МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

О.Л.Жиров

ПРИКЛАДНА СТАТИСТИКА

ПРАКТИКУМ

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за освітньою програмою «Системний аналіз і управління. Системи і методи штучного інтелекту»
спеціальності 122 Комп'ютерні науки, 124 Системний аналіз

Електронне мережне навчальне видання

Київ
КПІ ім. Ігоря Сікорського
2023

Рецензент *Бідюк П.І.*, д.т.н., професор кафедри ММСА,
ННІПСА НТУУ «КПІ імені Ігоря Сікорського»

Відповідальний
редактор *Данилов В.Я.*, д.т.н., професор кафедри ШІ

Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 8 від 02.06.2023 р.)
за поданням Вченої ради Навчально-наукового інституту прикладного системного аналізу
(протокол № 5 від 29.05.2023 р.)

В навчальному посібнику викладено методики виконання практичних завдань, передбачених силабусом з предмету «Прикладна статистика», та містить варіанти індивідуальних завдань. Посібник забезпечує проведення практичних занять з предмету «Прикладна статистика» для здобувачів ступеня бакалавр за спеціальностями 122 Комп'ютерні науки, 124 Системний аналіз.

Реєстр. № НП 22/23-742. Обсяг 4,32 авт. арк.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Перемоги, 37, м. Київ, 03056
https://kpi.ua
Свідоцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

© О.Л.Жиров
© КПІ ім. Ігоря Сікорського, 2023

ЗМІСТ

ВСТУП 4

Практичні заняття №1-2. Моделювання погодинної заробітної платні. Лінійна модель регресії з фіктивними змінними……. 6

Практичне заняття №3. Нелінійна модель погодинної заробітної платні 24

Практичне заняття №4. Моделювання залежності попиту від ціни методом інструментальних змінних 30

Практичне заняття №5. Аналіз часового ряду відношень курсів валют долара та євро. Одномірні моделі часових рядів 44

Практичне заняття №6. Аналіз часового ряду індексу цін на непродовольчі товари 55

Практичне заняття № 7. ARCH та GARCH моделі 65

Практичне заняття №8. Модель ціноутворення фінансових активів (CAPM) як модель регресії 74

Використана література 95

ВСТУП

Прикладна статистика є однією з фундаментальних дисциплін в освітній програмі бакалаврів з системного аналізу, оскільки багато бакалаврських робіт аналізує часові ряди фінансових і виробничих процесів. Прикладна статистика є складовою частиною системного аналізу, оскільки еквівалентною назвою прикладної статистики є статистичний аналіз даних.

В даному посібнику розглядаються ряд системних задач, які є важливими для більшості організацій і підприємств, незалежно від форм власності. Здатність вирішувати такі системні задачі, створювати відповідне програмне забезпечення для конкретних системних задач, засвоювати і використовувати новітні комп'ютерні системи для метою вирішення конкретних системних задач є перевагою випускників ІПСА. Посібник доповнює навчальний посібник [1].

Даний посібник і посібник [1] створені на основі першоджерел [3-8].

Індивідуальні дані для виконання практичних завдань обираються узгоджено з викладачем із статистичних сайтів європейських країн, які вказані в додатках до навчального посібника [1] .

Метою практичних занять є:

- вивчення та оволодіння методами збору та обробки статистичних даних у формі часових перерізів рядів;
- вивчення та оволодіння методами побудови статистичних і математичних моделей на основі функціонального підходу;
- вивчення критеріальної бази для дослідження якості даних і адекватності створених моделей;

- уміння аналізувати як структура даних впливає на розв'язання задачі статистичного моделювання, прогнозування та формування статистичних висновків;

- уміння перевіряти відповідність вхідних статистичних даних постановці задачі статистичного і математичного моделювання на основі отриманих даних;

- уміння розробляти статистичні моделі досліджуваних процесів в економіці та фінансах;

- уміння доводити коректність та ефективність побудованих статистичних і математичних моделей на розв'язуванні прикладних задач;

- уміння порівнювати різні методи моделювання, прогнозування, оцінювання ризиків та формування статистичних висновків.

Практичні заняття №1-2

Моделювання погодинної заробітної платні, лінійна модель регресії з фіктивними змінними

Мета роботи:

Засвоїти принцип найменших квадратів для оцінки простої (парної) і множинної моделі лінійної регресії. Набуття досвіду інтерпретації і порівняння моделей з різною кількістю факторів. Засвоїти використання фіктивниих змінних на прикладі врахування гендерних ефектів.

Зауваження. Всі практичні заняття мають підвищену складність і орієнтовані на студентів ІПСА. Оскільки на першому занятті викладач лише доводить план роботи на семестр, крім того, у студентів має бути час для первинного освоєння статистичного пакету Eviews, тому перше завдання розраховане на 2 практичних заняття. На восьми заняттях виконується сім завдань.

Вихідні дані (wages1.wf1):

SCHOOL – кількість років навчання

EXPER – досвід роботи після навчання.

MALE – фіктивна змінна, рівна 1 для чоловіків та 0 для жінок

WAGE – погодинна заробітна платня, доларів/год

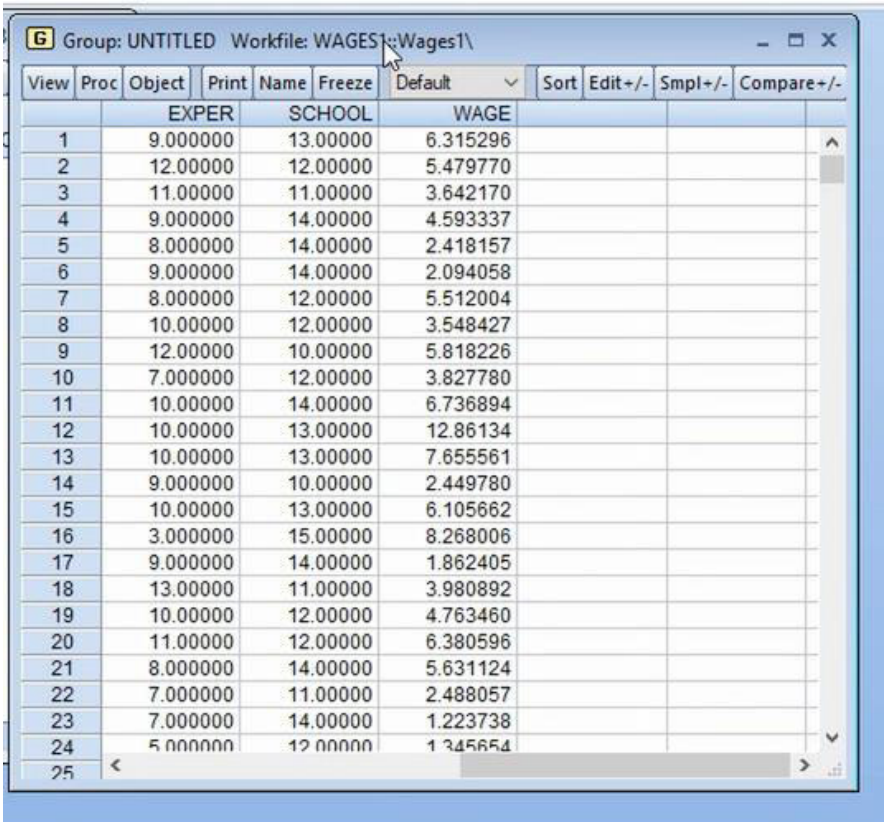
Завдання:

1. Визначити дескриптивні статистики і кореляційну матрицю для вихідних змінних.
2. Перевірити тест на рівність середніх значень погодинної заробітної платні для чоловіків та для жінок.
3. Оцінити модель вигляду WAGE=β_1+β_2 MALE+ε

4. Проінтерпретувати коефіцієнт детермінації, перевірити на значущість рівняння і коефіцієнтів рівняння.
5. Побудувати 95% довірчий інтервал для коефіцієнта регресії.
6. Оцініти модель впливу людського капіталу на заробітну платню в лінійній формі $WAGE=\beta_1+\beta_2 MALE+\beta_3 SCHOOL+\beta_4 EXPER+\varepsilon$
7. Провести тест Уальда на значення покращення оцінки за рахунок нових змінних.
8. По оцінці рівняння отримати розрахункові значення, побудуйте графіки розрахункових та фактичних значень заробітної плати, проведіть візуальне порівняння. Охарактеризувати точність розрахункових значень по відношенню помилки апроксимації MAPE.

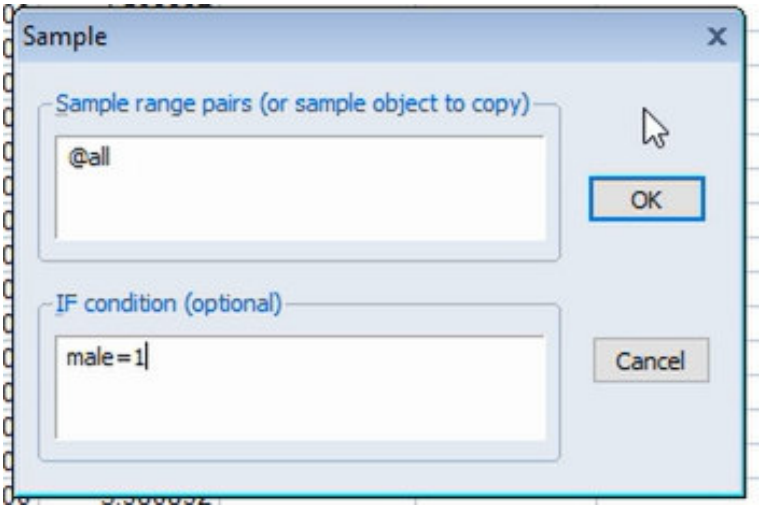
<u>Розв'язання:</u>

Об'єднаємо дані, що нас цікавлять: exper, school і wage, та відфільтруємо їх відносно чоловіків.



	EXPER	SCHOOL	WAGE
1	9.000000	13.00000	6.315296
2	12.00000	12.00000	5.479770
3	11.00000	11.00000	3.642170
4	9.000000	14.00000	4.593337
5	8.000000	14.00000	2.418157
6	9.000000	14.00000	2.094058
7	8.000000	12.00000	5.512004
8	10.00000	12.00000	3.548427
9	12.00000	10.00000	5.818226
10	7.000000	12.00000	3.827780
11	10.00000	14.00000	6.736894
12	10.00000	13.00000	12.86134
13	10.00000	13.00000	7.655561
14	9.000000	10.00000	2.449780
15	10.00000	13.00000	6.105662
16	3.000000	15.00000	8.268006
17	9.000000	14.00000	1.862405
18	13.00000	11.00000	3.980892
19	10.00000	12.00000	4.763460
20	11.00000	12.00000	6.380596
21	8.000000	14.00000	5.631124
22	7.000000	11.00000	2.488057
23	7.000000	14.00000	1.223738
24	5.000000	12.00000	1.345654
25			

group01



Group: GROUP01 Workfile: WAGES1::Wages1\

View | Proc | Object | Print | Name | Freeze | Default | Sor

	EXPER	SCHOOL	WAGE
1571	7.000000	14.00000	2.933289
1572	12.00000	11.00000	6.219131
1573	12.00000	12.00000	6.915378
1574	11.00000	11.00000	4.946563
1575	12.00000	12.00000	5.134564
1576	10.00000	12.00000	3.062224
1577	7.000000	11.00000	5.587216
1578	11.00000	12.00000	7.655561
1579	10.00000	14.00000	16.65645
1580	9.000000	15.00000	3.216882
1581	12.00000	12.00000	6.828760
1582	8.000000	11.00000	1.032880
1583	11.00000	12.00000	3.490936
1584	10.00000	12.00000	6.130999
1585	11.00000	13.00000	4.638444
1586	7.000000	15.00000	7.961783
1587	8.000000	15.00000	8.945824
1588	9.000000	13.00000	2.449780
1589	11.00000	12.00000	2.632941
1590	14.00000	8.000000	6.189372
1591	9.000000	12.00000	2.041483
1592	10.00000	11.00000	1.092794
1593	8.000000	14.00000	0.734934
1594	11.00000	12.00000	3.503185
1595			

Після фільтрації отримали вибірку в 1725 об’єктів

Далі розлянемо статистичні характеристики даної групи

Group: GROUP01 Workfile: WAGES1::Wages1\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

	EXPER	SCHOOL	WAGE
Mean	8.326377	11.44232	6.313021
Median	8.000000	12.00000	5.654321
Maximum	18.00000	16.00000	39.80892
Minimum	2.000000	3.000000	0.153480
Std. Dev.	2.359172	1.737319	3.498861
Skewness	0.260625	-0.475031	1.919730
Kurtosis	3.794166	4.240833	11.81645
Jarque-Bera	64.86015	175.5392	6646.368
Probability	0.000000	0.000000	0.000000
Sum	14363.00	19738.00	10889.96
Sum Sq. Dev.	9595.250	5203.511	21105.26
Observations	1725	1725	1725

Table: DESCRIPT_MALE Workfile: WAGES1::Wages1\

View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comments+

Date: 11/01/20 Time: 20:23
Sample: 1 3294 IF MALE=1

	EXPER	SCHOOL	WAGE
Mean	8.326377	11.44232	6.313021
Median	8.000000	12.00000	5.654321
Maximum	18.00000	16.00000	39.80892
Minimum	2.000000	3.000000	0.153480
Std. Dev.	2.359172	1.737319	3.498861
Skewness	0.260625	-0.475031	1.919730
Kurtosis	3.794166	4.240833	11.81645
Jarque-Bera	64.86015	175.5392	6646.368
Probability	0.000000	0.000000	0.000000
Sum	14363.00	19738.00	10889.96
Sum Sq. Dev.	9595.250	5203.511	21105.26
Observations	1725	1725	1725

Descript_male

Далі створимо матрицю кореляції.

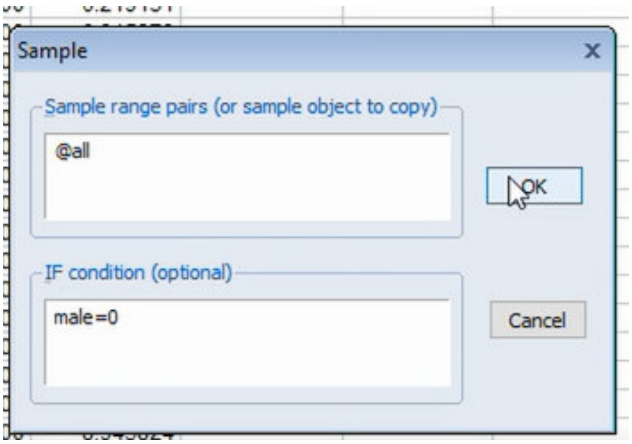
View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Covariance Analysis: Ordinary
Date: 11/01/20 Time: 20:25
Sample (adjusted): 1570 3294
Included observations: 1725 after adjustments

Correlation t-Statistic Probability	EXPER	SCHOOL	WAGE
EXPER	1.000000 ----- -----		
SCHOOL	-0.273282 -11.79259 0.0000	1.000000 ----- -----	
WAGE	-0.012973 -0.538547 0.5903	0.322171 14.12618 0.0000	1.000000 ----- -----

Correl_male

Тепер відфільтруємо відносно жінок. Та відтворимо ті самі дії, що і для чоловіків.



Sample

View | Proc | Object | Print | Name | Freeze | Default

	EXPER	SCHOOL	WAGE
1	9.000000	13.00000	6.315296
2	12.00000	12.00000	5.479770
3	11.00000	11.00000	3.642170
4	9.000000	14.00000	4.593337
5	8.000000	14.00000	2.418157
6	9.000000	14.00000	2.094058
7	8.000000	12.00000	5.512004
8	10.00000	12.00000	3.548427
9	12.00000	10.00000	5.818226
10	7.000000	12.00000	3.827780
11	10.00000	14.00000	6.736894
12	10.00000	13.00000	12.86134
13	10.00000	13.00000	7.655561
14	9.000000	10.00000	2.449780
15	10.00000	13.00000	6.105662
16	3.000000	15.00000	8.268006
17	9.000000	14.00000	1.862405
18	13.00000	11.00000	3.980892
19	10.00000	12.00000	4.763460
20	11.00000	12.00000	6.380596
21	8.000000	14.00000	5.631124
22	7.000000	11.00000	2.488057
23	7.000000	14.00000	1.223738
24	5.000000	12.00000	1.345654
25			

Після фільтрації отримали вибірку в 1569 об’єктів

Далі розлянемо статистичні характеристики даної групи

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

	EXPER	SCHOOL	WAGE
Mean	7.732314	11.83748	5.146924
Median	8.000000	12.00000	4.693262
Maximum	16.00000	16.00000	32.49740
Minimum	1.000000	5.000000	0.076556
Std. Dev.	2.171471	1.539355	2.876237
Skewness	-0.063330	-0.172562	1.975137
Kurtosis	3.211163	3.894340	13.95051
Jarque-Bera	3.963838	60.07664	8859.515
Probability	0.137805	0.000000	0.000000
Sum	12132.00	18573.00	8075.524
Sum Sq. Dev.	7393.572	3715.556	12971.66
Observations	1569	1569	1569

Table: UNTITLED Workfile: WAGES1::Wages1\

View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comments+

Date: 11/01/20 Time: 20:31
Sample: 1 3294 IF MALE=0

	EXPER	SCHOOL	WAGE
Mean	7.732314	11.83748	5.146924
Median	8.000000	12.00000	4.693262
Maximum	16.00000	16.00000	32.49740
Minimum	1.000000	5.000000	0.076556
Std. Dev.	2.171471	1.539355	2.876237
Skewness	-0.063330	-0.172562	1.975137
Kurtosis	3.211163	3.894340	13.95051
Jarque-Bera	3.963838	60.07664	8859.515
Probability	0.137805	0.000000	0.000000
Sum	12132.00	18573.00	8075.524
Sum Sq. Dev.	7393.572	3715.556	12971.66
Observations	1569	1569	1569

Descript_female

Далі розрахуємо кореляційну матрицю.

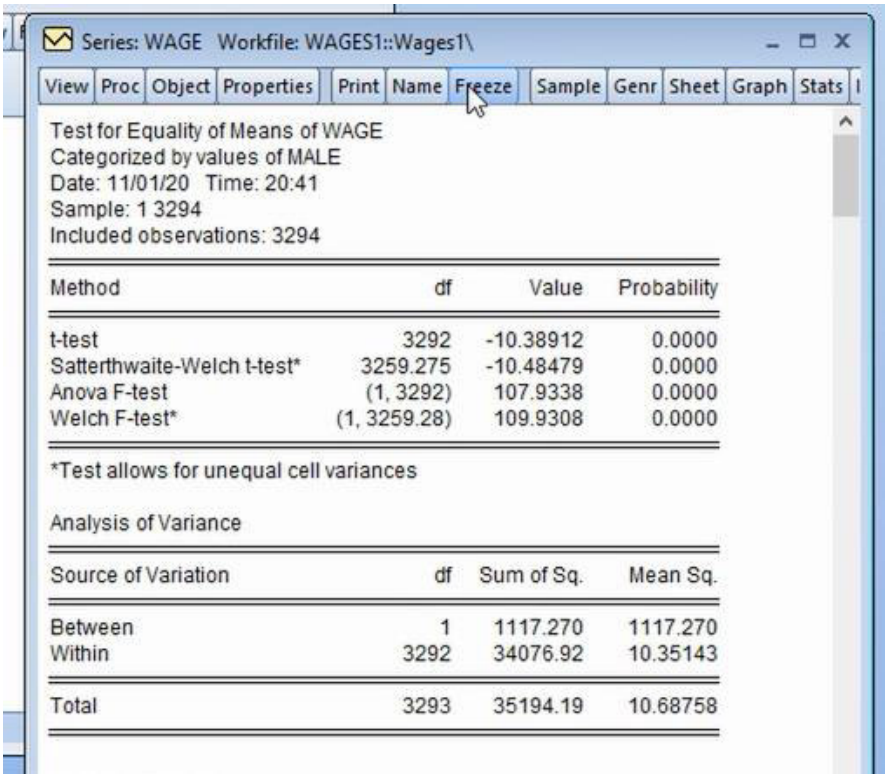
View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Covariance Analysis: Ordinary
Date: 11/01/20 Time: 20:34
Sample (adjusted): 1 1569
Included observations: 1569 after adjustments

Correlation t-Statistic Probability	EXPER	SCHOOL	WAGE
EXPER	1.000000 ----- -----		
SCHOOL	-0.056524 -2.241111 0.0252	1.000000 ----- -----	
WAGE	0.090110 3.581615 0.0004	0.294314 12.19046 0.0000	1.000000 ----- -----

Correl_female

Розглянемо спільну тенденцію заробітної плати.



Series: WAGE Workfile: WAGES1::Wages1\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats

Test for Equality of Means of WAGE
Categorized by values of MALE
Date: 11/01/20 Time: 20:41
Sample: 1 3294
Included observations: 3294

Method	df	Value	Probability
t-test	3292	-10.38912	0.0000
Satterthwaite-Welch t-test*	3259.275	-10.48479	0.0000
Anova F-test	(1, 3292)	107.9338	0.0000
Welch F-test*	(1, 3259.28)	109.9308	0.0000

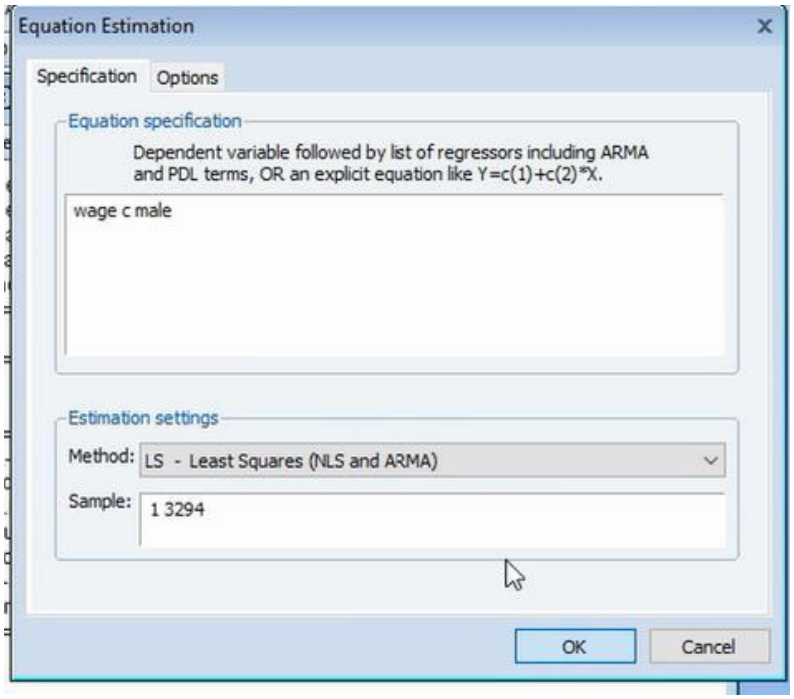
*Test allows for unequal cell variances

Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	1	1117.270	1117.270
Within	3292	34076.92	10.35143
Total	3293	35194.19	10.68758

Wage_test

Перевіримо статистичну значущість середніх значень заробітної платні чоловіків та жінок.



Equation Estimation

Specification | Options

Equation specification

Dependent variable followed by list of regressors including ARMA and PDL terms, OR an explicit equation like Y=c(1)+c(2)*X.

wage c male

Estimation settings

Method: LS - Least Squares (NLS and ARMA)

Sample: 1 3294

OK | Cancel

Table: UNTITLED Workfile: WAGES1::Wages1\

View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comments+/-

Dependent Variable: WAGE
Method: Least Squares
Date: 11/01/20 Time: 20:44
Sample: 1 3294
Included observations: 3294

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.146924	0.081225	63.36639	0.0000
MALE	1.166097	0.112242	10.38912	0.0000

R-squared	0.031746	Mean dependent var	5.757585
Adjusted R-squared	0.031452	S.D. dependent var	3.269186
S.E. of regression	3.217364	Akaike info criterion	5.175609
Sum squared resid	34076.92	Schwarz criterion	5.179313
Log likelihood	-8522.228	Hannan-Quinn criter.	5.176935
F-statistic	107.9338	Durbin-Watson stat	1.866250
Prob(F-statistic)	0.000000		

Table01

Equation: UNTITLED Workfile: WAGES1::Wages1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

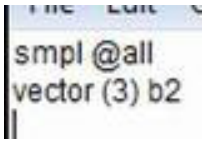
Dependent Variable: WAGE
Method: Least Squares
Date: 11/01/20 Time: 20:44
Sample: 1 3294
Included observations: 3294

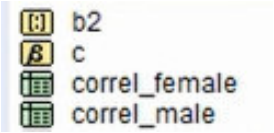
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.146924	0.081225	63.36639	0.0000
MALE	1.166097	0.112242	10.38912	0.0000

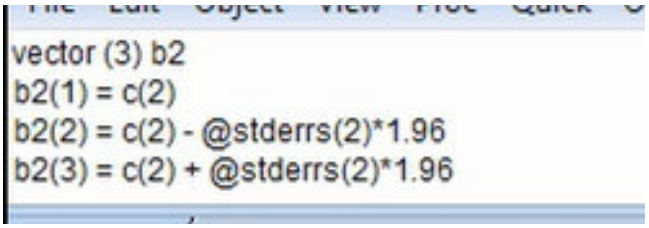
R-squared	0.031746	Mean dependent var	5.757585
Adjusted R-squared	0.031452	S.D. dependent var	3.269186
S.E. of regression	3.217364	Akaike info criterion	5.175609
Sum squared resid	34076.92	Schwarz criterion	5.179313
Log likelihood	-8522.228	Hannan-Quinn criter.	5.176935
F-statistic	107.9338	Durbin-Watson stat	1.866250
Prob(F-statistic)	0.000000		

Eq01

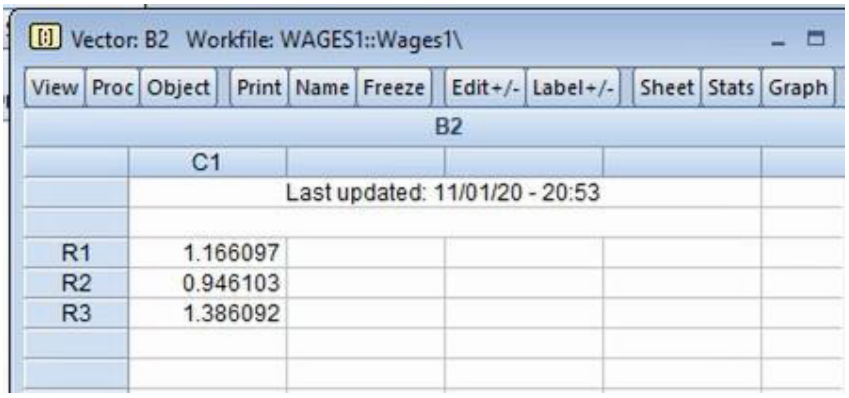
Щоб отримати інтервальну оцінку значень коефіцієнтів регресії ми створимо новий об’єкт: тривимірній вектор b2.



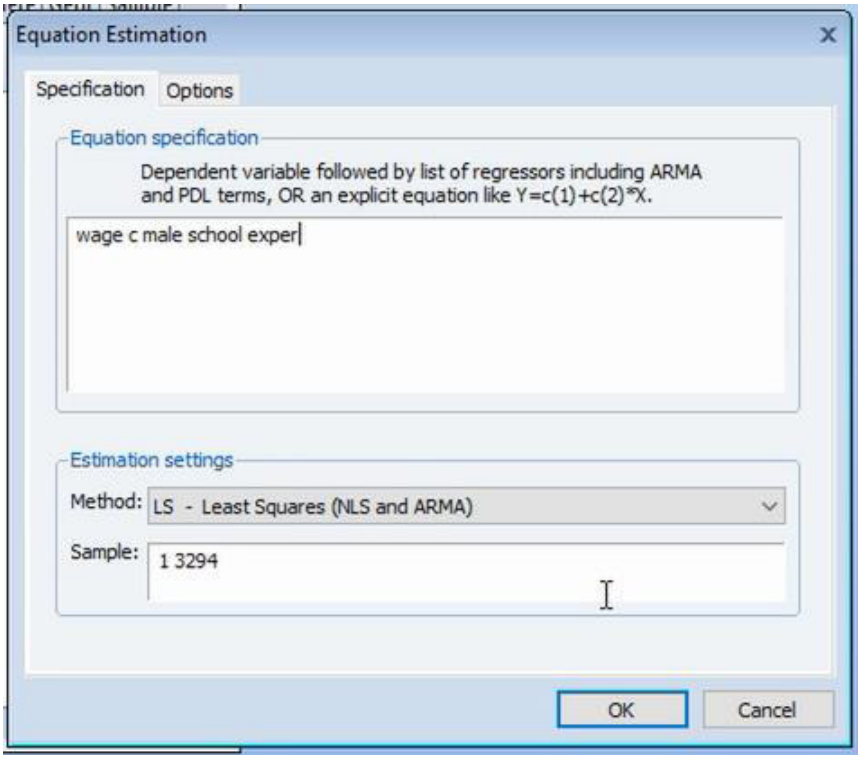




В результаті отримали вектор, що містить середнє значення заробітної плати, а також межі довірчого інтервалу.



Щоб включити в залежність також досвід роботи і отриману освіту, маємо:



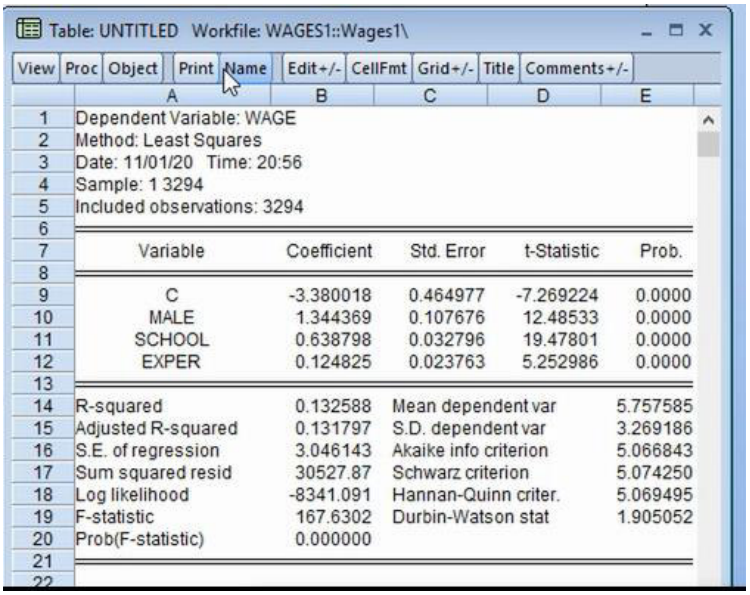


Table: UNTITLED Workfile: WAGES1::Wages1\

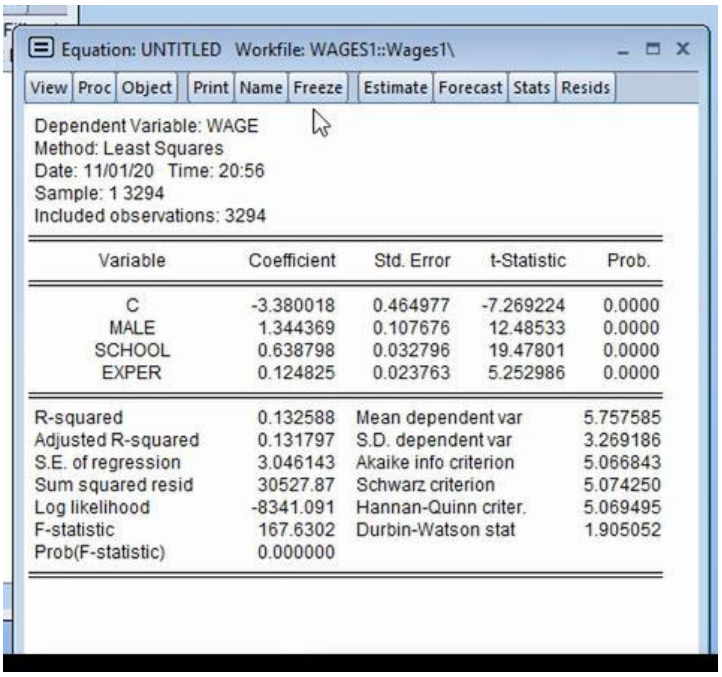
View | Proc | Object | Print | Name | Edit+/- | CellFmt | Grid+/- | Title | Comments+/-

Dependent Variable: WAGE
Method: Least Squares
Date: 11/01/20 Time: 20:56
Sample: 1 3294
Included observations: 3294

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.380018	0.464977	-7.269224	0.0000
MALE	1.344369	0.107676	12.48533	0.0000
SCHOOL	0.638798	0.032796	19.47801	0.0000
EXPER	0.124825	0.023763	5.252986	0.0000

R-squared	0.132588	Mean dependent var	5.757585
Adjusted R-squared	0.131797	S.D. dependent var	3.269186
S.E. of regression	3.046143	Akaike info criterion	5.066843
Sum squared resid	30527.87	Schwarz criterion	5.074250
Log likelihood	-8341.091	Hannan-Quinn criter.	5.069495
F-statistic	167.6302	Durbin-Watson stat	1.905052
Prob(F-statistic)	0.000000		

Table02



Equation: UNTITLED Workfile: WAGES1::Wages1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

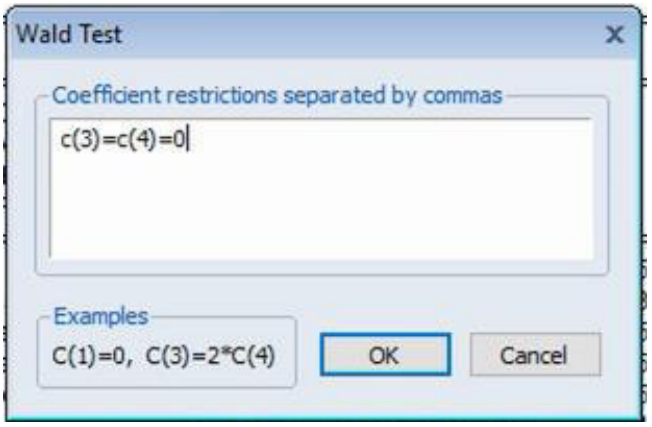
Dependent Variable: WAGE
Method: Least Squares
Date: 11/01/20 Time: 20:56
Sample: 1 3294
Included observations: 3294

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.380018	0.464977	-7.269224	0.0000
MALE	1.344369	0.107676	12.48533	0.0000
SCHOOL	0.638798	0.032796	19.47801	0.0000
EXPER	0.124825	0.023763	5.252986	0.0000

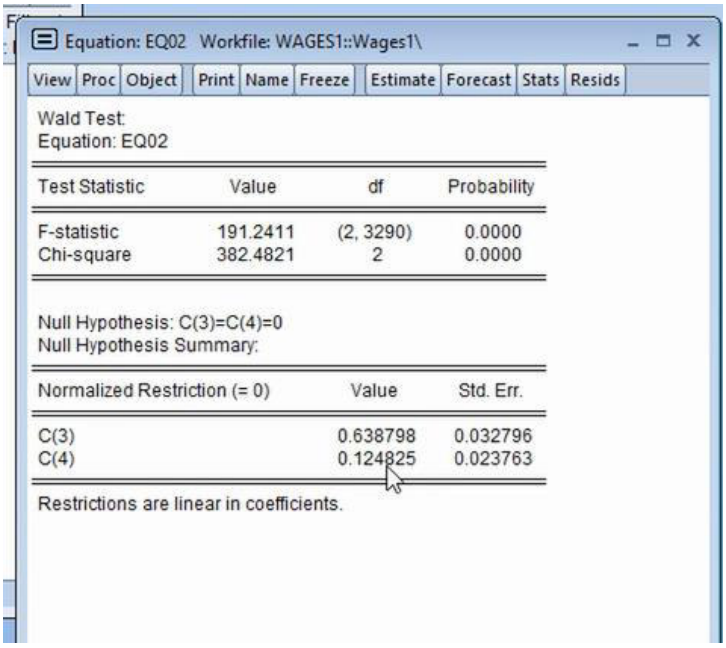
R-squared	0.132588	Mean dependent var	5.757585
Adjusted R-squared	0.131797	S.D. dependent var	3.269186
S.E. of regression	3.046143	Akaike info criterion	5.066843
Sum squared resid	30527.87	Schwarz criterion	5.074250
Log likelihood	-8341.091	Hannan-Quinn criter.	5.069495
F-statistic	167.6302	Durbin-Watson stat	1.905052
Prob(F-statistic)	0.000000		

Eq02

Тест Уальда:

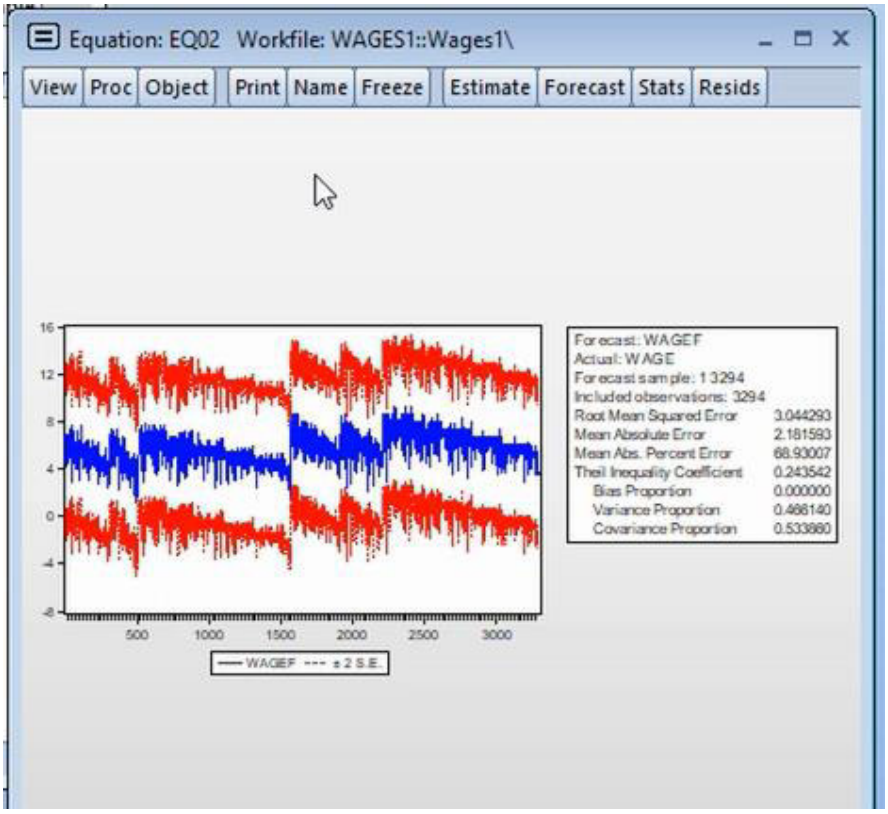


Це означає рівність коефіцієнтів при факторах school і exper.



Wald_eq02

Отримаємо розрахункові значення.



Wagef

Отримали графік розрахункового значення та його довірчого інтервалу.

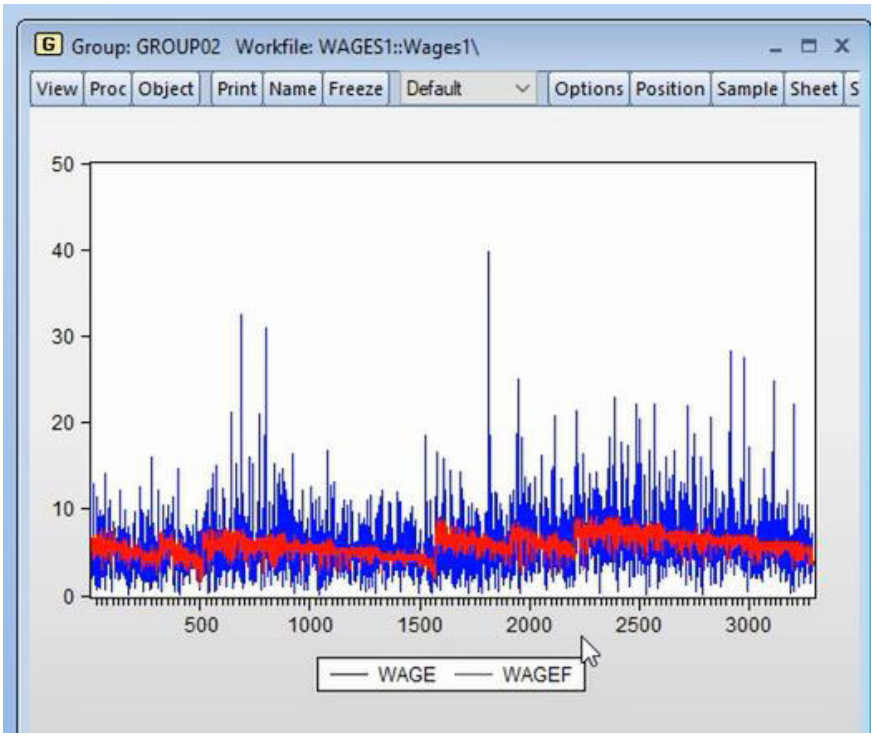
Побудуємо графіки розрахункових та фактичних значень заробітної платні для змінних wage і wagef.

Group: UNTITLED Workfile: WAGES1::V

View | Proc | Object | Print | Name | Freeze | D

	WAGE	WAGEF
1	6.315296	6.047781
2	5.479770	5.783460
3	3.642170	5.019836
4	4.593337	6.686579
5	2.418157	6.561753
6	2.094058	6.686579
7	5.512004	5.284158
8	3.548427	5.533809
9	5.818226	4.505864
10	3.827780	5.159332
11	6.736894	6.811404
12	12.86134	6.172606
13	7.655561	6.172606
14	2.449780	4.131388
15	6.105662	6.172606
16	8.268006	6.576424
17	1.862405	6.686579
18	3.980892	5.269487
19	4.763460	5.533809
20	6.380596	5.658634
21	5.631124	6.561753
22	2.488057	4.520535
23	1.223738	6.436928
24	1.345654	4.909681

Group02



Graph02

Всі зміни збережено в wages2.

Висновок

Для тих респондентів, зарплата яких тільки більш чим на одне стандартне відхилення відрізняється від середнього, різниця між фактичним та розрахунковими значеннями доволі велика. Розрахункові значення близькі до фактичних тільки для зарплат, які мало відрізняються від середніх, що, безумовно, говорить про дуже низьку якість моделі.

Індивідуальні дані для виконання практичного завдання №1-2

Якщо Ви народились n-го дня місяця, неважливо якого саме місяця і якого року, то:

1) Відкиньте перших 50xn із вимірів , на яких ми вивчаємо методику використання фіктивних змінних. Дані включають кризовий 2008 рік і відкидання перших змінних призводить до зміни якостей моделей, а відкидання значної кількості перших даних може призвести до суттєвого ускладнення обробки;

2) Повторіть самостійно виконання практичного завдання і викладіть на сайті групи в Ваш особистий каталог із звітами з прикладної статистики результати;

3) після заверщення курсу рекомендується залишити Ваш труд за весь семестр, бо ми з Вами протягом цього семестру навчимося будувати і досліджувати регресійні моделі різного типу складності, включно з авторегреійними моделями. Це буде базою для вивчення і застосування суттєво складніших регресійний моделей в бакалавроській роботі.

Практичне заняття №3

Нелінійна модель погодинної заробітної платні

<u>Мета роботи:</u>

Застосування нелінійних моделей, використання фіктивних змінних взаємодії і метода покрокового включення-виключення змінних, вивчення методів тестування специфікації рівняння регресії для вкладених і невкладених моделей

<u>Вихідні дані(wages1.wf1):</u>

EARNINGS=поточна погодинна заробітна плата в доларах;

EXPER=досвід роботи в роках;

MALE=фіктивна змінна, рівна 1 для респондентів-чоловіків;

S=число завершених років навчання;

EDUCBA=фіктивна змінна, рівна 1 для респондента з степенем бакалавра;

EDUCMAST=фіктивна змінна, рівна 1 для респондентів зі степенем магістра;

EDUCPHD=фіктивна змінна, рівні 1 для респондентів з докторською степеню;

Завдання:

1. Оцініть напівлогарифмічну модель(модель А)

LOG(EARNINGS) = $\beta_1 + \beta_2 MALE + \beta_3 EXPER + \beta_4 + \varepsilon^8$ (1)

2. Оцініть модель, що дозволяє відповісти на питання, чи однаково впливають досвід роботи і освіта на заробітну плату чоловіків і жінок, додавши до рівняння(3.1) члени, що відповідають за взаємодію пояснюючих факторів з фіктивною змінною, тобто

MALE*EXPER и MALE*S (рівняння 2)

LOG(EARNINGS)=$\beta_1 + \beta_2 MALE + \beta_3 EXPER + \beta_4 S + \beta_4 MALE * EXPER + \beta_5 MALE * S + \varepsilon$

(2)

Проінтерпретуйте коефіцієнти при змінних взаємодії, перевірте їх на значущість.

3. Використовуючи метод покрокового включення-виключення змінних, оцініть модель тільки зі значимими факторами. Виберіть найкращу модель за формальними критеріями. Проведіть тест Уальда на доцільність введення обмежень, що приводять від моделі з повним набором факторів (рівняння 3.2) до вкладеної моделі тільки зі значимими факторами.

4. Найдіть розрахункове значення заробітної плати по моделі, отриманій в п.3 для жінки і чоловіка з досвідом роботи 10 років і числом років навчання рівним 12.

5. Враховуючи результат п.3, проведіть більш детальний аналіз впливу рівня освіти на заробітну плату. Обмежтесь виявленням впливу досягнутої респондентом степені(бакалавр, магістр, PhD) на заробітну плату, враховуючи досвід і стать респондента.

Оцініть лінійну модель

EARNINGS=$\beta_1 + \beta_2 MALE + \beta_3 EXPER + \beta_4 EDUCBA + \beta_5 EDUCMAST + \beta_5 EDUCPHD + \varepsilon$

(4)

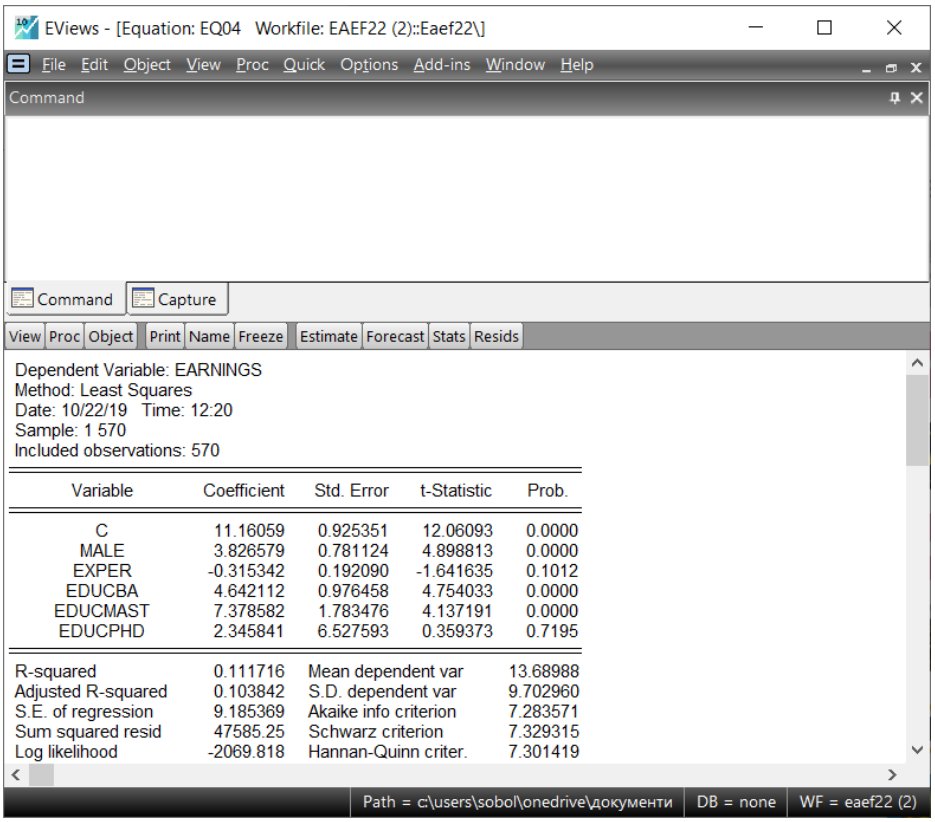
і напівлогарифмічну модель модель (5).

$$LOG(EARNINGS)=\beta_1 + \beta_2 MALE + \beta_3 EXPER + \beta_4 EDUCBA + \beta_5 EDUCMAST + \beta_5 EDUCPHD + \varepsilon \quad (5)$$

Порівняйте результати.

6. Виконайте PE-тестування моделі(3.4) проти моделі (3.5). Виконайте PE-тестування моделі (3.5) проти моделі (3.4). Який ваш висновок?

<u>Розв'язання:</u>



Dependent Variable: EARNINGS
Method: Least Squares
Date: 10/22/19 Time: 12:20
Sample: 1 570
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.16059	0.925351	12.06093	0.0000
MALE	3.826579	0.781124	4.898813	0.0000
EXPER	-0.315342	0.192090	-1.641635	0.1012
EDUCBA	4.642112	0.976458	4.754033	0.0000
EDUCMAST	7.378582	1.783476	4.137191	0.0000
EDUCPHD	2.345841	6.527593	0.359373	0.7195

R-squared	0.111716	Mean dependent var	13.68988
Adjusted R-squared	0.103842	S.D. dependent var	9.702960
S.E. of regression	9.185369	Akaike info criterion	7.283571
Sum squared resid	47585.25	Schwarz criterion	7.329315
Log likelihood	-2069.818	Hannan-Quinn criter.	7.301419

EViews - [Equation: EQ02 Workfile: EAEF22 (2)::Eaef22\]

File Edit Object View Proc Quick Options Add-ins Window Help

Command

Command Capture

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(EARNINGS)
Method: Least Squares
Date: 10/22/19 Time: 11:50
Sample: 1 570
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MALE	1.044247	0.260691	4.005696	0.0001
EXPER	0.001801	0.014422	0.124907	0.9006
S	0.130339	0.013661	9.541113	0.0000
MALE*EXPER	-0.025579	0.019510	-1.311068	0.1904
MALE*S	-0.050646	0.017394	-2.911671	0.0037
C	0.514849	0.203091	2.535070	0.0115

R-squared	0.276215	Mean dependent var	2.456463
Adjusted R-squared	0.269799	S.D. dependent var	0.541347
S.E. of regression	0.462591	Akaike info criterion	1.306522
Sum squared resid	120.6904	Schwarz criterion	1.352266
Log likelihood	-366.3588	Hannan-Quinn criter.	1.324370

Path = c:\users\sobol\onedrive\документи DB = none WF = eaef22 (2)

EViews - [Equation: EQ02 Workfile: EAEF22 (2)::Eaef22\]

File Edit Object View Proc Quick Options Add-ins Window Help

Command

Command Capture

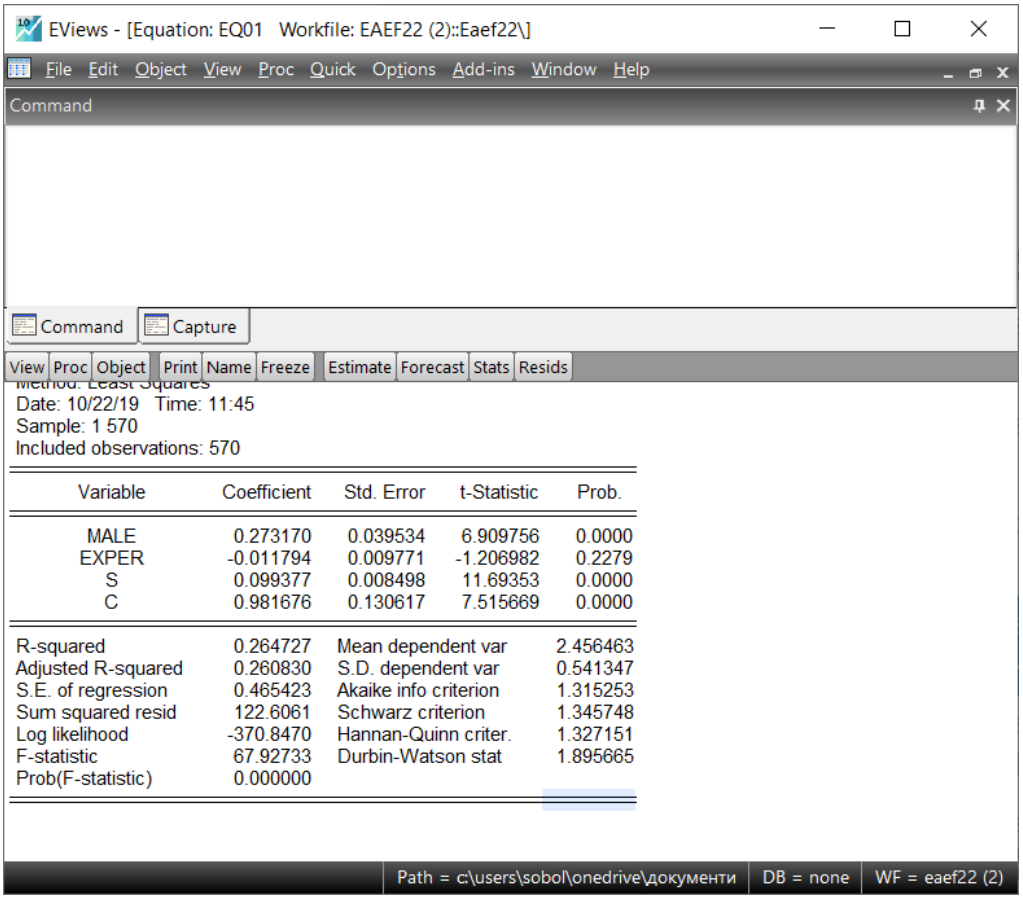
View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOG(EARNINGS)
Method: Least Squares
Date: 10/22/19 Time: 11:50
Sample: 1 570
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MALE	1.044247	0.260691	4.005696	0.0001
EXPER	0.001801	0.014422	0.124907	0.9006
S	0.130339	0.013661	9.541113	0.0000
MALE*EXPER	-0.025579	0.019510	-1.311068	0.1904
MALE*S	-0.050646	0.017394	-2.911671	0.0037
C	0.514849	0.203091	2.535070	0.0115

R-squared	0.276215	Mean dependent var	2.456463
Adjusted R-squared	0.269799	S.D. dependent var	0.541347
S.E. of regression	0.462591	Akaike info criterion	1.306522
Sum squared resid	120.6904	Schwarz criterion	1.352266
Log likelihood	-366.3588	Hannan-Quinn criter.	1.324370

Path = c:\users\sobol\onedrive\документи DB = none WF = eaef22 (2)



EViews - [Equation: EQ01 Workfile: EAEF22 (2)::Eaef22\]

File Edit Object View Proc Quick Options Add-ins Window Help

Command

Command Capture

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Method: Least Squares
Date: 10/22/19 Time: 11:45
Sample: 1 570
Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MALE	0.273170	0.039534	6.909756	0.0000
EXPER	-0.011794	0.009771	-1.206982	0.2279
S	0.099377	0.008498	11.69353	0.0000
C	0.981676	0.130617	7.515669	0.0000

R-squared	0.264727	Mean dependent var	2.456463
Adjusted R-squared	0.260830	S.D. dependent var	0.541347
S.E. of regression	0.465423	Akaike info criterion	1.315253
Sum squared resid	122.6061	Schwarz criterion	1.345748
Log likelihood	-370.8470	Hannan-Quinn criter.	1.327151
F-statistic	67.92733	Durbin-Watson stat	1.895665
Prob(F-statistic)	0.000000		

Path = c:\users\sobol\onedrive\документи DB = none WF = eaef22 (2)

<u>Висновок</u>

1. Результати оцінки Eq01 показує, що при інших рівних умовах (однаковий досвід роботи і кількість років навчання) чоловіки в середньому заробляють на 27% більше, ніж жінки. Чим більше досвіду та чим більша кількість років навчання, тим вища заробітна плата при однакових інших умовах.

2. Коефіцієнт при перехресному члені MALE*EXPER статистично не є значущим , отже, дана оцінка показує, що вплив досвіду роботи на заробітну плату незначуще відрізняється для чоловіків і для жінок. Але коефіцієнт при MALE*S значущий.

3. За формальними критеріями виключення факторів MALE і MALE*EXPER не погіршило якості оцінки, оскільки значення скоригованого коефіцієнта детермінації збільшилось, а значення інформаційних критеріїв Акайке і Шварца зменшились. Як показує результат тесту, відмінності в якості оцінки статистично не значущі, тому для подальшого аналізу вибираємо більш економну модель Eq03.

4. Таким чином, погодинна заробітна плата чоловіків в середньому вище погодинної заробітної плати жінок при однаковому досвіді роботи і рівні освіти. Однак, різниця в заробітній платі між чоловіками з різним рівнем освіти більше, ніж між жінками з такими ж характеристиками.

5. ніж респонденти без степені, респонденти, що мають степінь магістра, Як показує результат оцінки моделі в лінійній формі при інших рівних умовах респонденти, що мають степінь бакалавра, заробляють більше, більше, ніж респонденти без степені, респонденти, що мають степінь доктора (PhD), заробляють більше, ніж респонденти без степені. Згідно напівлогарифмічній моделі, респонденти, що мають степінь бакалавра, заробляють більше, ніж респонденти без степені. Ця оцінка збігається з оцінкою, отриманою згідно лінійної моделі.

6. Як показує результат оцінки, коефіцієнт при останньому доданку статистично незначущий, тому нульова гіпотеза не відхиляється і напівлогарифмічна модель достатня для пояснення впливу вибраних факторів на погодинну заробітну плату. Отриманий висновок в визначеній степені залежить від вибраного рівня значущості, однак нелінійна модель має кращі статистичні характеристики за лінійну, якщо враховувати також значення коефіцієнтів при факторах, що відображають рівень освіти.

Індивідуальні дані для виконання практичного завдання №3

Індивідуальні дані для нелінійної моделі в кожного студента ті ж самі, як в індивідуальних даних для попереднього завдання для побудови лінійної моделі з фіктивними змінними.

Практичне заняття №4

Моделювання залежності попиту від ціни методом інструментальних змінних

Мета: вивчення методу інструментальних змінних (ІЗ), використання двокрокового методу найменших квадратів для оцінки рівняння, вивчення методів тестування специфікації рівняння регресії з використанням тесту Дарбіна-Ву-Хаусмана.

АЛГОРИТМ ДВОКРОКОВОГО МЕТОДУ НАЙМЕНШИХ КВАДРАТІВ

Крок 1. Перевіряється кожне рівняння моделі на ідентифікованість. Якщо рівняння надідентифіковані, то для оцінки параметрів кожного з них можна використати оператор оцінювання:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y \\ X_1'Y \end{bmatrix}.$$

Крок 2. Знаходження добутку матриць поточних ендогенних змінних, які містяться у правій частині моделі, на матрицю всіх екзогенних змінних моделі, тобто $Y_1'X$.

Крок 3. Обчислення матриці $X'X$ і знаходження оберненої матриці $(X'X)^{-1}$.

Крок 4. Визначення добутку матриць всіх екзогенних змінних і ендогенних змінних у правій частині моделі, тобто $X'Y_1$.

Крок 5. Знаходження добутку матриць, які здобуто на кроках 2,3,4, тобто $Y_1'X(X'X)^{-1}X'Y_1$.

Крок 6. Визначення добутку матриць ендогенних змінних у правій частині моделі і екзогенних змінних, які внесені до даного рівняння, тобто $Y_1'X_1$.

Крок 7. Знаходження добутку матриць екзогенних змінних, які входять в дане рівняння, і ендогенних змінних правої частини системи рівнянь, тобто $X_1'Y_1$.

Крок 8. Визначення добутку матриць екзогенних змінних даного рівняння, тобто $X_1'X_1$.

Крок 11. Знаходження матриці, оберненої до блочної:

$$Q_s^{-1} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix}^{-1}.$$

Крок 10. Визначення добутку матриць $X'Y_1$, де X' — матриця всіх екзогенних змінних моделі, Y_1 — вектор залежної ендогенної змінної лівої частини рівняння.

Крок 11. Знаходження добутку матриць:

$$g_s = Y_1'X_1'X_1(X'X)^{-1}X'Y_1.$$

Крок 12. Визначення параметрів моделі:

$$\begin{bmatrix} \hat{a}_s \\ \hat{b}_s \end{bmatrix} = Q_s^{-1} g_s.$$

Крок 13. Обчислення *s*-ї залежної ендогенної змінної на основі знайдених параметрів a_s і b_s:

$$\hat{Y}_s = a_s Y_1 + b_s X_1.$$

Крок 14. Обчислення вектора залишків в *s*-му рівнянні системи:

$$u_s = Y_s - \hat{Y}_s.$$

Крок 15. Визначення дисперсії залишків для кожного рівняння:

$$\sigma_{u_s}^2 = \frac{1}{n-k-r+1} u_s' u_s$$

Крок 16. Знаходження матриці коваріацій для параметрів кожного рівняння:

$$asy\ \mathrm{var}\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \sigma_{u_s}^2 Q_s^{-1}$$

Крок 11. Знаходження стандартної помилки параметрів і визначення довірчих інтервалів:

$$S\begin{bmatrix}\hat{a}\\ \hat{b}\end{bmatrix} = \sqrt{\sigma^2_{u_s} Q_z^{-1}}$$

$$\hat{a} - t_{(\alpha)}S \le a \le \hat{a} + t_{(\alpha)}S;$$

$$\hat{b} - t_{(\alpha)}S \le b \le \hat{b} + t_{(\alpha)}S.$$

Очевидно, що використання двокроеового МНК є складною задачею, на щастя пакет Eviews містить інструментарій двокрокового МНК, що суттєво спрощує використання цього інструментарію.

Вихідні дані: Дані по продажу цигарок по 48 штатам США за 1995р. в цінах базового року

Завдання:

1. Оцінити еластичність попиту на цигарки від ціни методом найменших квадратів
2. Перевірити можливість використання змінних податків в якості інструментів за ендогенності ціни
3. Використовуючи метод ІЗ отрмати консистентну оцінку еластичності попиту по ціні
4. Оцінити еластичність попиту на циагрки від ціни і доходу методом найменших квадратів
5. Оцінити нелінійну модель двокроковим методом найменших квадратів
6. Повторно оцінити модель методом ДМНК, використовуючи вбудований алгоритм Eviews і порівняти результати
7. Провести тест Дарбіна-Ву-Хаусмана для порівняння оцінок коефіцієнтів по методу МНК та ІЗ і зробити висновки

Результати роботи

Імпортуємо з таблиці Excel дані по продажу цигарок в США за 1995р.

State	N	PACKPC	RAVGPRS	PERINC	RTAXSO	RTAXS
Alabama	1	115,8	1,135	29770,69	0,117	0,195
Arizona	2	104,5	1,587	26986,89	0,145	0,216
Arkansas	3	128,5	1,101	31888,56	0,113	0,205
Califonria	4	104,2	1,147	26906,31	0,119	0,198

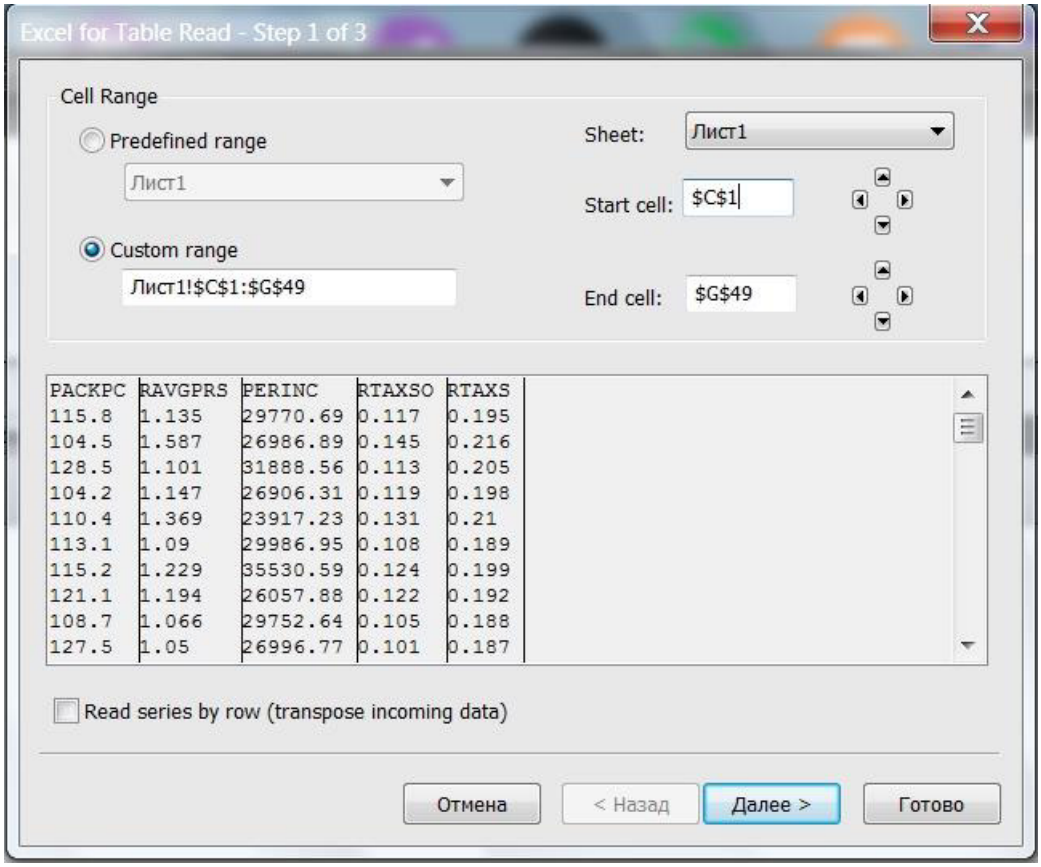
Colorado	5	110,4	1,369	23917,23	0,131	0,21
Connecticut	6	113,1	1,09	29986,95	0,108	0,189
Delaware	7	115,2	1,229	35530,59	0,124	0,199
Florida	8	121,1	1,194	26057,88	0,122	0,192
Georgia	9	108,7	1,066	29752,64	0,105	0,188
Idaho	10	127,5	1,05	26996,77	0,101	0,187
Illinois	11	119	1,137	28630,01	0,114	0,194
Indiana	12	120,8	1,218	33468,68	0,118	0,196
Iowa	13	121,3	1,565	31790,49	0,139	0,217
Kansas	14	108,5	1,275	26588,57	0,129	0,196
Kentucky	15	126,3	1,235	29664,68	0,125	0,193
Louisiana	16	124,3	1,145	29210,06	0,117	0,19
Maine	17	129,4	1,055	24807,54	0,103	0,187
Maryland	18	128,7	1,056	27407,5	0,102	0,187
Massachusetts	19	115,6	1,373	33289,99	0,13	0,206
Michigan	20	127,8	1,091	26032,08	0,111	0,192
Minnesota	21	118,8	1,211	35606,96	0,118	0,194
Mississippi	22	123,3	1,218	35483,2	0,12	0,193
Missouri	23	117,4	1,361	25277,34	0,134	0,209
Montana	24	127,8	1,163	26256,7	0,111	0,192
Nebraska	25	113,3	1,102	33815,83	0,109	0,189
Nevada	26	103,9	1,121	33768,68	0,114	0,191
New Hampshire	27	124,7	1,366	24846,03	0,132	0,204
New Jersey	28	109,3	1,122	35839,3	0,113	0,192
New Mexico	29	108,2	1,451	32694,46	0,137	0,211
New York	30	121,9	1,373	27521,65	0,131	0,2
North Carolina	31	108,9	1,205	28924,66	0,123	0,193
North Dakota	32	112,9	1,44	28792,35	0,136	0,19
Ohio	33	108,4	1,511	21878,45	0,138	0,215
Oklahoma	34	128,9	1,15	32359,94	0,116	0,189
Oregon	35	124,5	1,147	27407,19	0,12	0,191
Pennsylvania	36	101,7	1,568	23824,48	0,139	0,218
Rhode Island	37	129,6	1,184	36190,42	0,121	0,192
South Carolina	38	108,3	1,567	31039,36	0,136	0,218
South Dakota	39	107,3	1,572	33534,91	0,135	0,219
Tennessee	40	111	1,209	36380,18	0,119	0,199
Texas	41	109,2	1,389	22958,68	0,129	0,203
Utah	42	108,8	1,557	31101,32	0,137	0,216
Vermont	43	115,5	1,39	32848,36	0,13	0,203
Virginia	44	113,8	1,536	25679,71	0,138	0,217

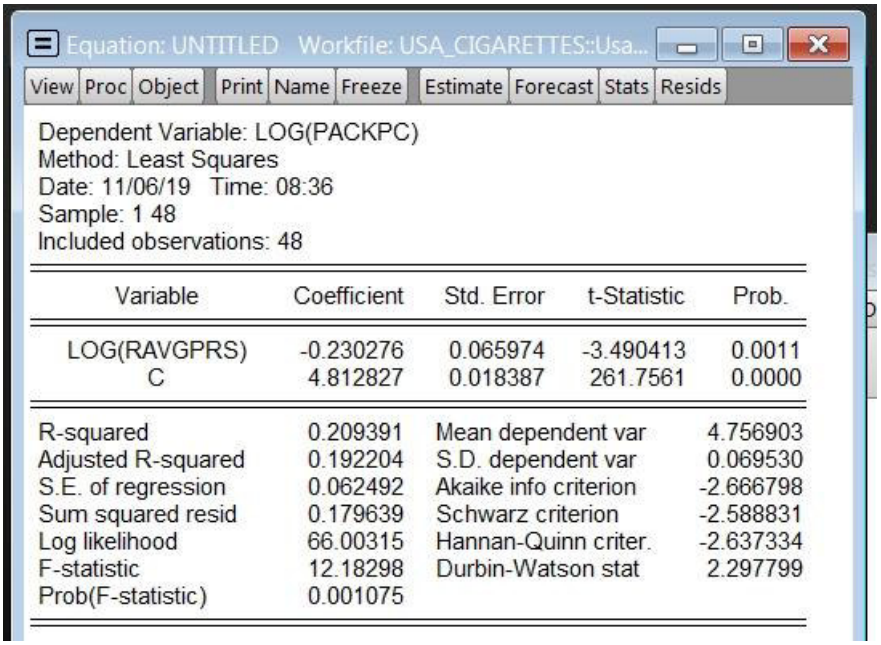
Washington	45	117,2	1,484	32074,01	0,12	0,209
West Virginia	46	122,8	1,108	28805,71	0,101	0,2
Wisconsin	47	119,4	1,568	22157,25	0,136	0,216
Wyoming	48	112,2	1,585	29891,64	0,14	0,22

1. Оцінку еластичності попиту проведемо звичайним методом найменших квадратів

Використаємо рівняння

$$LOG(PACKPC) = \beta_1 + \beta_2 LOG(RAVGPRS) + \varepsilon \qquad (4.1)$$





Equation: UNTITLED Workfile: USA_CIGARETTES::Usa...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PACKPC)
Method: Least Squares
Date: 11/06/19 Time: 08:36
Sample: 1 48
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(RAVGPRS)	-0.230276	0.065974	-3.490413	0.0011
C	4.812827	0.018387	261.7561	0.0000

R-squared	0.209391	Mean dependent var	4.756903
Adjusted R-squared	0.192204	S.D. dependent var	0.069530
S.E. of regression	0.062492	Akaike info criterion	-2.666798
Sum squared resid	0.179639	Schwarz criterion	-2.588831
Log likelihood	66.00315	Hannan-Quinn criter.	-2.637334
F-statistic	12.18298	Durbin-Watson stat	2.297799
Prob(F-statistic)	0.001075		

В цій моделі можуть бути порушені наступні передумови МНК:

- пропущена детермінанта попиту – дохід
- фактор ціна виміряний із похибками
- фактор ціна є ендогенною змінною і корелює із випадковою складовою
- стандартні похибки коефіцієнтів можуть мати зміщення, бо ми маємо справу із перехресною вибіркою

2. Побудуємо кореляційну матрицю для ціни за пачку цигарок і двох податків, перевіримо можливість використання їх як інструментальних змінних

Group: UNTITLED Workfile: USA_CIGARETTES::Usa_cigarettes\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Covariance Analysis: Ordinary
Date: 11/06/19 Time: 09:00
Sample: 1 48
Included observations: 48

Correlation t-Statistic Probability	RAVGPRS	RTAXS	RTAXSO
RAVGPRS	1.000000 ----- -----		
RTAXS	0.905905 14.50866 0.0000	1.000000 ----- -----	
RTAXSO	0.924613 16.46333 0.0000	0.809766 9.360144 0.0000	1.000000 ----- -----

Обидва фактори відповідають умові високої кореляції ендогенної змінної

Перевіримо відсутність кореляції інструментів із випадковою складовою

Для цього спочатку побудуємо ряд лишків регресії з **рівняння 1**

Series: RESID01_LS Workfile: USA_CIGARETTES::Usa_cigarettes\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/- | Adjust+/-

Last updated: 11/06/19 - 09:17

Residuals from equation with dependent variable LOG(PACKPC)

Modified: 1 48 // makeresids resid01_ls

1	-0.031802
2	-0.057288
3	0.065259
4	-0.134932
5	-0.036392
6	-0.064710
7	-0.018674
8	0.024620
9	-0.109517
10	0.046525
11	-0.004137
12	0.026722
13	0.088577
14	-0.070132
15	0.074438
16	0.041052
17	0.062411
18	0.057205
19	0.010306
20	0.057696
21	0.008700
22	0.047206
23	0.023736
24	0.072412
25	-0.060422
26	-0.143096
27	0.084904
28	-0.092223
29	

Розрахуємо коефіцієнти кореляції лишків зі змінними, які ми збираємось використовувати як інструменти

Group: UNTITLED Workfile: USA_CIGARETTES::Usa_cigarettes\

View | Proc | Object | Print | Name | Freeze | Sample | Sheet | Stats | Spec

Correlation

	RESID01_LS	RTAXS	RTAXSO
RESID01_LS	1.000000	-0.044171	-0.050643
RTAXS	-0.044171	1.000000	0.809766
RTAXSO	-0.050643	0.809766	1.000000

Обидва коефіцієнти, -0,04 і -0,05 не є статистично значимими, тому друга умова також виконується.

Виконання третьої умови слідує з того, що безпосереднього впливу податки на попит не мають, лише опосередковано, через ціну.

3. Проведемо оцінку **рівняння 1** методом найменших квадратів, використвуючи податок з продажів як інструмент

Для цього спочатку побудуємо МНК оцінку рівняння, в якому ціна буде залежною змінною, а податок з продажів – фактором

Equation: UNTITLED Workfile: USA_CIGARETTES::Usa...

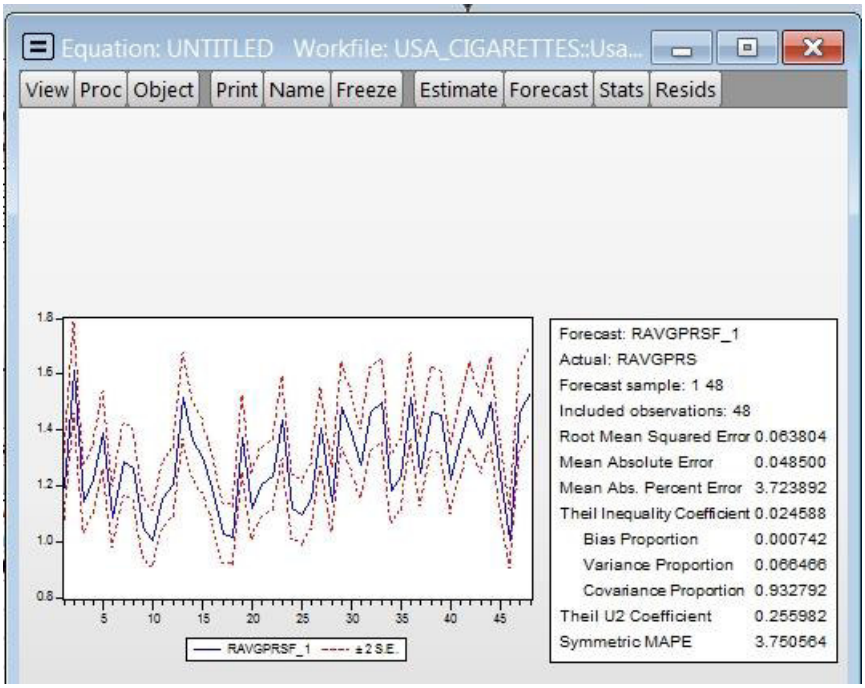
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(RAVGPRS)
Method: Least Squares
Date: 11/06/19 Time: 09:33
Sample: 1 48
Included observations: 48

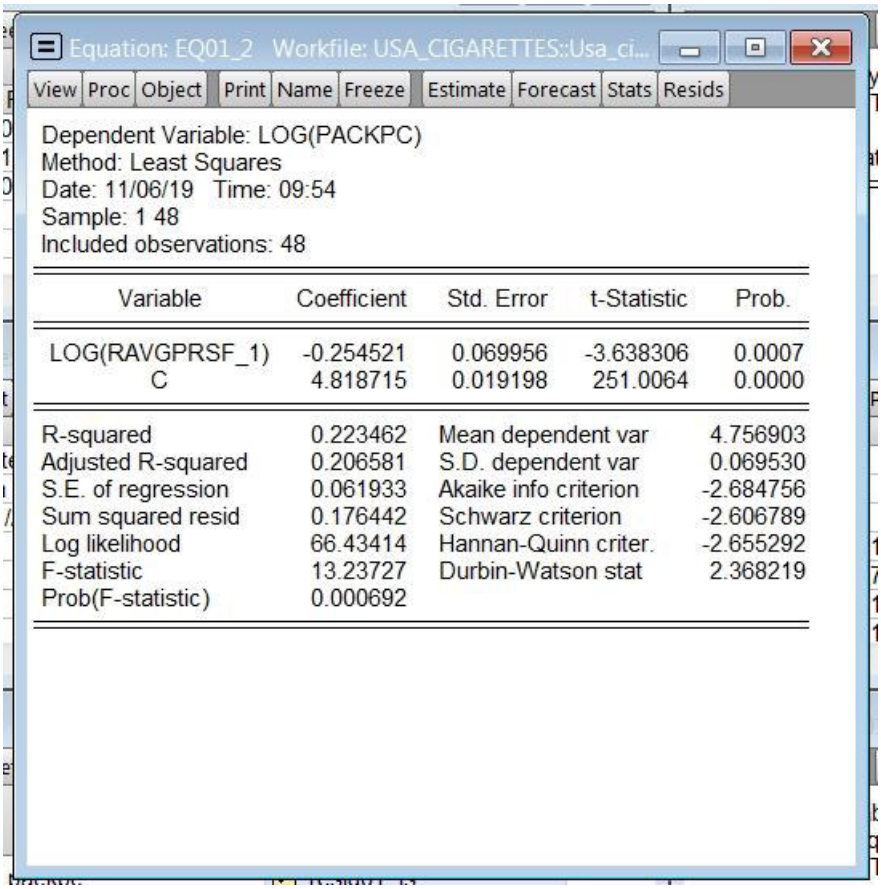
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RTAXSO	10.79817	0.605700	17.82759	0.0000
C	-1.085768	0.074870	-14.50199	0.0000

R-squared	0.873565	Mean dependent var	0.242857
Adjusted R-squared	0.870816	S.D. dependent var	0.138166
S.E. of regression	0.049660	Akaike info criterion	-3.126471
Sum squared resid	0.113440	Schwarz criterion	-3.048505
Log likelihood	77.03531	Hannan-Quinn criter.	-3.097008
F-statistic	317.8229	Durbin-Watson stat	1.196667
Prob(F-statistic)	0.000000		

Отримаємо розрахункові значення Forecast



І проведемо МНК оцінку рівняння із новим інструментом



Результат використання Двокрокового МНК до **рівняння 1**

4. Оцінка двофакторної моделі попиту на цигарки методом МНК

Equation: EQ02_LS Workfile: USA_CIGARETTES::Usa_c...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PACKPC)
Method: Least Squares
Date: 11/06/19 Time: 10:05
Sample: 1 48
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(RAVGPRS)	-0.237468	0.067758	-3.504630	0.0010
LOG(PERINC)	-0.037422	0.068150	-0.549117	0.5856
C	5.199379	0.704196	7.383429	0.0000

R-squared	0.214653	Mean dependent var	4.756903
Adjusted R-squared	0.179749	S.D. dependent var	0.069530
S.E. of regression	0.062971	Akaike info criterion	-2.631810
Sum squared resid	0.178443	Schwarz criterion	-2.514860
Log likelihood	66.16343	Hannan-Quinn criter.	-2.587614
F-statistic	6.149762	Durbin-Watson stat	2.307891
Prob(F-statistic)	0.004354		

Так як цигарки є специфічним товаром, попит на який не еластичний по доходу, то незначимість коефіцієнту при факторі доходу не протирічить очікуванням

5. Оцінимо нелінійну модель

$$LOG(PACKPC) = \beta_1 + \beta_2 LOG(RAVGPRS) + \beta_3 LOG(PERINC) + \varepsilon$$

Методом ДМНК використовуючи податки як інструментальні змінні

Для цього спочатку побудуємо МНК оцінку ендогенної змінної на всі екзогенні, в тому числі інструменти

Equation: EQ02_1 Workfile: USA_CIGARETTES::Usa_cigarettes\

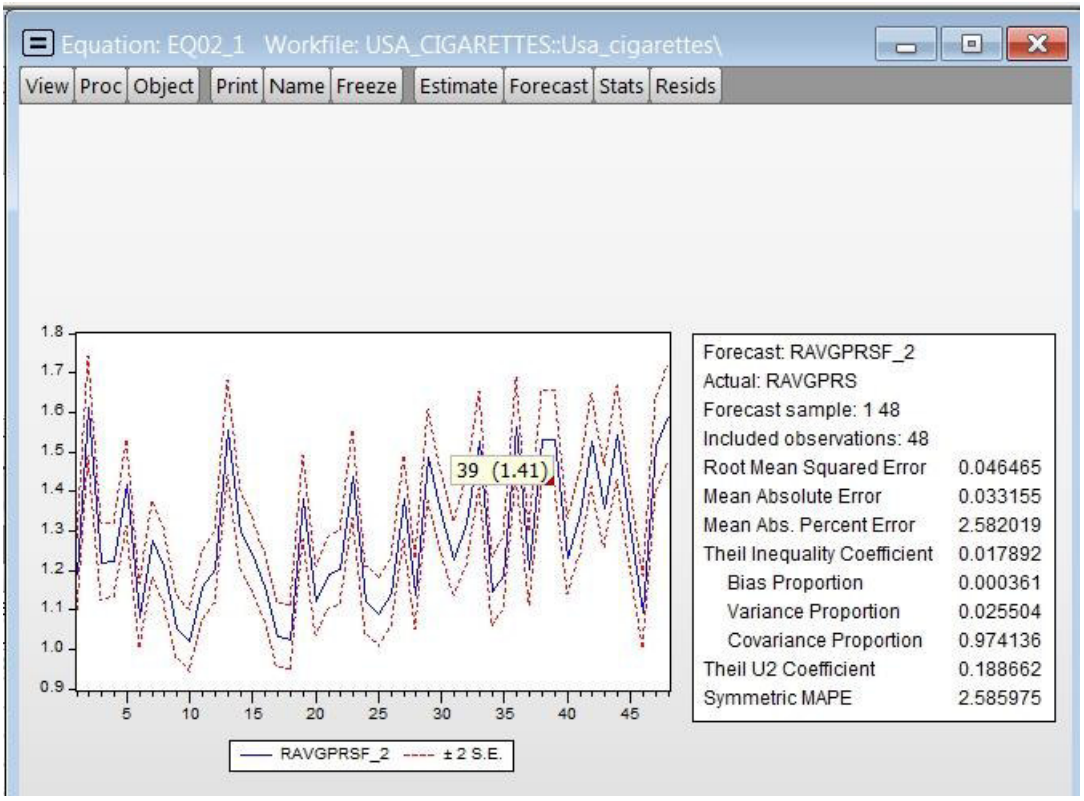
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(RAVGPRS)
Method: Least Squares
Date: 11/06/19 Time: 10:48
Sample: 1 48
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(PERINC)	0.006617	0.040436	0.163652	0.8708
RTAXSO	6.954862	0.775820	8.964528	0.0000
RTAXS	5.302314	0.863885	6.137752	0.0000
C	-1.743598	0.447609	-3.895364	0.0003

R-squared	0.931943	Mean dependent var	0.242857
Adjusted R-squared	0.927303	S.D. dependent var	0.138166
S.E. of regression	0.037253	Akaike info criterion	-3.662519
Sum squared resid	0.061062	Schwarz criterion	-3.506585
Log likelihood	91.90045	Hannan-Quinn criter.	-3.603591
F-statistic	200.8383	Durbin-Watson stat	1.733550
Prob(F-statistic)	0.000000		

Отримаємо для неї розрахункові значення Forecast



МНК оцінка **рівняння 2** з використанням розрахункових значень замість звичайних цін

Equation: EQ02_2 Workfile: USA_CIGARETTES::Usa_ci...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(PACKPC)
Method: Least Squares
Date: 11/06/19 Time: 10:52
Sample: 1 48
Included observations: 48

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(RAVGPRSF_2)	-0.262639	0.068964	-3.808330	0.0004
LOG(PERINC)	-0.042316	0.066961	-0.631945	0.5306
C	5.255809	0.692151	7.593447	0.0000

R-squared	0.243966	Mean dependent var	4.756903
Adjusted R-squared	0.210365	S.D. dependent var	0.069530
S.E. of regression	0.061785	Akaike info criterion	-2.669849
Sum squared resid	0.171783	Schwarz criterion	-2.552899
Log likelihood	67.07637	Hannan-Quinn criter.	-2.625653
F-statistic	7.260569	Durbin-Watson stat	2.355400
Prob(F-statistic)	0.001850		

6. Повторно оцінимо модель вбудованим методом ДМНК Eviews. Як інстрментальні змінні використовуватимемо дохід і податки.

Equation: EQ02_TSLS Workfile: USA_CIGARETTES::Us...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

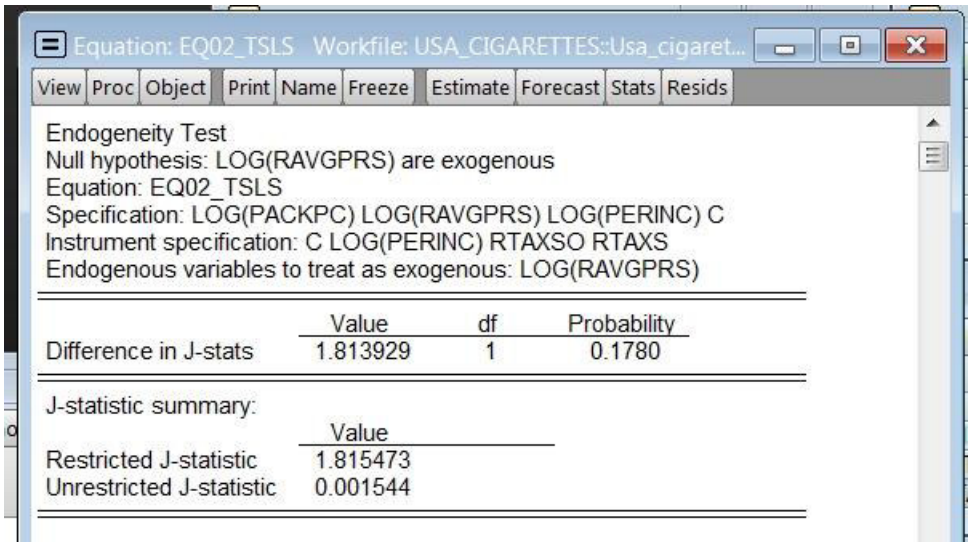
Dependent Variable: LOG(PACKPC)
Method: Two-Stage Least Squares
Date: 11/06/19 Time: 10:58
Sample: 1 48
Included observations: 48
Instrument specification: LOG(PERINC) RTAXSO RTAXS
Constant added to instrument list

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LOG(RAVGPRS)	-0.262639	0.070396	-3.730866	0.0005
LOG(PERINC)	-0.042316	0.068351	-0.619090	0.5390
C	5.255809	0.706522	7.438990	0.0000

R-squared	0.212245	Mean dependent var	4.756903
Adjusted R-squared	0.177233	S.D. dependent var	0.069530
S.E. of regression	0.063068	Sum squared resid	0.178990
F-statistic	6.968201	Durbin-Watson stat	2.299272
Prob(F-statistic)	0.002310	Second-Stage SSR	0.171783
J-statistic	0.001539	Instrument rank	4

В порівнянні із результатами використання ДМНК «вручну», в пункті 4, оцінки коефіцієнтів співпадаюсть, однак відрізняються їх стандартні похибки і статистичні характеристики, що показують якість оцінки рівняння. Можемо тако зробити висновок, що коефіцієнт при факторі ціни є статистично консистентним, а при доході – ні, в той час як рівняння в цілому є статистично консистентним.

7. Порівняємо оцінки, отримані методами МНК та ІЗ за допомогою тесту Дарбіна-Ву-Хаусмана із нульовою гіпотезою що ендогенності нема, як і причин для використання методу ІЗ.



Тест демонструє невизначеність, нульова гіпотеза не відкидається на рівні 15%, але відкидається на рівні 20%, що може бути слідством невеликих розмірів вибірки. В даному прикладі ми не можемо зробити висновок про більшу ефективність МНК-оцінки. На основі тесту і отриманих оцінок слід віддати перевагу мтоді інструментальних змінних для оцінки еластичності попиту по ціні.

Висновок

Під час виконання роботи ми набули досвіду суттєво спрощувати використння складного двокрокового МНК за допомогою програми Eviews . Також ми ознайомилися з поняттям інструментальної змінної, методами МНК, ДМНК та з тестом Дарбіна-Ву-Хаусмана. Набули досвіду та розуміння в яких випадках краще використовувати той чи інший метод при оцінці еластичності та специфікації рівняння регресії в цілому.

Індивідуальні дані для виконання роботи обираються згідно до посилань на сайти з прикладної статистики, вказані в посібнику до курсу [1] узгоджено з викладачем, що веде практичні заняття.

Практичне заняття №5

Аналіз часового ряду відношень курсу валют долара та євро. Одномірні моделі часових рядів

Метою практичного заняття є засвоєння методології проведення аналізу співвідношень курсу валют на прикладі співвідношення курсів євро та долару починаючи з середини 2002 до осені 2020 року.

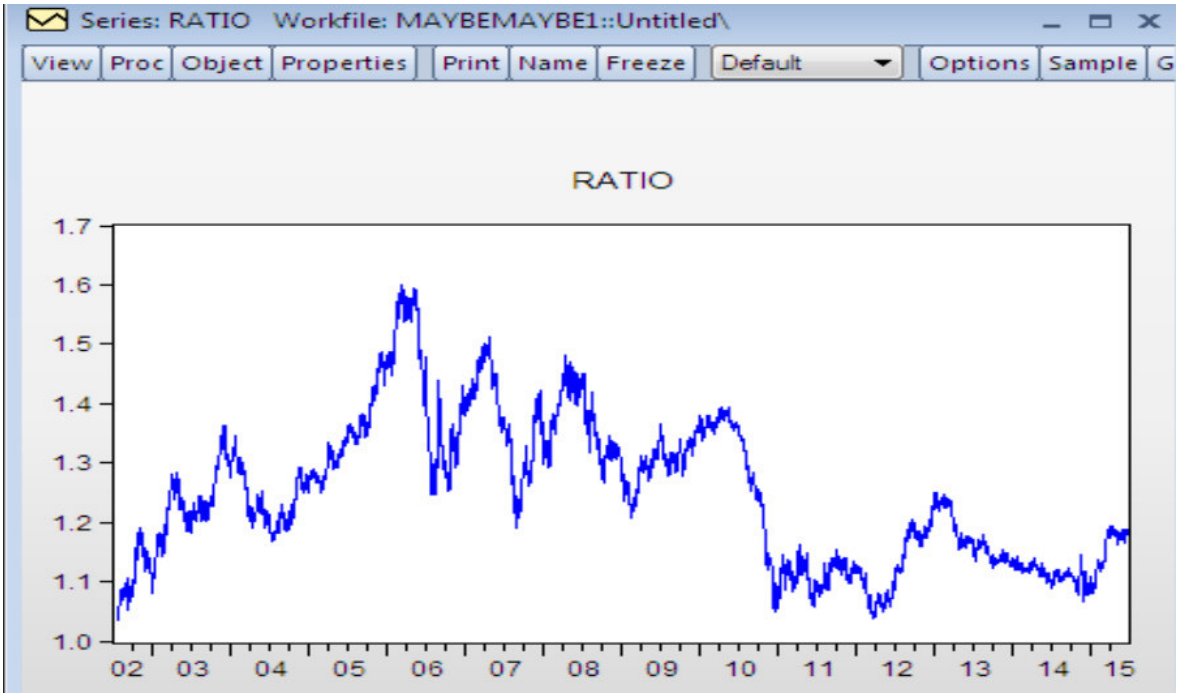
1. Побудувати графік часового ряду. Проаналізувати графік часового ряду , зробити попередні припущення про стаціонарність часового ряду.
2. Перевірити стаціонарність ряду за допомогою підходящої процедури. Визначити порядок інтегрованості ряду.
3. Оцінити автокореляцію та часткову автокореляцію функцію процеса. Надати їм опис та інтерпретувати, перевірити коефіцієнти на значимість.
4. Який тип базової моделі міг би описати ряд RATIO?
5. Оцінити параметри моделі з пункту 4.
6. Оцінити альтернативні моделі і порівняти їх за допомогою відповідних критеріїв.

Виконання роботи

1. Для роботи було обрано часовий ряд щоденних відношень курсу євро та долару починаючи з середини 2002 до осені 2020 року. Дані для аналізу були взяті з сайту https://www.banki.ru/products/currency/usd/. На основі отриманих даних був створений excel-файл, дані з якого було імпортовано в програму Eviews 7.

	Last up
	Imported from 'C:\
7/15/2002	1.0490
7/16/2002	1.0487
7/17/2002	1.0353
7/18/2002	1.0422
7/19/2002	1.0460
7/20/2002	1.0410
7/21/2002	1.0494
7/22/2002	1.0487
7/23/2002	1.0580
7/24/2002	1.0538
7/25/2002	1.0552
7/26/2002	1.0562
7/27/2002	1.0616
7/28/2002	1.0668
7/29/2002	1.0727
7/30/2002	1.0727
7/31/2002	1.0739
8/01/2002	1.0831
8/02/2002	1.0849
8/03/2002	

На основі цих даних був побудований графік, представлений нижче



З огляду на вигляд графіку можна запропонувати гіпотезу, що значення часового ряду коливаються навколо певного постійного значення С в районі від 1.2 та 1.3. Гіпотезу про стаціонарність, тобто незмінність параметру ряду можливо висунути з умов породжених цим рядом – це відношення курсів валют двох суб'єктів економічних сил - США та Європейського Союзу. Враховуючи, що за даний період відносини між цими двома суб'єктами були постійними, як і їх частки на світовому ринку можливо запропонувати, що закони відношення їх валют за цей період майже не змінювались.

2. Проведемо тест на одиничний корінь для даного ряду шляхом застосування розширеного тесту Діка-Фуллера.

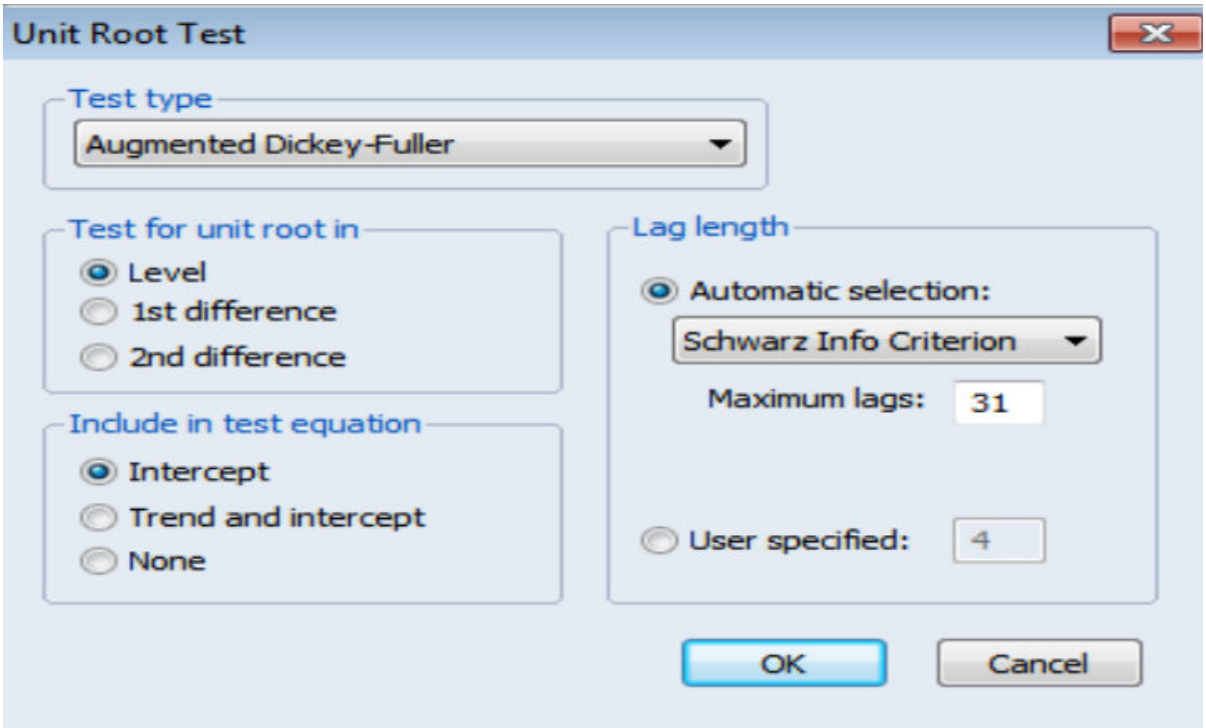
Наддамо пояснення поняттю «Одиничний корінь»

$$y_t = \beta_1 y_{t-1} + \varepsilon_t$$

Що можна переписати за допомогою деяких диференціальних перетворень як

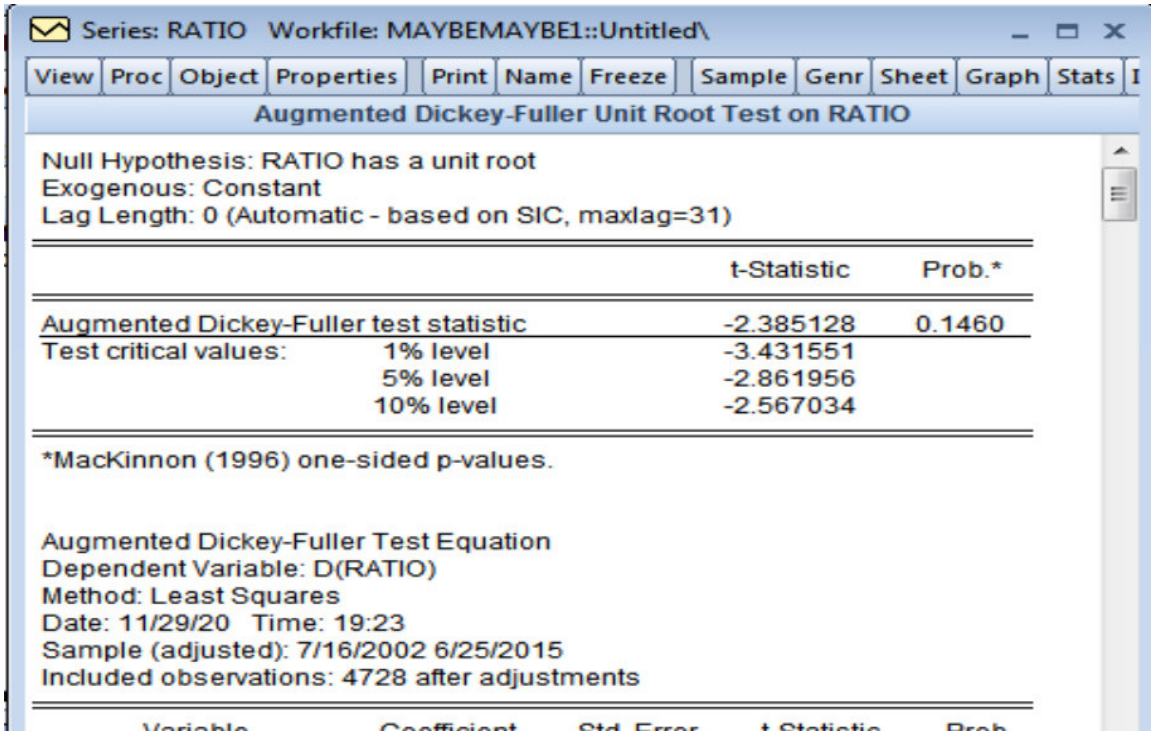
$$_{\Delta}y_t = (\beta_1 - 1)y_{t-1} + \varepsilon_t = \delta y_{t+1} + \varepsilon_t$$

Тест на одиничний корінь базується на висуванні гіпотези, що $\delta = 0$.



Умови тесту були задані так, щоб максимально уточнити значення на основі запропонованих та прогнозованих на основі графіку та теоретичних значеннях припущень. Наприклад, що коливання відбуваються навколо константи.

Результати тесту:



Series: RATIO Workfile: MAYBEMAYBE1::Untitled\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats

Augmented Dickey-Fuller Unit Root Test on RATIO

Null Hypothesis: RATIO has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=31)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.385128	0.1460
Test critical values:	1% level	-3.431551	
	5% level	-2.861956	
	10% level	-2.567034	

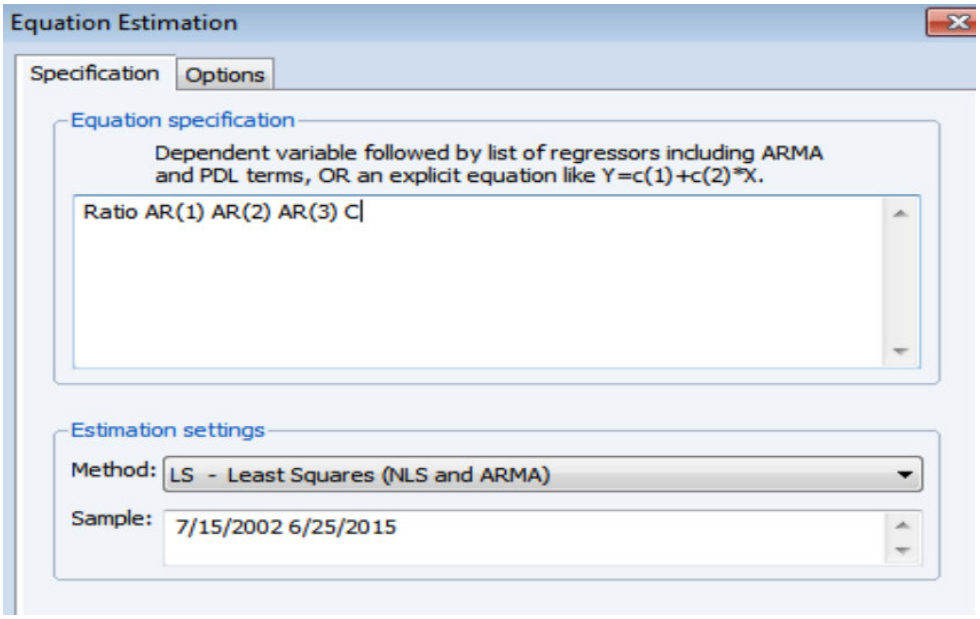
*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RATIO)
Method: Least Squares
Date: 11/29/20 Time: 19:23
Sample (adjusted): 7/16/2002 6/25/2015
Included observations: 4728 after adjustments

Ймовірність прийняття альтернативної гіпотези не дуже мала, проте в межах допустимого – 14,6%. Це дозволяє нам прийняти гіпотезу як про одиничний корінь так і про те, що значення коливаються навколо одного, постійного, значення.

З наведеного вище видно, що ряд має порядок інтегрованості рівний одиниці.

3. Згідно проведеній в пункті 2 процедурі (Тест на одиничний корінь)ми можемо стверджувати о коректності застосовувати до даного ряду гіпотез і тому ми можемо приступити до пошуку коефіцієнтів для рівняння, що описує його поведінку.



Скористаємося існуючими інструментами і спробуємо побудувати параметри залежності від параметрів

$$y_{t-2}, y_{t-3}$$

та приблизного знаходження константи.

Dependent Variable: RATIO
Method: Least Squares
Date: 11/29/20 Time: 19:40
Sample (adjusted): 7/18/2002 6/25/2015
Included observations: 4726 after adjustments
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.263899	0.051207	24.68227	0.0000
AR(1)	0.984922	0.014547	67.70549	0.0000
AR(2)	0.015649	0.020417	0.766452	0.4434
AR(3)	-0.002713	0.014543	-0.186517	0.8520

R-squared	0.996272	Mean dependent var	1.248935
Adjusted R-squared	0.996270	S.D. dependent var	0.122542
S.E. of regression	0.007484	Akaike info criterion	-6.951257
Sum squared resid	0.264479	Schwarz criterion	-6.945789
Log likelihood	16429.82	Hannan-Quinn criter.	-6.949335
F-statistic	420689.8	Durbin-Watson stat	1.999081
Prob(F-statistic)	0.000000		

Inverted AR Roots	1.00	.05	-.06

римані результати дозволяють нам прийняти гіпотезу про те що існує залежність від у y_{t-1}, проте залежність від інших параметрів $y_{t-2}, \ldots$ є більш ніж неочевидною в у випадку і для всіх параметрів нижчого ніж два y_t вкрай маловірогідною. Отримані параметри, підраховані за допомогою вбудованих засобів програми Eviews 7 вверсії.

Equation: UNTITLED Workfile: MAYBEMAYBE1::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

```
Estimation Command:
==========================
LS RATIO AR(1) AR(2) AR(3) C

Estimation Equation:
==========================
RATIO = C(1) + [AR(1)=C(2),AR(2)=C(3),AR(3)=C(4)]

Substituted Coefficients:
==========================
RATIO = 1.26389877792 + [AR(1)=0.984921662768,AR(2)=0.0156490069506,AR
(3)=-0.00271260771274]
```

З огляду на малість кожного з коефіцієнтів для AR(2), AR(3) можливо їх просто відкинути.

Пересвідчимося в правильності такого рішення та рішенні не брати AR(p),

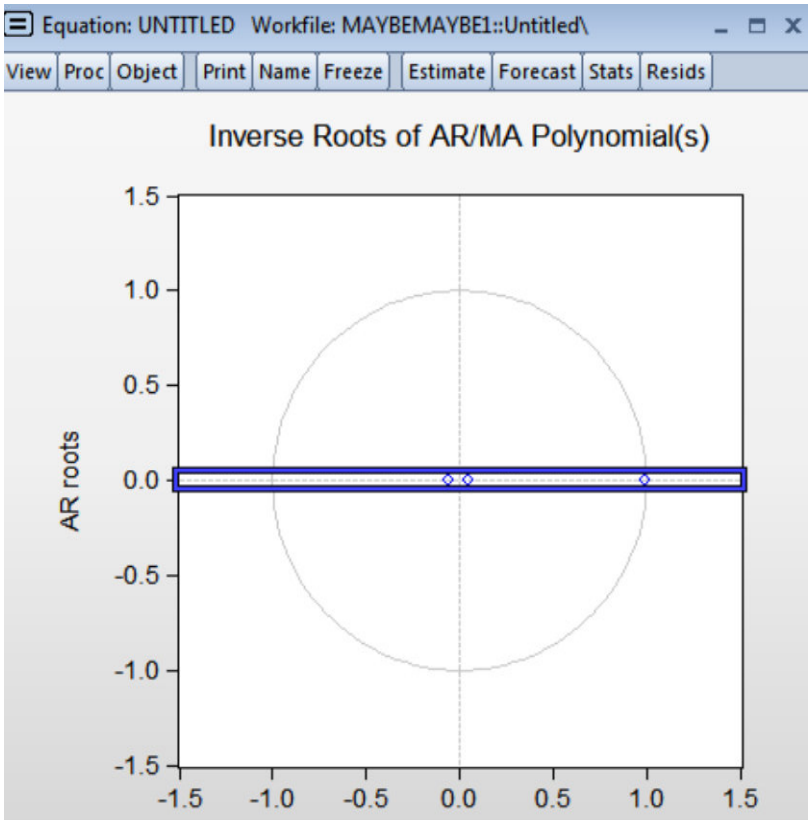
$$p \geq 4$$

За допомогою також вбудованого та доступного на даній версії вкладки побудови - Королелограми.

Date: 11/29/20 Time: 19:37
Sample: 7/15/2002 6/25/2015
Included observations: 4729

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.998	0.998	4711.5	0.000
		2	0.996	0.011	9403.9	0.000
		3	0.994	-0.011	14077.	0.000
		4	0.991	0.022	18732.	0.000
		5	0.989	-0.019	23367.	0.000
		6	0.987	-0.013	27983.	0.000
		7	0.985	0.007	32580.	0.000
		8	0.983	-0.007	37157.	0.000
		9	0.981	-0.017	41714.	0.000
		10	0.978	0.034	46253.	0.000
		11	0.976	0.025	50775.	0.000
		12	0.974	-0.027	55278.	0.000
		13	0.972	-0.003	59762.	0.000
		14	0.970	0.004	64228.	0.000
		15	0.968	-0.011	68676.	0.000
		16	0.966	-0.008	73105.	0.000
		17	0.964	-0.004	77515.	0.000
		18	0.962	-0.014	81905.	0.000
		19	0.959	0.010	86277.	0.000

Вигляд та розміщення трьох знайденних коренів. Вони всі лежать в середині комплексного кола радіуса 1, тому процес может бути коректно описаний та прогнозований.



Оскільки на основі наведених відомостей можливо однозначно стверджувати про значимість для першого і лише першого коефіцієнту то прогноз для моделі будемо будувати виключно з урахуванням знайденої постійної величини та залежності від $t-1$.

На основі наведених результатів отримуємо відповіді для пунктів 4, 5 та 6.

Перерахуємо значення для випадку одного AR(1) C значення критеріїв та поточнимо значення коефіцієнтів.

Equation: ONE Workfile: MAYBEMAYBE1::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RATIO
Method: Least Squares
Date: 11/29/20 Time: 20:15
Sample (adjusted): 7/16/2002 6/25/2015
Included observations: 4728 after adjustments
Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.262474	0.051723	24.40836	0.0000
AR(1)	0.997883	0.000888	1124.065	0.0000

R-squared	0.996274	Mean dependent var	1.248847
Adjusted R-squared	0.996273	S.D. dependent var	0.122590
S.E. of regression	0.007484	Akaike info criterion	-6.951623
Sum squared resid	0.264718	Schwarz criterion	-6.948890
Log likelihood	16435.64	Hannan-Quinn criter.	-6.950663
F-statistic	1263523.	Durbin-Watson stat	2.025911
Prob(F-statistic)	0.000000		

Inverted AR Roots 1.00

Equation: ONE Workfile: MAYBEMAYBE1::Untitled\

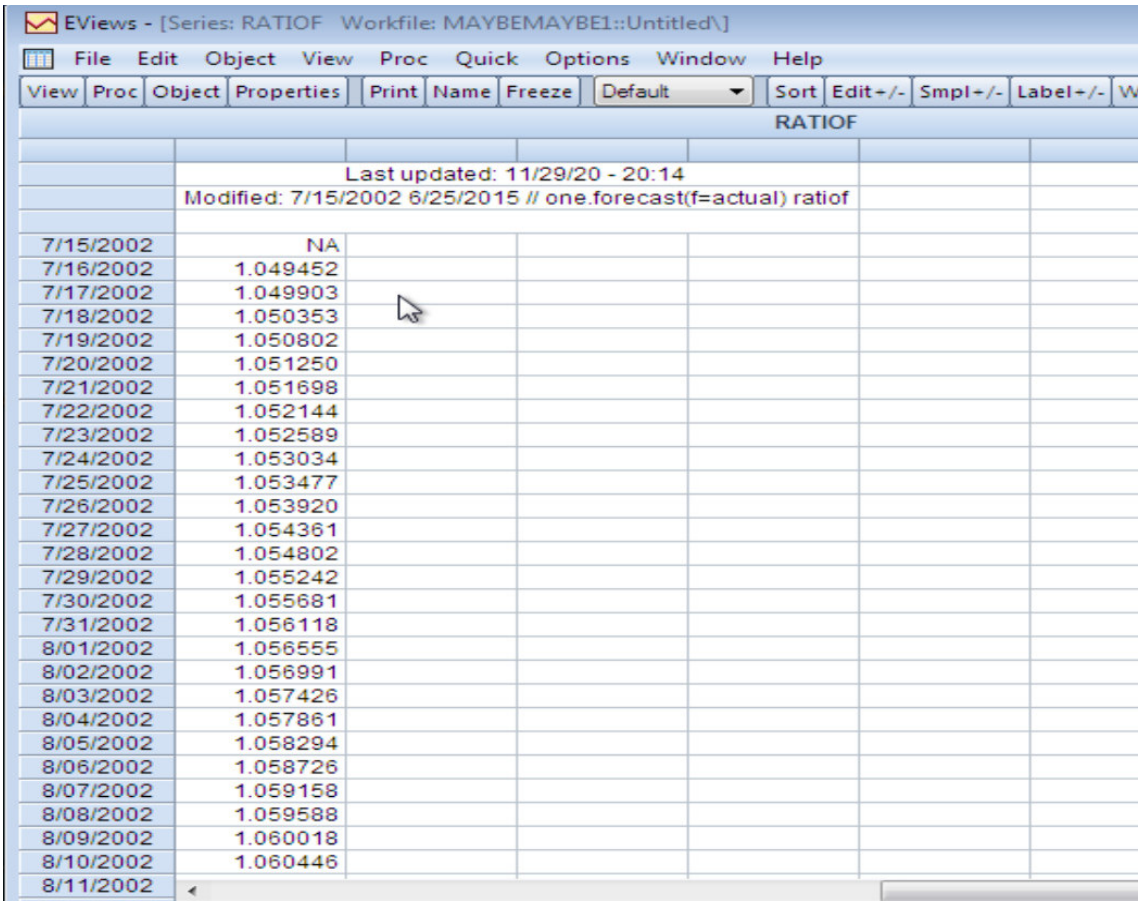
View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | R

Estimation Command:
==========
LS RATIO AR(1) C

Estimation Equation:
==========
RATIO = C(1) + [AR(1)=C(2)]

Substituted Coefficients:
==========
RATIO = 1.26247352134 + [AR(1)=0.997882616376]

Отриманли на основі цього прогноз:



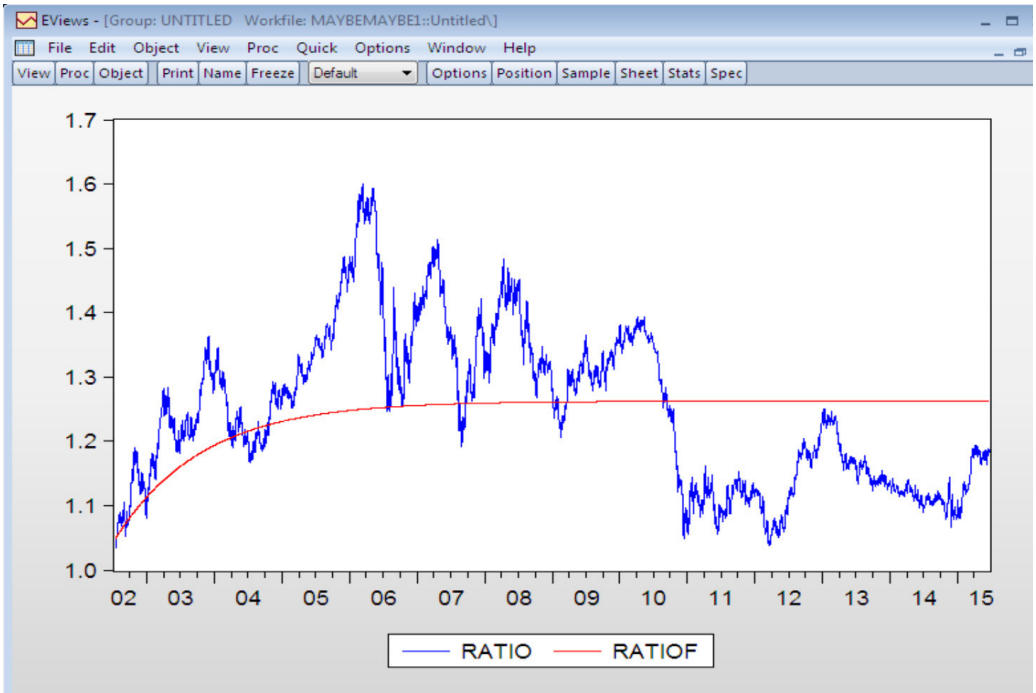
7/15/2002	NA
7/16/2002	1.049452
7/17/2002	1.049903
7/18/2002	1.050353
7/19/2002	1.050802
7/20/2002	1.051250
7/21/2002	1.051698
7/22/2002	1.052144
7/23/2002	1.052589
7/24/2002	1.053034
7/25/2002	1.053477
7/26/2002	1.053920
7/27/2002	1.054361
7/28/2002	1.054802
7/29/2002	1.055242
7/30/2002	1.055681
7/31/2002	1.056118
8/01/2002	1.056555
8/02/2002	1.056991
8/03/2002	1.057426
8/04/2002	1.057861
8/05/2002	1.058294
8/06/2002	1.058726
8/07/2002	1.059158
8/08/2002	1.059588
8/09/2002	1.060018
8/10/2002	1.060446
8/11/2002	

В порівнянні, скористаємося вбудованою функцією групування та виведемо графічно отримані таблиці в вигляді графіків.

Group: UNTITLED Workfile: MAYBEMAYBE1::Untitled\

View | Proc | Object | Print | Name | Freeze | Default | Sort | Tra

obs	RATIO	RATIOF1RO...
7/15/2002	1.0490	NA
7/16/2002	1.0487	1.049452
7/17/2002	1.0353	1.049903
7/18/2002	1.0422	1.050353
7/19/2002	1.0460	1.050802
7/20/2002	1.0410	1.051250
7/21/2002	1.0494	1.051698
7/22/2002	1.0487	1.052144
7/23/2002	1.0580	1.052589
7/24/2002	1.0538	1.053034
7/25/2002	1.0552	1.053477
7/26/2002	1.0562	1.053920
7/27/2002	1.0616	1.054361
7/28/2002	1.0668	1.054802
7/29/2002	1.0727	1.055242
7/30/2002	1.0727	1.055681
7/31/2002	1.0739	1.056118
8/01/2002	1.0831	1.056555
8/02/2002	1.0849	1.056991
8/03/2002	1.0823	1.057426
8/04/2002	1.0841	1.057861
8/05/2002	1.0818	1.058294
8/06/2002	1.0766	1.058726
8/07/2002		



З урахуванням потенційної високої хаотичності процесу таке, майже лінійне, наближення схоже на таке, що відповідає дійсності.

Висновок:

Часовий ряд – це ряд точок даних, проіндексованих в хронологічному порядку. В такому вигляді можливо описати або формалізувати більшість існуючих випадкових процесів в реальному світі. Така форма запису комфортна для аналізу та побудови прогностичних моделей.

В наведеному для даної роботи прикладі з курсами валют було виконані основні етапи роботи з аналізом деякого набору даних зібраного в вигляді ряду.

Користуючись можливостями програми та опираючись на теоретичні відомості, було підраховано приблизний вигляд рівняння, проведена перевірка на коректність виконаних процедур та створено прогностичну функцію.

Індивідуальні дані для виконання роботи обираються згідно до посилань на сайти з прикладної статистики, вказані в посібнику до курсу [1] узгоджено з викладачем, що веде практичні заняття.

Практичне заняття №6

Аналіз часового ряду індексу цін на непродовольчі товари

Мета: вивчення методів аналізу та моделювання поведінки часового ряду, виходячи з його попередніх значень.

Вихідні дані: індекси цін на непродовольчі товари в період з січня 2000 року по грудень 2005 року, щомісячні дані

Завдання:

1. Побудувати графік часосого ряду
2. Перевірити ряд на стаціонарність
3. Оцінити автокореляційну та частинну автокореляційну функції процесу. Зробити висновок про статистичну значимість коефіцієнтів
4. Визначити,який тип базової моделі може описати даний ряд
5. Оцінити параметри цієї моделі
6. Оцінити альтернативні моделі,порівняти їх
7. Знайти розрахункові значення індексу цін згідно з обраною моделлю, використовуючи динамічний та статистичний метод прогнозування

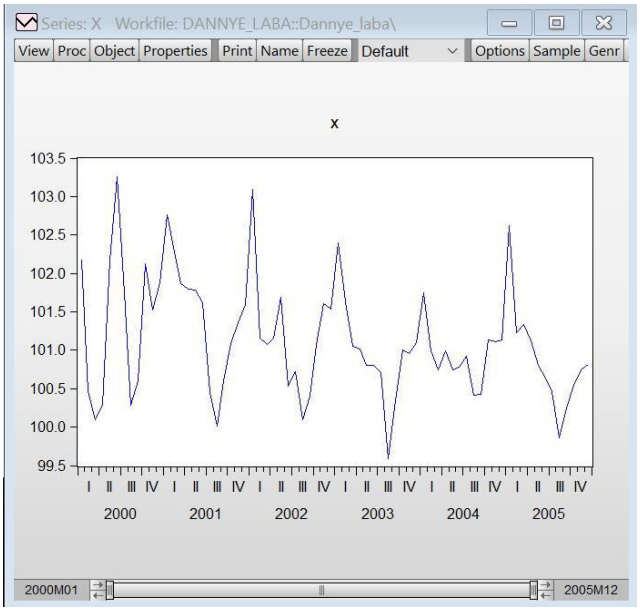
Результати роботи

1. Імпортуємо дані з таблиці Excel про індекси цін у EViews та побудуємо графік часового ряду

Date	x
01.01.2000	102,17
01.02.2000	100,46
01.03.2000	100,09
01.04.2000	100,30
01.05.2000	102,17
01.06.2000	103,26
01.07.2000	101,75
01.08.2000	100,29
01.09.2000	100,60
01.10.2000	102,13

01.11.2000	101,52
01.12.2000	101,88
01.01.2001	102,76
01.02.2001	102,28
01.03.2001	101,86
01.04.2001	101,79
01.05.2001	101,78
01.06.2001	101,62
01.07.2001	100,45
01.08.2001	100,01
01.09.2001	100,60
01.10.2001	101,09
01.11.2001	101,36
01.12.2001	101,60
01.01.2002	103,09
01.02.2002	101,16
01.03.2002	101,08
01.04.2002	101,16
01.05.2002	101,69
01.06.2002	100,53
01.07.2002	100,72
01.08.2002	100,09
01.09.2002	100,40
01.10.2002	101,07
01.11.2002	101,61
01.12.2002	101,54
01.01.2003	102,40
01.02.2003	101,63
01.03.2003	101,05
01.04.2003	101,02
01.05.2003	100,80
01.06.2003	100,80
01.07.2003	100,71
01.08.2003	99,59
01.09.2003	100,34
01.10.2003	101,00
01.11.2003	100,96
01.12.2003	101,10
01.01.2004	101,75
01.02.2004	100,99
01.03.2004	100,75
01.04.2004	100,99
01.05.2004	100,74
01.06.2004	100,78
01.07.2004	100,92
01.08.2004	100,42
01.09.2004	100,43
01.10.2004	101,14
01.11.2004	101,11
01.12.2004	101,14

01.01.2005	102,62
01.02.2005	101,23
01.03.2005	101,34
01.04.2005	101,12
01.05.2005	100,80
01.06.2005	100,64
01.07.2005	100,46
01.08.2005	99,86
01.09.2005	100,25
01.10.2005	100,55
01.11.2005	100,74
01.12.2005	100,82



2. Перевіримо ряд на стаціонарність, тобто здатність не змінювати свої характеристики з часом. Для цього використаємо розширений тест Дікі-Фуллера. На графіку видно, що зміни у ряда коливаються відносно деякої константи, тому у тестуюче рівняння включаємо константу.

Prob* - це гіпотеза про те,що рід не є стаціонарним. Ймовірність прийняття цієї гіпотези дуже мала - 0,0001, тому приймаємо альтернативну гіпотезу і робимо висновок, що даний ряд є стаціонарним.

Series: X Workfile: DANNYE_LABA::Dannye_laba\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Augmented Dickey-Fuller Unit Root Test on X

Null Hypothesis: X has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.854270	0.0001
Test critical values:	1% level	-3.525618	
	5% level	-2.902953	
	10% level	-2.588902	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X)
Method: Least Squares
Date: 10/30/19 Time: 16:30
Sample (adjusted): 2000M02 2005M12
Included observations: 71 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-0.496031	0.102184	-4.854270	0.0000
C	50.14390	10.33405	4.852297	0.0000

R-squared 0.254569 Mean dependent var -0.019014

Series: X Workfile: DANNYE_LABA::Dannye_laba\

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Augmented Dickey-Fuller Unit Root Test on X

10% level -2.588902

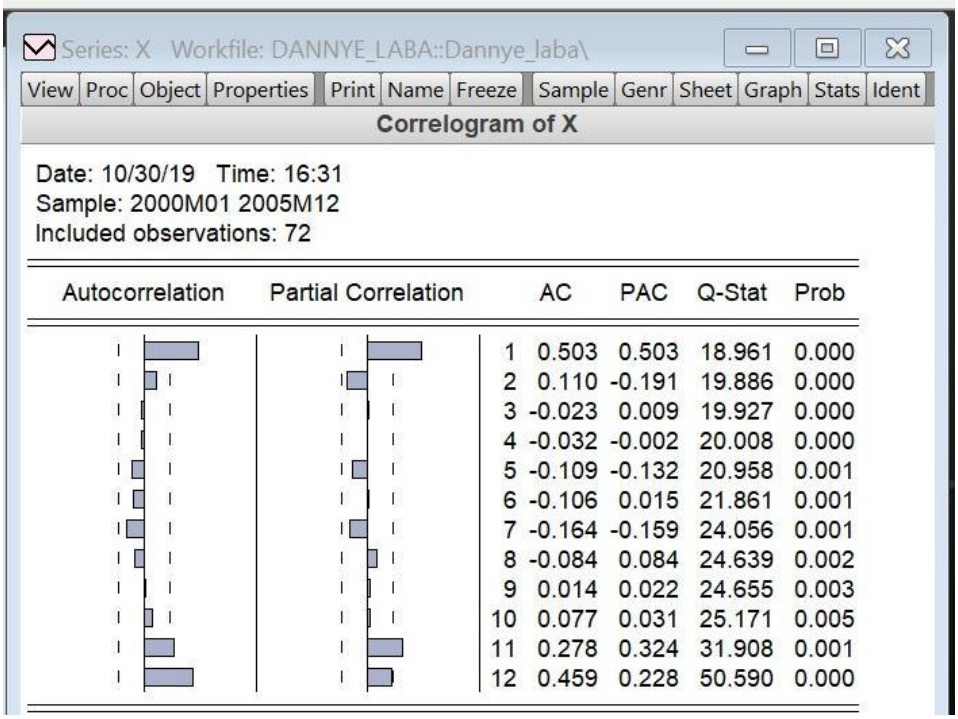
*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X)
Method: Least Squares
Date: 10/30/19 Time: 16:30
Sample (adjusted): 2000M02 2005M12
Included observations: 71 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	-0.496031	0.102184	-4.854270	0.0000
C	50.14390	10.33405	4.852297	0.0000

R-squared	0.254569	Mean dependent var	-0.019014
Adjusted R-squared	0.243766	S.D. dependent var	0.740749
S.E. of regression	0.644168	Akaike info criterion	1.986050
Sum squared resid	28.63172	Schwarz criterion	2.049788
Log likelihood	-68.50479	Hannan-Quinn criter.	2.011397
F-statistic	23.56394	Durbin-Watson stat	1.666128
Prob(F-statistic)	0.000007		

3. Побудуємо корелограму для визначення типу процесу. Довжину лагу задаємо 12, через те,що дані щомісячні.



З графіку частинної автокореляційної функції видно,що статистично значущими можуть бути лаги 1,3,6,8,11.

4. Проаналізувавши корелограму, можна зробити висновок про те, що даний ряд може бути описаний процесом AR(11). Оцінимо модель AR(11). На

першому кроці включимо лаги 1,3,6,8 та 11.

Equation: UNTITLED Workfile: DANNYE_LABA::Dannye_l...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: X
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/30/19 Time: 16:42
Sample: 2000M01 2005M12
Included observations: 72
Convergence achieved after 67 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	101.1201	0.273724	369.4236	0.0000
AR(1)	0.495923	0.104781	4.732954	0.0000
AR(3)	-0.097526	0.143695	-0.678700	0.4997
AR(6)	0.098819	0.107407	0.920043	0.3610
AR(8)	-0.094345	0.141095	-0.668662	0.5061
AR(11)	0.342776	0.119010	2.880232	0.0054
SIGMASQ	0.345755	0.058285	5.932130	0.0000

R-squared	0.375047	Mean dependent var	101.1243
Adjusted R-squared	0.317359	S.D. dependent var	0.749027
S.E. of regression	0.618862	Akaike info criterion	2.004688
Sum squared resid	24.89437	Schwarz criterion	2.226030
Log likelihood	-65.16876	Hannan-Quinn criter.	2.092805
F-statistic	6.501293	Durbin-Watson stat	2.031418
Prob(F-statistic)	0.000021		

5.

В результаті отримаємо, що лаги 3,6 та 8 не є статистично значущими. Залишимо лише значущі члени : 1 та 11. У результаті отримаємо, що 101.11 - оцінка середнього значення даного ряду.

Equation: UNTITLED Workfile: DANNYE_LABA::Dannye_l...

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: X
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 10/30/19 Time: 16:44
Sample: 2000M01 2005M12
Included observations: 72
Convergence achieved after 15 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	101.1100	0.278491	363.0633	0.0000
AR(1)	0.476519	0.089789	5.307081	0.0000
AR(11)	0.312244	0.106091	2.943156	0.0044
SIGMASQ	0.358612	0.058573	6.122489	0.0000

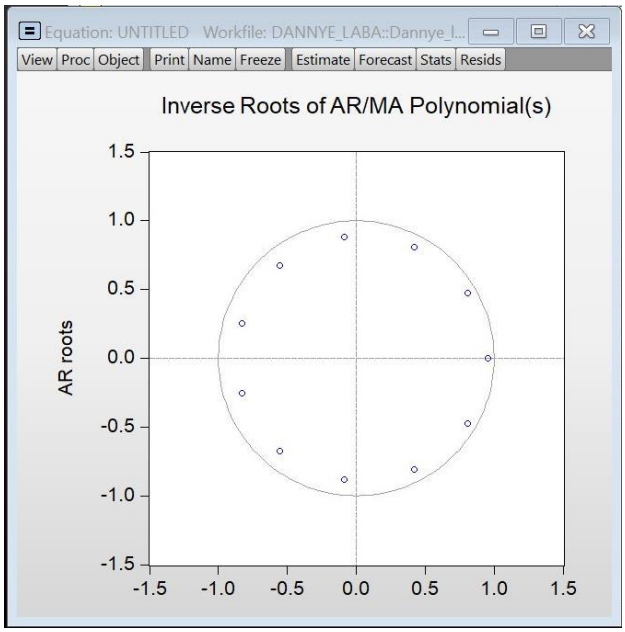
R-squared	0.351808	Mean dependent var	101.1243
Adjusted R-squared	0.323211	S.D. dependent var	0.749027
S.E. of regression	0.616204	Akaike info criterion	1.949337
Sum squared resid	25.82006	Schwarz criterion	2.075818
Log likelihood	-66.17613	Hannan-Quinn criter.	1.999690
F-statistic	12.30239	Durbin-Watson stat	1.968881
Prob(F-statistic)	0.000002		

Inverted AR Roots	.96	.81-.48i	.81+.48i	.42-.81i
	.42+.81i	-.09-.88i	-.09+.88i	-.55-.67i
	-.55+.67i	-.83+.25i	-.83-.25i	

6. Для діагностики отриманих рівнянь використаємо інструмент ARMA Structure.

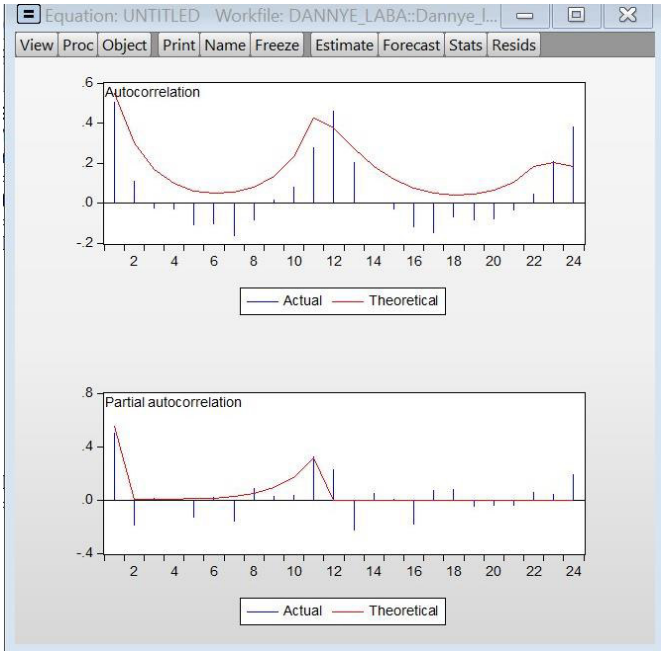
7.

Першим інструментом діагностики є розрахунок обернених коренів характеристичного поліному оператора зсуву, що служить для перевірки на стаціонарність отриманої AR моделі.



Для нашого рівняння помітимо,що усі корені лежать всередині одиничного кола, що означає, що процес є стаціонарним.

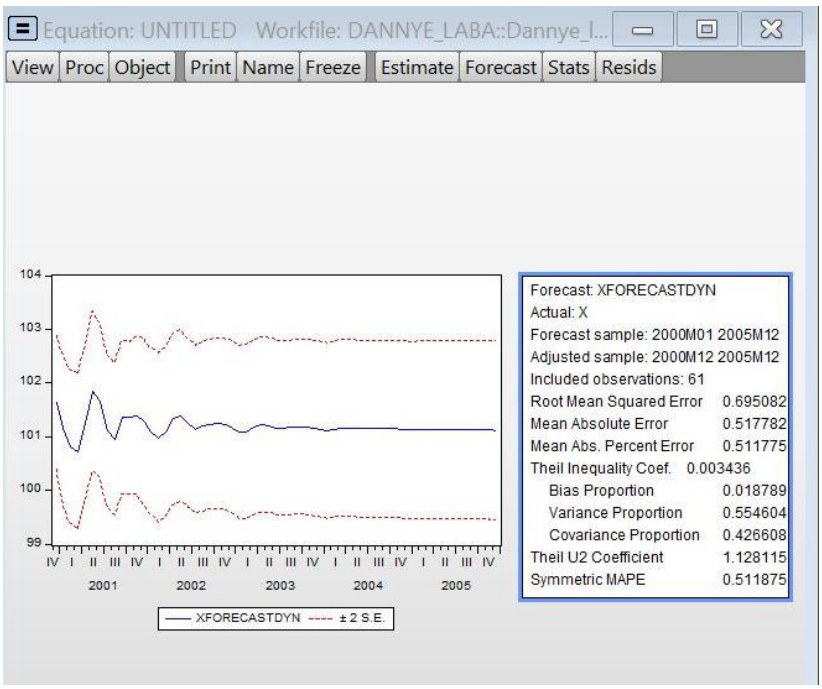
Оцінимо теоретичну та фактичну корелогами для нашого процесу:



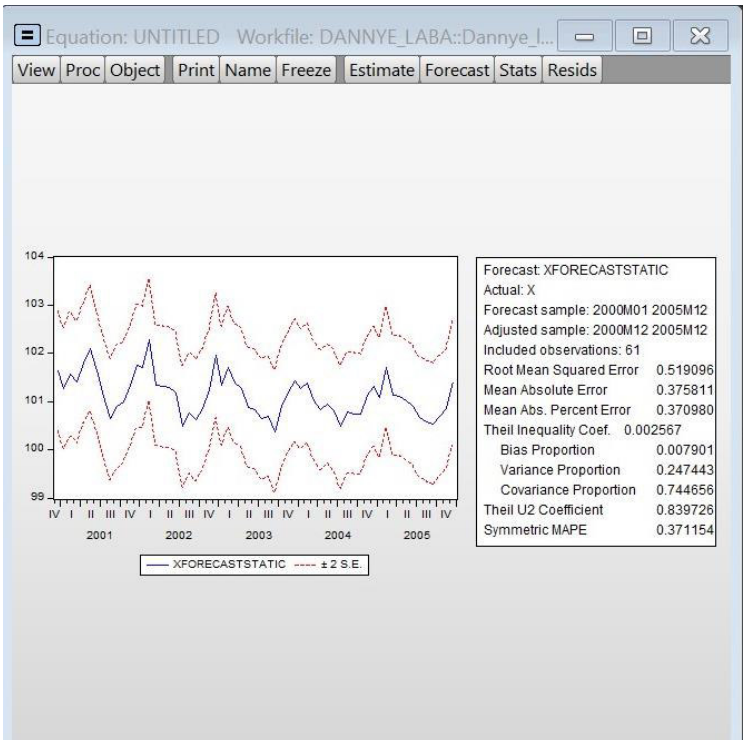
Корелограма процесу більше відповідає фактичній.

8. Отримаємо розрахункові значення індексу цін згідно з обраною моделлю, використовуючи динамічний та статистичний метод прогнозування.

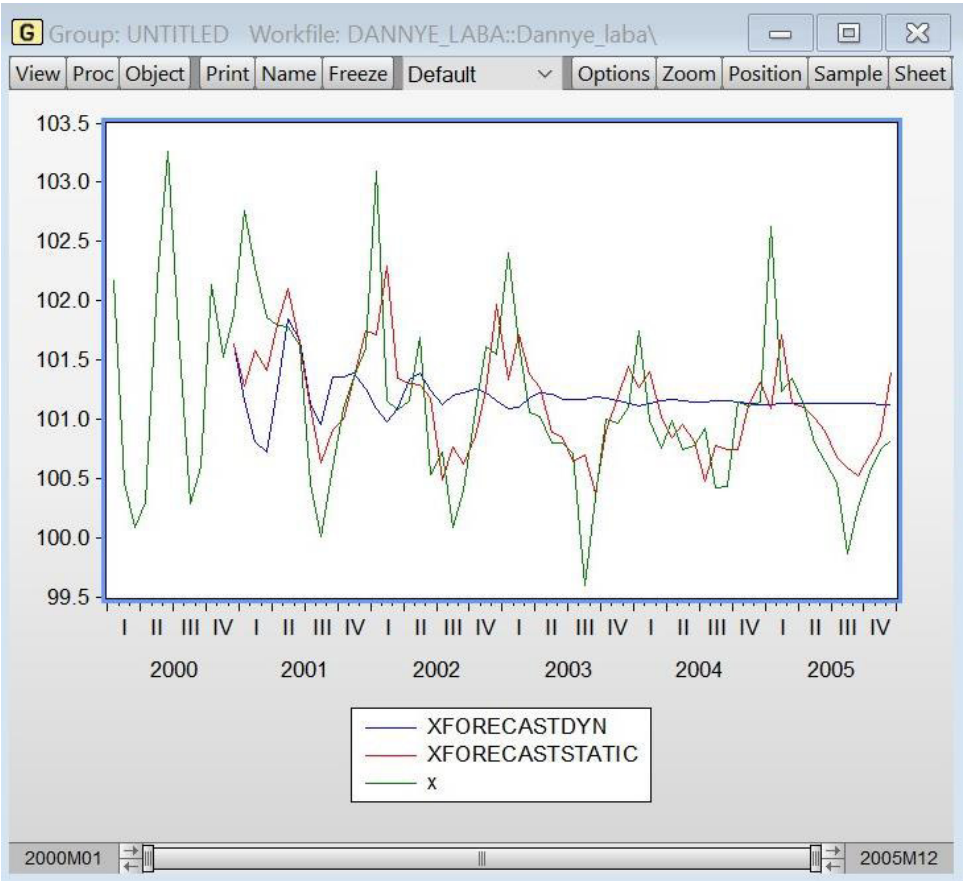
Результати методу динамічного прогнозування:



Результати статичного методу прогнозування:



Для порівняння,відкриємо дані ряди як групу та побудуємо їх графіки:



Бачимо, що отримання розрахункових значень статичним методом дає краще наближення фактичних та розрахункових значень, однак динамічний метод дозволяє виявити тенденцію та краще підходить для прогнозування.

Висновок.

Під час роботи ми ознайомились з програмою EViews та навчились працювати з нею, також ми ознайомилися з поняттям часового ряду, побудували його графік, перевірили ряд на стаціонарність; ми розглянули методи аналізу часового ряду та моделювання його поведінки,виходячи з його попередніх значень.

Індивідуальні дані для виконання роботи обираються згідно до посилань на сайти з прикладної статистики, вказані в посібнику до курсу [1] узгоджено з викладачем, що веде практичні заняття.

Практичне заняття № 7

ARCH та GARCH моделі

Розглядається період часу 2002-2010, який включає період світової кризи. Вже це дає привід думати, що обмежитись використанням стандартних підходів і методів прикладної статистики в даному випадку буде недостатньо.

У файлі wf1 → дані зміни курсу євро до долару за визначений період: 2002-2010 років.

Будемо використовувати дані закриття (close) ринку. Спочатку визначемо тип процесу та порядок інтегровності вихідного ряду.

1. Проаналізуємо графік зміни курсу за 8 років

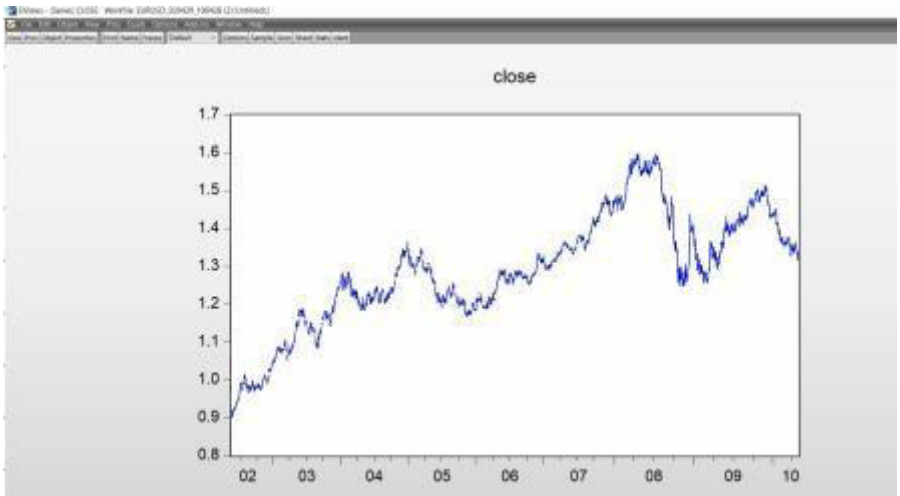
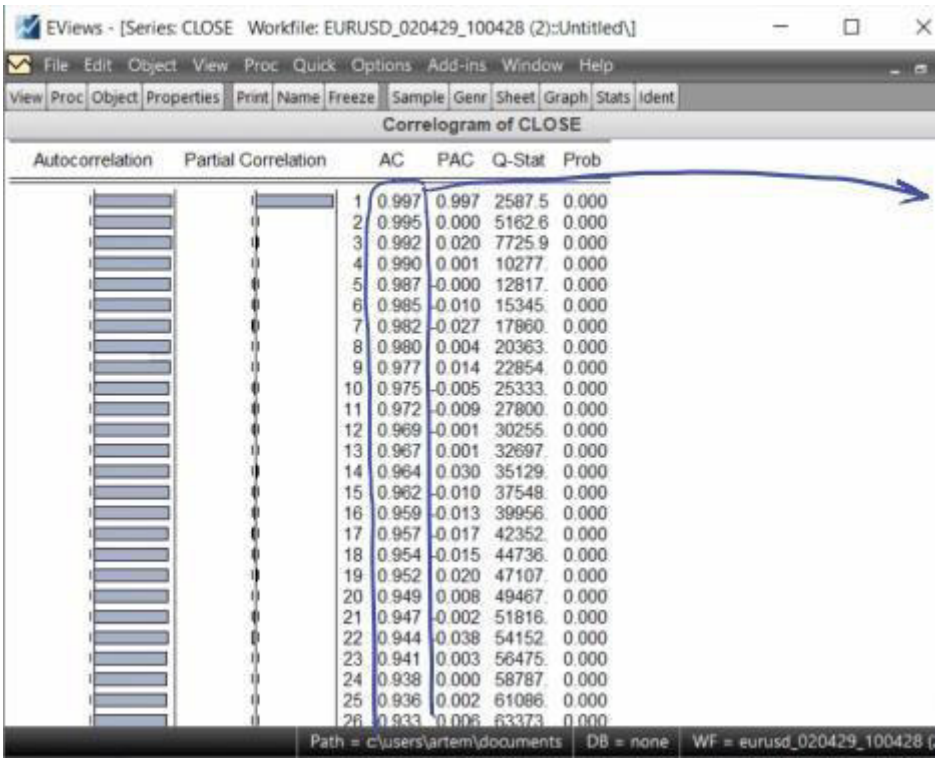


Рисунок 1. Графік нашого вихідного ряду

2. Будуємо корелограму (correlogram of close) ряду котирувань валют



	AC	PAC	Q-Stat	Prob
1	0.997	0.997	2587.5	0.000
2	0.995	0.000	5162.6	0.000
3	0.992	0.020	7725.9	0.000
4	0.990	0.001	10277.	0.000
5	0.987	-0.000	12817.	0.000
6	0.985	-0.010	15345.	0.000
7	0.982	-0.027	17860.	0.000
8	0.980	0.004	20363.	0.000
9	0.977	0.014	22854.	0.000
10	0.975	-0.005	25333.	0.000
11	0.972	-0.009	27800.	0.000
12	0.969	-0.001	30255.	0.000
13	0.967	0.001	32697.	0.000
14	0.964	0.030	35129.	0.000
15	0.962	-0.010	37548.	0.000
16	0.959	-0.013	39956.	0.000
17	0.957	-0.017	42352.	0.000
18	0.954	-0.015	44736.	0.000
19	0.952	0.020	47107.	0.000
20	0.949	0.008	49467.	0.000
21	0.947	-0.002	51816.	0.000
22	0.944	-0.038	54152.	0.000
23	0.941	0.003	56475.	0.000
24	0.938	0.000	58787.	0.000
25	0.936	0.002	61086.	0.000
26	0.933	0.006	63373.	0.000

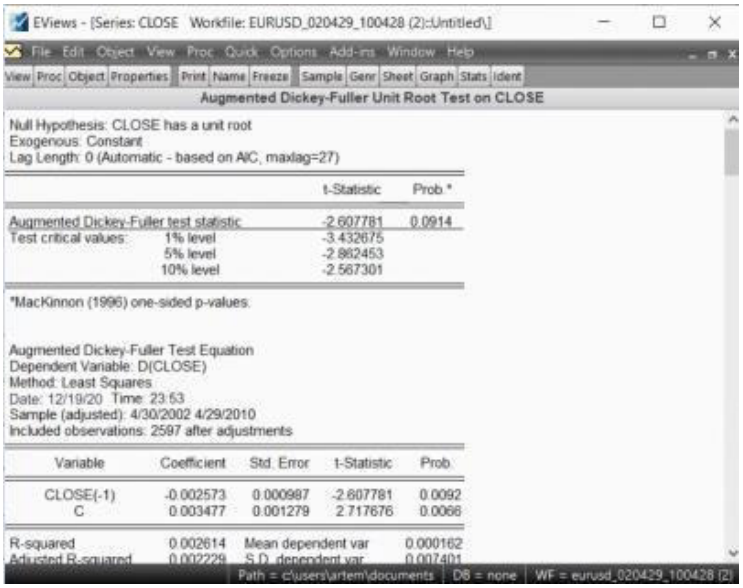
Бачимо, що коефіцієнт AC ≈ 1.

Отже, наш ряд може бути описаний процесом випадкового блукання.

Рисунок. Корелограма ряду котирувань валют

3. Проведемо тест Дікі-Фуллера та оцінимо ряд на стаціонарність.

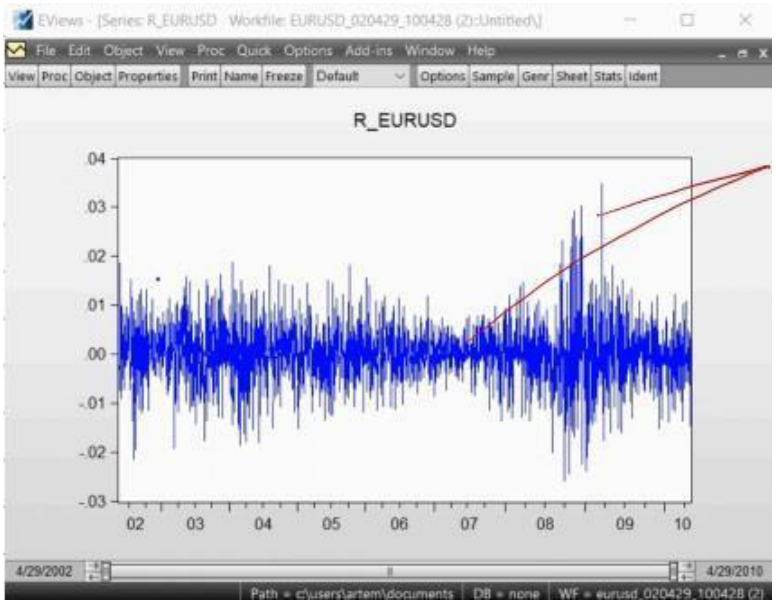
Оскільки всі коефіцієнти автокореляції збігаються до нуля, то порядок інтегрування ряду дорівнює 1, тобто різниця нашого ряду котирувань є стаціонарною.Рисунок . Результат тесту Дікі-Фуллера для ряду котирувань



валют

4. Побудуємо логарифмічну прибутковість r_eurusd ряда котірувань валют. Виконаємо команди:

Quick → Generate Series → R_EURUSD= LOG(close).



На графіку добре видно періоди з низькою та високою волатильністю

Рисунок 4. Графік ряду лог-прибутковості

5. Оцінимо рівняння. Виконаємо наступні команди:

Equation Estimation → ARCH → EURUSD C.

Збережемо наш результат як Eq01.

EViews - [Equation: EQ01 Workfile: EURUSD_020429_100428 (2)::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: R_EURUSD
Method: Least Squares
Date: 12/19/20 Time: 00:06
Sample (adjusted): 4/30/2002 4/29/2010
Included observations: 2597 after adjustments

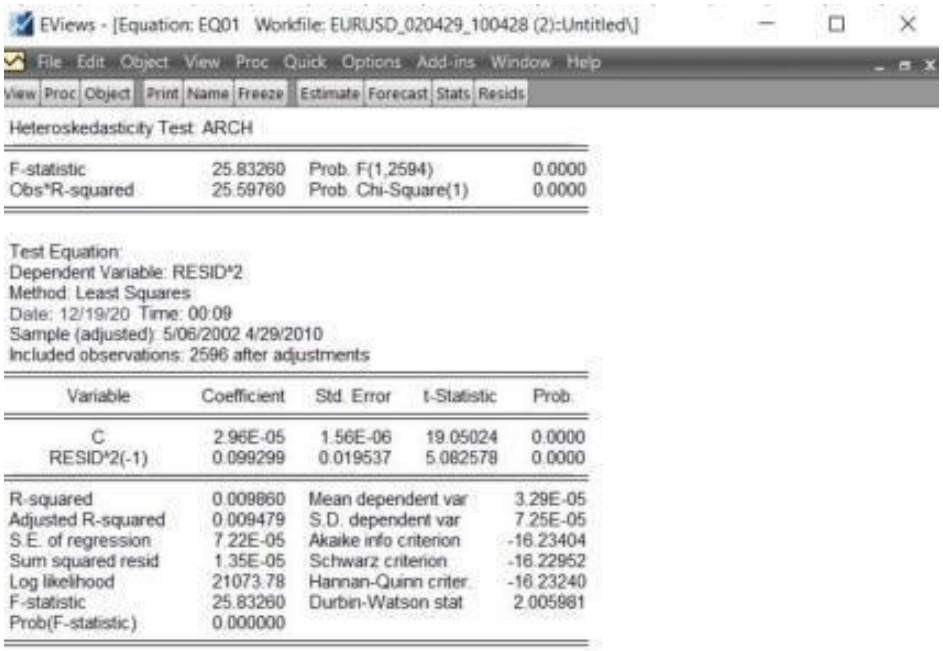
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000147	0.000113	1.308482	0.1908

R-squared	0.000000	Mean dependent var	0.000147
Adjusted R-squared	0.000000	S.D. dependent var	0.005736
S.E. of regression	0.005736	Akaike info criterion	-7.483564
Sum squared resid	0.085427	Schwarz criterion	-7.481307
Log likelihood	9718.408	Hannan-Quinn criter.	-7.482746
Durbin-Watson stat	2.035442		

Рисунок 5. Результат оцінки рівняння методом найменших квадратів

6. Тест на умовну гетероскедастичність. Виконаємо команди:

View → Residual Diagnostics → Heteroskedasticity Test → Arch (1)



EViews - [Equation: EQ01 Workfile: EURUSD_020429_100428 (2)::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Heteroskedasticity Test: ARCH

F-statistic	25.83260	Prob. F(1,2594)	0.0000
Obs*R-squared	25.59760	Prob. Chi-Square(1)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/19/20 Time: 00:09
Sample (adjusted): 5/06/2002 4/29/2010
Included observations: 2596 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.96E-05	1.56E-06	19.05024	0.0000
RESID^2(-1)	0.099299	0.019537	5.082578	0.0000

R-squared	0.009860	Mean dependent var	3.29E-05
Adjusted R-squared	0.009479	S.D. dependent var	7.25E-05
S.E. of regression	7.22E-05	Akaike info criterion	-16.23404
Sum squared resid	1.35E-05	Schwarz criterion	-16.22952
Log likelihood	21073.78	Hannan-Quinn criter.	-16.23240
F-statistic	25.83260	Durbin-Watson stat	2.005981
Prob(F-statistic)	0.000000		

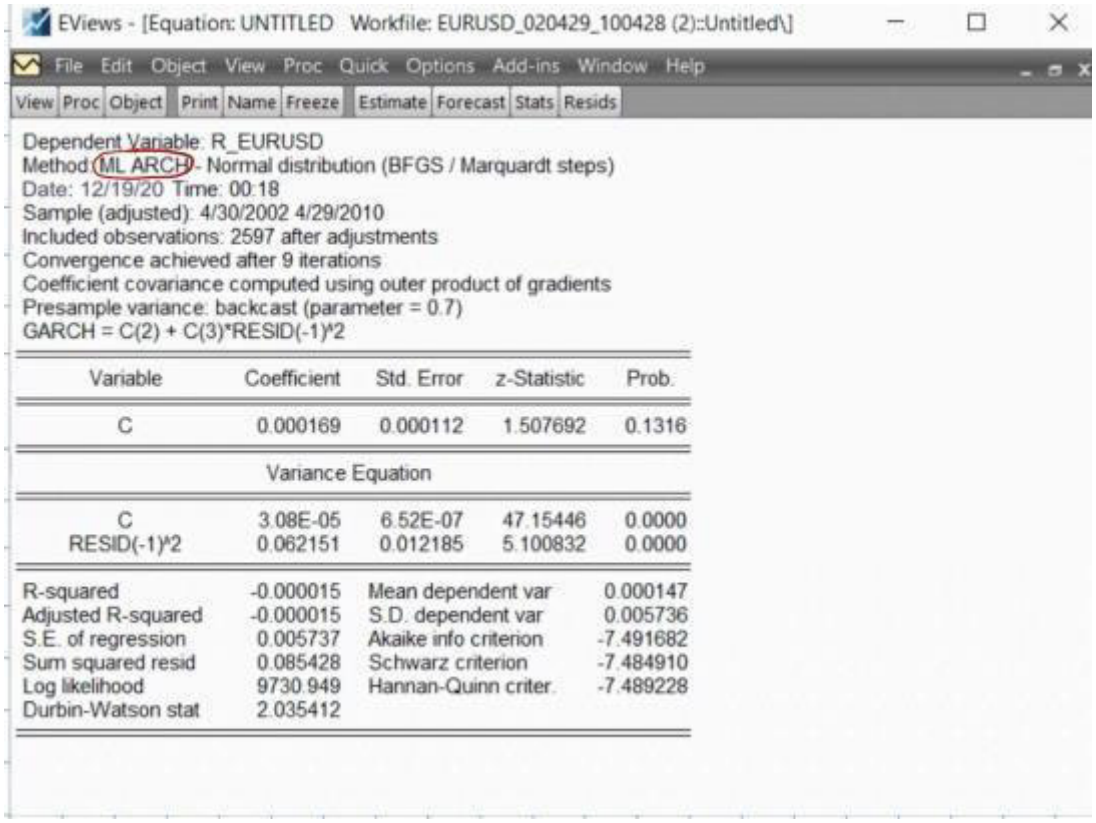
Рисунок 6. Тестування умовної гетероскедастичності в моделі

7. Оцінимо модель ARCH (1). Виконаємо команди:

Equation Estimation → Method ARCH → Order: ARCH -1, GARCH - 0.

Для оцінки моделі на умовну гетероскедастичність використали метод максимальної правдоподібності.

Збережемо результат у файлі та назвемо як Eq02



EViews - [Equation: UNTITLED Workfile: EURUSD_020429_100428 (2)::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: R_EURUSD
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 12/19/20 Time: 00:18
Sample (adjusted): 4/30/2002 4/29/2010
Included observations: 2597 after adjustments
Convergence achieved after 9 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000169	0.000112	1.507692	0.1316
Variance Equation				
C	3.08E-05	6.52E-07	47.15446	0.0000
RESID(-1)^2	0.062151	0.012185	5.100832	0.0000

R-squared	-0.000015	Mean dependent var	0.000147
Adjusted R-squared	-0.000015	S.D. dependent var	0.005736
S.E. of regression	0.005737	Akaike info criterion	-7.491682
Sum squared resid	0.085428	Schwarz criterion	-7.484910
Log likelihood	9730.949	Hannan-Quinn criter.	-7.489228
Durbin-Watson stat	2.035412		

Рисунок 7. Результат оцінки моделі ARCH (1)

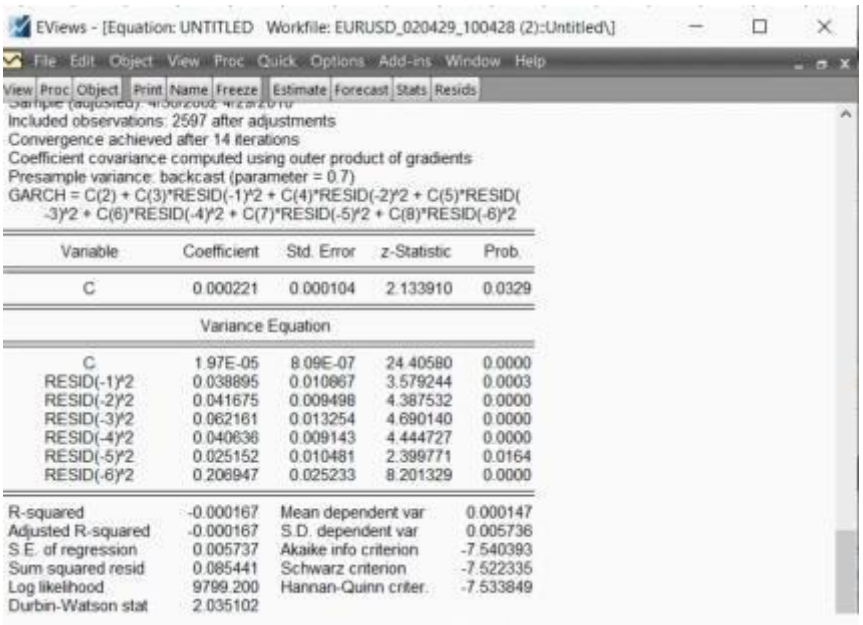
8. Перевіримо отриману модель на адекватність.

9. Виконаємо команди:

View→ Residual Diagnostics → ARCH LM Test (6)

(тобто кількість лагів рівна 6 у моделі)

Тест показав наявність ARCH - ефектів. Збережемо це рівняння як Eq03.



EViews - [Equation: UNTITLED Workfile: EURUSD_020429_100428 (2)::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Included observations: 2597 after adjustments
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*RESID(-2)^2 + C(5)*RESID(-3)^2 + C(6)*RESID(-4)^2 + C(7)*RESID(-5)^2 + C(8)*RESID(-6)^2

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000221	0.000104	2.133910	0.0329
Variance Equation				
C	1.97E-05	8.09E-07	24.40580	0.0000
RESID(-1)^2	0.038895	0.010867	3.579244	0.0003
RESID(-2)^2	0.041675	0.009498	4.387532	0.0000
RESID(-3)^2	0.062161	0.013254	4.690140	0.0000
RESID(-4)^2	0.040636	0.009143	4.444727	0.0000
RESID(-5)^2	0.025152	0.010481	2.399771	0.0164
RESID(-6)^2	0.206947	0.025233	8.201329	0.0000

R-squared	-0.000167	Mean dependent var	0.000147
Adjusted R-squared	-0.000167	S.D. dependent var	0.005736
S.E. of regression	0.005737	Akaike info criterion	-7.540393
Sum squared resid	0.085441	Schwarz criterion	-7.522335
Log likelihood	9799.200	Hannan-Quinn criter.	-7.533849
Durbin-Watson stat	2.035102		

Рисунок 8. Результат тестування додаткових ARCH - ефектів

10. Оцінка моделі GARCH - ряду умовної дисперсії. Виконаємо команди: Equation Estimation → Method ARCH → Order: ARCH -1, GARCH - 1.

EViews - [Equation: UNTITLED Workfile: EURUSD_020429_100428 (2)::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: R_EURUSD
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 12/19/20 Time: 00:33
Sample (adjusted): 4/30/2002 4/29/2010
Included observations: 2597 after adjustments
Convergence achieved after 30 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

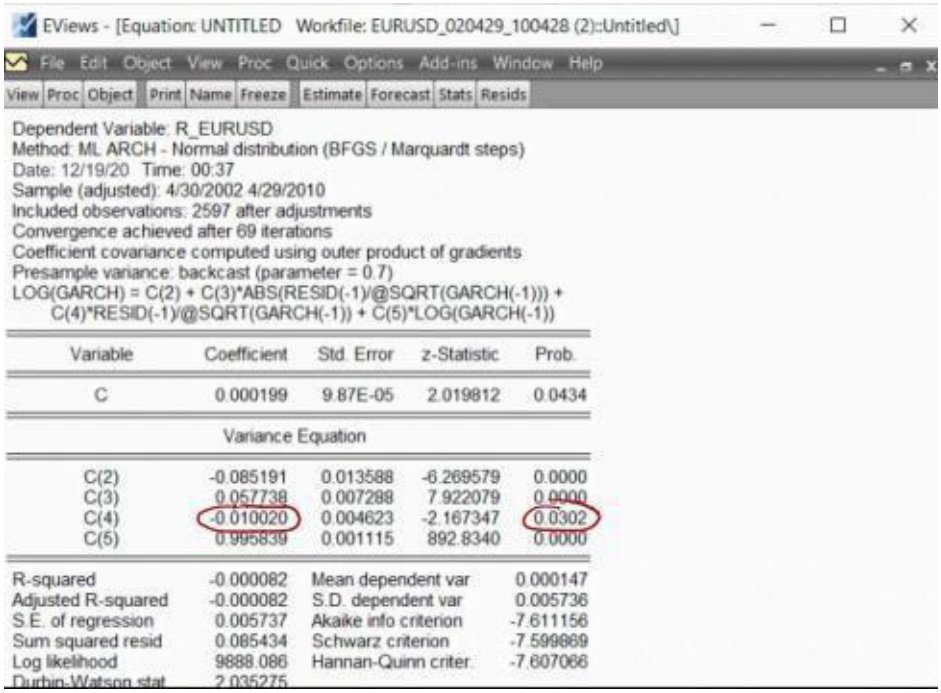
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000208	9.68E-05	2.147424	0.0318
Variance Equation				
C	8.88E-08	3.12E-08	2.841730	0.0045
RESID(-1)^2	0.024607	0.002949	8.343908	0.0000
GARCH(-1)	0.972446	0.002956	328.9794	0.0000

R-squared	-0.000111	Mean dependent var	0.000147
Adjusted R-squared	-0.000111	S.D. dependent var	0.005736
S.E. of regression	0.005737	Akaike info criterion	-7.617013
Sum squared resid	0.085436	Schwarz criterion	-7.607984
Log likelihood	9894.691	Hannan-Quinn criter.	-7.613741
Durbin-Watson stat	2.035216		

Рисунок 9. Результат моделі GARCH (1,1)

11. Оцінка асиметричної моделі EGARCH. Виконаємо команди: Equation Estimation → Method ARCH

→ Order: ARCH -1, GARCH - 1, Asymmetric order - 1 → r_eurusd c



EViews - [Equation: UNTITLED Workfile: EURUSD_020429_100428 (2)::Untitled\]

File Edit Object View Proc Quick Options Add-ins Window Help

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: R_EURUSD
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 12/19/20 Time: 00:37
Sample (adjusted): 4/30/2002 4/29/2010
Included observations: 2597 after adjustments
Convergence achieved after 69 iterations
Coefficient covariance computed using outer product of gradients
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(2) + C(3)*ABS(RESID(-1)/@SQRT(GARCH(-1))) + C(4)*RESID(-1)/@SQRT(GARCH(-1)) + C(5)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000199	9.87E-05	2.019812	0.0434
Variance Equation				
C(2)	-0.085191	0.013588	-6.269579	0.0000
C(3)	0.057738	0.007288	7.922079	0.0000
C(4)	-0.010020	0.004623	-2.167347	0.0302
C(5)	0.995839	0.001115	892.8340	0.0000

R-squared	-0.000082	Mean dependent var	0.000147
Adjusted R-squared	-0.000082	S.D. dependent var	0.005736
S.E. of regression	0.005737	Akaike info criterion	-7.611156
Sum squared resid	0.085434	Schwarz criterion	-7.599669
Log likelihood	9888.086	Hannan-Quinn criter.	-7.607066
Durbin-Watson stat	2.035275		

Асиметрія відсутня

Рисунок 10. Результат моделі EGARCH (1,1)

12. Для побудови ряду умовних дисперсій виконаємо команди: Proc → Make GARCH Variance Series - GARCH03.

Аналогічно побудуємо ряди для Eq04 та Eq05. Об‘єднаємо ряди в групу та побудуємо графік їх умовної волатильності.

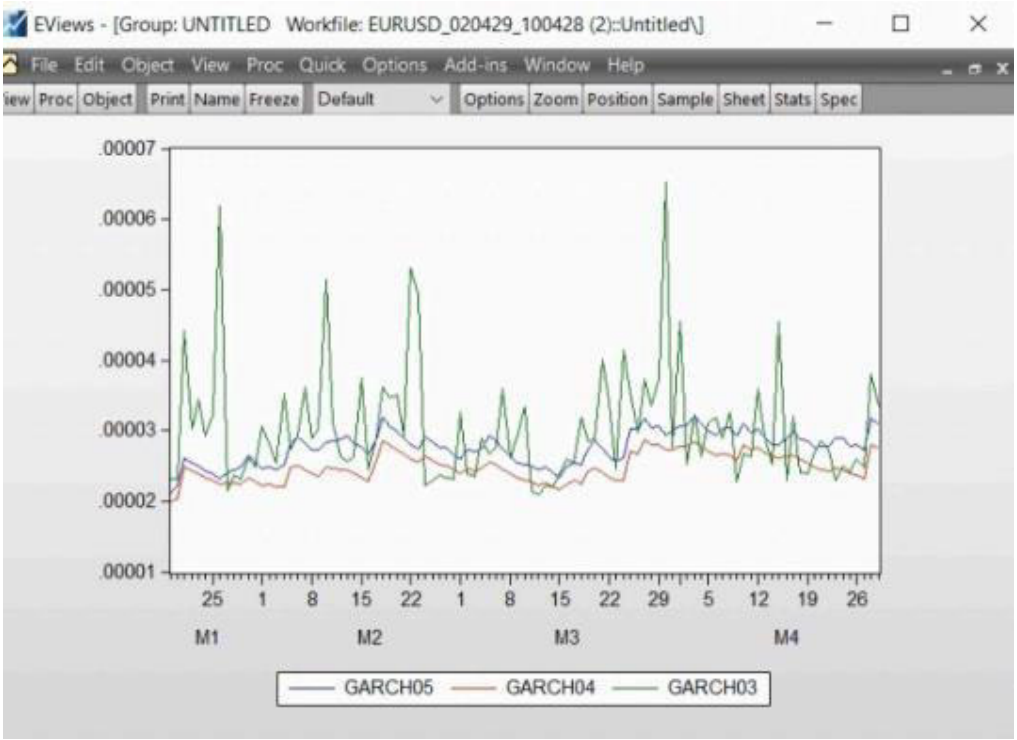


Рисунок 11. Графік умовної волатильності по моделям ARCH(6), GARCH(1,1) і EGARCH(1,1)

Висновок. Волатильність ARCH(6) є менш гладкою, ніж в інших моделях, зроблених на основі наших попередніх розрахунків, що видно з графіку вище.

Практичне заняття №8

Модель ціноутворення фінансових активів (CAPM) як модель регресії

Вихідні дані

За період з березня 2010 р по жовтень 2020 р. котирування акцій компаній (по закриттю біржі) на Нью-Йоркській фондовій біржі (NYSE):

https://finance.yahoo.com/quote/ABT/history?p=ABT (джерело даних для PFE, MRK, ABT, NYSE)
PFE (Pfizer Inc. –фармацевтична компанія)
MRK (Merck & Co., Inc –фармацевтична компанія)
ABT (Abbott Laboratoies – хіміко-фармацевтична корпорація)
NYSE (New York Stock Exchange) – Нью-Йоркській фондовий індекс

https://www.federalreserve.gov/data.htm (джерено даних для TB_3)
TB_3 (3-month Treasury bill secondary market rate discount basis) – дохідність по тримісячним казначейським векселям США, у відсотках на 360-денній базі
JAN – фіктивна змінна, що приймає значення 1 для січня і 0 для інших місяців.

	A	B	C	D	E	F
1	DATE	PFE	MRK	ABT	TB-3	JAN
2	01.03.2010	17.150000	37.349998	25.275864	0.15	0
3	01.04.2010	16.719999	35.040001	24.546568	0.16	0
4	01.05.2010	15.230000	33.689999	22.819288	0.16	0
5	01.06.2010	14.260000	34.970001	22.445044	0.12	0
6	01.07.2010	15.000000	34.459999	23.548584	0.16	0
7	01.08.2010	15.910000	35.160000	23.673332	0.16	0
8	01.09.2010	17.170000	36.810001	25.064753	0.15	0
9	01.10.2010	17.420000	36.310001	24.623337	0.13	0
10	01.11.2010	16.299999	34.470001	22.315498	0.14	0
11	01.12.2010	17.510000	36.040001	22.987219	0.14	0
12	01.01.2011	18.219999	33.169998	21.667768	0.15	1
13	01.02.2011	19.240000	32.570000	23.078381	0.13	0
14	01.03.2011	20.309999	33.009998	23.534189	0.1	0
15	01.04.2011	20.969999	35.950001	24.968792	0.06	0
16	01.05.2011	21.450001	36.750000	25.069550	0.04	0
17	01.06.2011	20.600000	35.290001	25.247076	0.04	0
18	01.07.2011	19.250000	34.130001	24.623337	0.04	0
19	01.08.2011	18.980000	33.099998	25.194298	0.02	0

Series: NYSE Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

NYSE

Last updated: 11/13/20 - 12:51

Imported from 'C:\Users\admin\Documents\KPI\3 курс\1 semester\...

2010M03	7447.800
2010M04	7474.400
2010M05	6791.570
2010M06	6469.650
2010M07	6998.990
2010M08	6704.150
2010M09	7281.070
2010M10	7513.350
2010M11	7430.940
2010M12	7964.020
2011M01	8139.160
2011M02	8438.550
2011M03	8404.980
2011M04	8671.410
2011M05	8477.280

Series: PFE Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

Last updated: 11/13/20 - 12:48

Imported from 'C:\Users\admin\Documents\KPI\3 курс\1 semester\...

2010M03	17.15000
2010M04	16.72000
2010M05	15.23000
2010M06	14.26000
2010M07	15.00000
2010M08	15.91000
2010M09	17.17000
2010M10	17.42000
2010M11	16.30000
2010M12	17.51000
2011M01	18.22000
2011M02	19.24000
2011M03	20.31000
2011M04	20.97000
2011M05	21.45000
2011M06	20.60000

Series: MRK Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

Last updated: 11/13/20 - 12:48

Imported from 'C:\Users\admin\Documents\KPI\3 курс\1 semester\...

2010M03	37.35000
2010M04	35.04000
2010M05	33.69000
2010M06	34.97000
2010M07	34.46000
2010M08	35.16000
2010M09	36.81000
2010M10	36.31000
2010M11	34.47000
2010M12	36.04000
2011M01	33.17000
2011M02	32.57000
2011M03	33.01000
2011M04	35.95000
2011M05	36.75000
2011M06	35.29000
2011M07	34.13000
2011M08	33.10000
2011M09	32.70000

Series: ABT Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

Last updated: 11/13/20 - 12:48

Imported from 'C:\Users\admin\Documents\KPI\3 курс\1 semester\...

2010M03	25.27586
2010M04	24.54657
2010M05	22.81929
2010M06	22.44504
2010M07	23.54858
2010M08	23.67333
2010M09	25.06475
2010M10	24.62334
2010M11	22.31550
2010M12	22.98722
2011M01	21.66777
2011M02	23.07838
2011M03	23.53419
2011M04	24.96879
2011M05	25.06955
2011M06	25.24708
2011M07	24.62334

Series: TB_3 Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

TB-3

Last updated: 11/13/20 - 12:48

Imported from 'C:\Users\admin\Documents\KPI\3 курс\1 semester\...

Original name: TB-3

2010M03	0.15
2010M04	0.16
2010M05	0.16
2010M06	0.12
2010M07	0.16
2010M08	0.16
2010M09	0.15
2010M10	0.13
2010M11	0.14
2010M12	0.14
2011M01	0.15
2011M02	0.13
2011M03	0.10
2011M04	0.06
2011M05	0.04
2011M06	0.04

Series: JAN Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

Last updated: 11/13/20 - 12:48

Imported from 'C:\Users\admin\Documents\KPI\3 курс\1 semester\...

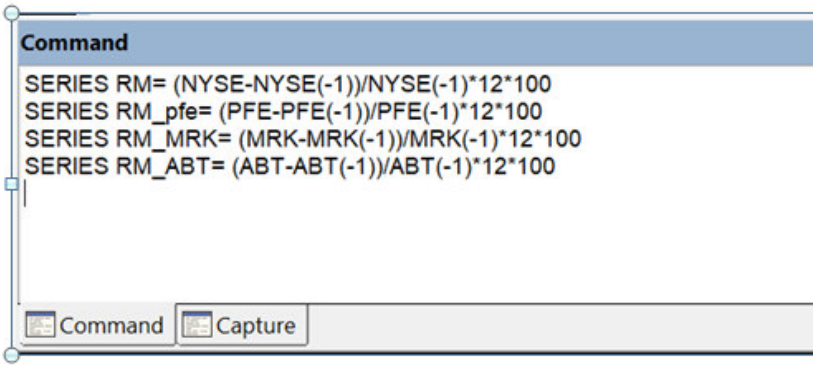
2010M03	0
2010M04	0
2010M05	0
2010M06	0
2010M07	0
2010M08	0
2010M09	0
2010M10	0
2010M11	0
2010M12	0
2011M01	1
2011M02	0
2011M03	0
2011M04	0
2011M05	0

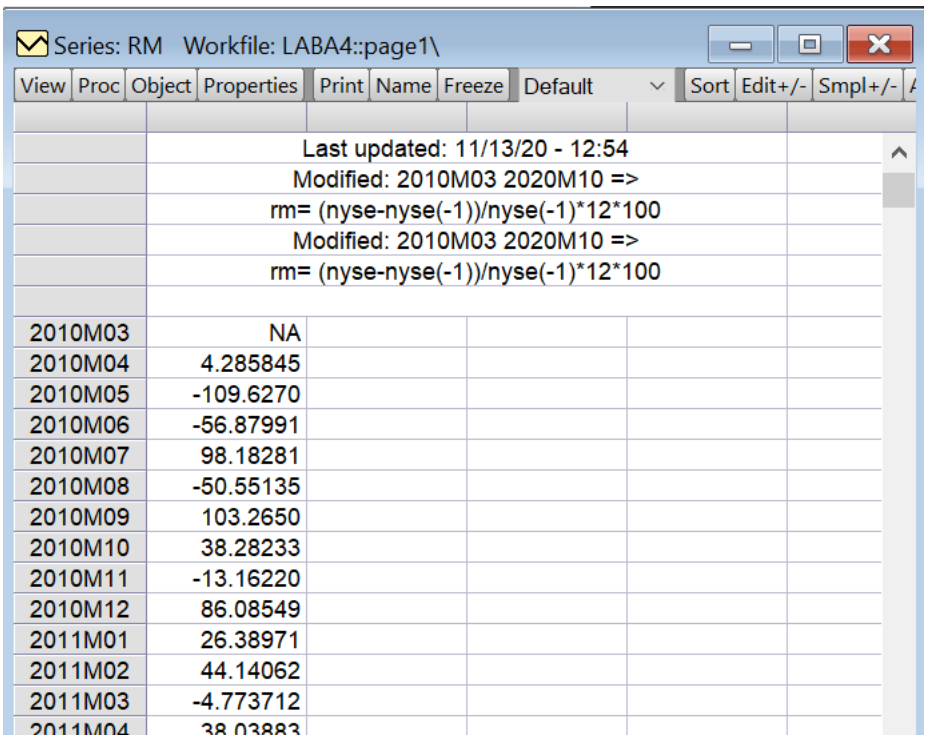
ЗАВДАННЯ

1. Побудувати ряди простих річних прибутковостей акцій та індексів за правилом

$$r_j = \frac{p_{jt} - p_{jt-1}}{p_{jt-1}} * 12 * 100$$

де p_j – ціна акцій компанії





2010M03	NA
2010M04	4.285845
2010M05	-109.6270
2010M06	-56.87991
2010M07	98.18281
2010M08	-50.55135
2010M09	103.2650
2010M10	38.28233
2010M11	-13.16220
2010M12	86.08549
2011M01	26.38971
2011M02	44.14062
2011M03	-4.773712
2011M04	38.03883

Series: RM_PFE Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

Last updated: 11/13/20 - 12:54
Modified: 2010M03 2020M10 =>
rm_pfe= (pfe-pfe(-1))/pfe(-1)*12*100

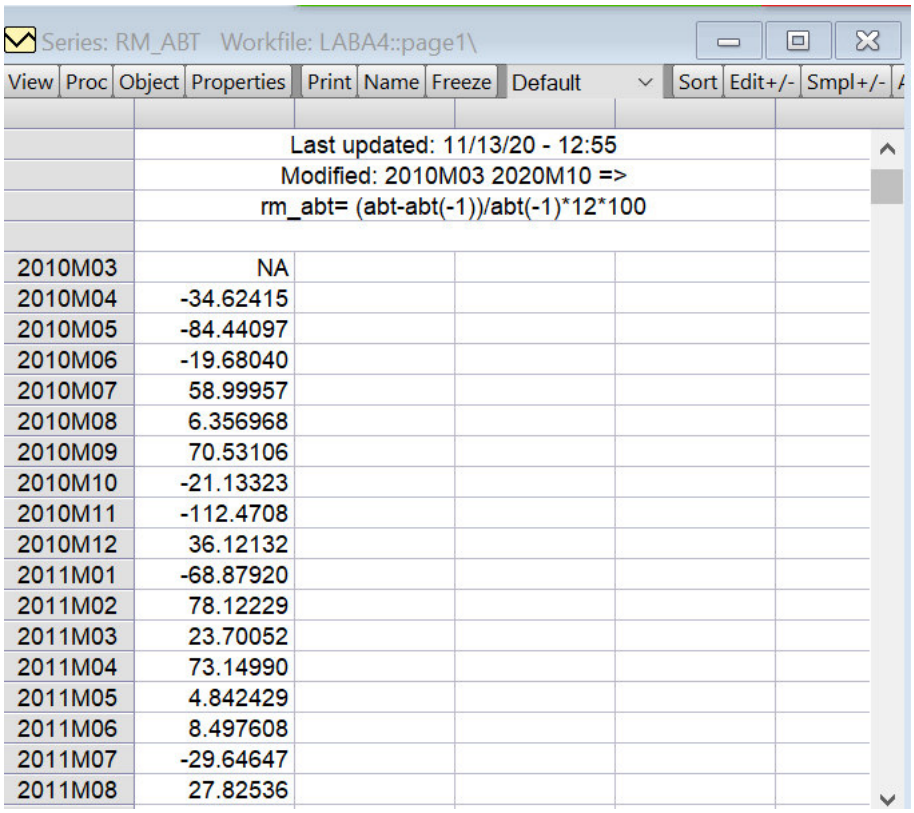
2010M03	NA
2010M04	-30.08753
2010M05	-106.9377
2010M06	-76.42810
2010M07	62.27209
2010M08	72.80000
2010M09	95.03457
2010M10	17.47234
2010M11	-77.15277
2010M12	89.07983
2011M01	48.65784
2011M02	67.17899
2011M03	66.73590
2011M04	38.99557
2011M05	27.46793

Series: RM_MRK Workfile: LABA4::page1\

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+/-

Last updated: 11/13/20 - 12:55
Modified: 2010M03 2020M10 =>
rm_mrk= (mrk-mrk(-1))/mrk(-1)*12*100

2010M03	NA
2010M04	-74.21678
2010M05	-46.23294
2010M06	45.59224
2010M07	-17.50078
2010M08	24.37612
2010M09	56.31403
2010M10	-16.29992
2010M11	-60.80969
2010M12	54.65622
2011M01	-95.56059
2011M02	-21.70629



2010M03	NA
2010M04	-34.62415
2010M05	-84.44097
2010M06	-19.68040
2010M07	58.99957
2010M08	6.356968
2010M09	70.53106
2010M10	-21.13323
2010M11	-112.4708
2010M12	36.12132
2011M01	-68.87920
2011M02	78.12229
2011M03	23.70052
2011M04	73.14990
2011M05	4.842429
2011M06	8.497608
2011M07	-29.64647
2011M08	27.82536

2. Оцінити модель виду (без вільного члена)

$$r_j - r_f = \beta_j(r_m - r_f) + \varepsilon_j$$

де r_j – прибутковість акцій компанії, j=1,2,3, rm – прибутковість ринку за NYSE, r_f – безризикова прибутковість.

Equation: EQ_PFE Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_PFE-TB_3
Method: Least Squares
Date: 11/13/20 Time: 12:59
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.731605	0.092389	7.918727	0.0000

R-squared	0.321592	Mean dependent var	7.938254
Adjusted R-squared	0.321592	S.D. dependent var	62.94530
S.E. of regression	51.84524	Akaike info criterion	10.74225
Sum squared resid	338679.1	Schwarz criterion	10.76464
Log likelihood	-681.1326	Hannan-Quinn criter.	10.75135
Durbin-Watson stat	2.026542		

Equation: EQ_MRK Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_MRK-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:00
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.455632	0.094279	4.832809	0.0000

R-squared	0.142018	Mean dependent var	7.422574
Adjusted R-squared	0.142018	S.D. dependent var	57.11670
S.E. of regression	52.90566	Akaike info criterion	10.78274
Sum squared resid	352675.2	Schwarz criterion	10.80514
Log likelihood	-683.7040	Hannan-Quinn criter.	10.79184
Durbin-Watson stat	2.137939		

Equation: EQ_ABT Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_ABT-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:01
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

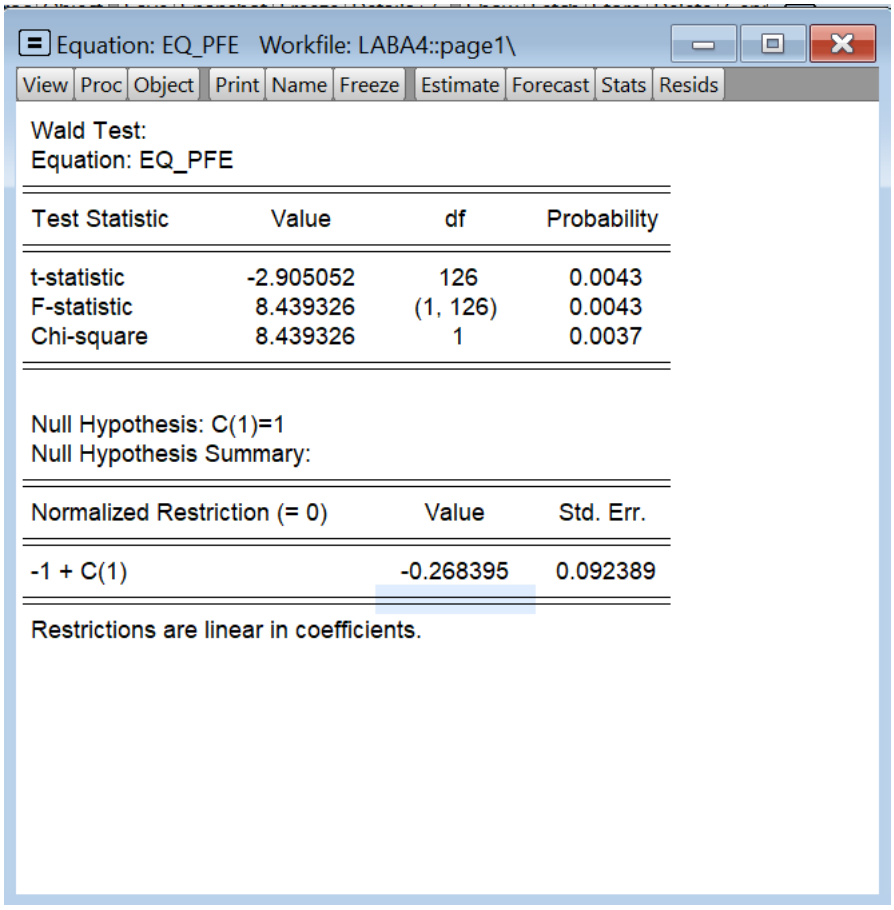
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.758000	0.093331	8.121651	0.0000

R-squared	0.307971	Mean dependent var	14.61371
Adjusted R-squared	0.307971	S.D. dependent var	62.95798
S.E. of regression	52.37364	Akaike info criterion	10.76253
Sum squared resid	345617.8	Schwarz criterion	10.78492
Log likelihood	-682.4204	Hannan-Quinn criter.	10.77163
Durbin-Watson stat	1.962964		

Компанія	Оцінка бета-коефіцієнта	Стандартна похибка оцінки бета-коефіцієнта
Pfizer Inc.	0.73	0.092
Merck & Co	0.46	0.094
Abbott Laboratoies	0.76	0.093

Отримані оцінки бета-коефіцієнтів відрізняються від 1 для всіх компаній, однак, щоб говорити, що ця відмінність статистично значима, треба провести тестування.

3. Протестувати на 5%-му рівні гіпотезу про рівність коефіцієнта βj = 1 для кожної компанії. Якщо гіпотеза відхиляється, протестувати гіпотези про "захисний/агресивний" характер акцій (провести односторонній тест Стьюдента на рівні 1%). Зробити висновки.



Wald Test:
Equation: EQ_PFE

Test Statistic	Value	df	Probability
t-statistic	-2.905052	126	0.0043
F-statistic	8.439326	(1, 126)	0.0043
Chi-square	8.439326	1	0.0037

Null Hypothesis: C(1)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(1)	-0.268395	0.092389

Restrictions are linear in coefficients.

Equation: EQ_MRK Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Wald Test:
Equation: EQ_MRK

Test Statistic	Value	df	Probability
t-statistic	-5.774022	126	0.0000
F-statistic	33.33933	(1, 126)	0.0000
Chi-square	33.33933	1	0.0000

Null Hypothesis: C(1)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(1)	-0.544368	0.094279

Restrictions are linear in coefficients.

Equation: EQ_ABT Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Wald Test:
Equation: EQ_ABT

Test Statistic	Value	df	Probability
t-statistic	-2.592927	126	0.0106
F-statistic	6.723269	(1, 126)	0.0106
Chi-square	6.723269	1	0.0095

Null Hypothesis: C(1)=1
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
-1 + C(1)	-0.242000	0.093331

Restrictions are linear in coefficients.

Табличне значення одностороннього критерія Стьюдента для 126 степенів свободи та 1% рівня значущості $t_{кр}$=2.615

Результати тестів на значення бета-коефіцієнтів:

Компанія	Тест Уальда $H_0: \beta_j = 1$ $H_1: \beta_j \neq 1$	Лівосторонній тест Стьюдента $H_0: \beta_j = 1$ $H_1: \beta_j < 1$	Правосторонній тест Стьюдента $H_0: \beta_j = 1$ $H_1: \beta_j > 1$	Характер акцій
Pfizer	H_0 відхиляється на рівні 1%	H_0 відхиляється на рівні 1%	-	захисний
Merck	H_0 відхиляється на рівні 1%	H_0 відхиляється на рівні 1%	-	захисний
Abbott	H_0 відхиляється на рівні 1%	-	-	варіюється

Отримані результати свідчать про захисний (або принаймні не суто агресивний) характер акцій компаній, що відповідає дійсності – великі корпорації (якими і є розглянуті компанії) прагнуть, в першу чергу, до стабільності.

4. Включити в модель вільний член:

$$r_j - r_f = \propto + \beta_j(r_m - r_f) + \varepsilon_j$$

Оцінити відповідні регресії. Протестувати гіпотезу про рівність нулю вільного члена.

Equation: EQ_PFE_1 Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: PFE-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:18
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.004928	0.012589	0.391442	0.6961
C	30.11252	0.626884	48.03526	0.0000

R-squared	0.001224	Mean dependent var	30.13874
Adjusted R-squared	-0.006766	S.D. dependent var	7.000521
S.E. of regression	7.024164	Akaike info criterion	6.752212
Sum squared resid	6167.360	Schwarz criterion	6.797003
Log likelihood	-426.7655	Hannan-Quinn criter.	6.770410
F-statistic	0.153227	Durbin-Watson stat	0.049089
Prob(F-statistic)	0.696137		

Equation: EQ_MRK_1 Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: MRK-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:21
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	-0.001290	0.027381	-0.047121	0.9625
C	55.83876	1.363462	40.95366	0.0000

R-squared	0.000018	Mean dependent var	55.83189
Adjusted R-squared	-0.007982	S.D. dependent var	15.21684
S.E. of regression	15.27745	Akaike info criterion	8.306255
Sum squared resid	29175.05	Schwarz criterion	8.351046
Log likelihood	-525.4472	Hannan-Quinn criter.	8.324453
F-statistic	0.002220	Durbin-Watson stat	0.035598
Prob(F-statistic)	0.962492		

Equation: EQ_ABT_1 Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: ABT-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:23
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	-0.005537	0.038521	-0.143741	0.8859
C	47.73534	1.918173	24.88583	0.0000

R-squared	0.000165	Mean dependent var	47.70587
Adjusted R-squared	-0.007833	S.D. dependent var	21.40924
S.E. of regression	21.49293	Akaike info criterion	8.988948
Sum squared resid	57743.26	Schwarz criterion	9.033738
Log likelihood	-568.7982	Hannan-Quinn criter.	9.007146
F-statistic	0.020661	Durbin-Watson stat	0.022372
Prob(F-statistic)	0.885937		

З отриманих результатів видно, що отримана оцінка змінної суттєво відрізняється від оцінки рівняння без вільного члену, і стає майже не значущою, в той час як ймовірність незначущості вільного члена стає мінімальною. З цього робимо висновок що запропонована авторами модель регресії не описує наші дані з належною точністю, й в реальній задачі доцільніше було б використовувати не лінійну по змінній модель. Для подальшого аналізу використовуватимемо модель з завдання 2.

5. Дати економічну інтерпретацію коефіцієнтів детермінації в розглянутих регресіях.

Значення коефіцієнтів детермінації

Компанія	Коефіцієнт детермінації (у рівнянні без вільного члена)
Pfizer	0,32
Merck	0,14
Abbott	0,30

Статистичний зміст коефіцієнта детермінації полягає в тому, що він показує, яка доля загальної дисперсії залежної змінної пояснюється рівнянням регресії. Оскільки дисперсія прибутковості є оцінкою ризику, то повний ризик можна

представити як суму систематичного (поясненого рівнянням регресії) і несистематичного (непоясненим рівнянням) ризику. Таким чином, в таблиці вище наведено долі систематичного ризику компанії, наприклад для Pfizer доля систематичного ризику становить 32%.

6. Перевірити наявність автокореляції першого порядку за тестом Дарбіна-Уотсона

Equation: EQ_PFE Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_PFE-TB_3
Method: Least Squares
Date: 11/13/20 Time: 12:59
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.731605	0.092389	7.918727	0.0000

R-squared	0.321592	Mean dependent var	7.938254
Adjusted R-squared	0.321592	S.D. dependent var	62.94530
S.E. of regression	51.84524	Akaike info criterion	10.74225
Sum squared resid	338679.1	Schwarz criterion	10.76464
Log likelihood	-681.1326	Hannan-Quinn criter.	10.75135
Durbin-Watson stat	2.026542		

Equation: EQ_MRK Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_MRK-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:00
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.455632	0.094279	4.832809	0.0000

R-squared	0.142018	Mean dependent var	7.422574
Adjusted R-squared	0.142018	S.D. dependent var	57.11670
S.E. of regression	52.90566	Akaike info criterion	10.78274
Sum squared resid	352675.2	Schwarz criterion	10.80514
Log likelihood	-683.7040	Hannan-Quinn criter.	10.79184
Durbin-Watson stat	2.137939		

Equation: EQ_ABT Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_ABT-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:01
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.758000	0.093331	8.121651	0.0000

R-squared	0.307971	Mean dependent var	14.61371
Adjusted R-squared	0.307971	S.D. dependent var	62.95798
S.E. of regression	52.37364	Akaike info criterion	10.76253
Sum squared resid	345617.8	Schwarz criterion	10.78492
Log likelihood	-682.4204	Hannan-Quinn criter.	10.77163
Durbin-Watson stat	1.962964		

Тестові статистики для тесту Дарбіна-Уотсона для всіх трьох компаній близькі до значення 2, тому можна припустити, що гіпотеза про відсутність автокореляції не відхиляється. Перевіримо: З рівнем значущості 0.05 та для 127 спостережень табличні значення статистики Дарбіна-Уотсона $d_L \approx 1.65$, $d_U \approx 1.69$. Для кожної з компаній статистика Дарбіна-Уотсона потрапляє в інтервал $(d_U; 4 - d_U) = (1.69, 2.31)$, отже на рівні значущості 5% гіпотеза про відсутність автокореляції приймається (як для рівняння з вільним членом, так і для рівняння без нього).

7. Перевірити гіпотезу про існування так званого “ефекту січня”, тобто припущення про те, що при інших рівних умовах прибутковість в січні відрізняється від прибутковості в інші місяці.

Equation: EQ_PFE_2 Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_PFE-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:57
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.736293	0.092899	7.925695	0.0000
JAN	-10.50525	16.48545	-0.637244	0.5251

R-squared	0.323788	Mean dependent var	7.938254
Adjusted R-squared	0.318379	S.D. dependent var	62.94530
S.E. of regression	51.96787	Akaike info criterion	10.75475
Sum squared resid	337582.4	Schwarz criterion	10.79954
Log likelihood	-680.9267	Hannan-Quinn criter.	10.77295
Durbin-Watson stat	2.019194		

Equation: EQ_MRK_2 Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_MRK-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:57
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.453919	0.094934	4.781432	0.0000
JAN	3.837766	16.84645	0.227809	0.8202

R-squared	0.142374	Mean dependent var	7.422574
Adjusted R-squared	0.135513	S.D. dependent var	57.11670
S.E. of regression	53.10584	Akaike info criterion	10.79807
Sum squared resid	352528.8	Schwarz criterion	10.84286
Log likelihood	-683.6777	Hannan-Quinn criter.	10.81627
Durbin-Watson stat	2.134110		

Equation: EQ_ABT_2 Workfile: LABA4::page1\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: RM_ABT-TB_3
Method: Least Squares
Date: 11/13/20 Time: 13:57
Sample (adjusted): 2010M04 2020M10
Included observations: 127 after adjustments

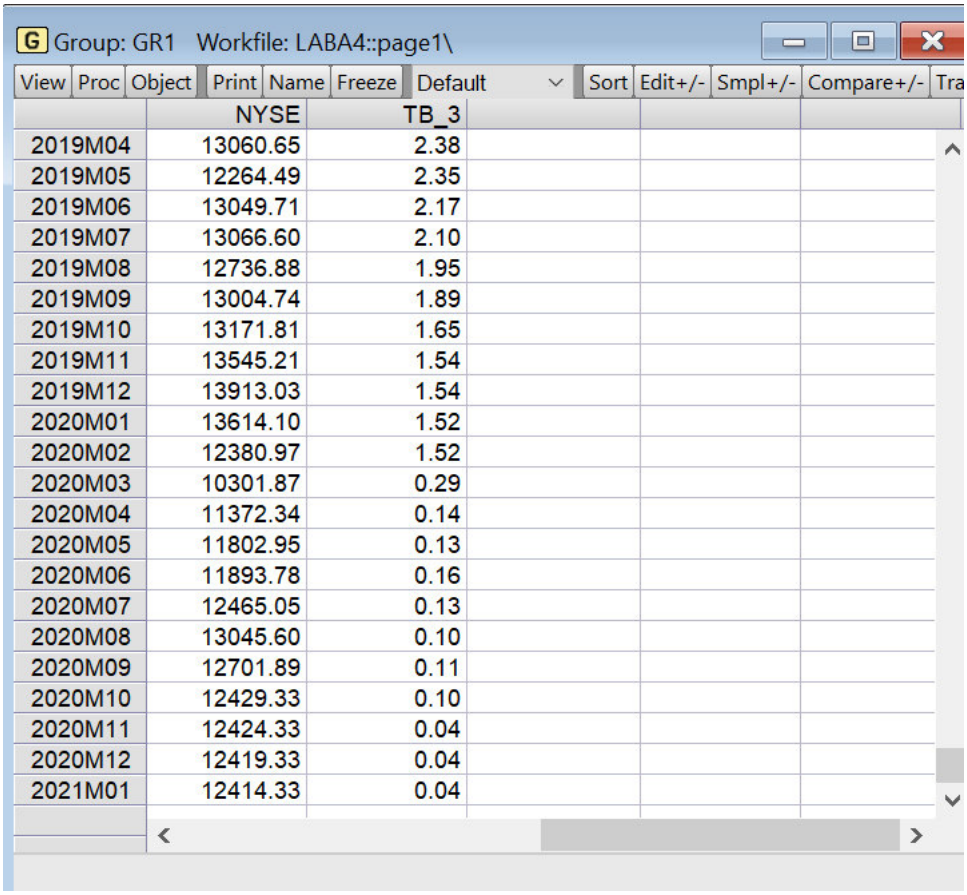
Variable	Coefficient	Std. Error	t-Statistic	Prob.
RM-TB_3	0.764903	0.093675	8.165526	0.0000
JAN	-15.46849	16.62302	-0.930546	0.3539

R-squared	0.312732	Mean dependent var	14.61371
Adjusted R-squared	0.307234	S.D. dependent var	62.95798
S.E. of regression	52.40153	Akaike info criterion	10.77137
Sum squared resid	343240.0	Schwarz criterion	10.81616
Log likelihood	-681.9821	Hannan-Quinn criter.	10.78957
Durbin-Watson stat	1.931787		

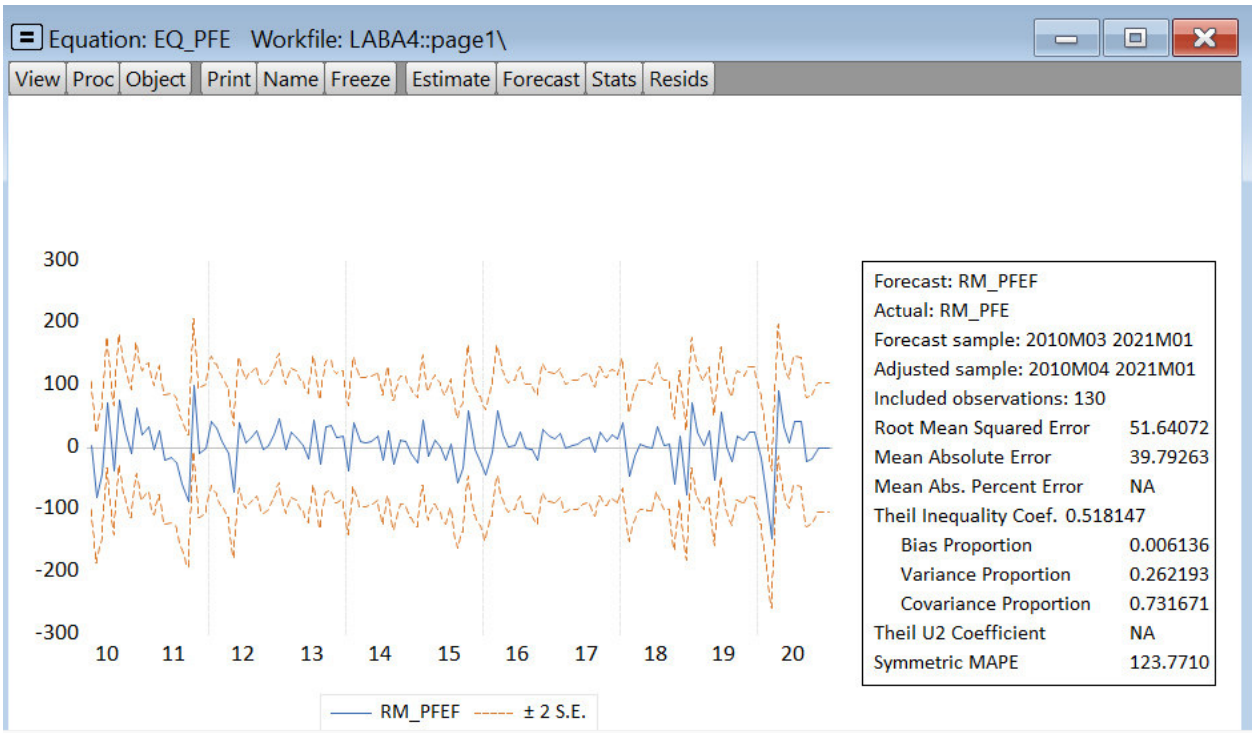
Компанія	Оцінка бета-коефіцієнта	Оцінка коефіцієнта при фіктивній змінній	Коефіцієнт детермінації
Pfizer Inc.	0.74 (0.00)	-10.51 (0.52)	0.32
Merck & Co	0.45 (0.00)	3.84 (0.82)	0.14
Abbott Laboratoies	0.76 (0.00)	-15.47 (0.35)	0.31

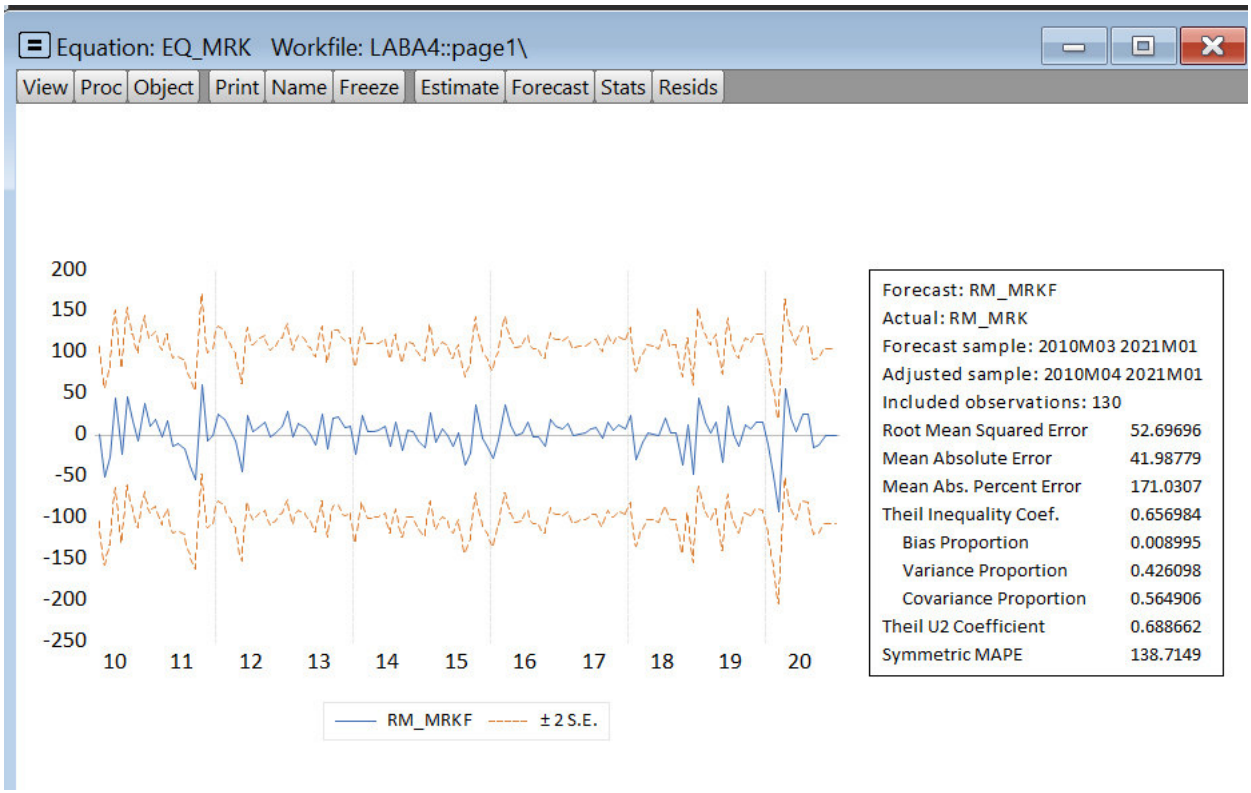
Результати оцінок показують, що ні для одної з компаній ми не можемо відхилити гіпотезу про відсутність ефекту січня на рівні значущості 5%, судячи по коефіцієнтам він має негативний напрямок – зниження прибутковості в січні для Pfizer Inc. та Abbott Laboratoies. У Merck & Co спостерігається зростання прибутковості, проте ця компанія найменш схильна до впливу серед розглянутих.

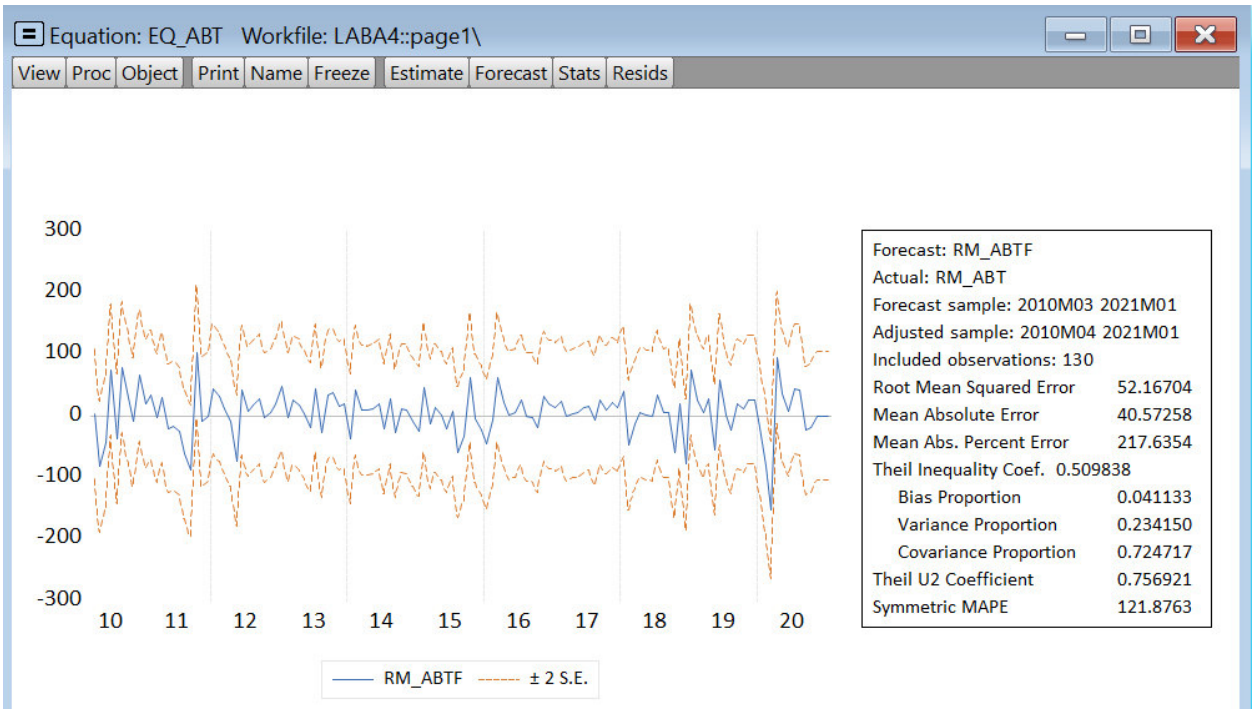
8. Зробити прогноз відносно очікуваної прибутковості розглянутих акцій на три місяці, якщо очікується пониження індекса NYSE на 5 пунктів в місяць, а безризикова дохідність збережеться на рівні 0,04% річних.



	NYSE	TB_3
2019M04	13060.65	2.38
2019M05	12264.49	2.35
2019M06	13049.71	2.17
2019M07	13066.60	2.10
2019M08	12736.88	1.95
2019M09	13004.74	1.89
2019M10	13171.81	1.65
2019M11	13545.21	1.54
2019M12	13913.03	1.54
2020M01	13614.10	1.52
2020M02	12380.97	1.52
2020M03	10301.87	0.29
2020M04	11372.34	0.14
2020M05	11802.95	0.13
2020M06	11893.78	0.16
2020M07	12465.05	0.13
2020M08	13045.60	0.10
2020M09	12701.89	0.11
2020M10	12429.33	0.10
2020M11	12424.33	0.04
2020M12	12419.33	0.04
2021M01	12414.33	0.04


Equation: EQ_PFE Workfile: LABA4::page1\
View Proc Object Print Name Freeze Estimate Forecast Stats Resids
300
200
100
0
-100
-200
-300
10 11 12 13 14 15 16 17 18 19 20
RM_PFEF
± 2 S.E.
Forecast: RM_PFEF
Actual: RM_PFE
Forecast sample: 2010M03 2021M01
Adjusted sample: 2010M04 2021M01
Included observations: 130
Root Mean Squared Error 51.64072
Mean Absolute Error 39.79263
Mean Abs. Percent Error NA
Theil Inequality Coef. 0.518147
Bias Proportion 0.006136
Variance Proportion 0.262193
Covariance Proportion 0.731671
Theil U2 Coefficient NA
Symmetric MAPE 123.7710





Результати прогнозних значень прибутковостей акцій за моделлю CAPM:

	Pfizer	Merck	Abbott
2020M11	-0,342431	-0,198172	-0,356229
2020M12	-0,342573	-0,198261	-0,356376
2021M01	-0,342715	-0,198349	-0,356523

ВИСНОВКИ

В ході виконання практичного завдання «Модель ціноутворення фінансових активів (CAPM) як модель регресії» навчального посібника [1] з використанням програмного забезпечення EViews було побудовано регресійні моделі ціноутворення фінансових активів для трьох великих фармацевтичних компаній - Pfizer Inc., Merck & Co., Inc та Abbott Laboratoies базуючись на даних за період з березня 2010 р по жовтень 2020 р. котирування акцій компаній на Нью-Йоркській фондовій біржі.

Зауважимо, що запропонована в даному завданні модель для цього класу задач не є найновішою чи найкращою, але може слугувати початком дослідження в бакалаврській, чи магістерській роботі. Після завантаження даних та побудови рядів простих річних прибутковостей, регресійні моделі були створені та протестовані на значущість бета-коефіцієнтів тестам Уальда та Стьюдента та визначено характер акцій для кожної з компаній. Наступним кроком була перевірка доцільності введення вільного члена, під час якої нові моделі було спростовано. Повернувшись до моделей без вільного члена, була надана економічна інтерпретація отриманих коефіцієнтів детермінації, здійснена перевірка на автокореляцію першого порядку за допомогою критерія Дарбіна-Уотсона, а також перевірена гіпотеза про наявність «ефекту січня» (різниці між звичайною прибутковістю та прибутковістю в січні), яка була підтверджена з різним характером для кожної з трьох компаній. Останнім кроком була побудова прогнозу відносно очікуваної прибутковості розглянутих акцій на три наступних місяці, що також було успішно реалізовано.

Індивідуальні дані для виконання роботи обираються згідно до посилань на сайти з прикладної статистики, вказані в посібнику до курсу [1] узгоджено з викладачем, що веде практичні заняття.

Використана література

1. Бідюк, П. І., Данилов В.Я., Жиров О.Л. Прикладна статистика [Електронний ресурс] : навч. посіб. для здобувачів ступеня бакалавра за спеціальностями 122 Комп'ютерні науки, 124 Системний аналіз ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 3,44 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2023. – 186 с. –

https://ela.kpi.ua/handle/123456789/54799

2. Бідюк П.І., Терентьєв О.М., Просянкіна-Жарова Т.І. Прикладна статистика. – Київ: НТУУ «КПІ», 2012. – 510 с.

3. Бідюк П.І., Меняйленко О.С., Половцев О.В. Методи прогнозування. – Луганськ: Альма-матер, 2008. – 305 с.

4. Бідюк П.І., Романенко В.Д., Тимощук О.Л. Аналіз часових рядів. – Київ: Видавництво «Політехніка», 2013. – 607 с

5. Бідюк П.І., Харрінгтон Т., Ткач Б.П. Математична статистика. – Київ: Видавничий Дім «Персонал», 2017. – 348 с

6. Коваленко І.П. Математична статистика. – Київ: Видавничий Дім «Слово», 2012. – 496 с.

7. Бобик О.І., Берегова Г.І., Копитко Б.І. Теорія ймовірностей і математична статистика. – Київ: Професіонал, 2007. – 560 с.

8. Лук'яненко І. Г., Краснікова Л . І. Л 84 Економетрика: П ідручник. — К.: Товариство "Знання", КОО, 1998. — 494 с. (з табл., граф.).

9. Бідюк П.І., Половцев О.В. Аналіз та математичне моделювання економічних процесів перехідного періоду, Київ, НТУУ «КПІ», 1999, 230с.

10. Gelman A., Carlin J.B., Stern H.S., Rubin D.B. Bayesian Data Analysis. – New York: CRC Press, 2004. – 670p. (надається в електронному вигляді).