The amount hold by Gree is almost the same as the sum of Midea and Haier. Table 15 shows the "cash and cash equivalents" of the three companies as a proportion of the current debt, current asset, total debt and total asset respectively.

Apparently, even the blue chip stocks like Haier, Midea and Gree, their bankruptcy probabilities are not zero, and the current bankruptcy probabilities are even not low, which fully shows that the bankruptcy risk needs constant attention. It's perhaps OK to change the name to default risk, credit risk and financial risk, but it is not right to

Table 15 The proportion of "cash and cash equivalents"

Company	Cash	Casset (%)	Cdebt (%)	Tdebt (%)	Tasset (%)
Midea	71,875.56	28.88	32.25	28.40	13.75
Gree	116,939.30	51.78	59.33	55.25	27.35
Haier	45,857.17	37.10	36.75	33.63	15.36

Cash = the amount of "cash and cash equivalents" (in million yuan); Casset = the proportion of "cash and cash equivalents" over current asset; Cdebt = the proportion of "cash and cash equivalents" over current debt; Tdebt = the proportion of "cash and cash equivalents" over total debt; Tasset = the proportion of "cash and cash equivalents" over total asset

ignore the bankruptcy risk. On the other hand, the relevant calculation results show that ZZ bankruptcy probability model is highly sensitive to the situation of companies and can fully measure or show the difference in bankruptcy risk among companies.

Based on the ZZ bankruptcy probability and cost model, the bankruptcy risk of a specific company at a specific time can be evaluated from the dimensions of bankruptcy probability and bankruptcy loss. These models are sufficient in flexibility and practicability, hence the bankruptcy probability and bankruptcy loss can be calculated for the current bankruptcy of the company in the short term or the overall bankruptcy of the company in the long term respectively.

The models are closed solution models sound in theory. Based on these models, financial ratios such as current ratio and leverage ratio can be converted into current and overall bankruptcy probabilities, hence enrich the implications of financial indicators and financial ratios in valuation and decision-making, and may herald some big progresses in financial (statement) analyses.

The model-based analyses show that, the short-term and long-term bankruptcy probability of all companies is greater than zero. No matter how healthy the company is at present, its overall and current bankruptcy probabilities are not as low as zero. Therefore, the bankruptcy risk is not only related to a few "unhealthy" companies. All companies, including healthy companies, should pay attention to the bankruptcy risk. External equity (stock) and debt (bond) investors, such as venture capitals and commercial banks, need theoretical and practical tools like the ZZ bankruptcy probability and bankruptcy cost model to measure and judge the bankruptcy risk of all their potential investment objects.

4 Firm Life Expectancy Prediction

Firm life expectancy is a major concern in both firm or equity valuation and risk management, but it draws very limited attention in academic research so far. The reason is obviously its toughness or infeasibility rather than unimportance.

4.1 General Understandings and Perspectives

(1) On the importance

No matter how famous a company is, it cannot escape the end of bankruptcy.

Lehman Brothers, a global diversified investment bank, founded in 1850, became the martyr with $613 billion in debt at the age of 158 on September 15, 2008 in the subprime mortgage crisis.

Eastman Kodak, founded in 1881, as the inventor of film, film, public cameras and digital cameras, has long been the leader and largest in the global market, went bankrupt at the age of 131 in January 2012.

Yahoo, founded in 1994, rising in the internet era, once an absolute overlord of the early internet. In July 2016, when it is 22 years old, Yahoo's core assets, were sold to us telecom giant Verizon at a price of US $4.83 billion.

This list can be endless, because a lot of companies, famous and unknown, are disappearing every year, every month and every day, and bankruptcy or shut down are the definite end over limited horizon for every company! Those famous companies are luckier than the unknown companies only in that they are known by more people before disappear.

There is no company with an infinite life span. The specific reasons may be different, but the final outcomes are the same. No matter how the company moves and transforms, it will be doomed to collapse in the storm of the market; No matter how many times they are rescued by M&A, they will pass away together with the buyers or acquirors.

If the survival and development are regarded as two major events for a company, survival is usually more important than development. This means that the discussion of the company's life expectancy is very important. If the end of a company were predicted, people may be more cautious in buying its stocks, bonds, products and jobs.

Unfortunately, similar to the bankruptcy, the research on firm life expectancy will encounter inexplicable resistance. For example, for a currently normal company, if the research conclusion is that the company is coming to an end, it may encounter the disgust or opposition of the existing shareholders, creditors, management, ordinary employees, etc.; If the research conclusion is that the company is safe and well, and the future is long, it will make people feel that the research conclusion is irrelevant and dispensable.

Anyway, the research concerning the life expectancy of a company are of great significance for many purposes and participants, such as the consideration of future business cooperation, future investment, future loans, and personal career development decisions of managers and employees.

The discussion on the Gordon growth model concerning stock value in Chapter "Stock and Equity Valuation: Where Discounting Does Not Work" shows also that it is necessary to have a more formal and serious discussion on the company's life, because this problem involves the longest period of earnings forecast, of course, it also involves the long-term growth rate of the company or its stock earnings.

However, the problem is very difficult, and previous academic research involves very limited. This is also why we did not specifically discuss on it in the previous chapters.

(2) On the methods

Different from people's life span, companies often have no certain "life span", and there may be no "signs" before they go to decline, which seems to have no rules to follow. This increases the difficulty of research. The previous literature focuses mainly on the descriptive statistics afterwards in this area. Although it has research feasibility, obviously, descriptive statistics afterwards have little to do with revelation of professional principles and enhancement of professional understanding.

It's worth noting that decisions are always future oriented. If the research of firm life is to have decision-making significance, it cannot stay in the statistics and description after the event. It must be further developed to the prior research of firm life expectancy. Perhaps, compared with afterwards statistics, prior estimation is much more difficult; However, the purpose and significance of scientific research lies in overcoming difficult problems, and we should not retreat from them.

It is not hard to imagine that the life span of a company is related to the risks it takes. The value and earnings of a company are in the process of ceaseless fluctuations (risks). But when it cannot pay the due debt even by sell all its assets, the company will be insolvent and go bankrupt, leading to the end of its life. The debt or leverage ratio hence is another factor in determining the life span of the company.

Therefore, the life of a company is related to its overall risk and debt ratio. Therefore, in order to calculate the life of a company, we need to know the quantitative relationship between the above two factors and the probability of bankruptcy.

Fortunately, the previous two sections of this chapter provide some theoretical models, which cast light on this issue or laid a certain theoretical foundation for solving this problem. In fact, according to the basic principle of queuing theory, given the bankruptcy probability, it is not difficult to find the life of the company or its equity.

Here we try to find a new way to use probability or queueing theory (rather than statistical description or regression) and financial professional methods as revealed in the previous sections to explore the prediction of corporate life expectancy. Similar analyses and calculations were published by Zhang [1, 2].

4.2 *The Bankruptcy Probability and Firm Life Expectancy*

Let us come straight to the point, consider directly the relationship between the bankruptcy probability and firm life expectancy. Assume bankruptcy represents the end of a company's life span, and the annual bankruptcy probability is p, which is constant in the future or represents the average probability over long term in the future. Then, the actual life span is determined by the year when the bankruptcy occurs, which belongs to forecasting or prediction to study.

There are many possible life-spans different in length for a company, and their weighted average is the firm life expectancy. The first or shortest possible life span is 1 year with probability of p, which means the company will go bankrupt in the first year; the second possible life span is 2 years with probability of (1 – p)p, which means the company will not go bankrupt in the first year but will go bankrupt in the second year; the third possible life span is 3 years with probability of $(1 – p)^2p$, which means the company will not go bankrupt in the first year and second year but will go bankrupt in the third year; and so on. Then the life expectancy (N) of the company is the expectation of all those possible life spans. Thus,

$$N = 1 * p + 2 * (1 - p) * p + 3 * (1 - p)^2 * p + 4 * (1 - p)^3 * p + \cdots \tag{9}$$

Then, divide two sides by p,

$$N/p = 1 + 2 * (1 - p) + 3 * (1 - p)^2 + 4 * (1 - p)^3 + \cdots \tag{10}$$

Then, multiply two sides by (1 – p),

$$N/p * (1 - p) = 1 * (1 - p) + 2 * (1 - p)^2 + 3 * (1 - p)^3 + \cdots \tag{11}$$

Then, Eq. (10) minus Eq. (11),

$$N/p - N/p * (1 - p) = 1 + (1 - p) + (1 - p)^2 + (1 - p)^3 + \cdots \tag{12}$$

Then, multiply two sides by (1 – p),

$$[N/p - N/p * (1 - p)] * (1 - p) = (1 - p) + (1 - p)^2 + (1 - p)^3 + \cdots \tag{13}$$

Then, Eq. (12) minus Eq. (13),

$$\left[N/p - N/p * (1 - p)\right] - \left[N/p - N/p * (1 - p)\right] * (1 - p) = 1 \tag{14}$$

Then, divide two sides by N,

$$\begin{aligned} 1/N &= \left[1/p - 1/p * (1 - p)\right] - \left[1/p - 1/p * (1 - p)\right] * (1 - p) \\ &= \left[1/p - 1/p * (1 - p)\right]\left[1 - (1 - p)\right] \\ &= \left[1/p - 1/p * (1 - p)\right]p \\ &= 1 - (1 - p) \\ &= p \end{aligned} \tag{15}$$

That is,

$$N = 1/p \tag{16}$$

Therefore, the life expectancy of a company is the reciprocal of its annual bankruptcy probability. We have solved the quantification of the annual bankruptcy probability, that is Eq. (8). Therefore, the firm life expectancy can be hopefully determined by combining Eqs. (8) and (16), that is,

$$\mathrm{N} = 1/\mathrm{p} = 1/\left\{1 - \left[1 - \mathrm{N}\left(\frac{\ln L}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}\right)\right]^{1/\mathrm{T}}\right\} \tag{17}$$

Now we find a fundamental solution to the problem of the estimation of firm life expectancy. Similar to all the fundamental solutions in this book, the solution is a closed form model (Eq. 17), the form and variables of the model are all determined by strict concepts and strict logic reasoning, rather than choose subjectively or data processing based on a sample data. For consistence and convenience, Eq. (17) can be referred to as ZZ firm life expectancy model.

We have differentiated two kinds of bankruptcies in previous sections, the current bankruptcy and the overall bankruptcy. Apparently, the overall bankruptcy rather than the current bankruptcy determines the life of a company. The annual bankruptcy probability, p, hence is the annual bankruptcy probability in the overall bankruptcy.

For instance, we have worked out the cumulative (overall) bankruptcy probabilities of Midea, Gree and Haier are 0.0234537%, 0.0000126%, 0.0131535% respectively (Table 11); their corresponding annual bankruptcy probabilities then are 0.02607436%, 0.00001509%, 0.01536367% respectively. Now, based on the ZZ firm life expectancy model, Eq. (17), their corresponding firm life expectancies are,

$$\text{Midea: } 1/0.02607436\% = 3836\ \text{(year)}^{4};$$

$$\text{Gree: } 1/0.00001509\% = 6{,}626{,}676\ \text{(year)};$$

$$\text{Haier: } 1/0.01536367\% = 6509\ \text{(year)}.$$

The results seem too long to believe, because it is uncommon for a company to last over 1000 years, do not mention the 6,000,000 years (the result to Gree). The reason is that we use a constant annual bankruptcy probability, and this annual bankruptcy probability is estimated based on the situations of the three companies during their prosperous or flourishing period or golden age, but such a period or age cannot last forever, or obviously, cannot last till the ends of these companies.

Put it another way, the annual bankruptcy probabilities are not representative for the long enough future; or they cannot represent the average level in the future. For simplicity, assume the annual bankruptcy probabilities are 1% when the three companies in their worse or fading period; take the simple average of this 1% and the above annual bankruptcy probabilities as their average annual bankruptcy probabilities over

[4] Round up rather than round off.

future, then,

$$\text{Midea: } (1\% + 0.02607436\%)/2 = 0.51303718\%;$$

$$\text{Gree: } (1\% + 0.00001509\%)/2 = 0.50000755\%;$$

$$\text{Haier: } (1\% + 0.01536367\%)/2 = 0.50768183\%.$$

Recalculate their life expectancies based on the above average annual bankruptcy probabilities, then,

$$\text{Midea: } 1/0.02607436\% = 195\ (\text{year});$$

$$\text{Gree: } 1/0.00001509\% = 200\ (\text{year});$$

$$\text{Haier: } 1/0.01536367\% = 197\ (\text{year}).$$

Apparently, these life expectancies are more believable.

Therefore, attention should be paid on the change hence the representativeness of the annual bankruptcy probability for the calculation of the firm life expectancy. At least, the annual bankruptcy probability estimated based on the current situation cannot be used directly in the calculation. Of course, the above simple and arbitrary processing does not represent a good or final solution, and how to make a better estimation of the annual bankruptcy probability is an issue worth to explore deeply.

Based on the ZZ firm life expectancy model, the relatively long life expectancies of Midea, Gree and Haier, mainly comes from their lower volatilities and leverages as well as short debt maturities. Other companies, with different volatilities, leverages and debt maturities, will of course have different life expectancies. It is perhaps more interesting to calculate the life expectancies of the ordinary or average companies.

As a base case, consider an average company, its volatility is 20%, the leverage is 50%, the debt maturity is 3 years. Then, based on the ZZ firm life expectancy model,

$$\mathrm{N} = 1/\left\{1 - \left[1 - \mathrm{N}\left(\frac{\ln 50\%}{20\%\sqrt{3}} + \frac{20\%\sqrt{3}}{2}\right)\right]^{1/3}\right\} = 87.8\ (\text{year})$$

The firm life expectancy is very sensitive to all the three influential variables, the volatility, the leverage and the debt maturity. The firm life expectancy will change significantly for a slight change in the three variables. As an average company has life expectancy about 88 years or more generally 80–90 years is down to earth or makes sense to a great extent. This implies that the values of the three influential variables assumed in the base case are probably right as the typical or average values.

Table 16 Volatility, leverage and the firm life expectancy (year)

	Leverage	20–40%	40–60%	60–80%
Volatility	Midpoint	30%	50%	70%
10–20%	15%	900,539	537	27
20–30%	25%	579	35	10
30–40%	35%	64	14	7

Leverage = the range of leverage; Volatility = the range of volatility; Midpoint = midpoint value of the risk class

Take volatility 10–20% as low risk, 20–30% as moderate risk, and 30–40% as high risk. Take leverage 20–40% as low leverage, 40–60% as moderate leverage, and 60–80% as high leverage. Combine the high, moderate and low classes of volatility and leverage, take the midpoint or midvalue as the value of the relevant risk class and put it into the calculation, the firm life expectancies of the nine risk levels are shown in Table 16.

The firm life expectancies in Table 16 are significantly different from each other. The extreme values, such as life expectancy of 900,539 years or 7 years seems not so realistic. This confirms again that companies will move between the risk levels over long run, and implies that the transition of risk level over time must be factored in the estimation of firm life expectancy. Put it another way, the ZZ firm life expectancy model can only be used with the adjustment of the annual bankruptcy probability.

4.3 The Firm Life Expectancies Based on Moody's Rating Data

As the world's largest (default) risk rating company, Moody's has recorded and accumulated large database about firm life cycle processes. Table 17 shows Moody's cumulative default probabilities of companies with various risk levels calculated based on the data accumulated during 1970–2010. Although not the same in concept, Moody's default probability can be roughly regarded as bankruptcy probability.

Note that Table 17 shows the cumulative bankruptcy probability, that is, the probability of bankruptcy once over some years. According to the ZZ firm life expectancy model, the reciprocal of an average annual bankruptcy probability is the company's life expectancy. Therefore, the following focuses on how to calculate the reasonable or effective annual bankruptcy probability.

With the cumulative bankruptcy probability, it is not difficult to get the average annual bankruptcy probability. Based on Eq. (2) and Table 17, the average annual bankruptcy probabilities are calculated as shown in Table 18.

An average annual bankruptcy probability can be derived from each of the cumulative bankruptcy probabilities in Table 17 by using Eq. (2); the results are the figures in Table 18. Obviously, these average annual bankruptcy probabilities are not equal,

Table 17 Cumulative default probability (%): based on Moody's 1970–2010

Maturity	1	2	3	4	5	7	10	15	20
Aaa	0.000	0.013	0.013	0.037	0.104	0.244	0.494	0.918	1.000
Aa	0.021	0.059	0.103	0.184	0.273	0.443	0.619	1.260	2.596
A	0.055	0.177	0.362	0.549	0.756	1.239	2.136	3.657	6.019
Baa	0.181	0.510	0.933	1.427	1.953	3.031	4.907	8.845	12.411
Ba	1.157	3.191	5.596	8.146	10.453	14.440	20.101	29.702	36.867
B	4.465	10.432	16.344	21.510	26.173	34.721	44.573	56.345	62.693
Caa	18.163	30.204	39.709	47.317	53.768	61.181	72.384	76.162	78.993

Source Moody's rating data
Maturity = the time horizon the cumulative bankruptcy probability strides across
Caa: includes all grades below Caa

Table 18 Annual bankruptcy probability (%): based on Moody's 1970–2010

Maturity	1	2	3	4	5	7	10	15	20	Average
Aaa	0.000	0.007	0.004	0.009	0.021	0.035	0.050	0.061	0.050	0.026
Aa	0.021	0.030	0.034	0.046	0.055	0.063	0.062	0.084	0.131	0.059
A	0.055	0.089	0.121	0.138	0.152	0.178	0.216	0.248	0.310	0.167
Baa	0.181	0.255	0.312	0.359	0.394	0.439	0.502	0.615	0.660	0.413
Ba	1.157	1.608	1.901	2.102	2.184	2.203	2.219	2.322	2.273	1.997
B	4.465	5.360	5.775	5.875	5.888	5.911	5.730	5.376	4.810	5.466
Caa	18.163	16.456	15.521	14.804	14.298	12.644	12.074	9.117	7.505	13.398

The same as Table 17

but it is hard to judge which one is more representative. Therefore, we calculate the simple average of these annual bankruptcy probabilities in the last column of Table 18. It can be considered that this average is more representative.

However, in the long run, this average is not representative enough. Even if we don't pursue more details, a change in the future of the company can't be ignored. That is, in the long run, the risk level of the company will change. In other words, in a long enough time, a company may change from its current risk level to any other six levels, so that the average annual bankruptcy probability will be different. Although there may be little difference in the average annual bankruptcy probability, the effect of small difference accumulated over the years cannot be ignored. Therefore, for any company at a specific risk level, the average annual bankruptcy probability in Table 18 is not qualified as the long-term applicable average annual bankruptcy probability.

In order to get the long-term applicable average annual bankruptcy probability, it is necessary to consider the change of risk level on the basis of the average annual bankruptcy probability in Table 18. That is to say, in the long run, what is the probability that a company of one level will rise or fall to other levels. It is necessary to discuss the probability of a company's specific time varying between different risk levels in combination with the specific situation. In order to focus on the general ideas and methods of estimating the life expectancy of a company, we would not discuss further on this topic here, and simply assumed that in the long run, the probability of the company changing to the first, second, third, fourth, fifth and sixth nearest level every year is 6%, 5%, 4%, 3%, 2% and 1% respectively; deducting the total probability of change to each level from 100%, the difference is the probability of remaining in the current level.

For example, for a company currently in AAA level, the probability of changing to AA level is 6%, the probability of changing to A level is 5%, the probability of changing to Baa level is 4%, the probability of changing to Ba level is 3%, the probability of changing to B level is 2%, and the probability of changing to Caa level is 1%; The sum of these probabilities is 21%. Therefore, the probability of remaining at the current level is 79%. At present, for companies in AA level, the probability of changing to AAA and A level is 6%, the probability of changing to Baa level is 5%, the probability of changing to Ba level is 4%, the probability of changing to B level is 3%, and the probability of changing to Caa level is 2%; The sum of these probabilities is 26%. Therefore, the probability of remaining at the current level is 74%. By analogy, the long-term applicable average annual bankruptcy probability can be obtained as shown in Table 19.

Based on the reciprocal of the long-term applicable annual bankruptcy probability in Table 19, the firm life expectancies of various risk level are shown in Table 20.

According to Table 20, at present, companies with A-level or above are expected to have a life span of more than 100 years, and companies with AA level are expected to have a life span of more than 200 years; Companies from B to baa are expected

Table 19 Average annual bankruptcy probability: long term applicable value

Grade	Number	Original (%)	Adjusted (%)	Calculation
Aaa	(1)	0.026	0.352	(1) × 79% + (2) × 6% + (3) × 5% + (4) × 4% + (5) × 3% + (6) × 2% + (7) × 1%
Aa	(2)	0.059	0.588	(1) × 6% + (2) × 74% + (3) × 6% + (4) × 5% + (5) × 4% + (6) × 3% + (7) × 2%
A	(3)	0.167	0.869	(1) × 5% + (2) × 6% + (3) × 71% + (4) × 6% + (5) × 5% + (6) × 4% + (7) × 3%
Baa	(4)	0.413	1.232	(1) × 4% + (2) × 5% + (3) × 6% + (4) × 70% + (5) × 6% + (6) × 5% + (7) × 4%
Ba	(5)	1.997	2.452	(1) × 3% + (2) × 4% + (3) × 5% + (4) × 6% + (5) × 71% + (6) × 6% + (7) × 5%
B	(6)	5.466	4.998	(1) × 2% + (2) × 3% + (3) × 4% + (4) × 5% + (5) × 6% + (6) × 74% + (7) × 6%
Caa	(7)	13.398	11.035	(1) × 1% + (2) × 2% + (3) × 3% + (4) × 4% + (5) × 5% + (6) × 6% + (7) × 79%

Original = Annual bankruptcy probability (%) based on Moody's cumulative one; Adjusted = Long-term applicable average annual bankruptcy probability, after the adjustment of risk level transition

Table 20 Firm life expectancy (years) at various risk levels

Risk level	Aaa	Aa	A	Baa	Ba	B	Caa
LABP (%)	0.352	0.588	0.869	1.232	2.452	4.998	11.035
Elife (year)	283.962	170.152	115.125	81.159	40.783	20.007	9.062

Risk level = the risk classification of Moody's; LABP (%) = the long-term applicable annual bankruptcy probability; Elife (year) = the firm life expectancies

to have a life span of 10–100 years; The average life span of companies below CAA level is 9 years; As you can imagine, many companies at this level go bankrupt after surviving for about 10 years.

As short as 10 years or so, as long as 200 years or so. This makes sense because it is basically the same as the firm life spans in reality. This also reflects the rightness or reliability of the above analysis in two aspects. On one hand, the Moody's data is large enough to reflect the overall situation; On the other hand, the above analysis and calculation based on Moody's data is correct or reasonable. In other words, the calculation method from default probability to firm life is correct or reasonable.

Obviously, although the above calculation process is to obtain the life distribution of companies at all risk levels, it is also applicable to calculate the life expectancy of a single company. As far as a company is concerned, it is necessary to find out the risk level of the company and the cumulative default probability in adequate years. With these basic data ready, we can go through the above process to calculate the average annual bankruptcy probability and the long-term applicable average annual bankruptcy probability, and then calculate the life expectancy of the company.

From the cumulative default probability to the average annual bankruptcy probability, there are objective model (Eq. 2) can be used; but from the average annual bankruptcy probability to the long-term applicable average annual bankruptcy probability, the calculation is based on a subjective and perhaps reasonable assumption. The assumption hence the calculation is not necessarily right or the best, and may need to be discussed more deeply. Anyway, we will not further this discussion here, and leave the topic to the subsequent application research.

4.4 The Firm Life Expectancies Based on Empirical Volatilities

Previous section estimates the average life expectancies for companies in various risk levels based on Moody's data. This is by no means the firm life expectancies can only be estimated based on a large quantity of data. With the ZZ firm life expectancy model in hand, we now try to do the same estimation based on very limited data.

According to the long-term empirical data of the U.S. stock market, the volatility of stocks in traditional industries is mainly distributed in the range of 20–40%. As more and more high-tech companies, internet companies and various start-up companies join in the stock market, the volatility is enlarged to some extent, for example, most stocks are between 10 and 50%. The volatilities of the companies on the top of their industries, such as companies rated as AAA, may be close to 10%; the volatilities of those poor companies without decent business or promising future, such as those rated as below grade C, may be close to 60% or even as high as 70%.

The volatility in the ZZ firm life expectancy model represents the overall risk of the company, that is, the standard deviation of the return on total assets of the company. The stock volatilities just represent part or main part of the company volatilities. Another part volatilities come from the debt, which are significantly lower than that of the stocks in the same company. According to the discussion in Sect. 1 in Chapter "Capital Asset Pricing: An Easy and Unified Solution" or as summarized in Table 5 in Chapter "Capital Asset Pricing: An Easy and Unified Solution", most debts (bonds) volatilities are significantly lower than that of stock, around 5% of stock volatilities in the same company, ranges from 0.0 to 3.8%. Therefore, the company volatilities are mainly determined by the stock volatilities and leverages, and are significantly lower than that of the stocks when the leverage is not very high.

For simplicity, assume the leverage is 50% and remains constant. Imitating the risk classification in Moody's ratings in previous section, we now divide the possible volatilities into seven degrees. Based on the previous analyses and findings, taking 6.0% and 42.0% as the group median of the lowest and highest volatility respectively, this range is divided into seven equidistant intervals, namely 3.0–9.0%, 9.0–15.0%, 15.0–21.0%, 21.0–27.0%, 27.0–33.0%, 33.0–39.0%, 39.0–45.0%; the risk groups were named by A, B, C, D, E, F and G, and the corresponding median values were 6.0%, 12.0%, 18.0%, 24.0%, 30.0%, 36.0% and 42.0%, respectively. Imitating the calculation based on Moody's data, along with the debt maturity changes from 1 to 20 years, the cumulative bankruptcy probability of each risk group is calculated according to the ZZ bankruptcy probability model, i.e. Eq. (7), as shown in Table 21.

Based on Table 21, calculate the annual bankruptcy probability of each risk group under each time horizon, as shown in Table 22.

Table 21 The cumulative bankruptcy probability of risk groups

Grade	Median (%)	Debt maturity (year)								
		1	2	3	4	5	7	10	15	20
A	6.0	0.000	0.000	0.000	0.000	0.000	0.001	0.019	0.207	0.716
B	12.0	0.000	0.003	0.062	0.282	0.716	2.146	5.083	10.401	15.309
C	18.0	0.008	0.472	1.935	4.046	6.414	11.173	17.538	25.923	32.327
D	24.0	0.282	3.057	7.220	11.428	15.309	21.943	29.673	38.937	45.654
E	30.0	1.537	7.757	14.138	19.621	24.263	31.689	39.886	49.377	56.127
F	36.0	4.046	13.417	21.189	27.335	32.327	40.071	48.418	57.926	64.597
G	42.0	7.488	19.215	27.790	34.267	39.416	47.283	55.654	65.070	71.570

Grade = risk class; Median = mean of volatility; Debt maturity = consideration over time horizon

Table 22 The annual bankruptcy probability of risk groups (%)

Grade	Median (%)	Debt maturity (year)									Average
		1	2	3	4	5	7	10	15	20	
A	6.0	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.014	0.036	0.006
B	12.0	0.000	0.002	0.021	0.071	0.144	0.309	0.520	0.729	0.827	0.291
C	18.0	0.008	0.236	0.649	1.027	1.317	1.678	1.910	1.981	1.933	1.193
D	24.0	0.282	1.540	2.467	2.988	3.269	3.477	3.459	3.235	3.003	2.636
E	30.0	1.537	3.957	4.954	5.314	5.406	5.299	4.962	4.437	4.036	4.433
F	36.0	4.046	6.950	7.631	7.673	7.512	7.053	6.406	5.608	5.059	6.438
G	42.0	7.488	10.120	10.285	9.958	9.537	8.740	7.810	6.772	6.095	8.534

Similarly, this annual bankruptcy probability is not the long-term applicable annual bankruptcy probability. Considering the transition of risk level, based on the same assumption in the calculation in previous section (Table 19), adjust the annual bankruptcy probabilities into the long-term applicable annual bankruptcy probabilities, as shown in Table 23.

Based on the long-term applicable average annual bankruptcy probability in Table 23, the firm life expectancy in each risk group is calculated, as shown in Table 24.

Table 24 shows that the firm life expectancy can also derived without so large a data base as in Moody's. The results in Table 24 are derived by the ZZ bankruptcy probability model, while the results in Table 20 are derived based on Moody's default probability. The two tables are calculated via the same steps, and the results are similar to each other. Note that in this part of the calculation, although the companies

Table 23 Average annual bankruptcy probability: long term applicable value

Grade	Number	Original (%)	Adjusted (%)	Calculation
A	(1)	0.006	0.534	(1) × 79% + (2) × 6% + (3) × 5% + (4) × 4% + (5) × 3% + (6) × 2% + (7) × 1%
B	(2)	0.291	0.961	(1) × 6% + (2) × 74% + (3) × 6% + (4) × 5% + (5) × 4% + (6) × 3% + (7) × 2%
C	(3)	1.193	1.758	(1) × 5% + (2) × 6% + (3) × 71% + (4) × 6% + (5) × 5% + (6) × 4% + (7) × 3%
D	(4)	2.636	2.861	(1) × 4% + (2) × 5% + (3) × 6% + (4) × 70% + (5) × 6% + (6) × 5% + (7) × 4%
E	(5)	4.433	4.190	(1) × 3% + (2) × 4% + (3) × 5% + (4) × 6% + (5) × 71% + (6) × 6% + (7) × 5%
F	(6)	6.438	5.730	(1) × 2% + (2) × 3% + (3) × 4% + (4) × 5% + (5) × 6% + (6) × 74% + (7) × 6%
G	(7)	8.534	7.497	(1) × 1% + (2) × 2% + (3) × 3% + (4) × 4% + (5) × 5% + (6) × 6% + (7) × 79%

Original, Adjusted, the same as Table 19

Table 24 Firm life expectancy (years) at various risk levels

Risk level	A	B	C	D	E	F	G
LABP (%)	0.534	0.961	1.758	2.861	4.190	5.730	7.497
Elife (year)	187.191	104.109	56.870	34.958	23.865	17.451	13.339

Risk level, LABP, Elife, the same as Table 20

are also divided into seven groups based on risk, they are simply divided into equal distance groups based on the value of volatility, which may be different from Moody's grouping, such as the proportion of each group in the overall. In the case of differences in classification, the results are similar and roughly in line with the actual situation. This implies that both the two methods have preliminary rationality and reliability.

Strictly speaking, the bankruptcy probability calculated by ZZ bankruptcy probability model may be slightly different from Moody's default probability data. In reality, when the company is in danger of default or bankruptcy, it will take a variety of self-help or rescue measures, that is, to strengthen risk management. For the former low-risk companies such as those rated as AAA, this kind of effort may be particularly effective, so that the company can avoid danger. On the contrary, for companies with higher risk level, there may be an external overreaction, which will deteriorate rapidly if there is a slight disturbance, and it is useless to strengthen risk management. Moody's default probability, as a statistical data, naturally includes the effect of such artificial efforts and external reactions. Therefore, the life span of relative safe companies such as AAA and AA may be longer; the riskier companies such as Caa have a shorter life span.

This further explains the small difference between the calculation results based on ZZ bankruptcy probability model and Moody's default probability data. In other words, if we consider the changes of volatility when companies with different risk levels are in bankruptcy risk, the results of the two methods will be closer. In any case, it can be roughly said that the calculation based on ZZ bankruptcy probability model shows that the overall life of companies with various risk levels is as short as 10 years or so and as long as 200 years or so, which is basically consistent with the actual situation of the company in reality and the data results of Moody's company. This confirms the correctness and effectiveness of the method used, or the ZZ bankruptcy probability model.

Of course, although the above calculation process is for the companies in a whole group to get the company life distribution of each risk level, it is also applicable to calculate the life of any single company. As far as a single company is concerned, it is necessary to estimate the value of the three variables of the company's volatility (σ), debt ratio (L) and inspection period (T), and then use ZZ bankruptcy probability model to calculate the average annual bankruptcy probability and long-term applicable average annual bankruptcy probability, and then the company's life expectancy is just the reciprocal of the long-term applicable average annual bankruptcy probability.

The findings in firm life expectancy are worth considering and applying in the aspects of company prospect, income forecast, valuation, risk rating and risk management. Taking the company or its stock valuation as an example, in the conventional valuation, the discounted cash flow (DCF) method is the most reasonable method in theory. The DCF method refers to convert all future returns (earnings or cash flows) into present value, and then work out the asset value by summing these present values. Literally, the returns should be discounted are all the returns over the future years, but it is too difficult to forecast these returns year by year and the effort input this way cannot be justified by the accuracy of the final valuation. It is generally accepted that a better way is to estimate a rationalized initial return and an annual growth rate, then all future earnings can be expressed based on the initial returns and the growth rate. Obviously, whether the total number of years in the future over which the returns earned, or the annual growth rate, its estimation or forecasting is relied on the company's life span.

Two efforts are needed to measure the life expectancy of an actual company correctly. One is the right method or model; the other is the basic data needed by the method or model. This part focuses on the method (model and its application) in estimating the company's life expectancy, so we emphasize the soundness in theory and the strictness in logic. The estimations of basic data in the calculation are reasonable, but they are not necessarily the best or perfect processing; readers had better not to copy them in application. The reason is that although the basic data is equally important, it is not the subject of this book. In most cases, we seek a simple method to reduce the space and to avoid deviating too much from the main topic; so how to estimate the basic data need to discuss further by more subsequent research.

Anyway, the discussions in this chapter may extend readers' understanding on relevant issues and the methods in previous chapters, and can hopefully bring beneficial enlightenment to the risk management and related decision-making of commercial banks, guarantee companies, insurance companies as well as various industrial and commercial companies and financial institutions.

References

1. Zhang Z. Volatility and company life expectancy—research based on ZZ bankruptcy cost model. Res Financ Issues. 2018:91–9.
2. Zhang Z. Modeling corporate bankruptcy risk and its application. Mon J Finance Account. 2021;11:1–6.
3. Altman EI, Haldeman R, Narayanan P. Zeta analysis: a new model to identify bankruptcy risk of corporations. J Bank Finance. 1977;1(1): 29–53.

Further Readings

4. Beaver WH. Financial ratios as predictors of failure. J Account Res. 1966;4:71–111.

5. Beaver WH. Alternative accounting measures as predictors of failure. Account Rev. 1968;43:113–22.
6. Altman EI. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Finance. 1968;189–209.
7. Agarwal V, Taffler RJ. Twenty-five years of the Taffler z-score model: does it really have predictive ability? Account Bus Res. 2007;37:285–300.
8. Ashraf S, Félix EGS, Serrasqueiro Z. Do traditional financial distress prediction models predict the early warning signs of financial distress? J Risk Financ Manage. 2019;12:55–66.
9. Koh HC, Tan SS. A neural network approach to the prediction of going concern status. Account Bus Res. 1999;29:211–16.
10. Sigrist F, Hirnschall C. Grabit: gradient tree-boosted Tobit models for default prediction. J Bank Finance. 2019;102:177–92.
11. Hull JC. Risk management and financial institutions. 3rd ed. Wiley; 2012. p. 249.

Further Discussion: A Novel—The Falling Apples

Contents

A novel fabricated based on Newton's biography

> A nation cannot stand on the peak of science without theoretical thinking for a moment.
>
> — Friedrich Engels

Foreword

There are countless science fiction novels in ancient and modern times, and there are also many masterpieces that make people remember over long time and talk about it. However, there are few scientific enlightenment novels. If scientific fantasy can ignite people's scientific enthusiasm, then scientific enlightenment can awaken people's scientific or theoretical thinking. No matter how hot the enthusiasm is, it cannot replace sober and correct thinking. Obviously, scientific enlightenment is equally important as scientific fantasy.

Due to various reasons, such as catering to reviewers' preferences, etc., today's financial research has taken a devious road and has stagnated over decades, and even confused the difference between this discipline and other disciplines—the basic duty or function of this discipline in the scientific system. In this case, effective scientific

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_13

thinking in today's finance related area is particularly precious. It is not only necessary for scholars to do scientific research, but also necessary for our ordinary people to make rational decision, and to distinguish truth, goodness and beauty from false, evil and ugliness.

The following is a story fabricated based on Newton's biography. Please allow me here to express my most ashamed apology to the great Mr. Newton. Dear Mr. Isaac Newton, I am very very sorry from the bottom of my heart for use your name in such abrupt and offensive manner; but this is the only or most efficient way I can think of to enlighten scientific thinking in the finance related area.

1 The Carefree Childhood

In the autumn of 1655, Newton was 12 years old. As always, when autumn comes, there would be endless farm work. Newton did not like farm work. But he was a filial child. "With so many heavy and dirty jobs, I can't push them all to my parents." He had grown to manhood, and he wanted do as much as he can for his family.

That's what Newton thought and just what he did. One day, the breeze was gentle and the sun was warm. Newton helped his dad stab corn cobs all morning and followed his mother to harvest sweet potatoes in the afternoon. Newton's family did not grow many sweet potatoes, and the farm work was finished in the mid of the afternoon. After working continuously for hours, Newton felt tired and thirsty. He thought of a good place to go: Uncle Sam's orchard near his home.

"Really, I had not visited Uncle Sam's orchard for a long time. The apples, pears and grapes there might be near ripe."

Then, Newton asked his mother for leave and ran straight to Uncle Sam's orchard. Uncle Sam was a very kind person; he liked Newton very much. He saw Newton running all the way to him, and his face was already full of smiles. He took Newton's little hand and went straight to his "house", a "shack" built in the orchard. As soon as he entered the house, Newton saw two baskets of apples standing against the wall. "Oh, apple, that's great. I want to eat it."

"No problem, it's all reserved for you, and I know you're a little greedy cat." Uncle Sam picked up a large, round red apple from the basket, and handed it over. "Eat slowly, don't worry, everything in this basket was left for you." Newton ate the big apple with great enjoyment and felt that his physical fatigue was instantly compensated.

Finished the apple, Newton said he was going to play in the orchard. Uncle Sam said, "Play as you like, boys should play. But do not eat everything you see. I have no medicine for you if you get a tummy upset by the bad fruit." Newton ran out of the "house", like a bird fly into the forest. He got into the grapevine for a while, and climbed to the one he always likes to climb. The biggest apple tree, and then the biggest walnut tree, had a great time. Playing around, Newton felt a little sleepy. Unconsciously, he fell asleep under an apple tree.

Newton was sleeping soundly. In a daze, he felt as if someone had given him a head-on blow. Newton suddenly woke up and found an apple rolling along the slope on the right side of his foot. Oh, it turned out that an apple that the farmer missed when harvesting was falling because of over ripe and hit him.

Although the carefree childhood was full of hardships, all the memories left in Newton's young mind were wonderful. Somehow, the falling and rolling apples from the apple tree that day seemed to leave an indelible impression in his mind.

After going to school and knowing some scientific knowledge, Newton was so fascinated by science that he often forgot to eat and sleep, and he was tinkering with experiments, calculations, etc. almost every day. He made a lot of gadgets that his peers never thought of. Such as a special windmill to trick a mouse, an alarm clock to wake up by dripping water, a kite with a lantern, a folding lantern, and so on.

Nobody knew whether there were immortals of foresight in the world, but even if there were no immortals, every ordinary person around Newton could see that Newton came naturally for scientific research.

2 Initial Involvement in Scientific Research

Newton graduated from high school with excellent grades at the age of 18 and entered Trinity College, Cambridge University in 1661. There, Newton was systematically and fully exposed to the ideas of philosophers such as Aristotle and Descartes, as well as the theories of astronomers such as Galileo, Copernicus, and Kepler. Copernicus's heliocentric theory excited him so much that he made up his mind to take physics and mathematics as his lifelong research direction.

In 1665, Newton graduated from the University of Cambridge. Focusing on physics and astronomy for a long time, and with his talent in mathematics, Newton quickly deduced the Law of Universal Gravitation. It was not difficult for any real expert to understand that it was an extraordinary discovery, and it can be said to be the most significant astronomical and physical discovery after Copernicus. Newton was overjoyed by the discovery and spent two overnight writing an 8000-word dissertation explaining the new discovery.

Newton went to the post office and spent two dimes to send the article to a professional journal. A couple of months passed, no response from the journal. It took Newton two more dimes to mail the article to another professional journal. Unexpectedly, another few months passed, and Newton heard nothing from the journal. Newton modified the article, and spent again two dimes to mail his article to another professional journal. Once again, a couple of months passed, nothing came as Newton expected. Newton could not believe the result and could not figure it out. So he revised the article again, and chose another professional journal, and spent again two dimes to mail the article out. As always, no any response and nothing happened to his article.

In this way, Newton went to the post office time and time again and mailed his article to one professional journal agency after another again and again and again. Seeing that two years had passed, no professional journal agency gave him good news, except some expressed curtesy thanks. Newton was wondering, perhaps I needed to run to a nearest professional journal to see what happened and to consult personally.

The journal agency answered Newton's question. Newton understood that the professional journal itself did not decide the selection of manuscripts. The professional journal agency employed peers in physics to review and determine the acceptance of the manuscript. Newton asked, "Would you let me meet with a reviewer?" The editor of the journal said, "Sorry, that won't work at all. What place do you think here is? Who do you think the reviewers we employed are? Who do you think you are by yourself? Do you think they should see you when you wanted to see them? What a ridiculous request! We are all embarrassed for that request! Let me tell you, our system and standard are high and strict, and the articles received are all reviewed and selected anonymously by peer experts."

"I could not publish my thesis, and I could not see and consult with the reviewers. But what could I do by that?" Newton returned home, puzzled. In desperation, he seemed to find some ideas, that was, to see how those published papers were written, which might be helpful to pass the anonymous review? So, Newton spent some dimes and bought many professional journals to read.

Guess what? Reading those published articles, Newton suddenly felt his brain was widely opened, and his intelligence and wisdom were inspired and enhanced.

Newton saw the dropping or falling of apples in his childhood, and based on this, he deduced the law of gravitation that was difficult to publish; he was surprised to find that other scholars also had other "abundant" and easy-to-publish research findings based on the same phenomenon. This falling apple phenomenon had repeatedly appeared in the articles of these journals, and had become the current (then) research hotspot in physics.

Some scholars study the relationship between apple size and the probability of falling; Some scholars study the relationship between temperature and the probability of falling apples; some study the relationship between precipitation and the probability of falling apples; some study the date distribution of apples falling; some study the regional distribution of apples falling; some study Altitude distribution of apples falling; some study the relationship between apple falling and fertilization; some study the relationship between apple fall and watering; some study the speed of apple fall; some study the rolling distance of the apple from the apple tree after falling; still others study the rolling route of apple after falling (Straight line or curve, quadratic or cubic or higher curve).

Heck, the questions were so "rich", too "funny"! Too "imaginative"!

Newton, who was born and eager to learn, thought and asked himself, "why haven't I come across so many "state-of-the-art" research results before? So, holding these "leading-edge research findings" that passed the anonymous review, he had his nose in those journals except eating and sleeping.

Studied intensively for two months, Newton found that it was not easy to finished a major part of so many “leading-edge results” in a short period. Before finishing of reading and understanding the “Findings” of the previous issue, the “new Findings” of the next issue are coming. For example, some scholars studied the relationship between apple size and falling probability in the previous issue; the next issue may further study the relationship between apple shape and falling probability, or the relationship between apple sugar content and falling probability, or in the atmosphere relationship between PM250 concentration[1] and drop probability. In the previous issue, someone researched “the rolling distance of apples after falling”; the next issue may further study “the rolling distance of apples based on the ground slope”; and in the next issue, they may further study “the apples rolling distance based on different ground hardness and humidity”; and the next issue may further study “the rolling distance of apples after falling based on different obstacles on the ground (such as cow dung, shit, etc.)”. Along with the new issues of the journals coming one by one, the length of the article had also increased gradually.

What made Newton’s shameful and self-insufficient was that these articles all stuffed full of data and statistics or data processing. Unlike those articles, he reached conclusions based on logical or mathematical painstaking reasoning, and ultimately could not come up with data to confirm his conclusions.

3 Some Advice from John

Newton was distressed. One day, reading the professional articles from 8 am to 8 pm, he did not remember whether he had dinner or how many meals he had had. He just felt dizzy and swollen, and he was in a bad mood when he saw that his clothes were loosening in the mirror. So, he thought he should talk to the warm-hearted neighbor, Teacher John, to resolve the long-term backlog of depression.

By the way, Newton was now a lecturer in physics at the Cambridge University in London and had not lived in his hometown any longer. He lives in dormitory of the university and John was a colleague in the physics department.

After hearing from Newton, Teacher John expressed sympathy and understanding. He reassured Newton. “Everything was difficult at the beginning. At the beginning, people can’t find the door, and slowly or soon, you will adapt.”

“But there are so many professional articles that you can’t read completely or even half of them. What should I do?”

“Oh, this was easy to handle. You could not study everything. You can just grab the niche that you are interested in, just look at the articles in this niche, and only write articles in this niche. In this way, time was enough.”

“But there are endless essays in such a small area as apple falling.”

[1] There is no PM2.5 in that time; PM250 consists mainly the broken leaves.

“No, no, the topic of apple falling was too big. It can be studied from countless angles such as climate, soil, moisture, pests and diseases. You should focus on more narrow areas; it was impossible to make innovations and breakthroughs without further narrowing down.”

“What the hell was that?”

Seeing the anxiety of this new colleague, after a moment hesitation, John explained patiently.

“You see, for instance, some scholars had studied the problem of obstacles on the ground. Then we can divide the ground obstacles into living things and non-living things; they can be further divided into animals and plants; animals can also be divided into meat- eating and plant-eating (herbivorous) animals and omnivores animals. It can also be divided according to volume or weight; plants can also be divided according to woody and herbs and moss, and also can be divided according to the depth of the root system or the size of stems and leaves. Among them, woody can be further divided according to trees, shrubs, broad leaves, needles, etc.”

“Can the animals be divided according to the length width ratio of their faces?”

“It’s a novel idea. Maybe you can try it. But, it is estimated that animals with different facial length and width ratios may have the same thinking and behavior habits; while animals with the same aspect ratio may also have different thinking and behavior habits.”

“Can we use animal feces as a classification criterion?” An article Newton had just read flashed by in his mind.

“Ah, right. You can study the effects of animal dung, and you can further narrow it down to cow dung or dog shit.”

“In this way, you only need to read less articles. For example, you need read only those related to “shit”. Then, you had time to do further “innovative” research, and of course you can focus on writing those related to “shit”.”

“How can we go deeper and get “innovative findings”? Newton seemed a little curious, but could not figure it out by himself.”

“For example, you specialize in researching the impact of dog shit on the falling and rolling distance of apples. You can delve into the impact of differences between the old and new, dry and wet, and the composition of the shit.”

“Can this be regarded as “innovation”?”

“Of course, remember, if others have not studied it, and you studied it, you can declare that this is your innovation. According to academic terms, other scholars are studying under the “perfect” assumption that no pig poop, dog shit, sheep poop, etc. on the ground; but you had come to a conclusion in case of dog shit on the ground, and you contribute an innovation by “relaxing” the “condition” of “no dog shit”. If others scholars are studying under the “perfect” assumption that only common dog shit on the ground; but you had come to a conclusion in case of wet or dry shit on the ground, and you contribute an innovation by “relax” the “condition” of “common dog shit”. Every time you “relax” one “condition” in the assumption, that will be an innovation.”

“So what if all these have been researched by others?” Newton didn’t seem very self confident yet.

"That's okay. As there is no research draws a conclusion on the weight and volume of shit, so you can "relaxing" the condition of "regardless of shit weight and volume", and then you can contribute "innovation"."

"Is there any more ideas?" Newton thought it was better to prepare more alternative directions or topics.

"As long as you are willing to use your brain, there would always be. For example, do dog shits of different genders had different effects? Do dog shits with colds, fevers, and dog shits with heart disease, diabetes, arthritwas, or tuberculosis had different effects?"

John himself didn't expect to be able to help Newton come up with so many wonderful topics, and he had to admire his own inspiration—this was also the result of hard work everyday and hard work over years—this was what called well-trained!!

"May I study the difference between poodles, pug, husky, golden retriever, Akita; or short haired, long haired, hairless and silky haired dogs?" Newton finally learned how to find the "train of thought". When Newton talked about "hairless dogs", he and John both looked at each other in unison, and there was a strange feeling in their hearts.

"Yes, my brother, you finally understand. If you know what to do, then work hard. I had a few more articles to read."

Newton knew that John had always been very tight in schedule, and it was time to say good bye to John. So he thanked him and leaved quickly.

Newton breathed out when he left John's quarters. He found the conversation with John very rewarding, and the depression cumulated these days had disappeared a lot, and he found a workable direction.—"Yes, just read and write "shit"".

With the way forward, Newton worked hard again and again, finding out the papers about the impact of shit on the fall of apples in the professional journals he bought and he found in the library. Another few months had passed, Newton sorted out the three-foot-thick material. He was pleased to find that there were not many papers on the effect of dogs with different hairs on the rolling distance of apples after falling. In this way, his topic should not be outdated and his writing should be easy to innovate.

Everything was ready except "undertake writing". Newton felt much more relaxed.

This ease reminded him of a problem, and he felt awaking from a dream. Why are he working so hard? In fact, he wanted to publish his "law of gravity." His efforts actually started two years ago. "But what am I doing now?"

All he did now seemed to be reading "shit", writing "shit", and sending "shit" paper.[2] All these were nothing to do with the "law of universal gravitation", because it seemed very difficult to incorporate the "law of gravity" into the "shit paper". If it could not be incorporated into the "shit paper", would the "law of gravity" still not be published? Without publishing the "law of gravity", busy in writing a bunch

[2] By the way, the professional or research articles were no longer called articles, they used to be called as essay, thesis or dissertation. Now the prevailing name is "paper". Newton did not adapt to this term for a long time. Every time he heard or talked about it, he felt like "waste paper" or "toilet paper".

of “shit papers”, even if it was published, what’s the point? Isn’t this contrary to the original intention?

Newton was once again in deep distress.

4 The Data Business of Little Sam

In confusion, Newton decided to return to his hometown during the autumn vacation to help his father and mother for harvesting of crops. He also wanted to drop by Uncle Sam’s orchard by the way. To brush up on the falling of the apple after all these years, maybe it would generate new inspiration.

Newton didn’t know until he returned home that the autumn harvesting was finished. To Newton’s relief, his parents were healthy and fine, but only older than before. So, after having breakfast the next day, Newton went to Uncle Sam’s orchard. Parents had told him that Uncle Sam’s orchard had changed a lot. Newton also wanted to find out as soon as possible what the orchard’s new appearance looks like.

Arriving at the orchard, Newton was stunned by the sight appeared before his eyes. Where was the shadow of the orchard? It was clearly a layered parking lot. This was a huge three-story building with the same footprint as the original orchard. A sign hung on the gate side of the first floor, writing “Sam Red Apple Data Co., Ltd”. Banners were located on the outside of the first, second, and third floors respectively. The text on the banners showed “large sample database”, “big database”, and “cloud database” respectively.

Newton approached the entrance, and saw a young girl sitting inside the door, he asked tentatively, was Uncle Sam “working” here? The girl looked at Newton, not like a bad person, but not like a rich person, so she said coldly that Uncle Sam was not “working” here. Newton asked, “Where are all the original apple trees here?”

The girl obviously didn’t wanted to answer such a tuneless question. Ask directly, “What’s problem are you looking for Uncle Sam?”

“We have a lot of catch up to do, we are old friends.” Newton felt a glimmer of hope.

“Uncle Sam was old and returns to his home in Wales to enjoy his retirement. His son, Little Sam, was now the boss here.”

“It’s also good to meet with Mr. Little Sam, then,” Newton thought. Since he’s here, it’s better to enter the “Orchard” and have a look.

The girl dialed the boss’s internal telephone, explaining Newton’s intentions. There was a clear voice on the other side of the phone. “Okay, yes, let him come up. I happen to had a ten-minute gap.”

The girl took off the phone and repeated it. “Go up to the second floor in front of you, go through the “assumption gate”, take the “regression corridor”, and the president’s office is on your right hand. Just go in. The boss had only ten minutes free for you.”

Both the “assumption gate” and the “regression corridor” are convenient and easy to find and to go, and Newton soon came to the “President Office”. Entering into the president office of Sam, Newton felt really bright and spacious, and the area was three or four times as large as his current dormitory room. No, it might be five or six times as large because Newton accidentally found that there was another inner room connected with this room.

Newton was seated before a tea table, and the secretary brought the tea. Newton simply explained his “friendship” with Uncle Sam, and directly threw his biggest doubts to the new boss. “How did the orchard become like this now?”

“Old man’s old thinking was outdated and he could not seize new opportunities in the new era. Now was the era when data was king. How many data can be produced by apple trees in a year and how much money can they make? So, I did not produce natural apples anymore, and I changed the orchard to produce data directly and massively.” President Little Sam’s words span a lot, and the amount of information was also very large.

“How do you produce data?” Newton felt puzzled.

“You should had seen that my first floor of these workshops produced sample data, of course, large sample data, not small sample data; second floor of these workshops produced big data; third floor of these workshops produced cloud data.”

“How do you think of production data?” Newton didn’t really understand what President Little Sam’s words meant, such as big data or cloud data, etc.

“This was my business sense of smell,” said Little Sam, complacently. “When I graduated from college, I felt that the emphasis on data in the academic field would go to extreme, to the stage of worship and superstition of data. Scholars would become one by one the slaves of data; those logics, reasonings, majors and professions would be all obsolete. Without data and statistics, you could not play scientific research. Only data and statistics are useful; and millions of scholars worldwide are trapped in this data processing game and cannot extricate themselves from such a useless game; and this stage would last a long time.”

“At such a stage, data rules the world. No matter what the data was, whether it was true or false, whether it represents accidental or inevitable, whether it represents right or wrong, and whether the data-based research can solve problems, everything is fine so long as it is a data-based research. Whether the conclusion had practical significance, whether the research question was valuable or layman or even ridiculous, everything was fine as long as there is data in your paper; with data you can travel all over the world, and without data, you cannot move one step further. So I declined the teaching position at the university after graduation; I could hardly wait to start my own data company.”

Newton nodded again and again. Although he had not fully understood Sam Red Apple Data’s business, Newton had truly felt the brilliance of Little Sam’s career from his short “speech”. Considering that he couldn’t even publish a small scientific finding “law of universal gravitation”, Newton felt that he was really too incompetent and too humble.

It seems difficult for Little Sam to stop halfway in telling his successful story.

“Data was divided into fields, so which field should I choose to produce data for? I chose to produce data on the falling of Apple. Now it is proved that my choice was very correct. Apple falling are now a persistent hot spot in academic research. But in fact, this did not fully represent my personal vision and ability. It should also be attributed to the ancestral virtue or ancestor’s inheritance. Because this was originally an apple orchard, the falling of apples are the first and easiest for me to think of.”

“But the apple trees are gone. Where did the data on the falling of the apples come from?” Newton apparently failed to grasp the points introduced by President Little Sam just now. Mr. Little Sam was talking about producing data, not harvesting data.

Facing such a slow and persistent “layman” like Newton, Sam had to patiently explain.

“Our data production includes data collection and data processing. I cut down and cleared all the apple trees and replaced them with artificial intelligence desktops. We are a data center, and many apple orchards over the country are our members; Each artificial intelligence desktop is responsible for collecting data on the Apple’s falling in the member orchard in one aspect, for the apple itself, like size, weight, color, sugar content, etc.; for the natural environment, like geographical location, ground fluctuation, air temperature, humidity, soil characteristics; for the social and cultural environment, like population status, education level, income level, pet ownership rate, etc. of course, the most important is the Apple’s falling data, Including the number, time, weather, altitude and ground obstacles.”

“Our business not only creates revenue and profits for ourselves, but also contributes to win-win results to various parties. For example, our upstream, our orchard members can not only get the harvest and income of apple, but also get additional income from data providing; For example, our downstream, i.e. universities, scientific research institutes and other scientific research institutions, get systematic and timely Apple falling data from us, and promote the research of various scientific research topics accordingly. Not to mention, our company and partners have created a lot of employment opportunities for the society. Of course, our business is growing rapidly, with an average annual growth rate of 300% in the past three years. At present, our business directly and indirectly provides 100000 jobs for the society.”

“Oh, that’s it.” Newton finally seemed to understand the core products and processes of Sam Red Apple Data Co., Ltd. Unexpectedly, Uncle Sam’s orchard became a base for academic research in the hands of his visionary and prosperous offspring. Newton suddenly realized that the company might be very important to himself.

Time flies. Newton’s visit has lasted for 20 minutes. The Secretary opened the door to remind that the next group of visitors had been waiting downstairs. President Sam looked at his watch and politely suggested, “sorry, let’s talk for another two minutes.” A wit flash by in Newton’s mind and he asked a key question, “If I study the effect of “shit” on the fall and rolling of apple and need your data, how much need I pay for the data?”

President Little Sam paused for a moment. “Considering our neighbor relationship, according to internal price calculations, large sample data costs 10,000 pounds, big data costs 50,000 pounds, cloud data costs 100,000 pounds. But you should understand that when others writing an article based on big data or cloud data, you can hardly publish an article with sample data. Anyway, based on enough data, your paper proceeds without hindrance; based on insufficient data, your paper cannot proceed even just one step!"

The price was so high, it was totally beyond Newton's expectation. Newton's one-year salary plus all piecemeal subsidies, the total amount was only 10,000 pounds. It turned out that he might not afford to "write" a "dog shit" article at all.

Coming out of Sam Red Apple Data Co., Ltd, Newton was more determined to publish the "law of gravity", because he has no other choice or there was no other way to go. But there was another question that made him wonder:

How could other teachers "afford" to write the "shit" paper?

5 Further Advice from John

Newton felt that he should share his new discoveries with John and consult with him. Once again after dinner, Newton carefully knocked on John's door. Newton felt more acquainted with John and hence more relaxed. So he came straight to the point, he told John his new discoveries in Sam Red Apple Data Co., Ltd. and his doubts.

John obviously had the relevant experience. He did not feel new to Newton's findings and doubts, but he thinks Newton deserves more sympathy and help. Because in the current (then) academic environment, people who are so down-to-earth and do not seek fame and wealth to do scientific research are really too rare, and should belong to "endangered species" which worth protecting and saving!

He told Newton, "To do scientific research, you must apply for a research fund. Without a research fund, any findings or discoveries are not counted and recognized, you can achieve nothing except your empty hands after all the efforts you put on the scientific research!"

John's words hit the nail on the head and made Newton's stare open mouthed for half a minute.

Are you serious? Can it not be counted or recognized if I spend my own money to do the scientific research? Only those researches consume the money from various scientific research funds or from the government are counted and recognized. Was this the case? How could there be such a strange thing in the world!"

"It is absolutely true! And the more you consume the money from various scientific research funds, the better your scientific research is rated or evaluated! Your research performance evaluation had nothing to do with whether you had scientific discoveries, whether your scientific discoveries had actual value, and whether they are big or small discoveries! Of course, you can boast or exaggerate your findings or contributions when the findings are unimportant or the results have no much actual value."

"Anyway, if you get the same findings as other scholars, you pay for it yourself, other scholars pay for it by the research funds, or you pay for it even also by the research funds, but you spend less than others, then their scientific research will be rated better than you."

"Even," John felt that he still didn't express it thoroughly. "Even if you have important research discoveries, and others have no important discoveries or no meaningful discoveries at all, but you just spent your own money and did not spend the money from the research funds or government, and the others spent the money from the research funds or government, then their researches are evaluated better than yours."

"But why they adopt and enforce such bad rules? It is too ridiculous, and it is too unfair!" Newton felt that the evaluation rule is too unacceptable.

"The evaluation rule and its unfairness are not your business, and are all beyond what you can control! As a matter of fact, they always stress that this rule (the more you spend government money, the more you are right and glory) should be further and continuously reinforced rather than replaced or even a little adjusted."

"You just need to know that the evaluation rules are like this. The rest is to spare no effort or do everything you can to apply for the fund."

"Trust me, if you could not get research funds, no matter how big or how many your findings are, they are just regarded as a shit! By the way, anonymous reviews of professional journals also "recognize research funds", and usually give priority to papers supported by research funds." So, applying for the fund is the most critical step for your research.

Newton saw John with a sincere expression on his face and knew that John was helping himself seriously. He nodded again and again and repeated, "I will try, I will do it."

"When you are success in applying scientific research funds, you then have money to buy "shit" data, and then you can afford to write "shit" papers, and it would be easy to publish them further. In this way, everything goes naturally and smoothly."

Mr. John made the rules of scientific research clear and thorough. However, Newton's problem was still not solved. Since the law of universal gravity had already been derived at his own expense, how can he ask the government for money again?

John meditated for a while. "Well, this is not important. The key is that the law of gravity as a topic does not conform to the current research "paradigm" and it is not easy to get success in the competition of application. It is necessary to rely on "shit" topics to apply for scientific research funds."

It seemed that because it was not easy to get research fund, this law of gravity was difficult to publish, and it was not recognized. However, Newton still believes that this was the most important discovery in the field of science (at that time), and it should be published, even if it was not recognized for the time being. It was even more difficult for Newton to generate interest in applying for funding and publishing "shit" paper.

Thinking about it, Newton decided to try his best to publish the law of gravity. In this way, after a long journey, Newton returned to the starting point and did the same thing as before, that was, sending the article to one professional journal after another.

6 The Reviewing Comments from Professional Journals

Unlike in the past, due to the further improvement of the professional journal review system, submissions[3] could often be responded after two or three months. Although what Newton had received so far were all rejections, there was still a glimmer of hope.

A professional journal responded with the following review comments:

"This research is too backward. We have already studied the problem after the fall of apple in our field, but this article is still limited to the study of the cause of the fall. It is recommended that the author read the papers of the current professional journal and learn about the latest domestic and international cutting-edge research." So, Newton spent a few more months, deleted or simplified the elaboration of the law of gravity, and added an analysis of the possible route and distance of the apple after it fell.

A few months later, another professional journal responded with the following review comments:

"This study is too lacking in specific local characteristics. We have studied the relationship between apple fall and various climates, soils, moisture, and farming habits in our country, and the research in this article was limited to the study of general causes of apple fall. It is recommended that author read more the papers in the current professional journal to learn about the latest research in this field." So, Newton spent a few more months properly adding an analysis of the various possible causes of the apple fall under British climate and farming habits.

A few months later, another professional journal responded with the following review comments:

"This research is too backward and inhuman. The role of human activities are mentioned nothing in the falling of apple, which is like the research of 30 years ago. It is recommended that the author read more papers in current journals to learn the latest international and domestic issues research and cutting-edge research." So, Newton spent a few more months adding an analysis of the possible effects of human behavior on apple fall, including the effects of people inhaling oxygen, exhaling carbon dioxide, and children throwing stones, etc.

A few months later, another professional journal responded with the following review comments:

"This research is too irregular. The author is simply a layman in academic research. Where are the research assumptions, where is the research design, where is the sample data? Where are the control variables? Why do you not get the model without statistical regression? This research is far from the standard of publishing! The author is advised to read more papers in current journals and learn basic research routines first. "So, Newton spent several months to revise his article and tried to move close to the popular "routines".

[3] For unknown reasons, as a prevailing expression, sending articles now was called submitting manuscripts.

A few months later, another professional journal responded with the following review comments:

"The literature review of this research is too incomplete. The cited literatures are too old, and many of them are from decades ago. Important papers on the falling of apple published in recent years, such as AB authoritative papers and CD authoritative papers, are not mentioned. It is recommended that authors read more papers in current journals to understand the latest international and domestic research, especially those cutting-edge research." So, Newton spent several months intensively reading the "latest research", especially those "authoritative" papers, and replenish a lot of new and "authoritative" papers in the review of the literature in his papers.

In this way, struggled again and again and again among the review comments, working hard for another two years, Newton did not make even one step further. Now, Newton really felt a little exhausted. At the same time, Newton seemed to understand the truth, no matter how he modified or improved his thesis, the reviewer could always give a "reason" to "kill" his paper. In fact, what matters was the intention of the reviewer, not his paper; and the reviewer's intention on the law of gravity was determined before reading his thesis, because his law of gravity and the "shit" paper were essentially different. The review expert could not accept or would not understand his law of gravity.

Gradually, Newton understood that as time went by, the hope to publish the law of gravity was not getting bigger, it was getting slimmer. Because the professional journal's review system was becoming more and more stringent; the review experts had been selected as review experts just because they had published large number of "shit papers" in accordance with the "rules." Under the "anonymous and stringent review" of these reviewers, that was, "closely monitored by stealth", it was almost impossible for a paper that did not meet the "shit" paradigm to slip through the net by chance and got published.

Newton thought, since this was not the way, could he find another way to solve the problem?

7 The Reviewing Comments from Scientific Research Funds

Later, another colleague quietly told Newton that in fact, some professional journals could publish for a fee; and some scholars pay this fee by using the research funds. This fee was called the "page fee" and could alleviate the poverty of the professional journal. Newton felt a glimmer of hope at once. Since the regular path was not accessible, this informal approach was worth to try. Obviously, the key to this method was "money."

Unfortunately, what Newton lacked was money. He must be one of the poorest scholars if not the poorest one.

Newton had reached the age of marriage. He was definitely a smart and handsome young man and had stable and decent work, but attracted to no girl. In addition to Newton's natural weakness in talk with girls, a key reason was that Newton's salary was meager and there was no additional "scientific research funds". In a city like London with high prices and high consumption, especially the rocketing home prices, Newton's meager salary was merely enough to get three meals a day. He could neither buy a house nor support his family. Where was a girl willing to marry the poor Newton, without a house or a car and maybe even possible freezing or hunger one day in the future?

Of course, this was just the "vulgar thoughts" of outsiders. What in Newton's mind at that time was just how to publish his law of gravitation. As for whether any girl was interested, Newton did not mind at all. Now, the key to publish was "money".

Thinking back and forth, there was no way out, because it was impossible for Newton to save the money for professional journal pages by not eating. It seems that the only way left is to apply for scientific research funds. As well known among scholars, he must give up the law of gravitation if he wanted to apply the scientific research funds.

However, Newton still did not willing to apply for research funds under the topic of "shit" and felt that his conscience would not be able to handle that, which was a waste of state money. Therefore, he decided to apply for scientific research funds in the name of the further research or applied research of the law of gravity.

In Newton's time, university teachers could apply for a total of three "national" scientific research funds. There were also some "local" funds. But Newton judged that those local funds were more likely unable to understand the law of gravity; so he focused on the three "national" scientific research funds. Because the application time of these three projects was almost evenly spaced in a year, with Newton's diligence and energy, he "turns into battle" one by one every year in these three "national" scientific research funds.

Although Newton was full of confidence, but the reality was very cruel. All the hard work and efforts were finally exchanged for nothing after "fought" three national funds one by one over eight years, and Newton received a wide range of negative review opinions, and even more terrible than the rejections from the professional journals, there are quite a lot opinions ambiguous and vague or unintelligible. Some opinions with clear meanings are as follows.

"The applicants obviously overestimate his abilities. Comparing his research results with those of Copernicus and Kepler shows that the applicant lacks the humility and prudence, which are the basic research qualities that a scholar should have."

"The subject of the application is too large to have research feasibility. The applicant is only a lecturer, had no influence in the academic field, and could not mobilize more research resources; even highly respected professors had difficulties in doing such a subject."

"As a lecturer, it is commendable to apply for such a major topic, but obviously the applicant did not had the corresponding ability. The content of the application is mostly self-exaggerating. It is recommended that the applicant do what he could and choose an appropriate topic to apply."

"The topic is outdated, vague and lacks innovation."

"Applicants and their research team are weak."

"The content of the application is mostly self-judgment and lacks objective basis."

……

In this way, from 1671 to 1678, Newton's best and most precious years spent on the road of applying for scientific research funds, a full eight years! Every year during the eight years, Newton had been the denominator in the application for scientific research funding. Various scientific research funding organizations did not respond, or replied similarly like, although your application was not bad, but chose the best from the best, we had to regret you and welcome you to apply again next year. In the end, all the other things were delayed, such as scientific research and teaching, as well as his personal marriage!

During these eight years, Newton always believed that he would win in the end as long as he keeps trying. Therefore, he repeatedly fought and repeatedly defeated; repeatedly defeated and repeatedly fought; never gave up when facing the difficulties and failures for eight years. However, after eight years, he understood finally that he was bound to failure, because the anonymous review "circle" in the "gateway" of the professional journal and the anonymous review "circle" in the "gateway" of scientific research funds are the same "circle". Since he had spent ten years proving that this "circle" would not "allow" his papers to be published, how could this "circle" "allow" his research to be funded?

To make things even worse, as time passed by, other scholars upgraded as associate professors and professors, while Newton was still a "lecturer", and the "relative qualifications" for applying for scientific research funds and publishing papers decreased year after year. Put it another way, for Newton, the difficulty of applying for scientific research funds and publishing papers had increased year after year.

But Newton, a farmer by nature, was born with a "bullishness". More than two decades passed, he still insists on publishing his law of gravity. It was a period of painful meditation and mental suffering. Newton was still thinking: since this road was not feasible, could he find another way to solve the problem?

8 The Alchemy

Newton, who by nature likes scientific exploration, did not like doing business and was not good at stock speculating. But, neither business nor stocks, where could he make money to publish his law of gravity? Newton heard that gold was very valuable. He then thought that if he had enough gold in his hands, he wouldn't had to worry about money. However, why was gold valuable? Just because it is extremely scarce and not easy to obtain. So how could you get gold casually or easily?

One day, Newton accidentally read from a book that an oriental wisdom holds that Tai Chi gives birth to the two opposing forces, yin and yang (negative and positive), the yin and yang give birth to four phenomena, the four phenomena generate the eight trigrams, and the eight trigrams create everything in the world. This means that everything in the world, including earth, soil, and gold, comes from the same source. Since sources of soil and gold are the same, could gold be produced from soil in some way?

Newton, who was good at studying and exploring, quickly found out from the tremendous amount of literature materials that the ancient sages had similar ideas. Many people had been working on such studies in mysterious east, but they had not succeeded until now (then). The method of producing or turning into gold from other substances had never been found, but it had already been named as "Alchemy."

Newton, who was driven mad, couldn't wait to plunge himself into the exploration of "alchemy". In the following years, Newton read over all the alchemy literature materials, consulted various seemingly true and fake alchemists, visiting and discussing time and time again, and racked his brains to think and calculate, and dreamed every day the final success to turn the soil into gold and the golden mountain surround him. But the earth's soil could not be changed into gold at the end; on the contrary, he himself seemed to be tortured as incoherent, delusional, as well as auditory and visual hallucinations.

People around noticed Newton's abnormal behavior. Seeing his come and go independently, deliberation and devotion every day, people call him "mad cow" affectionately; others call him "mad old man" more affectionately.

9 The Models of Scientific Research

On Wednesday, the Department of Physics held a conference to recognize advanced models, stipulating that any in-service faculty must attend without special circumstances. At the meeting, three "excellent youth" were promoted to professors and doctoral supervisors because of their outstanding research performances.

The first excellent youth studied the relationship between the sun and the earth. After ten years of persistent observation and recording of the data, the new professor found that except for days with cloudy rain and severe PM250 pollution, the sun always rises from the east and sets in the west, and revolves around the earth every day. Although only ten observation points were arranged evenly in the university campus, he reasoned that the same results could be observed at every observation corner of the world. He further inferred that, on those cloudy days or days with severe PM250 pollution, although the sun could not be seen, after complex multivariate statistical regression estimation and inference, the sun should also rise from the east and set in the west. The conclusion then is obvious, big data proved that the sun revolved around the earth, which manifested that Copernicus was wrong. This later research finding had been enthusiastically sought after by many in the physics community, and its research results had been ranked among the best in the “citation rate” in major scientific journals for two consecutive years.

The second excellent youth studied the relationship between apple fall and apple leaf fall. In order to obtain big data, the new professor mobilized all his relatives and friends. For five years, he persistently observed and recorded the relationship between apple fall and leaf fall in many orchards in many villages in his hometown. The basic conclusion is that the apple leaf drop cycle was longer than the apple drop cycle. He also had one of the most proud discoveries, that was, the leaves of apple trees with diseases and pests do not necessarily fall earlier, or even later. His explanation for this was that after the leaves had been eaten by the insects, the leaves are lighter and less likely to fall. It was said that the “citation rate” of this study in various major professional journals had skyrocketed. Newton felt a little embarrassed when he heard it, and whispered to a colleague next to him, was this “discovery” different from “common sense”? The teacher apparently knew that Newton did not understand “academic research”, and dismissed Newton with a disdainful look, which could be regarded as a reply.

The third excellent youth spent three years studying the relationship between the length of women’s skirts and the apple fall. This research had been widely praised by scholars in the physics for its humanity and research difficulty. Humanization was obvious, at least it could be seen that this excellent youth was concerned about women; compared with Copernicus squatting in the attic of the church one night after another to observe the cold celestial bodies, this was a huge progress. The difficulty of research could be understood. In the past three years, the excellent youth insisted on going to London’s busiest street every day during the summer season, at least in June, July and August, to observe closely the length of the skirt worn by every passing woman. Then during the autumn season, at least in August, September and October, the excellent youth insisted on going every day to the apple orchard nearest the city of London to observe the apple’s falling. The most commendable thing was that in the hot August, the excellent youth observed the length of past ladies’ skirts on the street every day during the day, and went to the orchard at night to observe the falling of apples, and slept only 2–3 hours a day. This spirit of research and action had moved many old physics experts and set an example for newcomers in the physics.

The third excellent youth was the youngest, but obviously his experience had caused the most repercussions. As soon as he stepped down from the podium, many young and even older scholars chased him for academic consultation and communication. Of course, many of them are master’s and doctoral students, who expressed that they were deeply inspired by his research and were willing to follow him and instructed by him to do physics research.

The consulting and discussing were very hot. “Could I study the relationship between the thickness of women’s skirts and the falling of apples?” “Of course, but the thickness of women’s skirts is not easy to observe, the data collection would be more difficult for this study.” “Could I study the relationship between the proportion of women who wear skirts and the falling of apples?” “Because some ladies do not wear skirts?” “Of course, yes, the data collection for this study is less difficult and more feasible.” “Excuse me, could I study the relationship between women’s hair length and the falling of apples?" "Yes, this is a good idea, and innovative as well." "Could I study the relationship between certain characteristic of men and apple fall, such as the size of their noses or faces?" A small girl finally seized the opportunity to interject. "Yes, very good, this idea was better, more innovative, and also had the social significance of gender equality." The new professor praised the little girl.

Seeing so many young students interested in scientific research, the three young and new professors smiled with satisfaction, relief, and happiness.

After the department conference, Newton came back to his room, and suddenly felt that he was old. Yes, since his graduation from Trinity College, Cambridge University, he had been at the position for more than 20 years. He talked to himself in the mirror, "you have always claimed to love physics and love scientific research, but what have you achieved? And what are you doing every day now? You have always been proud that the law of gravity could be compared with Copernicus' discoveries, but you failed to publish it after so many years; what more ridiculous is the "Heliocentric Theory" of Copernicus was easily proved to be "wrong" by young scholars, and the key was that it is a data based conclusion!"

Newton's eyes were dull, thinking again and again, puzzled. One conclusion was clear, that was, "I am stupid—not ordinarily stupid, but extremely stupid."

10 The First Ray of Dawn

As the wheel of seasons turning, countless number of cycles of spring, summer, autumn and winter passed by, the time arrived at 1685. After repeated revisions, a big good news came to Newton, a colleague from his department helped him by persuading a professional journal agreed to give a special consideration to Newton's law of gravity for publishing. Due to the revision again and again and the addition of new literature, the length of the paper increased from the original 8000 words to 38,000 words now.

Newton got the news at ten o'clock in the evening and was so excited that he burst into tears. "I'm hit, I'm hit, I got it, I got it—" Newton danced and shouted in the little room he lived for decades. Newton's shout broke the tranquility of the night and disturbed the neighbors. Someone upstairs protested, "What time is it? What's your name, you silly cow?" Newton stopped shouting, but he was not angry to the rude words from upstairs, still happy, just felt his heart was sweet like he had eaten honey.

This professional journal agency was located not far from the university, about 10 miles away. After having breakfast the next day, Newton couldn't wait to go straight to the professional journal agency. He wanted to finalize the details of the paper as soon as possible, and also wanted to thank the journal agency face to face for their kindly consideration.

Newton walked wildly for three hours, and finally came to the professional journal agency. The editors in the journal agency office got a surprise.

"We originally thought that the lecturer author was a young man. From the perspective of encouraging young scholars, we were willing to publish this paper, which is dissatisfied with two anonymous reviewers. But you look not so young. Are you sure you are the lecturer author?" The editors said that in this case, they needed to reconsider this paper and their decision.

The idea of the professional journal was completely beyond the expectation of Newton. In a hurry, Newton said that he was indeed a young man when he completed the law of gravity, but "grew" old during the submission process. Newton was speaking in complete honesty, but the editors were amused by Newton's "street smart" and "humor", saying that they would consider Newton's "special circumstances."

Back from the professional journal agency, Newton felt tired. Looking at himself in the mirror, with so many wrinkles on his face and so much white hair on his head, Newton felt more tired than he was sitting under Uncle Sam's apple tree when he was a child. Perhaps because of his age, Newton felt he was not so eagerly looking forward to a positive reply from the professional journal he just visited.

As the saying goes, heaven never seals off all the exits. Mr. John had an old classmate who went to a publishing house as editor in chief. John introduced Newton's law of gravity to him, the editor-in-chief was willing to talk with Newton about the publication of his work. To make a long story short, after many discussions, the editor-in-chief finally agreed to dispense with the anonymous review process and publish a book on astrophysics written by Newton.

Both John and Newton knew that the publication of the book was not helpful for Newton to upgrade his professional title (from lecturer to professor), because the University of Cambridge follows the academic convention: only papers were recogonized and counted; books were not recogonized or counted. This "rule" is even more ridiculous than the other "rules". A book is much more difficult to write, did not mention that a book with many innovative discoveries like the law of gravity. But rules were rules. Submission experience in past decades had proved that Newton's law of gravity was impossible to pass anonymous reviews and be published in professional journals. Therefore, this was a rare chance for Newton. Otherwise, the law of gravity might be gone with wind after Newton passed away.

Newton finally found an exit to display his talent and all the scientific discoveries cumulated over decades: he won the chance to write a book on physics of his own without various harasses from various reviewers. Newton waited for such a moment from young to old! This was the fortunate part out of all the misfortunes for Newton as well as for the law of gravity, for the physics or the whole science, for all of us, because science is always general or extensive benefits from heaven, the discoveries of Newton could benefit the development of physics as well as every common person in the world like us.

In the next two years, Newton was completely focused and devoted himself to the writing of "Mathematical Principles of Natural Philosophy". With the help of John and the kind editor, Newton's Principles were published as promised in 1687.

During the first few decades after publication, "Mathematical Principles of Natural Philosophy" did not receive much substantial response, and the citation rate in academic journals was pitifully low. According to (at that time) popular academic evaluation standards, books were basically not counted as academic discoveries, and only papers in professional journals, especially "advanced" journals, i.e. journals rated as A or B, are considered as academic discoveries. Newton hence had been regarded as no academic discoveries, and naturally he did not upgrade to a senior title until he retired.

Over time, those papers with high citations had been gradually forgotten, more and more people realized that Newton's law of universal gravitation was the true discovery and true progress of physics as well as the whole science.

Epilogue

A sound of alarm clock awoke me from a dream, and pulled me back from Britain in the seventeenth century. I sat up and calmed down, wondering why I had such a complicated and depressed dream. I am a person engaged in science of finance; and never do research in natural science or physics. How could I had such a dream?

Looking out the window, the rising sun had illuminated the world; people with various dreams are already on their roads; the streets and alleys that had rested for a night are crowded and bustling again. It turned out that I had a nightmare; and all the torments and struggles in the nightmare have gone and disappeared along with the sunrising.

As far as I know, Newton didn't suffer from the "torment" of publishing paper. Anyway, now, more than 300 years have passed, society must have progressed, and people's consciousness must have improved. It must be commonly recognized that scientific discoveries were the ladder of human development. Is there anyone in our society indeed like to hinder the publication and dissemination of real scientific discoveries?

I am so proud of the progress of human society, I am so proud of today's publication of scientific discoveries, and I am so proud of today's granting of scientific research funds! Thinking of those, a smile of satisfaction, relief, and happiness appeared on my face.

Thinking Questions

Witnessing the fall of apple, Newton deduced the law of gravitation, and some scholars thought of other problems, even endless problems. What is scientific thinking? What is the fundamental difference between non-scientific and scientific thinking? If possible, consider this problem based on the situation of finance.

Comparing with Newton's law of universal gravitation, why those papers on topics like "the relationship between the length of women's skirt and the apple fall" are more welcomed by review experts and easy to pass the anonymous reviews and be published in the professional journals? If possible, consider this problem based on the situation of finance.

Further Discussion: Why Financial Theory Stagnates Over Decades?

Contents

After the World War II, with the recovery and development of the economies of main countries, the research of financial theory had also boomed over about 30 years and obtained a lot of theoretical discoveries. These theoretical discoveries not only solve relevant financial problems and promote the understanding of relevant issues, but also promote financial research from experience-based level to the height of science and theory, becoming milestones one after another in the development of financial theory.

For various reasons, in recent decades since 1980s, the financial research and findings went back to experience-based or even common sense level, such as information asymmetry theory, behavioral finance theory, pecking order theory, etc. Perhaps most people have not realized that those new and popular "theories"are just well known common senses being renamed; and the new stylized financial research does not intend originally to solve decision problems in finance, but just to describe various and endless phenomena in finance and to make some explanations based on subjective guesses.

Z. Zhang, *Fundamental Problems and Solutions in Finance*,
Contributions to Finance and Accounting,
https://doi.org/10.1007/978-981-19-8269-9_14

Related to the common sense level in research, after decades of research in full swing, the remained financial problems have basically not been further solved, and there is even no progress in the relevant understanding. However, the problem is, why is financial research has been stayed at the common sense level or even lower than common sense level? The reasons behind this phenomenon deserve our deep thinking and reflection.

1 Financial Theory: 1950–1980

The major theories of finance, including portfolio theory (1952), MM models I and II (1958 and 1963), capital asset pricing model (1964) and Black–Scholes model (1973), were established before 1980. This section briefly reviews the finding process of these financial theories, hoping that readers can gain a deeper understanding of financial theory from the 30 years glory.

1.1 Portfolio Theory

Harry Markowitz was born in a Jewish family in Chicago in 1927. During high school, Markowitz developed a strong interest in physics and philosophy. After receiving his bachelor's degree from the University of Chicago, Markowitz chose economics to continue his postgraduate studies at the University of Chicago. During this period, he was guided by some economists, such as Milton Friedman, Tjalling Koopmans, Jacob Marschak and Leonard Savage.

Markowitz chose the application of mathematical methods to analyze stock investment decisions as the research direction of his dissertation, which was affirmed and encouraged by his mentor Jacob Marschak. In 1952, Markowitz published his research findings in a paper. For the first time, he accurately defined return and risk as the mean and variance of expected return, and introduced accurate mathematical methods into the research of investment (stock) portfolio. Because of the originality of basic concepts and mathematical applications, this paper was later regarded as the first work of modern financial theory, because it laid a foundation for finance and financial analysis, that is, to comprehensively consider risks and returns.

In 1955, Markowitz passed the defense of his dissertation on portfolio theory, and finally obtained his doctorate in economics and graduated from the University of Chicago. Interestingly, due to the "too strong" innovation of the paper, the defense process was also questioned by Milton Friedman and other famous economists. In 1959, Markowitz systematically collated his own research findings, completed the book writing of "Portfolio Selection: Efficient Diversification of Investments", and provided a systematic and complete portfolio theory for the financial field.

Markowitz's research involves the optimization of portfolio, and obtains the Markowitz efficient frontier. Through the trade-off analysis of mean variance, the calculation method for solving the optimal portfolio problem is derived. According to Markowitz's method, the optimal portfolio can be determined by comprehensively considering the risk, return of securities and the risk preference of investors. In the process of studying this method, Markowitz found that through the portfolio optimization, the expected return can be improved with the same or even lower risk. This is known as the "only free lunch" in the financial market.

Portfolio theory laid the basic logic of finance and financial theory, and it was also the foundation of capital asset pricing model. In 1990, Markowitz won the Nobel Prize for his outstanding contribution to portfolio theory and finance.

1.2 MM Model

Like Markowitz, Merton Miller was born in a Jewish family in Boston in 1923. In 1952, Miller received his doctorate in economics from Johns Hopkins University, and then went to the London School of economics for his first job as a university assistant.

Unlike Markowitz and Miller, Franco Modigliani's experience is more legendary. He was born in 1918 in a Jewish family in Rome, Italy. In 1935, Modigliani entered the University of Rome by skipping the senior high school. His family hoped that he would continue his father's career to study medicine, but he was bored or even afraid of it, so he chose the law as his major that he thought was the most far away from medicine. However, Modigliani was not interested in law. He completed his undergraduate study in the way of "skipping classes and cramming for help" and obtained a bachelor's degree in law from the University of Rome in 1939. Later, because of his Jewish background and anti-fascist ideas, Modigliani left Italy and went to Paris; Soon, he went to the United States. While making a living, Modigliani had a strong interest in economics. In 1942, Modigliani became a lecturer in economics and statistics at Bard College.

In 1952, Modigliani became a professor of economics and industrial management at Carnegie Mellon University (then known as Carnegie Institute of Technology). There, he met Miller, and their regular discussion and cooperation had produced interesting findings that have an important impact on financial science. Among which, the most important fandings are MM models I and II on capital structure.

At first, Modigliani and Miller were arranged to teach "financial management" courses. Due to the lack of teaching and practical experience in this field, Modigliani and Miller discussed various issues related to corporate finance together. They found that the explanations on financing and capital structure in relevant works were mostly vague and confusing, so they decided to sort out a clue through in-depth thinking and discussion.

In 1958, Modigliani and Miller published their findings in a paper, which is the famous MM model I today. MM model I proves through the principle of no arbitrage market equilibrium that, in the case of no corporate income tax and bankruptcy problems, the total value of the company has no relationship with the proportion of debt capital in total capital of the company, that is, the amount of debt (financial leverage) the company uses does not affect the value of the company and the weighted average capital cost of the company.

The discovery of Modigliani and Miller casts light and hope to the chaotic study of capital structure at that time. Financial scholars feel deeply inspired. However, their conclusion is derived on the premise that the company has no income tax and bankruptcy problems. Because this premise is far from the reality, their conclusions do not represent the final solution of the optimal capital structure problem. Of course, their findings make the thinking and direction of capital structure research more clear.

In 1963, Modigliani and Miller made another breakthrough and published their second paper on capital structure, which is today's famous MM model II. In this paper, Modigliani and Miller took corporate income tax into account and improved MM model I. Unfortunately, due to the fact that the bankruptcy risk was still not considered quantitatively, MM model II could not take the bankruptcy cost into account, and then could not solved the problem of optimal capital structure.

Nevertheless, Modigliani and Miller established the basic framework and direction for quantitative analyze and solve the problem of capital structure, which is the most important findings of capital structure research so far, and also created a precedent for the application of non-arbitrage equilibrium analysis in financial analysis. Modigliani and Miller won the Nobel Prize in 1985 and 1990 respectively.

1.3 Capital Asset Pricing Model

William Sharpe was born in Boston, Massachusetts in 1934. After 1940, he moved to Texas with his parents and then to California. In 1951, Sharpe entered the University of California, Berkeley. In Berkeley, Sharpe originally planned to study medicine, but one year later, he found that he was not interested in medicine, so Sharpe transferred to the University of California, Los Angeles to study business management. During this period, the microeconomic theory aroused his great interest.

Sharpe joined the army after graduation. A short time later, he joined the Rand Corporation in 1956. At the same time, Sharpe studied for a Ph.D. at the University of California, Los Angeles. Following the guidance of his tutor, Sharpe read Markowitz's theory and was deeply attracted. Fortunately, Markowitz, who founded the portfolio theory at that time, also happened to work for Rand Corporation. Sharpe then had chances to discuss many issues of interest with Markowitz. In fact, Sharpe's doctoral thesis was completed after listening to Markowitz's suggestions and guidance, and he successfully received his doctoral degree in 1961.

In 1961, Sharpe moved to Seattle to teach at the Foster School of Business, University of Washington. The stability of life and work enables Sharpe to rethink further on the basis of his doctoral thesis. By the autumn of 1961, Sharpe had a new understanding of the factors affecting securities returns. In january1962, Sharpe reported this finding at the University of Chicago. Soon thereafter, he wrote and submitted a paper based on the finding. Due to the negative opinions of the reviewers of the journal and the reorganization of the editorial team, the publication was delayed until September 1964. This is today's famous capital asset pricing model (CAPM).

The capital asset pricing model describes the quantitative relationship between risk and return for the first time, which is related to the determination of capital cost, investment return and discount rate. It has vast application potential in Securities and industrial investment decision-making, financing and risk management of various companies, and occupies a core position in financial theory. Sharpe won the 1990 Nobel Prize for this contribution.

When it comes to the capital asset pricing model, what needs to mention is that Jack Treynor (1961, 1962), John Lintner (1965) and Jan Mossin (1966) have made similar discoveries in the first half of 1960s. Although Sharpe's paper were not published smoothly, it was finally published fortunately. However, Treynor's two papers in 1961 and 1962 were not published, and of course, he lost the opportunity to win awards for the similar contribution.

1.4 Black–Scholes Model

Fischer Black was born in Washington, D.C., in 1938. In 1964, Black received his doctorate in Applied Mathematics from Harvard University. Shortly afterwards, Black joined Arthur D. Little, where he met Treynor, who discovered the capital asset pricing model. Influenced by Treynor, Black became very interested in capital asset pricing and finance.

In 1971, Black taught at the University of Chicago; Later, he joined MIT Sloan School of Management. In 1984, Robert Rubin, the business director of Goldman Sachs, hoped that Black would join him. Black also felt that Goldman Sachs was more conducive to his creativity than universities, so he joined Goldman Sachs at the invitation of Robert Rubin.

Due to the influence of Treynor, Black shifted his interest and efforts to financial issues. Therefore, Black's initial research on financial problems was to test the consistency between the capital asset pricing model and the actual data, and try to obtain new findings. It is also in this kind of research that Black met his later research partner Myron Scholes.

Scholes was born in 1941 in Ontario, Canada. Influenced by his family, Scholes got involved in business and investment affairs very early. When he was in high school, his family opened a special stock account for him; so, he got acquaintance to investment quite young.

In 1962, Scholes completed his bachelor's degree in economics at McMaster University in Hamilton, his hometown. During his college years, one of his teachers introduced the discoveries of George Stigler and Milton Friedman, both were famous economists of the University of Chicago, who later won the Nobel Prize in economics in 1982 and 1976 respectively. Scholes applied for graduate school at the University of Chicago and finally succeeded.

In 1964, Scholes received an MBA from the University of Chicago. Subsequently, he successfully applied for the doctor's degree of Eugene Fama and Merton Miller (one of the founders of MM model), and obtained the doctor's degree through thesis defense in 1969. Fortunately, Scholes stayed at the University of Chicago and became colleagues with Michael Jensen, Richard Roll and others famous economists. He could continue to consult Eugene Fama and Merton Miller on financial issues.

More fortunately, Scholes met Black, who also taught at the University of Chicago at that time, and both of them were interested in option pricing. Option pricing is one of the most challenging problems in finance over very long time. It has attracted the attention and efforts of economists and mathematicians in the past century, including Paul Anthony Samuelson (who won the Nobel Prize in 1970 and was at MIT). From 1968 to 1971, Black and Scholes were fascinated by the problem of option pricing, and devoted all available time to thinking, discussing and researching on option pricing. At the same time, one of Samuelson's doctorate students also showed great interest in it and made the same efforts. That was Robert Merton.

Merton was born in New York in 1944. He received a Bachelor of Science (majoring in Engineering Mathematics) from the school of Engineering and Applied Sciences of Columbia University; He obtained a Master of Science degree from the California Institute of Technology. Later, he joined Paul Samuelson as a doctorate student at Massachusetts Institute of Technology. In 1970, he received his doctor's degree. Merton's intelligence, mathematical ability and keen insight into financial problems impressed Samuelson deeply. Samuelson praised him as Newton in the field of finance. Dr. Merton naturally stayed in the MIT Sloan School of Management after graduation, and continued his research on option pricing.

On the basis of previous studies and through long-term unremitting efforts, Black, Scholes and Merton have seen the dawn of solving the option pricing problem. By 1973, Black and Scholes finally published their paper on option pricing, deriving an option pricing model which was sound in theory and convenient in application. Merton also published some papers revealed more insights of option pricing and its application. Scholes and Merton won the Nobel Prize in 1997. Black failed to win the prize because of his early death.

The Black–Scholes model solved the problem of European option pricing, and cast light for the solution of more financial problems in theory and practice. Readers can also see its extraordinary functions from the previous chapters of this book, such as solving the problem of optimal capital structure and determining the discount rate based on total risks, etc. Theoretically, discount method and option pricing method are two complementary methods that are indispensable in valuation and financial decision-making.

2 Reasons for Stagnation of Financial Theory

Before 1980, many financial scientists and their research findings solved important financial problems one by one, forming a brilliant 30-year development of financial theory. However, when we move the historical lens back to recent decades, it is difficult to see such dazzling financial scientists (not necessarily the Nobel Prize winner) and financial research findings! Why?

Literally, the bounden duty of scientific or academic research is to improve theories and to solve problems. However, the financial research in the recent decades did not meet this basic requirement. The major financial theories have no substantial progress and the fundamental problems that were outstanding decades ago remain unsolved. The financial research findings during recent decades can seldom add to the financial textbook, because most (if not all) of the findings are not to solving problems but to describe phenomena. Therefore, based on the finance related textbooks, viewing from the point of problem-solving, almost no progress can be found.

Such as the valulation of regular bond and common stock; the determination of the fair or bubble free P/E, P/B and P/S ratio; the determination of discount rate considering total risk for debt, equity and total capital; the loan pricing or the determination of loan interest rate and loan decision in commercial banks; the impact of debt financing or leverage on firm value; the determination of the optimal capital structure; the impact of current ratio and debt ratio on the bankruptcy risk of a company; and so on. In mainstream finance (rather than in this book), obviously, these fundamental problems in theory and practice remaining unsolved.

Therefore, financial theory has stagnated for decades! Revealing, thinking and discussing such issues may make our financial scholars feel ashamed. However, this is an objective fact, and its existence cannot be denied by hiding or covering it. On the contrary, with a minimum sense of scientific spirit and social responsibility, it is not difficult to understand, if we do not reveal, think or discuss such issues now, we will be more ashamed of the times, and even harm the next generation or several generations of financial scholars and financial research.

Thus, why has financial theory made no substantial progress over decades? This problem is worth pondering by every financial scholar today, and also worth alerting and warning by every financial scholar of the future generations. Perhaps it is worth to mention here, to find the reasons why the financial theory has stagnated for recent decades, this section may point out various deviations and errors in the financial research during recent decades. The purpose of the analysis is to change the status quo of financial theory, and please do not consider from the narrow personal interests. By the way, the references to "before or after 1980" in this section is a convenient expression. The change in financial research is a gradual process, and "1980" is not a clear-cut boundary, or not a specific year.

2.1 Some Specious Reasons

The author has asked various insiders about the stagnancy of financial theory on various formal and informal occasions, and collect many opinions about this issue. Before further discussion, it is interesting to analyze simply some of those specious answers.

1. Insufficient funds?

A seemingly reasonable explanation is the insufficient funds for financial research. Although the funds in financial research are far less in amount than those for natural science and engineering projects. It is obviously not tenable to explain the poor results of financial research in recent decades. Since 1980, the world economy has continued to growth, and more and more resources have been invested in financial research. No one will doubt this!

Just imagine that Markowitz was only 25 when he published his portfolio theory in 1952. He was still a student of the University of Chicago. How much research funding could he have? At that time, Modigliani and Miller had to sit down and seriously thought about the decision-making of capital structure because they were afraid of lack of experience because they were giving courses to "experienced" students for the first time. Where could they find founds to support their financial research?

Similarly, the solutions to the relevant fundamental financial problems derived in previous chapters are seldom supported by research funds. An interesting question is: which are more qualified to represent the progress of financial theory, those supported by research funds and published in top financial journals or the solutions in this book?

As a matter of fact, since 1980, financial research has been supported by various funds more frequently. Even many professors have sufficient funds, but why cannot those sufficient funds promote more good research findings? Will the money support be counterproductive?

2. Less efforts?

If we say that funds will be counterproductive, it is in line with an old Chinese saying, that is, "adversity leads to genius". In other words, the more hardships, tribulations and poverties you experienced in past, the greater your achievements.

However, "adversity reverals genius" is often associated with "being adversity but steady" and "being adversity but diligent". If in addition to financial support, scholars also steady in will and diligent in work, it should not be impossible to achieve decent results.

Then, since 1980, have scholars in the field of finance work hard enough?

Obviously yes. This can be confirmed by the rejection rate of submissions from professional journals. Since 1980, the rejection rate of financial journals has risen sharply. So far, the rejection rate of general core journals has reached more than 90%. According to reliable information, this tends to reach more than 99% for the top level journals (such as journal of finance). This implies that scholars have done much more research than before.

In fact, in the current financial field, on one hand, less proportion of papers got published, on the other hand, a minority among the scholars have become "high-yield writers" in publishing papers. Of course, most scholars have no chance to publish their findings. Moreover, every paper today is often much longer than it was before 1980. This also shows the increase of efforts of scholars.

However, with financial support and more efforts input, it seems that the (real) research findings should be much more and the progress of the financial theory should be expected. But why can't such a decent result come true? It is conceivable that with the improvement of living and nutrition, the intelligence of scholars will certainly rise, at least not decline.

3. Poor technical conditions?

In a sense, scholars' intelligence is the most important factor that determines the progress of the financial theory. With the economic development of most countries in the world, living and medical conditions have been continuously improved. Obviously, except for some special cases, scholars' intelligence generally is unlikely to decline.

On the other hand, some experts think of the technical conditions or technical equipment for financial research, which should be at least a complementary factor to scholars' intelligence. Although the technical conditions for financial research may be much humble comparing with the research in natural science and engineering projects, they have been continuously improved, and are much better than the financial research during 1950s–1970s.

Specifically, since 1980, the software and hardware of computers have developed rapidly. Before 1980, it was difficult for financial research to have computer support; Today's technical conditions allows one person to use multiple computers at the same time in financial research. Not only that, with the development of computers, the network has also developed rapidly, which greatly facilitates scholars' access to literature, data as well as computing.

However, this makes it even more difficult to understand: with better research conditions, why does financial theory stagnate over decades? Why "human intelligence plus artificial intelligence" cannot produce better "research findings" than "single human intelligence" decades ago?

4. Low standards?

At present, various universities and scholars have joined the torrent of scientific research competition. As a scholar in a university or research institution, you are subject to more and more strict assessment. If you fail to pass the assessment, you will be punished or even dismissed.

The publication of papers and the application for scientific research funds are also subject to strict review. It is often heard that only the best can be selected from the best in the screening of professional journals and scientific research funds. It is conceivable that in such an environment, scholars themselves should not dare slack off. When the requirements in all aspects are raised, the scholars will naturally work harder than before.

Now it is even more difficult to understand: under the pressure of higher assessment standards and strict requirements, why no high-level theories or findings turn up? Professional journals claim to choose the best from the best. Why cannot they select the solutions to those fundamental financial problems or the theories that can add in the new textbooks?

5. Diminishing marginal return?

Diminishing marginal utility is a universal law, which is manifested in the law of diminishing marginal return in production. Some people speculate that there should also be a law of diminishing marginal return in financial research. So it is normal that the returns from the effort input during the recent years are not as good as those decades before.

However, the diminishing marginal return implies the return will be diminishing gradually. How come there were more than n major theoretical findings in those past decades, and suddenly no one came out over recent decades? This is clearly a vertical drop rather than a marginal decline.

If we go a little deeper, there is an important premise for the law of diminishing marginal return, that is, the technical conditions remain unchanged. When technology advances, for example, from the historical process of economy and society, the marginal return to people's labor input is obviously increasing rather than decreasing.

This is especially true in the development of science and technology. For example, the progress of computers is accelerated rather than decelerated. What is financial theory? Please do not forget our discussion in Chapter "Finance and Its Fundamental Problems": financial science is also a science! As a social science, how can the development of financial science casually have "the diminishing marginal return"?

6. Financial theory reached its peak around 1980?

Some people speculate that the problems that are easy to study and solve have been studied and solved in the past decades. The problems that can be studied in the recent decades are extremely difficult problems, which are not easy to study and solve. Put it another way, they speculate that maybe the financial theory has reached its peak around 1980, and the major problems worth studying and solving have been studied and solved, and there are no major problems to study in the recent decades.

If you have read the previous chapters of this book, or some of them, you will certainly not agree with such a guess. For example, how to value a regular bond or a common stocks, the simplest and most common securities, based on their return and risk, how to estimate the theoretically reasonable P/E ratio and P/B ratio of stocks or the market, and so on. These problems are not difficult problems in finance, but they have not been solved so far, if the solutions in this book are not counted.

If you have read some financial books or journals, you certainly know that even in the current academic environment with fierce competition, high standards, after screening the best from the best, mistakes in concepts, formulas and methods are still common in mainstream financial textbooks and top financial journals. Especially, most of these mistakes are not difficult to find and to correct.

Then, did financial theory reach its peak in around 1980 or recently? Of course not.

Finally, why financial theory stagnates over decades? It is a problem! It is a problem not so easy to answer; it is a problem worth to ponder for every scholar in finance area!

2.2 Financial Research Content in Recent Decades

Why has the financial theory stagnated for decades? You may feel more confused about it. Yes, with more research resources, better research conditions, more research efforts, more strict assessment, and higher paper evaluation standards, how can there be no substantial progress in financial theory over decades? What is worse, simple mistakes are common in financial textbooks and top financial journals without correction during long time or decades.

This implies that the reasons about the stagnation of financial theory are not in the inputs of resources, conditions, efforts, assessments, etc. We need to dig deeper.

Of course, previous analyses also provide some clues. Since the reasons do not lie in the factors mentioned, it may lie in the direction (content) or method of research efforts. That is, the contents and methods of financial research in recent decades are worth to be examined. No matter what the ultimate reasons are, it is very important for financial academic research and theoretical development to understand the reasons!

Before starting the discussion, it is necessary to review the nature and attributes of the finance discipline, because the nature and attributes of a discipline determine whether the research contents, methods and objectives are appropriate or correct. According to the discussion in Chapter "Finance and Its Fundamental Problems", financial science is an application-oriented decision-making science, and its essence is the asset value; Financial science is a decisional science, not a descriptive science; The research goal should be to improve the decision-making, not to describe or explain data or the actual decision-making results; Finance is a quantitative science, which should support and improve practical decision-making by providing relevant quantitative methods or models.

As to the content of financial research, since the 1980s, financial research has focused on market efficiency, behavioral finance and agency theory. Most of these contents or theories were put forward in the 1970s and flourished after the 1980s.

1. Market efficiency hypothesis

Efficient markets hypothesis (EMH) means that the securities market is sensitive enough to relevant information; In fact, it was put forward on the basis of the early "random walk" hypothesis of stock prices. In 1970, Eugene Fama, a professor at the University of Chicago, deepened and standardized the market efficiency hypothesis. Since then, the market efficiency hypothesis has aroused extensive discussion and research in financial field.

To be exact, market efficiency refers to the efficiency of relevant information reflected by the market price of securities (stocks). Fama has divided three kinds or degrees of efficient securities markets according to the range of information reflected by securities prices.

The first degree is the weak efficient market, which means that the securities prices can fully and timely reflect historical price information. In such weak efficient market, it is impossible to find undervalued securities by analyzing historical price information which is normally referred to as technical analyses in finance, so conceptually or in terms of expected return, if a market reaches the weak form efficiency, it is impossible to obtain abnormal returns from investment based on technical analyses. Therefore, there is no need for investors to care about historical price information.

The second degree is the semi-strong efficient market, which means that the securities price can fully and timely reflect not only the historical price information, but also other public information of the company, such as the relevant information in financial reports and news reports. This kind of information is now often collectively referred to as fundamental information. In the semi-strong efficient market, it is impossible to find undervalued securities by analyzing fundamental information, which is known as fundamental analyses; so it is impossible to obtain abnormal expected returns from investment based on fundamental analyses. Therefore, in the semi-strong efficient market, there is no need for investors to care about fundamental information.

The third degree is the strong efficient market, which means that the securities price can not only fully and timely reflect all kinds of public information, but also reflect the inside information that is not public or disclosed so far. It is conceivable that when the market reaches strong form efficiency, the undervalued securities cannot be found by any information analysis, and thus abnormal returns cannot be obtained. Therefore, investors do not need to care about any information.

After the hypothesis of market efficiency was put forward, it was hotly discussed and studied because it seemed to be a problem of practical value. For example, if the securities market has reached the weak form efficiency, there is no need to analyze the price trend, so there is no need to do technical analysis. If the market has reached semi-strong form efficiency, there is no need to analyze public information, so there is no need to do fundamental analysis.

According to the above logic, the understanding of market efficiency determines how to analyze the stock market and how to make investment decisions. Up to now, most of the research found that the stock market has reached weak form efficiency or semi-strong form efficiency, but not strong form efficiency. However, it is a little strange that these conclusions have little influence on the analysis and decision-making of the stock and securities market; technical and fundamental analyses are still popular, and investors still invest based on these analyses.

Therefore, although the research on market efficiency has been booming in academic research in the past decades, from the perspective of practical effect, these studies have not played the role envisaged in academic research. It is worth thinking for scholars interested in market efficiency.

One thing is certain, that is, although market efficiency has been proposed and studied as a research hotspot for decades, it can only be regarded as a hypothesis, not a theory, because it cannot withstand the minimum scrutiny. In fact, this hypothesis only expresses half of a "closed-loop" logic. If we supplement the follow-up part of the logic, we can find that this is actually just a logical paradox about the securities market or the formation of securities prices.

According to the hypothesis of security market efficiency, the security price has fully and timely reflected the corresponding information, and the analysis based on the relevant information is invalid; Then, investors do not have to care about and analyze the corresponding information. Take the weak form efficiency as an example.

In such a market, the security price has fully and timely reflected the information of the past price, and the analysis based on the past price is invalid; Then, investors do not have to care about and analyze the information of past prices.

Let's continue this logic reasoning. Since investors do not need to care about the information of past prices, so they are unable or unnecessary to make trading decisions based on the past prices. If so, how can the information of past prices be reflected in the current securities prices? Since the information of past prices cannot be reflected in the securities prices, the technical analysis based on the past prices should work, i.e., trading based on the technical analysis can obtain abnormal returns. This means that the securities market has not reached the weak form efficiency, that is, the securities market is invalid.

From valid to invalid, and from invalid to valid. This is obviously a circular logical paradox. Therefore, the "hypothesis" of market efficiency put forward in the 1970s is only a half logical paradox. In that case, how much theoretical or practical significance does it have to prove whether this half paradox is true or not true through research?

It is conceivable that if we continue to study according to the current research method, the conclusion is nothing more than: according to some sample data, the market has reached weak form efficiency; According to some sample data, the market has reached semi-strong efficiency; Basically, no market has reached strong form efficiency. Obviously, we can know the conclusion without any research.

Some scholars state that such research helps to improve the efficiency of the market. But can we really improve market efficiency? According to the complete logic, improving efficiency will lead to market inefficiency, and the conditions of efficient market is just market inefficiency. In this case, how can we improve market efficiency? Of course, the research on market efficiency has little help because every market is experiencing the circulation of efficiency and inefficiency; no certain conclusion for a certain market.

In fact, because of the differences in product and demand and supply, the security market has been more efficient than the product market; it is easy to reach or approach the weak form efficiency. In this case, it may be more meaningful to study how to improve the standardization (different from efficiency) of the securities market. Or, it may be more meaningful to study how to improve the efficiency of the product market (rather than the financial market).

On the other hand, it seems that the efficient markets hypothesis neglect something very important. Consider the situation that a good (or bad) news about a stock is released in the market. Investors may be able to judge that the direction of the stock price will move up (or down), but it is very hard to estimate the precise amount of the change. That is, the market or the stock price may be very sensitive or efficient to the new information, but the fundamental analyses are still necessary because investors cannot derive the target price by simple judgement or intuition. Put it another way, the target price is determined based on the valuation of the stock.

Therefore, it can only conclude that the stock price is a random walk, but this cannot extend further to whether the market is efficient, let alone to what extent of the efficiency and the data-based statistical test

2. Behavioral finance theory

In the 1970s, Amos Tversky, Daniel Kahneman and some cognitive psychologists found that people are often irrational when making judgments and decisions under uncertain conditions, proving that the hypothesis of human rationality in traditional theories is wrong, and that people's deviation in making decisions is regular. In 1979, Kahneman and Tversky found through experimental comparison that most investors did not conform to the rational assumptions of the standard financial theory, and their behavior was not always rational or risk averse, so they put forward the expectation theory.

Kahneman and Tversky's paper has aroused a wide response in the field of economics and finance. Scholars in these fields have listed various irrational performances of investors and decision makers. Thus, after the 1980s, a research upsurge of behavioral economics and behavioral finance has formed. For example, Thaler (1987, 1999) of the University of Chicago, who was more influential, studied the time series of stock returns and investor psychology. Shiller (1981, 1990) of Yale University mainly studied the abnormal fluctuation of stock prices, the "herding behavior" in the stock market, and the relationship between speculative prices and popular mentality. But strictly speaking, behavioral finance has not come up with theories and methods that can solve any theoretical and practical problems.

In 2002, the Royal Swedish Academy of Sciences awarded Kahneman the Nobel Prize in economics, for that he laid the foundation for "a new research field. Kahneman's main discovery is to demonstrate that under uncertain conditions, human decision-making will systematically deviate from the predictions made by the standard economic theory. However, whether the theory of individual choice can be extended to group decision-making and how the results of laboratory experiments can be applied to the real world situation are still far from conclusive."

Behavioral finance originates from the query that the standard financial and financial conclusions deviate from the reality, and the pursuit is that the conclusions conform to the reality. However, is it possible to improve decision-making when the conclusion conforms to reality becomes the highest pursuit? In fact, financial theory is a science of decision-making. What should be pursued is to improve people's decision-making, not to describe or conform to people's decision-making. Therefore, behavioral finance violates the fundamental characteristics of finance, and misunderstands finance as descriptive science rather than decision science.

As a matter of fact, the main function of economics and finance is not to predict the future, nor to describe the past; instead, it pursues to draw the correct decision conclusion needed at present; this decision conclusion can be used as the basis, standard or guidance for people's decision-making; This correct decision conclusion has nothing to do with the degree of human rationality, and there is no need to make assumptions about the degree of human rationality. Apparently, only the correct decision conclusion can work as the guidance; on the contrary, the description conclusion that conforms to people's actual decision results has no such a function, and should not be the pursuit of economics or finance. In other words, the so-called behavioral economics or behavioral finance is actually a conceptual and logical disorder, a misunderstanding of real (rather than traditional) economics and finance.

In this regard, there are a number of interesting issues worth considering. For example, before the emergence of behavioral economics or behavioral finance, did people really think their decisions were completely rational? The irrationality in people's decision-making is actually a common sense. Why does this common sense need to be "discovered" by establishing a new discipline such as behavioral economics or behavioral finance? Compared with "traditional" economics or finance, does the emergence of behavioral economics or behavioral finance really help to reduce irrationality in decision-making? What is its working principle? To talk about irrationality in decision-making, we actually need to know the standard of rational decision. How can behavioral economics or behavioral finance find the standard of rational decision-making for various decision problems without "traditional" economics or finance? However, if "traditional" economics or finance have provided the right answer to a decisional problem, why do we need further to know other one or more incorrect answers?

3. Agency theory

The agency theory, namely the principal-agent theory, was first put forward by Jensen and Meckling in 1976. Later, this theory developed into corporate governance theory, that is, how to coordinate the internal relations of the company, motivate and control the agents at all levels within the company, such as the board of directors, management at all levels and employees, to strive to achieve the fundamental goal of the company, that is, to maximize the wealth of shareholders.

Agency theory is based on asymmetric information. According to the agency theory, the asymmetry of information may occur before the parties sign the contract (ex ante) or after the parties sign the contract (ex post), which is easy to cause adverse selection and moral hazard respectively. The content of agency theory has been introduced in more detail in various books, we do not intend to repeat it here.

To some extent, the agency theory has the same problem as behavioral finance, that is, it takes human behavior as the research object to describe or analyze people's (work) attitude, which is obviously inconsistent with the financial discipline in the research object. As revealed in Chapter "Finance and Its Fundamental Problems", the fundamental object of finance is asset value or valuation, which concerning the return and risk of the asset.

It is certain that that agency theory may be conducive to "corporate governance", but it is absolutely not conducive to the solution of financial theory and decision-making problems. Just imagine that, can the stock valuation problem be solved by applying the agency theory? Can the problem of determining discount rate be solved by applying the agency theory? Or, can the problem of optimal capital structure be solved by applying the agency theory?

As we have already known, finance aims at the right decision conclusion. But both behavior and agency problems involve the incorrect or non-optimal decisions. In fact, the function of the agency theory focuses on explaining phenomena. If we intend to use the agency theory to solve the problems of financial theory and the relevant decision, it is absolutely impossible within recent decades, 100 years or 1000 years. Focusing on the attitude of various agents, the agency theory can even hardly be regarded as a quantitative science. So how can it become the content of finance?

To sum up, since the 1980s, the content of financial research either has little practical significance, or violates the essential attribute of the financial discipline, or does not belong to the subject of finance. This is obviously one of the important reasons for the lack of progress in financial theory over decades. Perhaps readers will think that the above explanation is very simple; why cannot the financial scholars around the world think of that? The direct cause of this situation is the insufficient understanding of the essential characteristics of finance in academic research. From this, we can see the necessity and significance of discussing the nature and attributes of finance in Chapter "Finance and Its Fundamental Problems" of this book. There are of course deeper reasons some of which may be revealed in the following discussions.

2.3 Financial Research Methods in Recent Decades

Besides the problems in content, are there any problems in research methods? For example, if the research on the above topics does not help to solve the financial problems remained unsolved, then at least the obvious mistakes in the original textbooks should be corrected as decades passed by. Why are there so many simple mistakes that cannot be corrected?

The methods of financial research have been significantly changed over recent decades. During the first three decades after born of finance, most of the financial research are theoretical research and applied research. Some of the scholars focus on the derivation of the theoretical models, such as the Gordon growth model; some of the scholars focus on the application of the theoretical models, such as how to use the Gordon growth model to value a stock. In recent decades, empirical research rose and became the overwhelming method in published papers. At first glance, empirical research attempts to prove a certain viewpoint or model through "data", while theoretical and applied research relies on logical reasoning to draw conclusions. However, the difference between empirical research and theoretical and applied research is not so simple and clear.

The rise and prosperity of empirical research is highly consistent with the stagnation of financial theory in time. Is this a historical coincidence, or is there an inherent necessity? Is there any internal relationship between the empirical research and the stagnation of financial theory? To understand the relationship between empirical research and stagnation of financial theory, we need to make more in-depth analyses of empirical research.

1. What is empirical research

Empirical research originated from Bacon's empirical philosophy and Newton Galileo's natural science research. French philosophers Condorcet (1743–1794), Saint Simon (1760–1825) and Conde (1798–1857) advocated that the spirit of natural science demonstration should be implemented in the study of social phenomena, and that the study of social phenomena should be refined and accurate by means of procedural, operational and quantitative analysis starting from experience. Economics first introduced empirical research methods in the 1930s and 1940s. The empirical research on finance basically rose in the 1970s and 1980s and spread rapidly.

Early empirical studies emphasized the objectivity and universality of scientific conclusions, emphasized the use of empirical observation data and experimental research to reveal general conclusions, and required such conclusions to be verifiable under the same conditions. Generally speaking, empirical research is a research method to summarize the essential attributes and laws of things from individual to general through a large number of observations, experiments and investigations. Therefore, the pursuit of empirical research is to describe phenomena.

From the early origin, empirical research originated from the experience summary of natural science research. However, when trying to transfer the research experience of natural science to the research of social science, the ancient philosophers did not consider the important difference between social science and natural science, the significant difference between decisional science and descriptive science in social science, or the fact that the observation data in social science is different from the observation data or experimental data in natural science.

Phenomena in natural science are often stable in a certain period, that is, there is often no difference between the past data and the future data. However, social science is likely to have timeliness. Observation data in Social Sciences vary with time, place and conditions, and different sample data show different characteristics and laws. In this case, empirical research draws a conclusion based on a set of past sample data, but can we say that this conclusion is in line with the universal law, or even how is this conclusion different from the universal law? Obviously, the reality of one time and place may not be the reality of another time and place.

As far as social science is concerned, if reliable and representative sample data can be obtained, the conclusions of empirical research may be in line with the more common reality. But even so, it is impossible to meet the requirements of decision-making science (such as most branches of economics and management disciplines), because decision-making is always future oriented. The conclusion of empirical research can at best meet the reality of the past. Although knowing the past helps to predict the future, but the future is indeed different from the past.

The future can be predicted based on the past, but predicting the future cannot directly replace decision-making. For example, it is predicted that the annual growth rates of earnings of stocks A and B over the next 10 years will be 10% and 20% respectively. Can such prediction result directly explain the decision of whether to buy A or B? Obviously not. Only the decisional science like finance can explain how to choose between A and B. For example, if A is undervalued by 20% and B is overvalued by 30% according to the current market price. Obviously, we should buy A instead of B. How can we know whether A and B are undervalued or overvalued and how much they are undervalued or overvalued? This still depends on financial science to answer. Neither describing the past nor predicting the future can answer such questions.

Due to the complexity of social and economic phenomena, when it is difficult to get a clear idea, especially the quantitative relationship, empirical research may be used to find some clues. In another word, empirical research is just an amateur or adjunctive method, the professional method in finance is discounting and option pricing, which is featured as quantitative trade off between risk and return. Of course, to solve problems in finance, professional method is the first choice, and empirical research at best is the second choice. It is not necessary to do empirical research when the problem can be solved by the professional method.

However, the situation in reality is that, after entering into financial fields in the 1970s and 1980s, the empirical research not only spread, but also tried to exclude other research methods. Under the exclusion and control of empirical research, traditional theoretical research and applied research are considered as "low-level" research because they do not meet the "norms" and "requirements" of empirical research, while empirical research apparently below the theoretical level is considered as "high-level" research; More and more professional journals refuse to accept the traditional theoretical research and applied research. Since empirical research emphasizes "knowing what is" rather than "knowing why", such research is certainly not conducive to the understanding of financial problems, which will naturally lead to no fundamental problem being solved and no substantial progress in financial theory over long time.

As a matter of fact, the exclusion of other studies is not in line with the idea of empirical research put forward by the previous philosophers. However, the current empirical research problems are not only like this. Different from the assumptions of the previous philosophers, the current empirical research in financial research has already formed a fixed "routine", that is, to test a guessed relationship or model (research hypothesis) based on a set of data. The purpose of the research is to test whether the guessed relationship is tenable or to obtain the coefficient of the regression model, rather than to solve the corresponding financial decision-making problems. From the research method to the structure of the article has been programmed and mechanized. Unfortunately, stylization and mechanization may help to improve industrial production efficiency and promote mass and large-scale production, but they are not suitable for academic research, because innovation is the necessary feature of scientific research, so does the financial research.

Based on the current empirical research, it has the following basic characteristics: the purpose of the research is to test a guess (hypothesis), not to solve fundamental financial problems or contribute financial theories; research problems and contents come from subjective guesses rather than practice, so they may be completely useless; the premises of a set of data are not clear, the conditions for the conclusion are not known, so even there is a clear conclusion, it is difficult to apply; As a guess-based research, it is difficult to get a conclusion better than common sense. Therefore, even the research problem has practical significance, the conclusion is difficult to be meaningful. Unfortunately, this kind of empirical research excludes the traditional theoretical research and applied research. How can financial science make progress or avoid stagnation if it does not study decision-making problems but test conjectures?

2. Advantages of empirical research

Anyway, empirical research has been widely welcomed once it is launched, which implies that it must have some unique advantages. According to the current empirical research in finance, the advantages of empirical research are summarized as follows.

(1) Controllable difficulty

Empirical research is the method-oriented research, the relationship among concepts and variables are unconstrained by professional logic and quantitative relations in finance; In other words, the empirical research of other disciplines is similar to that of finance. The basic routine is to put forward the hypothesis of guessing, select and define variables, data processing (i.e. run on the computer with statistical software), draw conclusions and give simple explanations.

It would have been more or less difficult to measure variables. But in fact, up to now, there is no strict right or wrong standard for how to measure the variables of subjective choice. No matter how difficult it is to measure variables, such as agency cost, beauty degree, etc., researchers can always find an indicator to represent it, so measuring is not a difficult problem.

In this way, different empirical studies not only have similar routines, but also have similar difficulties. Moreover, the difficulties in empirical studies are those of statistics rather than the difficulties of finance, such as difficulties in measuring or trading off risk and return.

(2) The research can be finished before the problem solved

The purpose of empirical research is not to solve problems, but to use a set of data to test a subjective guess, and there is basically no strict reasoning process. Therefore, it has nothing to do with whether the relevant financial decision-making problems are solved or not. In this way, the task of "research" can be completed without solving the problem, that is, research is completely equal to writing articles. Therefore, if scientific research is defined as research aimed at solving scientific problems, empirical research is not scientific research, but simply "writing articles".

Such research has nothing to do with the difficulty of the decision-making problem itself, and whether it can solve the relevant theoretical and practical problems. It is very convenient to complete "article writing". Therefore, if the number of published articles is used to evaluate the scientific research performance of scholars, this kind of research has overwhelming superiority and will be exclusively favored. Because it is obvious that in the research aimed at solving problems, even if the most difficult financial problems that have not been solved in the world for decades, such as option pricing or capital structure, it is enough to write one or two articles to solve them; However, empirical research can write hundreds or thousands or millions of articles without the annoyance for solving the problem.

The problem is: do we really need such kind of research? Is it qualified as a scientific research?

(3) Speaking with data

In reality, there are two ways for people to draw conclusions. One is to draw conclusions based on data; the other is to draw conclusions based on logical reasoning. For various reasons, most people believe data speaks louder than logic reasoning. In fact, this may only suitable to simple problems, such as problems in daily life. For example, the increase in life expectancy may indicate that, on average, people's health has improved. From simple life experience, many people naturally believe that in scientific research, conclusions based on data are also more reliable than conclusions based on logical reasoning. Empirical research draws conclusions based entirely on data and data processing, so it is naturally easy to be accepted and trusted by more people.

Data is the result of statistics of phenomena. For scientific problems, it is easy to draw a wrong conclusion simply from data, because the purpose of scientific research is to reveal the deeper nature of things. For example, if we count the running time, distance and speed of all the students in a school, a possible conclusion may be, speed = distance/time + 1. Although this conclusion is supported by data, it cannot be correct. On the contrary, it can be judged that there must be something wrong in the data or data processing. For another example, if we study and compare the falling speed of rocks, apples and cotton affected by gravity, the experiment and observation data show that stones are faster than apples and apples are faster than cotton at the same height and falling the same distance; However, such data only represent the effect of air resistance and cannot accurately reflect the effect of gravity. For another example, according to the observation year after year, the sun rises in the east and sets in the west every day. Maybe you will conclude that the sun revolves around the earth. But this is obviously based on your wrong understanding, draw the wrong data, and then draw the wrong conclusion.

The quality of observation data in social sciences is far inferior to that of experimental or observation data in natural sciences. It may be that there has been little discussion on the difference between social sciences and natural sciences in history. People have not been fully aware of this, so they particularly like and trust "talking with data". After introduced into finance, the empirical research takes the name of "speaking with data" and expands rapidly. This is an important reason why empirical research is welcomed and trusted both within and beyond the academic circle.

In fact, there are many ways to "speak with data", and "speaking with data" is not equal to empirical research, and empirical research may not be an efficient way to "speak with data". In particular, data can "say different things". In the empirical study of relevant issues, different or even opposite conclusions will be drawn because of the difference in sample data or data processing. In this case, which conclusion should be believed? Therefore, if we do not explore the deep nature of things, the surface "data" is not so reliable.

The object of empirical research may be sample data, large sample data, or big data. However, in terms of data, the object of logic-based research is actual complete or total data. Why we do not believe "speed = distance/time + 1", even the equation backed by large sample data, because we know a conclusion of "speed

= distance/time", which is backed by complete or total data, i.e. logic reasoning. Obviously, "speed = distance/time" is correct; "speed = distance/time + 1" is incorrect.

Therefore, no matter in or out of the field of scientific research, do not be confused by surface data. Do not forget: scientific research should explore the essence of things, not the surface appearance or data. Financial science is also science! If we only deal with past data, or even become the slaves of data, how can financial theory move forwards?

(4) Multiple other advantages

Empirical research papers have fixed structure and formats such as hypothesis guessing, literature review, variable selection, data processing, conclusion or model and suggestions. If you do not know much about finance, you can also imitate the fixed structure and formats and complete a empirical research paper that meets the specifications. Empirical research papers can be mechanized and standardized production, which greatly improves the productivity of scholars. The empirical research papers are also easy to understand and write review opinions according to the general format, which also saves the reviewers' time. In addition, the starting point of empirical research is "data", and the process is "mechanized operation", so the readers have no way to check the data and the processing and cannot refute the conclusions. There are also many other advantages of empirical research, which will not be listed one by one here. With so many advantages, it is no wonder that empirical research is so widely accepted and so popular in academic circle.

3. Shortcomings of empirical research

Everything has its pros and cons, and empirical research will also have its drawbacks. Based on the current empirical research, there are mainly the following problems.

(1) Level of subjective guess

Since the conclusion and difficulty of empirical research depend on the initial guess, the level of guess also determines the level and value of the empirical research. However, there are inevitably two extremes in such a guess-based research.

First, professionals' guess level is not higher than that of ordinary people, and their guess may not be higher than common sense. This means that the conclusions of the study may not be higher than common sense. As a result, a large number of empirical studies are entangled in the common sense of "whether there are lice or not on bald heads". What is the significance of a conclusion that is not higher than common sense?

Second, long-term guess-based research will inevitably "enhance" the imagination of scholars and make all kinds of senseless or ridiculous "guesses". For example, when no efficient way to evaluate the value of stocks has been found for a long time, some people wonder whether there is a certain relationship between "the length of women's skirts and the rise and fall of stock prices"; When the optimal capital structure cannot be solved for a long time, some people wonder whether there is a certain relationship between the company's debt ratio and the gender ratio of the company's managers, or the length width ratio of managers' faces, etc.

Of course, no matter how senseless or ridiculous the guess is, empirical research can always finish papers. The question is, can such papers promote the progress of financial science even if the guess is not completely senseless? For example, the guess in the pecking order theory can be regarded as one of the best works in empirical research. Can financial problems (such as the optimal capital structure) be solved by this kind of guess-based research? Can financial theory make substantial progress by guessing of scholars all over the world?

(2) What can empirical research prove

According to the current situation, empirical research can be justified by adjusting "assumptions", "variables" and "data" until the conclusions are easy to explain. It may be used to test a relationship. For example, are companies with fast earnings growth more generous in dividends payment or companies with slow earnings growth? Or we can get a regression model, for example, the company's stock price = a * earnings per share + b * net assets per share. From the perspective of process, the starting point of empirical research is objective data, and the middle process is mechanized operation. The conclusion seems naturally impeccable.

The problem is that a set of data is data at a specific time, place and condition, and no one can rule out the contingency. How credible is the conclusion based on a set of contingent data? For example, according to a certain set of data, it is found that companies with fast earnings growth pay generous dividends. Do you then believe that companies with fast-growing earnings will pay generous dividends? According to a set of data, the test result is: stock price $= 9.29 \times$ EPS $+ 1.16 \times$ net assets per share. Do you believe that the stock price should conform to this model? In fact, researchers themselves may not believe the conclusions of their empirical research. Because obviously, the results of test or regression vary across sets of data, and even the conclusions and the coefficient signs in the model are possible to be opposite.

The advantage of "speaking with data" is that there is no need to reason or understand. When the researchers themselves do not understand the reasons behind a phenomenon, they can draw a conclusion, and the conclusion is irrefutable. This is a good way to persuade or deceive "laymen". However, based on the scientific attitude of seeking truth from facts, we should not be superstitious about "speaking with data", but should really think about it: is strict logic more persuasive? Or is a set of data at a specific time, place and condition more convincing?

(3) Quantitative research or qualitative research

Empirical research does not need the professional logic of the financial discipline, nor does it conform to the basic characteristics of the financial discipline. Both the test and the regression of the guessed hypothesis or guessed model belong to statistical research, not financial research. In terms of content, finance studies how to make financial decisions rather than studies the past or the guess based on the past. At best, empirical research may have predictive significance for the future. But prediction is not decision-making after all, nor can it replace decision-making.

In fact, empirical research can hardly be regarded as quantitative research strictly. Despite the form and process look like quantitative research, no one actually believes the quantitative part of the conclusion. For example, suppose "stock price = 9.29 × earnings per share + 1.16 × net assets per share" is a regression model obtained through empirical research. For such a conclusion, most people will believe that there is a positive correlation between stock price and earnings per share and net assets per share, because it is intuitive. As for the specific quantitative relationship between changes in earnings per share and stock price, which is 1:9.29 in the regression model, and the relationship between changes in net assets per share and stock price, which is 1:1.16 in the regression model, no one will believe it exactly, because they will apparently vary across data. Put it another way, the quantitative conclusion is not equally important as the conclusion in direction or the sign of the coefficients in the regression model.

If empirical research is uncertain in terms of qualitative or quantitative research, it is apparently not financial research, because finance is absolutely a quantitative subject. A research that is not financial at all has become the mainstream research in finance, and even rules and excludes other real financial research, that is the apparent reason why the financial theory has stagnated for decades.

(4) Hard to innovate

As discussed in Chapter "Finance and Its Fundamental Problems" of this book, innovation is the focus and core of scientific research. Any research should advocate innovation, and empirical research nominally emphasizes innovation. However, a problem is: how does empirical research innovate?

Theoretical research has always emphasized innovation. In theoretical research, what is innovation and how to innovate are very clear. When a problem unsolved before can be solved by a research, that is innovation; Solving problems in a simpler or more reasonable or more efficient way is also innovation; Although there is no final solution to the problem, it is also innovation to go into the deeper level of the problem through reasoning and reveal the deeper essence of the problem. MM model is a typical example of this kind of innovation without solving problems.

Then, how does empirical research innovate? The purpose of empirical research is not to solve problems, but to use data to test a guess and realize self justification. In this case, it is impossible to achieve innovation by solving or better solving problems. So, can innovation be realized by revealing deeper essential problems? Unfortunately, this is also difficult for empirical research, because the guess in empirical research is unlikely going above the common sense. In order to get rid of the dilemma of being unable to innovate, today's empirical research takes the finding of a new influencing factor to a target variable (such as net profit, stock price, etc.) as innovation. Therefore, the research on discount rate or reasonable rate of return is constantly innovative; Similarly, the research on the capital structure is also constantly innovative. However, the problem is that similar innovations are published endlessly, but the problem solving model is still not available. We do not know how to determine the discount rate by considering total risks, nor how to calculate the optimal debt rate of a company.

An obvious question is, if these are also innovations, what are the specific purposes and uses of such innovations?

In fact, from a philosophical point of view, everything in the world is connected. Even if others have found 100, 1000 or 10000 influencing factors, you can certainly find a new one. The question is, what's the use of finding another factor that is probably insignificant? In particular, the influencing factors of a target variable may be hierarchical, and theoretical research can reveal its direct influencing factors, such as four direct factors. However, each of these 4 factors may have four direct influencing factors respectively. Furthermore, the 16 influencing factors may have 64 direct factors; and then, 128 factors, and then 256 factors, etc. However, the problem is that empirical research relies on conjecture to find influencing factors. How can they distinguish which are direct factors and which are indirect factors that affect target variables? Without distinction, of course, more and more influencing factors will be found, and the relationship between these factors and the target variables will become more and more confused, and the related understanding will become more and more confused too.

From one perspective, empirical research actually reveals the statistical laws of data. What is the data? It is the phenomenon, not the essence. Therefore, empirical research focuses on the law of phenomenon, and it is difficult to reveal the deeper essence. From another perspective, empirical research is guess-based research. The quality of an empirical study is largely determined by the guess. What is a guess? It is simple imagination, not rigorous and in-depth reasoning. How can simple imagination go deep into the essence?

Therefore, as a research to test a guess with data, it is not easy for empirical research to innovate. Because of the incompetence of innovation, the financial theory has to stagnate over decades when this method became popular and dominating.

(5) Is data qualified as a test rule?

For financial problems, theoretical research focuses on the general model to solve decision-making problems; application research solves the application problem of the theoretical model. Then, what is the unique role of empirical research?

From the early research, such as Black's empirical research on capital asset pricing model, empirical research attempts to prove whether a theoretical model is right or wrong, or how much deviation there is. However, to test the right or wrong or deviation of something, it needs a universally applicable standard. What are the standard for empirical research? Data, of course.

However, as a standard, in order to measure the right or wrong or deviation of something, it must remain stable or constant. However, data are the surface characteristics of phenomena. Will they remain constant? Obviously, the data does not have such characteristics. For example, one of the commonly used data in finance is the market price of stocks. Even if there is no clear change of influencing factors, the price of stocks will change every minute. Just think, can such a changeable phenomenon data be used as a standard to test the capital asset pricing model or stock valuation model?

Of course, today's empirical research has developed beyond testing theoretical models, but to replace theoretical and applied research, and to directly draw theoretical models or application models based on data. The question is, in the case of thousands of sets of data and new data continuously generated or come up, each set of data may have hundreds of data processing methods, is the theory or application model based on one processing method of one set of data deterministic and universal? What is its value and significance?

If a theoretical model is derived based on strict logic reasoning, which is equivalent to prove based on total or complete data, it is definitely a more rigorous and reliable proof than the proof based just on one set of data. Does it make sense to further proved a logic based theoretical model once again by a sample data?

In sum, the financial research after 1980, viewing from research topics and methods, has seriously deviated from the essential characteristics of the financial discipline. This explains why financial theory has stagnated over decades.

3 Relevant Policy Suggestions

Finance is a field full of elites in most or all countries in the world. Readers may have questions. The above reasons revealed for the stagnation of financial theory are easy to understand. Do the elites in the financial field not know or understand them? It's actually not the case. In fact, many scholars in the financial field may know that the the mainstream research has some serious problems. The following two evidences may be good examples.

One evidence is that decision makers in the field of financial practice have basically stopped reading various financial academic journals. Although it seems that there are a large number of readers of every financial academic journals, a closer look at these readers can confirm that there are no practitioners who come to consult those journals for decision problem solving. The vast majority of readers are students and scholars in the field of finance. They are forced by the writing requirements of graduation thesis and the pressure of professional title evaluation to read the papers in the relevant journals and to write the similar paper following the same pattern in order to publish their papers and meet the research performance standard. From this, it can be judged that the decision-makers in the field of practice should have understood how much "practical value" the empirical research as the mainstream financial research has.

Another evidence is that scholars in the financial field have been reluctant to discuss financial issues. It seems that the academic activities of finance are very active, various domestic and international "seminars" have emerged one after another. However, these "seminars" are nothing more than "exhibition" and "walk show"; nothing more than that the "masters" and "rookies" are coming down from a high position to make a brilliant debut, display and show off. No time allowed for equal and in-depth "discussion"; no one cares about no solution to financial problem! It can be judged that scholars have deeply understood the "mainstream rules" of current financial research and are no longer willing to spend time on "meaningless" things such as solving theoretical and applicational problems in finance.

Apparently, the "mainstream content" and "mainstream method" are determined by the "mainstream rules". In order to change the drawbacks of current financial research and release the inherent development vitality of financial science, we must consider improving the "mainstream rules". As we all know in the financial academic field, the so-called "mainstream rules" mainly include research performance assessment and paper or manuscript review. This section attempts to discuss policy suggestions for improvement from these two aspects.

3.1 Research Performance Assessment

The research performance assessment here includes research performance assessment and rewards and punishments on this basis. In the 1970s and 1980s, universities and research institutions chose and gradually strengthened two assessment criteria: first, how much research funds in your account, and second, how many your papers published. The more funds you get and more papers you publish, the better your research assessment will be; Otherwise, your research assessment is worse. Scholars' awards and promotions are linked to the assessment.

Later, this assessment standard was "improved", that is, it pays more attention to "quality" in addition to quantity. How to measure the quality of scientific research? Western Universities and research institutions have chosen level or grade or rank as an indicator to measure the quality of scientific research. In other words, the scientific research funds and the professional journals are artificially classified into different levels or ranks, such as level ranked as A, B and C. An article published in "high-level" journals is higher in quality than that published in "low-level" journals. Some journals have no level; papers published in no level journal are not counted in performance assessment. Similarly, research backed by high (low) level research funds represents higher (low) quality research; research without support of research funds (research done at one's own expense) does not count.

Such scientific research assessment criteria seem obviously unreasonable. For instance, if Einstein's theory of relativity was published in non-level journals without the support of research funds, then, such research findings are treated as nothing under the "improved" or "strengthened" assessment criteria. But why do these universities choose and adhere to such criteria? In the decades after that, except for the discussion on how to further strengthen those criteria, there has been seldom discussion on why to choose those unfair and unreasonable criteria. It is one of the biggest mysteries today in scientific research that has lasted for decades!

This research performance assessment system can be summarized as look for and check out "the level or rank, the number of papers, the amount of money". Anyone with a little common sense knows that in our world, resources are limited; Wasting or occupying resources should not be advocated. At present, the basis of research performance assessment—scientific research funds and journal pages, especially the "high level or high rank" scientific research funds and journal pages, is extremely scarce and seriously in short supply. According to such standards, research performance assessment seems to advocate wasting or occupying resources, focusing on input rather than output. The greater the cost, the better; as to solving problems or innovations, they are totally unimportant. Unfortunately, such a doubtful or obvious wrong assessment criteria spread gradually to all over the world from western universities and institutions!

Under such an assessment, scholars cannot devote themselves to teaching and scientific research; they even need to devote most of the time and energy to things other than teaching and scientific research. First of all, under the criteria of "the level or rank, the number of papers, the amount of money", the advantages of empirical research become increasingly outstanding. For improving research efficiency, most scholars choose empirical research; scholars interested in traditional theoretical and applied research become less and less. Secondly, in case of only papers published in level journal can be counted, most of scholars devoted themselves into the public relations with journals and the anonymous reviewers. In academic field, a saying became popular, "knowing something is not as good as knowing someone". So, scholars are very busy writing empirical research papers, writing funds applications, extending and maintaining public relations. As to how to solve the unsolved financial problems, how to revise the mistakes in financial textbooks, how to pull the financial theory out of the mire of stagnation, etc., who care? In fact, when most or all scholars rush into empirical research, it is not easy to see and correct the mistakes in the financial textbooks, because identifying and correcting these mistakes depends on professional logic, that is, the theoretical research; it is the same for solving financial problems.

As the saying goes, forgetting history means betrayal. We might as well briefly recall history. Which important scientific discoveries in history can attribute to the pressure of assessment? The assessment based on "the level or rank, the number of papers, the amount of money" may be able to "press out" empirical research papers, but difficult to "press out" solutions to theoretical and application problems. Even on the contrary, while the majority of scholars strive to satisfy the assessment "quota", they will have no time and energy to study the real important financial theory and problem. Therefore, there is an important premise to promote scientific research through strengthening assessment, that is, the assessment standard must be correct, otherwise, it will backfire. In fact, the stagnation of financial theory over decades coincides with the strengthening of research performance assessment. Apparently, the wrong performance assessment is a key factor for the stagnation of financial theory over decades.

Therefore, the way to improve the management or assessment of financial scientific research is very clear and simple, that is to give up the research performance assessment criteria of "the level or rank, the number of papers, the amount of money"; instead, choose the number of theoretical and practical problems solved and the importance of the problems, or other more correct standards. Even if it is impossible to establish new effective assessment criteria, the prevailing criteria should be abolished, because the prevailing criteria are completely wrong and form a reverse incentive for scholars. Such an assessment is not as good as none. For example, the traditional ranking in order of seniority is obviously better than such reverse incentives.

At present, empirical research is often regarded as “high-level” research in the financial field, while traditional theoretical and applied research is “low-level” research. A considerable number of financial journals only or mainly accept empirical research papers. In general, empirical research draws conclusions based on past experience (data); or it is a summary of past experience. Experience or summary of experience is obvious lower than theory and its application, because normally, theory can be obtained by upgraded the relevant experience or summary of experience. According to the current research performance assessment, because one empirical research papers are relatively easy to be accepted by high level journal, one empirical research paper is equivalent to more than one theoretical research paper in terms of research performance.

Literally, the purpose of theoretical research is to solve problems; The purpose of empirical research is to describe phenomena. As to a financial problem, one theoretical research paper may be enough to study and provide a solution. This paper may take years or even decades to complete because the problem is too hard to derive a solution. However, for empirical research to deal with the same problem, it is impossible to meet with formidable or insurmountable difficulties, because empirical research does not intend to solve the problem. Therefore, there is no problem that research conclusions cannot be drawn, and hundreds or thousands or even millions of empirical research papers can be completed without a solution to the problem. Therefore, in scientific research assessment, even if one empirical research paper is equivalent to one theoretical research paper, it is extremely unfair to the theoretical research, or it is too or extremely in favor of empirical research.

Obviously, the rapid rise and overwhelming proportion of empirical research is not because the development of theory or practice needs empirical research, but because of the reverse incentive from the unreasonable assessment. The more strengthened or stringent the unreasonable assessment is, the harder for the financial theory to make progress. Therefore, this reverse incentive assessment must be corrected or abolished as soon as possible.

Actually, scientific research does not have to be assessed; It’s better not to assess without appropriate assessment criteria. Doing scientific research is different from doing business. Assessment, rewards and punishments may work in business activities, but they are not applicable for scientific research. Major scientific research discoveries in history have nothing to do with assessment. Even under the good intentions and correct criteria, research performance assessment has little significance for scientific development; However, under the wrong criteria, the assessment must be harmful, that is, it will form a reverse incentive, and waste scholars’ energy and time for research, restrict academic and disciplinary progress, and cause the stagnation of theoretical innovation and disciplinary development.

Once the wrong research performance assessment was abolished, scholars do not have to strive for increasing the number and rank of their papers; If scholars do not pursue the number of papers under the pressure of assessment, they do not need to finish papers without inspiration and theoretical innovation; In this way, both the number of submissions and the rejection rate of submissions to professional journals will decline significantly; Thus, the real research innovation is not difficult to publish. Therefore, scholars can engage in scientific research rather than put their limited intelligence and energy into public relations. Subsequently, professional journals can also save a lot of review costs, and the "publication cycle" of papers can be greatly reduced. The development of financial theory can be restored and promoted.

In sum, the research assessment over recent decades formed a serious reverse incentive. Scholars were encouraged to apply more and more research funds, to publish more and more unnecessary papers, especially the empirical research papers, and to participate too much public relations activities. The overflow of unnecessary empirical research seizes and controls too much of research resources, such as research funds, journal pages, seminar or discussion time, etc., hinders other more valuable research, especially the theoretical and applicational research. This reverse incentive has led to the stagnation of financial theory for decades, and large number of fundamental problems have not been studied and solved.

3.2 Anonymous Review of Manuscripts

Since the research performance is assessed based on "the level or rank, the number of papers, the amount of money", the review of funds application and paper submission is very critical. Any scholar who wants to publish papers (rank and number of papers) and obtain research funds (amount of money) must pass the review. In fact, the review of manuscript submissions and the review of funds applications are similar, and they are basically controlled by the same group of scholars (experts). Therefore, we would not distinct between these two reviews in the following discussions.

There is normally no open standard for review, or the reviewer can make his or her own judgement in private or secret even there is a nominal review standard. The standards implemented by various journals and scientific research funds are similar. Taking the review of manuscript submissions as an example, the main influence or review factors include the following aspects:

(1) Whether the author is supported by research funds. Papers without research funding support are reviewed as poor.
(2) The number of times the author's previous papers have been cited. Authors with papers cited less are often reviewed as poor.
(3) The norm of the manuscript. Manuscripts that do not conform to the "norm" of empirical research are reviewed as poor.
(4) Literature review. The literature review is incomplete, especially those omit recent papers are reviewed as poor.
(5) Whether the conclusion conforms to the sample data. The conclusion without actual data verification is reviewed as poor.

(6) Promote marginal innovation. The papers are reviewed as good if the author is modest, but the papers with major innovations or breakthroughs are reviewed as poor, because the author seems not modest enough.

As for whether the article has solved or better solved the relevant problem, whether the conclusion or model is sound in (financial) theory and convenient or reliable in application, and whether there are basic concepts and logical errors in the process and result, those are not the focus of the review. The review is particularly suitable for and tolerant of empirical research; what important for theoretical research, such as the cleverness, efficiency and innovation of problem solving, as well as the correctness in concepts and logic reasoning are not important in the review. Therefore, the empirical research has become the dominant research method soon; some fundamental problems in finance have remained unsolved; and some basic concepts and logic mistakes have remained in financial journals and textbooks.

What is more serious is that the current review system has been "improved" to anonymous review, and there is actual no need to explain the reasons for rejection. Even if individual journals have open review criteria or feedback reasons for rejection, the review opinion is irrefutable or undiscussable. As we all know, both the author and the reviewer are financial scholars, and may be interested in the same topic. This means that the reviewers are "both referees and athletes". Another aspect far more serious than the referee system in the field of sports is that the judgement of referee in financial research is not subject to any public supervision. The review behavior is a complete dark box operation. It is not restrained in advance, not supervised in the process, and not investigated or punished afterwards.

In addition to the reviewer acting as both "referees and athletes", and no supervision, the review opinion is undiscussable, no matter how absurd or invert justice. That is, the secret judgment does not allow discussion. Is that scientific research? Obviously, if discussion is not allowed, that is the officer command or order in military rather than scientific research. In fact, scientific development history tells us, all major scientific discoveries are "found" by researchers, not "made" by the review of reviewers. On the contrary, papers with real theoretical innovation are often difficult to be understood and uneasy to pass the review, and the review process can hardly play a positive role except causing publication delay. For example, Black and Scholes' option pricing papers have been rejected several times; Treynor's paper on capital asset pricing was not published; Sharpe's paper was published after being delayed for two years because the journal changed its judges.

Anyway, maybe we should all be grateful. If those major discoveries happened in recent years, or if the manuscript review in those early years were strengthened as perfect as today, and the empirical research was as popular as today, I'm afraid we wouldn't have chance to see the Black–Scholes model, MM model and Sharpe capital asset pricing model, etc.; they have no way to be published under the close surveillance of nowadays anonymous review. The author found that Mr. Merton still insists on doing some research with theoretical and applied value, but these research findings can only be published in "non advanced" or "low level" journals.

However, in addition to being grateful, we should also ask: what is the reason for such a design and implementation of the review system? What is the intention? This problem deserves serious consideration by every university, every proessional journal and every financial scholar. A fact that cannot be ignored is that the recent decades stagnancy of financial theory coincide with the strengthening of review standards and assessment systems in financial research. Even if the review standard is correct and execute properly, the review may not benefit scientific development, because authors rather than reviewers are the driving force for financial theory progress. Under incorrect standards, review is bound to be harmful to scientific research. Limiting and controlling academic research through anonymous and secret review is only in line with the wishes of a few scholars, which will inevitably frustrate the enthusiasm and creativity of the majority of scholars and cause the stagnation of financial theory.

Since the current review system has seriously hindered the progress of financial science, all forms of anonymous reviews should be abolished, and transparency and fairness should be inhanced in various review occasions, and the review should be supervised by editors, authors and readers, and give authors and applicants sufficient "right to response or discuss". If there are no qualified reviewers with both ability and virtue, the control of review should be gotten rid of and give all manuscripts equal opportunities to publish, such as all manuscripts are put online so that readers can read and evaluate them; all scholars for scientific research should be given equal funds. Perhaps this belongs to equalitarianism, but in any case, equalitarianism is much better than reverse incentive. In this way, the inherent development power of the financial discipline can be released, and the hope for financial theory to make progress may come back.

Of course, if necessary, the review system can also be improved through some reward and punishment policies. For example, if the review recommends papers with significant innovation and theoretical contributions, awards or honors will be granted; If papers with significant innovation and theoretical contributions are rejected by mistake, some punishment will be imposed. Major innovations and theoretical contributions cannot be judged in a short time. Therefore, perhaps the simplest way is to abandon the review system. The reason is very simple. If scholars only write papers when they are inspired and understand or solve a professional problem that others have not solved; Then, the author is likely to surpass other scholars on the topic, including the reviewers. Such an article is certainly worth publishing and does not need expert review. Perhaps the editors of the journal can improve the manuscript by polishing the words, but the so-called peer review opinions are superfluous.

Note that in the absence of research performance assessment pressure, financial scholars do scientific research with intellectual hunger and professional consciousness, and will not irresponsibly "mass produce" papers. If scholars do not write papers without inspiration, innovation and problem solving, the number of submissions to professional journals and thus the rejection rate of submissions will decrease significantly, and the supply of journal pages may be enough to satisfy or even exceeds the demand. Since empirical research belongs to statistical research, empirical research papers should be submitted to various statistical journals. Professional financial journals should mainly accept and publish financial theory and application papers. Then, the number of journals may be enough to satisfy the number of submissions, and there is no need for review to screen. Obviously, the above suggestions are fully feasible.

3.3 Looking Forward to "Dawn Light"

Due to various reasons, a scientific research system of "emphasizing assessment and manuscript review but despising research" has gradually formed in the research of financial science. This system seriously violates the inherent law of scientific research and development. Looking back at human history, we can see that all major scientific discoveries in history were made by research, not by assessment, nor by manuscript review. Such scientific research system provides reverse or negative incentive, hinders scientific research and its publication, led to huge wastes of research resources, such as scientific research funds, professional journal pages, as well as time and energies of scholars. That is the main reason why the financial theory has stagnated over decades and many fundamental financial problems remain unsolved.

The current financial scientific research assessment and manuscript review system came from the choice made decades ago. After the choice, in addition to continuous strengthening, there is no necessary reflection, discussion and correction. However, the "paper number, rank and money" emphasized in the current assessment are actually irrelevant or even contrary to scientific research achievements and scientific development; The current manuscript review system is only conducive to ensuring that a few people "act as both referees and athletes, secretly judge and have supreme power", which is not conducive to mobilizing the enthusiasm and creativity of the majority of scholars. Such a performance assessment and manuscript review system greatly encourage guess based research; The prosperity of financial research in recent decades is actually a flood of guess based research. This leads to the unprecedented prosperity of financial research in decades on the surface, the extremely limited theoretical progress in essence, and even the basic decision-making nature of the financial discipline is becoming more and more ambiguous, and many mistakes and omissions in the financial textbooks cannot be corrected.

Empirical research is based on the conclusion or model guessed, which is tested or regressed with sample data. There is no rigorous logic behind these guesses, and they have to be proved their correctness in line with the reality. However, conforming to reality is the pursuit of descriptive science, not the pursuit of decisional science—decisional science should seek to improve decision-making, not conform to the result of past decision-making. For decisional science, what should be emphasized is that the premise conforms to reality, not that the conclusion conforms to reality. Therefore, this "mainstream" research is fundamentally wrong in the "main category" of social science. For decisional science, if we want to conform to reality, reality itself is the most realistic. Why do we have to study or research on it? On the other hand, only certain results may be obtained based on rigorous logical reasoning, while countless hypotheses can be proposed based on guessing and actual data. Therefore, after the prevalence of empirical research, the research literature is overwhelming, and the research conclusions are diverse. Unfortunately, due to the wrong subject "category" and not subject to professional logic constraints, researchers' guesses are becoming more and more random and bizarre. For example, the ratio of face length to width and the number of meetings in a year have been guessed or assumed to be the factors affecting the company's performance; For another example, human characteristics such as beauty, ugliness, obesity and thinness have been guessed or assumed to be factors affecting stock analysis. Understandingly, such research is basically no longer scientific research.

Obviously, the reverse or negative incentive system of scientific research assessment and submission review are seriously harmful and destructive to financial research and financial theory, and must be abandoned. Even if we can't find a perfect assessment and review system at the moment, it should be abandoned as soon as possible, because such a system is always generating reverse or negative incentives, and it's worse than no performance assessment and submission review. Scientific research management is different from business management. In the field of scientific research, abandoning assessment, especially the current unreasonable assessment, will release the scientific research vitality of scholars and promote the improvement of scientific research performance. At the same time, scholars will no longer irresponsibly produce large quantities of junk papers, so the pages of financial journals will be released and be adequately supplied. At the same time, there is no need for submission reviews. In this way, scientific research will return to normal order. Financial research will hopefully get out of the quagmire of stagnation over decades, and major breakthroughs in financial theory will be just around the corner.