<u>Лабораторна робота №3</u>

Методичні вказівки до виконання

Тема: “Оцінювання параметрів регресійних рівнянь за допомогою пакету Eviews”

<u>Мета роботи:</u> *Навчитись обчислювати за допомогою пакету Eviews оцінки коефіцієнтів рівнянь регресії, авторегресії, авторегресії із ковзним середнім та множинної регресії на основі вибірок статистичних даних. Пояснити призначення та використання параметрів (статистик), за допомогою яких визначається ступінь адекватності моделі процесу.*

Для зручної навігації по документу використовуйте гіперсилки зміста, зажати кнопку Ctrl + натиснути ліву кнопку миші.

ЗМІСТ

1. Описові статистики часових рядів системи Eviews

До описових статистик часового ряду відносять:

– *<u>середнє</u>* (mean) – значення математичного сподівання;

– *<u>мода</u>* (mode) – це значення елемента ряду, яке зустрічається найчастіше;

– *<u>медіана</u>* (median) – це значення елемента ряду, який є більшим або дорівнює, і одночасно меншим або дорівнює половині інших елементів ряду; наприклад, для ряду значень 4, 4, 5, 7, 10 медіаною є 5, для ряду 5, 4, 11, 4, 7 медіаною також є 5;

– *<u>стандартне відхилення</u>* (standard deviation) – корінь квадратний з дисперсії:

$$\sigma = \sqrt{D(y)} = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}\left[y(k)-y\right]^2},$$

де $N -$ довжина (потужність) ряду; $y -$ середнє значення ряду, $D(y) -$ дисперсія;

– *<u>коефіцієнт асиметрії</u>* (skewness) – характеризує симетричність (хвостів) розподілу і розраховується за виразом:

$$S = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{y(k)-y}{\sigma}\right]^3;$$

якщо $S > 0$, то правий хвіст розподілу довший,

при $S < 0$ довшим є лівий хвіст розподілу;

якщо $S = 0$, то розподіл симетричний (в реальному житті симетричних хвостів не буває);

– *<u>ексцес</u>* (kurtosis) – характеризує відмінність форми розподілу від нормального (Гауса) і розраховується за виразом:

$$K = \frac{1}{N}\sum_{k=1}^{N}\left[\frac{y(k)-y}{\sigma}\right]^4;$$

$K = 3$ для нормального розподілу;

якщо $K > 3$, то форма розподілу буде “гострішою” від нормального;

при $K < 3$ форма розподілу буде “плоскішою” від нормального;

– *статистика Жак-Бера* (Jarque-Bera) – тестова статистика, яка показує наскільки близьким є ряд до нормального розподілу; фактично це різниця між значеннями S і K для досліджуваного ряду та нормального розподілу:

$$JB = \frac{N-n}{6}\left[S^2 + \frac{1}{4}(K-3)^2\right],$$

де $n -$ число коефіцієнтів, використаних для побудови моделі ряду; при нуль-гіпотезі щодо нормальності розподілу статистика Жак-Бера має розподіл χ^2 (хі квадрат) з двома степенями свободи. Ймовірність, зв’язана із статистикою Жак-Бера, показує ймовірність справедливості нуль-гіпотези. Мала ймовірність свідчить про те, що нуль-гіпотезу щодо нормальності розподілу необхідно відхилити.

2. Приклад виводу описової статистики в Eviews

Крок 1. Створіть робочий файл програми **Eviews** (**WorkFIle**) з нерегулярним типом даних (**Undated or irregular**) зі 100 значень (**Start date: 1** та **End date: 100**).

Крок 2. Завантажте файл **File–>Import–>Read Text-Lotus-Excel** з тестовими даними ***example_data.txt*** (файл знаходиться в папці *ATS_lab_03*). Задайте ім’я ряду ***“y”***.

Крок 3. Подвійним кликом лівої кнопки миші, по часовому ряду ***“y”***, відкрийте вікно значень ряду.

Крок 4. У діалоговому вікні **Series: Y Workfile: Untitled** натисніть на кнопку ***View***, а потім на ***Descriptive Statistics*** *і* ***Histogram and Stats***., як показано на рис. 1. **View–> Descriptive Statistics–> Histogram and Stats.** В результаті отримаєте таблицю з описовою статистикою введеного ряду, рис. 2.

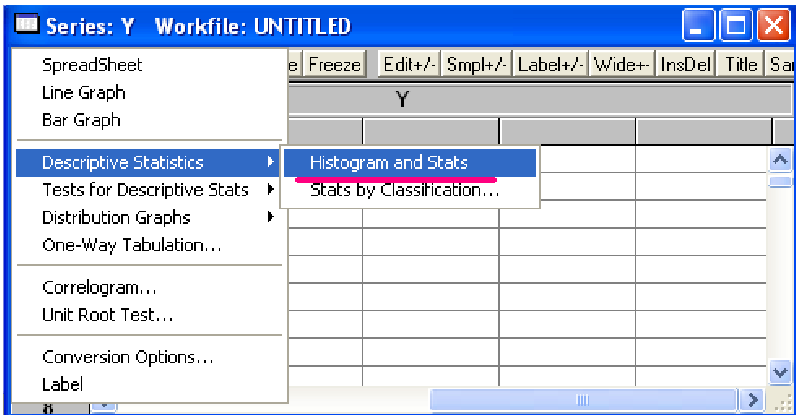


Рис. 1. View–> Descriptive Statistics–> Histogram and Stats

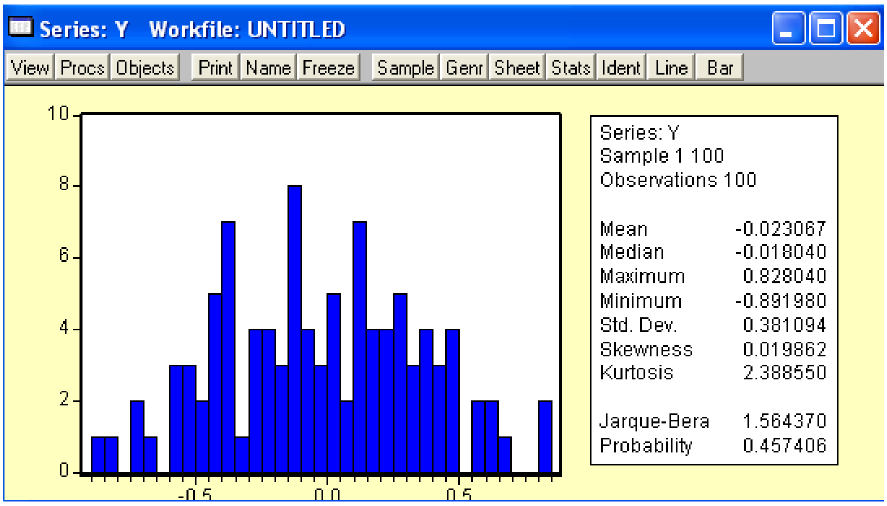


Рис. 2. Описова статистика часового ряду

Додаткові відомості.

1. Для повернення до таблиці завантажених значень часового ряду виконайте послідовність команд **View–>SpreadSheet.**

2. В одному вікні можна дивитися описові статистики декількох часових рядів.

Приклад:

2.1. Згенеруйте додатковий ряд даних. В командній строчці **Eviews** введіть

series y2=y*y

2.2. В діалоговому вікні робочого файлу утримуючи кнопку **shift** виділіть ряди ***"y"*** та ***"y2"***. Натисніть праву кнопку миші у списку, що з'явився, оберіть опції **Open–>as Group**. Для більшого розуміння дивіться рис. 3.

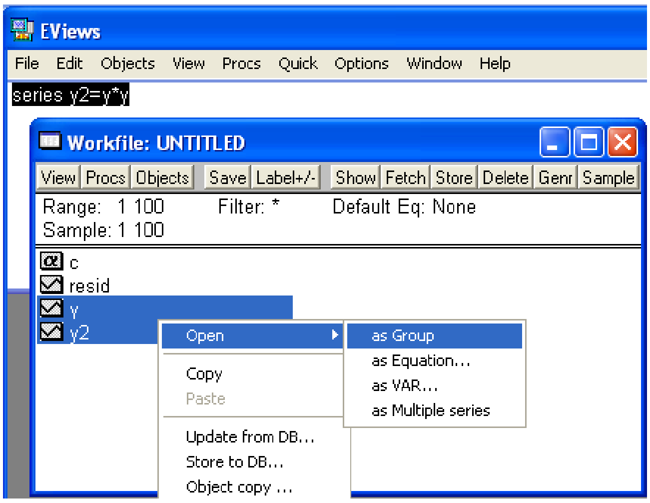


Рис. 3

2.3. У вікні значень рядів ***"y"*** та ***"y2"***, що з'явилося, виконайте команди **View–>Descriptive Statistics–>Common Sample**.

3. Оцінка коефіцієнтів економетрикчних рівнянь в Eviews

Крок 1. Створіть робочий файл програми **Eviews** (**WorkFIle**) з нерегулярним типом даних (**Undated or irregular**) зі 100 значень (**Start date: 1** та **End date: 100**).

Крок 2. Завантажте файл **File–>Import–>Read Text-Lotus-Excel** з тестовими даними ***example_data.txt*** (файл знаходиться в папці *ATS_lab_03*). Задайте ім’я ряду ***“y”***.

<u>Постановка задачі дослідження.</u>

Будемо вважати, що часовий ряд даних ***“y”*** описується авто регресійним рівнянням першого порядку $y(k) = c(1) + c(2) \cdot y(k-1) \;\; AP(1)$ і нам необхідно оцінити коефіцієнти $c(1)$ та $c(2)$.

3.1. Командний режим

Крок 3.

Функція **LS** (Least Squares) системи **Eviews** дозволяє обчислювати коефіцієнти економетричних рівнянь за методом найменших квадратів.

В командній строчці системи **Eviews** введіть команду

<u>Варіант №1.</u>

```
ls y=c(1)+c(2)*y(-1)
```

<u>Варіант №2.</u>

```
ls y c y(-1)
```

<u>Варіант №3.</u>

```
equation eq_my1.ls y=c(1)+c(2)*y(-1)
```

В останньому випадку створюється об’єкт системи **Eviews** з ім’ям **“eq_my1”** в якому зберігається результат виконання функції **ls**.

<u>Варіант №4</u>

```
ls y c ar(1)
```

Аналог попередніх варіантів, в цій специфікації **ar(p)** – позначає авто регресійну складову з лагом p.

Також в описані специфікації команди **ls** можна одразу задавати ковзне середнє з $q-$м лагом параметром **ma(q)**:

ls y c ar(1) ma(1)

3.2. За допомогою діалогових вікон

<u>Варіант №1.</u>

В головному меню системи **Eviews** натисніть кнопку **Quick**, потім **Estimate Equation**, дивись рис. 4. У діалоговому вікні **Equation Specification** введіть специфікацію **y=c(1)+c(2)*y(-1)** або **y c y(-1)**, дивись рис. 5 та 6 відповідно. Натисніть кнопку “**OK**”.

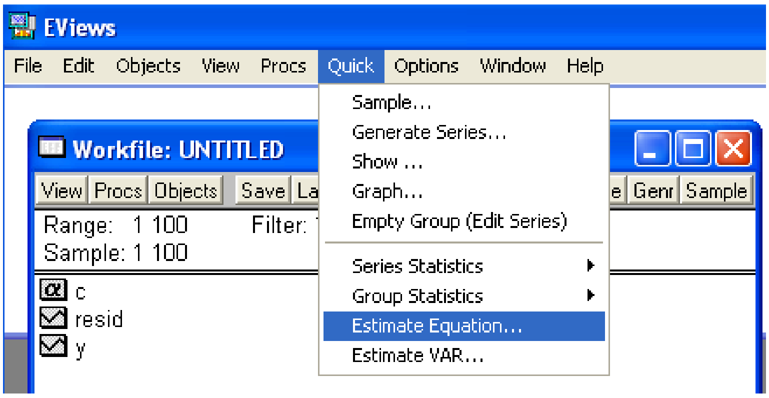


Рис. 4

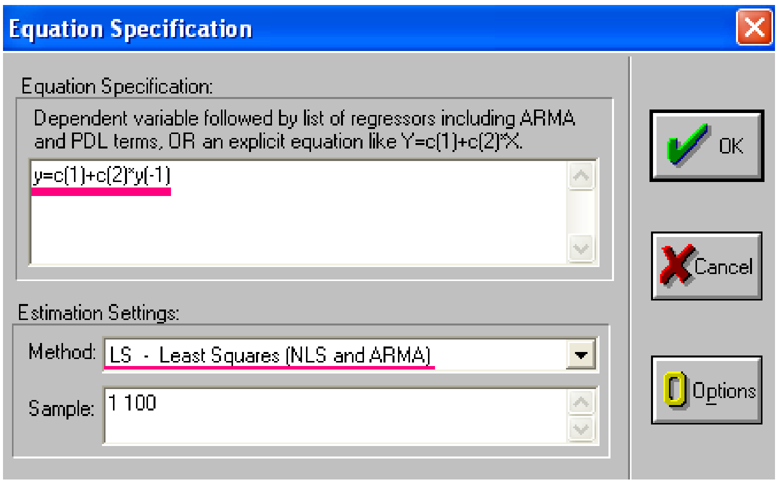


Рис. 5

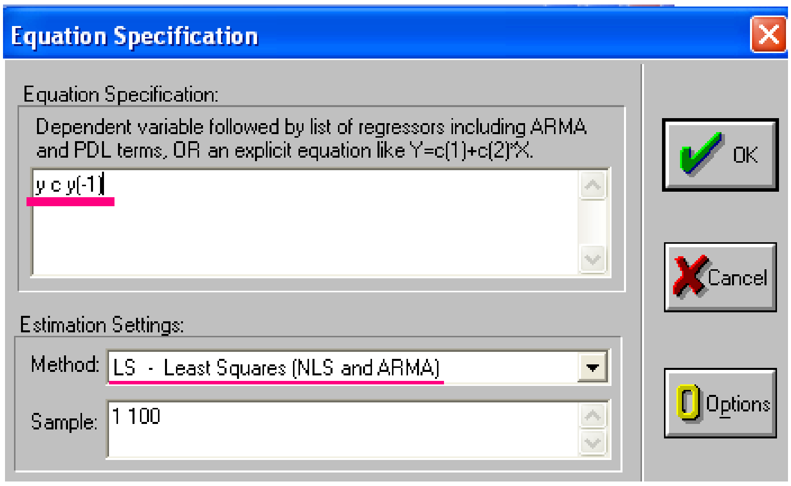


Рис. 6

<u>Варіант №2.</u>

В головному меню системи **Eviews** натисніть кнопку **Objects**, потім **New Object**, дивись рис. 7. У діалоговому вікні New Object оберіть тип об’єкту **Equation** (він встановлений по замовченню) та ім’я (по замовченню **Untitled**), дивись рис. 8. Натисніть кнопку “**OK**”. У діалоговому вікні **Equation Specification** введіть специфікацію **y=c(1)+c(2)*y(-1)** або **y c y(-1)**, дивись рис. 5 та 6 відповідно. Натисніть кнопку “**OK**”.

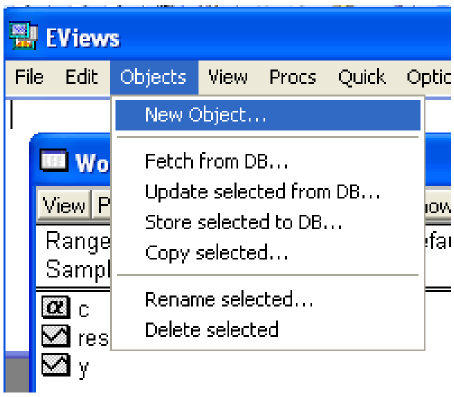


Рис. 7

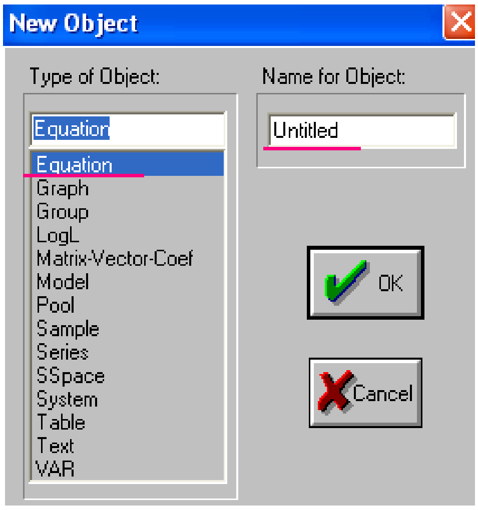
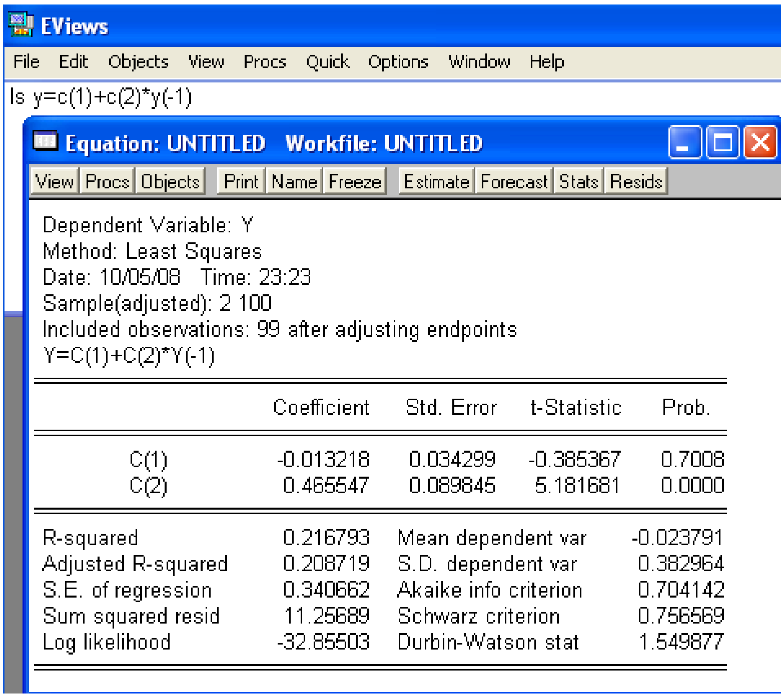


Рис. 8

3.3. Аналіз отриманих результатів оцінювання

В результаті буде отримане вікно з результатами наведене на рис. 9.



EViews

File Edit Objects View Procs Quick Options Window Help

ls y=c(1)+c(2)*y(-1)

Equation: UNTITLED Workfile: UNTITLED

View Procs Objects Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/05/08 Time: 23:23
Sample(adjusted): 2 100
Included observations: 99 after adjusting endpoints
Y=C(1)+C(2)*Y(-1)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.013218	0.034299	-0.385367	0.7008
C(2)	0.465547	0.089845	5.181681	0.0000

R-squared	0.216793	Mean dependent var	-0.023791
Adjusted R-squared	0.208719	S.D. dependent var	0.382964
S.E. of regression	0.340662	Akaike info criterion	0.704142
Sum squared resid	11.25689	Schwarz criterion	0.756569
Log likelihood	-32.85503	Durbin-Watson stat	1.549877

Рис. 9 Результат оцінювання коефіцієнтів

Тобто часовий ряд даних ***“y”*** описується авто регресійним рівнянням першого порядку $y(k) = c(1) + c(2) \cdot y(k-1) = -0{,}0132 + 0{,}4655 \cdot y(k-1)$.

Робочий файл (**Workfile**) зберігає в службовому векторі-змінній **C** коефіцієнти моделі, а в векторі **RESID** – знаходяться залишки або похибки моделі. Вектор **RESID** являє собою різницю між фактичним значенням часового ряду ***“y”*** та отриманим за моделлю: $RESID = \varepsilon = y - \hat{y}$ або $RESID = y - X \cdot \theta$, де X – матриця вимірів, а θ – коефіцієнти моделі.

Введемо декілька важливих змінних. Якщо в загальному виді рівняння процесу можна представити у вигляді $y = X \cdot \theta + e$ (де e – шумова складова), то далі **<u>під значенням T буде розумітися кількість строк, а під k кількість стовбців в матриці вимірів X.</u>**

<u>R-squared</u>

Коефіцієнт детермінації обчислюється за формулою $R^2 = \dfrac{\mathrm{var}(\hat{y})}{\mathrm{var}(y)}$, де $\mathrm{var}(\hat{y})$ – дисперсія залежної змінної, оціненої за допомогою побудованої математичної моделі, а $\mathrm{var}(y)-$ дисперсія залежної змінної, оціненої за допомогою її фактичних значень. Для гарної моделі $R^2 \to 1$.

Adjusted R-squared

Уточнений коефіцієнт детермінації обчислюється за формулою:

$$adjusted\,R^2 = 1-\left(1-R^2\right)\cdot\frac{T-1}{T-k-1}.$$

Цей коефіцієнт може приймати від'ємні значення але він завжди менший або дорівнює коефіцієнту детермінації.

S.E. of regression

S.E. of regression – Standard Error of the Regression.

СКП – середньоквадратична похибка або корінь із квадрата середньої похибки (в літературі RMSE – root mean square error) в загальному випадку обчислюється за формулою:

$$СКП(y,\hat{y}) = RMSE(y,\hat{y}) = \sqrt{MSE(y,\hat{y})} = \sqrt{E\left((y-\hat{y})^2\right)} = \sqrt{\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{N}}.$$

де $N-$ розмір вибірки, $y-$ фактичні значення, а $\hat{y}-$ значення побудовані за моделлю.

Для нашого випадку:

$$SE\,of\ regression = s = \sqrt{\frac{RESID^{Tp}\cdot RESID}{T-k}},$$

де $N-$ розмір навчальної вибірки, а $p-$ загальна кількість регресорів в моделі (при цьому константи не враховуються).

Sum square resid

Сума квадратів похибок моделі (SSE – sum of square error) в загальному випадку обчислюється за формулою

$$SSE = \sum_{t=1}^{N} \left(y - \hat{y} \right)^2 ,$$

де $N -$ розмір вибірки, $y -$ фактичні значення, а $\hat{y} -$ значення побудовані за моделлю.

Для нашого випадку: $SSE = RESID^{Tp} \cdot RESID$.

Необхідно відмітити, що окрім RMSE та SSE, ще існує середньоквадратична похибка (MSE – mean square error) яка обчислюється за формулою

$$MSE = E\left((y - \hat{y})^2 \right) = \frac{\sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2}{N} .$$

<u>Log likelihood</u>

Значення логарифмічної функції правдоподібності (log likelihood function) застосовується для перевірки гіпотези нормального розподілу похибок моделі (resid):

$$l = -\frac{T}{2} \cdot \left(1 + \log(2 \cdot \pi) + \log\left(\frac{RESID^{Tp} \cdot RESID}{T} \right) \right) .$$

Для кращої моделі $l \to \min$.

<u>Mean dependent var</u>

Mean of the Dependent Variable

Математичне сподівання $\hat{y}$.

<u>S.D. depend var</u>

Standard Deviation of the Dependent Variable

Середньоквадратичне відхилення $\hat{y}$ (корінь квадратний з дисперсії).

Akaike info criterion

Інформаційний критерій Акайке (AIC) обчислюється за формулою:

$$AIC = -\frac{2 \cdot l}{T} + \frac{2 \cdot k}{T}$$

де $l -$ значення логарифмічної функції правдоподібності. Для кращої моделі $AIC \to \min$.

Schwarz criterion

Критерій Шварца (SC), аналог критерію Акайке:

$$SC = -\frac{2 \cdot l}{T} + \frac{k \cdot \log(T)}{T}.$$

Для кращої моделі $SC \to \min$.

Необхідно зауважити, що окрім критеріїв Шварта та Акайте ще існують інші, серед досить популярних відмітимо байєсівський інформаційний критерій (BIC):

$$BIC = -2 \cdot \ln(l) + k \cdot \ln(T)$$

де $l -$ значення логарифмічної функції правдоподібності. Для кращої моделі $BIC \to \min$.

Durbin-Watson stat

Коефіцієнт Дурбина-Уотсона (DW – Durbin-Watson) показує адекватність побудованої моделі та обчислюється за формулою

$$DW = \frac{\sum_{t=2}^{N} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{N} \varepsilon_t^2},$$

де $\varepsilon -$ вектор залишків (різниця між значеннями отриманими за моделлю та фактичними), при цьому $DW \in [0;4]$. Для найкращої моделі $DW \to 2$, це означає що залишки моделі між собою не автокорелюють.

Std. Error

Оцінка стандартної похибки обчисленого коефіцієнта моделі. Вказує достовірність коефіцієнта – чим більше значення оцінки стандартної похибки тим більше флуктуаційних шумів в оцінці коефіцієнта.

Флуктуація – випадкове відхилення від середнього значення величини.

t-Statistic

Статистика Стьюдента, обчислюється як відношення коефіцієнта до стандартної похибки. Застосовується для перевірки гіпотези, щодо рівності коефіцієнта моделі нулю.

$$t - Statistic = \frac{Corfficient}{Std.Error}$$

Даний критерій був розроблений Вільямом Госсеттом для оцінки якості пива в компанії Гіннесс. У зв'язку із зобов'язаннями перед компанією про нерозголошення комерційної таємниці (застосування цього статистичного апарата для аналізу), стаття Госсетта вийшла в журналі "Біометрика" під псевдонімом "Student" (Студент).

Probability.

Вказує ймовірність виключення відповідного коефіцієнта з рівняння моделі. Фактично це значення ймовірності справедливості нуль гіпотези про рівність відповідного коефіцієнта рівняння моделі нулю.

4. Теоретичні відомості про кореляційні функції

4.1. Загальні відомості про кореляційні функції

Коефіцієнт кореляції, а в загальному випадку кореляційна функція, дозволяють встановити зв’язок між змінними. Кореляція може бути лінійною або нелінійною в залежності від типу залежності, яка фактично існує між змінними. Досить часто на практиці розглядають тільки лінійну кореляцію (взаємозв'язок), але більш глибокий аналіз потребує використання для дослідження процесів нелінійних залежностей. Складну нелінійну залежність можна спростити, але знати про її існування необхідно для того, щоб побудувати адекватну модель процесу.

Кореляційна матриця дозволяє встановити зв'язок **залежної** (**ендогенної**) змінної з **незалежними** (**екзогенними**). Розглянемо кореляційну матрицю вимірності 3×3, яка обчислюється для трьох змінних x, y, z:

$$R=\begin{bmatrix} r_{yy} & r_{xy} & r_{zy} \\ r_{yx} & r_{xx} & r_{zx} \\ r_{yz} & r_{xz} & r_{zz} \end{bmatrix}, \text{ де } r_{yx}=r_{xy},\ r_{yz}=r_{zy},\ r_{xz}=r_{zy}. \qquad (4.1)$$

Нехай $y-$ ендогенна змінна; $x, z-$ екзогенні змінні, які впливають на залежну змінну. Тобто, необхідно встановити існування залежності виду: $y=f(x,z)$, яка може бути представлена у формі регресії змінної y на незалежні змінні $x,\ z$:

$$y(k)=a_0+a_1x(k)+a_2z(k)+\varepsilon(k), \qquad (4.2)$$

де $k-$ дискретний час (наприклад, в секундах, хвилинах, годинах, днях, тижнях, місяцях і т.д.); $\varepsilon(k)-$ випадкова змінна, введення якої в модель пояснюється наступними причинами:

– часто буває неможливо визначити всі незалежні змінні, які впливають на залежну змінну, а тому рівняння (4.2) описує процес з похибкою;

– можуть існувати такі незалежні змінні, які неможливо виміряти і включити в математичну модель, а тому їх розглядають як збурення і вважають, що їх спільний вплив на залежну змінну описується випадковою змінною $\varepsilon(k)$;

– в регресійне рівняння (4.2) можуть бути включені пояснюючі (екзогенні) змінні, які фактично на y не впливають, але формально між ними і y існує ненульова кореляція;

– будь-який метод оцінювання коефіцієнтів рівняння регресії вносить свою методичну похибку, яка також повинна бути врахована в моделі.

Вважають, що сукупний вплив всіх вказаних факторів можна з деяким припущенням описати випадковою змінною $\varepsilon(k)$. Оскільки вона не вимірюється, то оцінити її значення (похибку моделі або остаток) можна тільки після оцінювання коефіцієнтів моделі, тобто

$$\hat{\varepsilon}(k) = e(k) = \hat{y}(k) - y(k),$$

де $\hat{y}(k) -$ оцінка змінної $y(k)$, отримана по моделі; $y(k) -$ фактичні вимірювання.

Для обчислення елементів матриці R необхідно мати синхронні в часі вибірки значень всіх трьох змінних y, x, z. Формула для обчислення коефіцієнтів кореляції має вигляд:

$$r_{yx} = \frac{1}{N-1} \frac{\sum_{k=1}^{N} \{[x(k) - x][y(k) - y]\}}{\sigma_x \sigma_y},$$

де $N -$ довжина вибірки даних; $x, y -$ середні вибіркові x, y; $\sigma_x, \sigma_y -$ стандартні відхилення, тобто корені квадратні з їх дисперсій. Наприклад,

$$\sigma_y = \sqrt{\sigma_y^2} = \left[\frac{1}{N-1} \sum_{k=1}^{N} [y(k) - y]^2 \right]^{1/2},$$

де $N -$ число вимірів змінної y; $y -$середнє значення ряду $\{y(k)\}$, яке обчислюється за формулою:

$$y = \mu_y = \frac{1}{N} \sum_{k=1}^{N} y(k).$$

Коефіцієнти кореляції показують на степінь взаємозв’язку між змінними. Очевидно, що перед формальним обчисленням коефіцієнтів кореляції необхідно виконати аналіз процесу та визначити присутність (чи відсутність) логічного зв'язку між змінними. Це дозволяє ввести в подальшому в розгляд тільки ті змінні, які дійсно впливають на залежну змінну. Очевидно, що для правильного вибору змінних необхідно достатньо глибоко аналізувати процес, який моделюється.

На основі значень коефіцієнтів кореляції приймається рішення про включення незалежних змінних (або регресорів) в рівняння регресії

$$y(k) = a_0 + b_1 x(k) + b_2 z(k) + \varepsilon(k),$$

яке може бути представлене в загальному вигляді як

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + a_3 x_3(k) + ... + a_{p-1} x_{p-1}(k) + \varepsilon(k). \quad (4.3)$$

Можна показати, що між коефіцієнтами регресії b_1, b_2 і коефіцієнтами кореляції r_{yx}, r_{yz} існує однозначний взаємозв'язок.

Рівняння (4.3) представляє собою **<u>лінійну регресію</u>** $p-$го порядку, але досить часто приходиться застосовувати більш складні нелінійні моделі. Характерним представником нелінійної відносно змінних регресії є поліноміальна регресія порядку $p-1$:

$$y(k) = a_0 + a_1 x \ (k) + a_2 x^2(k) + a_3 x^3(k) + ... + a_{p-1} x^{p-1}(k) + \varepsilon(k). \quad (4.4)$$

В рівняння (4.4) включена тільки одна незалежна змінна $x(k)$, але очевидно, що воно може бути розширене будь-якими іншими змінними.

Для визначення необхідності включення в рівняння регресії авторегресійної складової необхідно обчислити та дослідити автокореляційну функцію змінної $y(k)$. Рівняння з авторегресійною складовою має вигляд:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + b_1 x(k) + b_2 z(k), \quad (4.5)$$

тобто, в рівняння регресії введена авторегресійна (АР) складова другого порядку.

Порядок авторегресії визначається за допомогою автокореляційної функції. Число коефіцієнтів автокореляційної

функції, які відмінні від нуля в статистичному смислі, і буде порядком авторегресії.

Коефіцієнти вибіркової автокореляційної функції обчислюють за формулою:

$$r_y(s) = r_{y(k)y(k-s)} = \frac{1}{N-1} \frac{\sum_{k=s+1}^{N} \{[y(k) - y][y(k-s) - y]\}}{\sigma_y^2}, \qquad s = 1,2,3,... \quad (4.6)$$

де σ_y^2 – вибіркова дисперсія змінної $y(k)$. Число коефіцієнтів АКФ, відмінних від нуля в статистичному смислі, вказує на порядок авторегресійної частини моделі.

4.2. Теоретична автокореляційна функція

Якщо вибіркова АКФ обчислюється з використанням значень вимірів часового ряду, то теоретична АКФ обчислюється за допомогою рівняння, яке описує процес. Розглянемо для прикладу стохастичний процес АР(1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k).$$

Дисперсія цього процесу визначається так:

$$Var[y(k)] = Var[a_1 y(k-1)] + Var[\varepsilon(k)].$$

Оскільки для стаціонарного процесу (тобто процесу, статистичні характеристики якого не змінюються в часі) дисперсія залишається постійною, то

$$\mathrm{var}[y(k)] = \mathrm{var}[y(k-1)].$$

Таким чином, можемо записати

$$\mathrm{var}[y(k)]\,(1 - a_1^2) = \mathrm{var}[\varepsilon(k)],$$

або

$$\mathrm{var}[y(k)] = \frac{\sigma_\varepsilon^2}{1-a_1^2} = \gamma_0$$

при умові, що $|a_1| < 1$ (умова стаціонарності).

Знайдемо теоретичне значення коваріації для сусідніх значень залежної змінної, тобто для $y(k)$ та $y(k-1)$ (значення зміщення $s=1$). Оскільки $\mathrm{cov}[y(k)y(k-s)] = \mathrm{cov}[y(k)y(k+s)]$, то можна записати

$$y(k+1) = a_1 y(k) + \varepsilon(k) = a_1[a_1(k-1) + \varepsilon(k)] + \varepsilon(k+1) =$$

$$= a_1^2 y(k-1) + a_1\varepsilon(k) + \varepsilon(k+1).$$

Тепер можна знайти коефіцієнт коваріації при $s=1$:

$$\mathrm{cov}[y(k)y(k+1)] = E\{[a_1 y(k-1) + \varepsilon(k)][a_1^2 y(k-1) + a_1\varepsilon(k) + \varepsilon(k+1)]\} =$$

$$= a_1^3\,\mathrm{var}[y(k-1)] + a_1\,\mathrm{var}[\varepsilon(k)] =$$

$$= a_1^3 \frac{\sigma_\varepsilon^2}{1-a_1^2} + a_1\sigma_\varepsilon^2 = \frac{a_1\sigma_\varepsilon^2}{1-a_1^2}.$$

По аналогії для довільного значення s знайдемо:

$$E\{[y(k)-\mu][y(k-s)-\mu]\} = \frac{\sigma_\varepsilon^2 a_1^s}{1-a_1^2} = \gamma_s.$$

За визначенням, значення коефіцієнтів теоретичної автокреляційної функції знайдемо як відношення:

$$\rho_s = \frac{\gamma_s}{\gamma_0}.$$

Таким чином, $\rho_0 = 1,\ \rho_1 = a_1,\ \rho_2 = a_1^2,\ \ldots,\ \rho_s = a_1^s.$

4.3. Часткова автокореляційна функція

Уточнити порядок авторегресійної складової можна за допомогою часткової автокореляційної функції (ЧАКФ), яка обчислюється за виразами:

$$\Phi_{11} = r(1), \quad \Phi_{22} = \frac{r_2 - r_1^2}{1 - r_1^2};$$

$$\Phi_{ss} = \frac{r_s - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_{s-j}}{1 - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_j}. \quad (2.7)$$

ЧАКФ чіткіше відображає порядок АР-моделі завдяки відсутності впливу проміжних коефіцієнтів кореляції на вибрані значення змінної. Тобто, коефіцієнт Φ_{11} характеризує степінь взаємозв'язку між сусідніми (в часі) значеннями змінної, а Φ_{22} характеризує взаємозв'язок між значеннями змінної, які розділені в часі двома періодами дискретизації.

Значення коефіцієнтів вибіркової ЧАКФ можна наближено визначити за допомогою експериментальних даних наступним чином. Коефіцієнт a_{11} моделі $y(k) = a_{11} y(k-1)$ можна поставити у відповідність коефіцієнту ЧАКФ $a_{11} \approx \Phi_{11}$, а коефіцієнт a_{22} моделі $y(k) = a_{22} y(k-2)$ приблизно дорівнює коефіцієнту Φ_{22}. Коефіцієнти a_{11}, a_{22} оцінюють методом найменших квадратів.

Відмінність коефіцієнтів автокореляційної функції від нуля в статистичному смислі означає, що існує деякий вираз (формула), яка дозволяє підтвердити або спростувати цей факт. Одним із загальноприйнятих підходів до визначення цього факту є обчислення статистичного параметра (або просто статистики) Л'юнга-Бокса $Q(r_k)$, який визначається за формулою:

$$Q(r_k) = N(N+2) \sum_{k=1}^{s} r_k^2 / (N-k),$$

де $N-$ довжина вибірки даних змінної, для якої знайдено значення АКФ r_k; $s-$ число коефіцієнтів АКФ, які досліджуються на суттєву відмінність від нуля. Знайдене значення статистики порівнюють з табличним і таким чином приймається чи спростовується початкова гіпотеза.

5. Команда movav

Синтаксис:

@movav(x,n).

Команда *movav* призначена для формування простого ковзного середнього по попереднім n значенням часового ряду з ім'ям ***"x"***.

$$MA(i)=\frac{x(i)+x(i-1)+...+x(i-n+1)}{n}=\frac{\sum_{j=1}^{n}x(i-j+1)}{n}$$

Приклади використання:

Series MV=@movav(x,3)

$$MA(i)=\frac{x(i)+x(i-1)+x(i-2)}{3}.$$

В цьому випадку КС буде обчислюватися по попереднім 3 значенням часового ряду.

Series MV=@movav(x+1,3)

$$MA(i)=\frac{x(i+1)+x(i)+x(i-1)}{3}.$$

Зміщення на одиницю значень часового ряду в даному випадку дозволяю при обчисленні отримати середнє значення КС.

Series MV=@movav(x+2,3)

$$MA(i) = \frac{x(i+2) + x(i+1) + x(i)}{3}$$

В цьому випадку КС буде обчислюватися по майбутнім 3 значенням часового ряду.

Зауваження.

Існує безліч варіантів обчислення вагових коефіцієнтів КС. Але для обчислення вагових коефіцієнтів КС пропонується використовувати найбільш розповсюджену формулу:

$$w_i = (1-\alpha)^{n-i+1};$$

тобто $w_1 = (1-\alpha)^n$; $w_2 = (1-\alpha)^{n-1}$; …; $w_n = (1-\alpha)^2$; $w_n = (1-\alpha)$, де $\alpha = \frac{2}{n+1}$, а $n-$розмір вікна ковзного середнього. В даному випадку при аналізі економетричних даних більшу цінність має більш свіжа інформація щодо процесу який досліджується. На рис. 35 наведена гістограма вагових коефецієнтів.

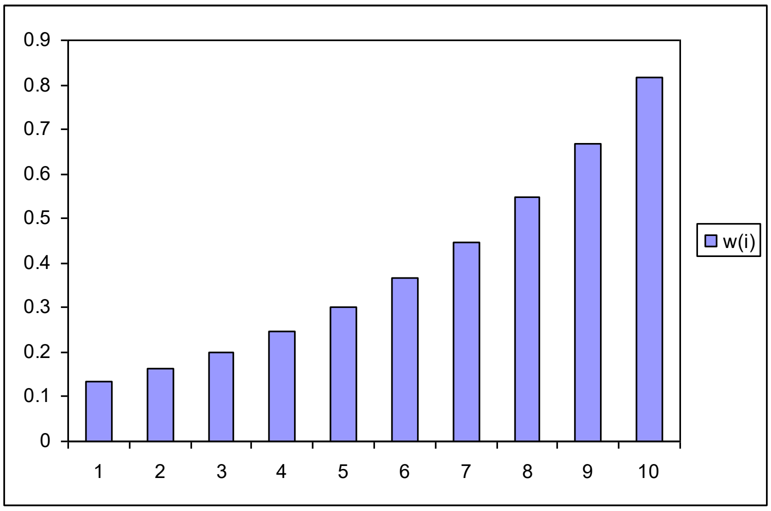


Рис. 35 Випадок коли ваги ЕКС зменшуються з віддаленістю від поточного періоду

Приклад функції на MatLab для обчислення вагових коефецієнтів ЕКС.

```
% size_n - размер окна скользящего среднего
% alpha - коеффициент
% w - вектор весовых коэффициентов
function ec;
size_n=10;
alpha=2/(size_n+1);
alpha,
for i=1:size_n
   w(i)=(1-alpha)^(size_n-i+1);
end;
w',
plot(w);
```

6. Послідовність побудови рівняння авторегресії із ковзним середнім

Необхідно чітко розуміти що АРКС може будуватися двома способами, які за своє глибинною суттю відрізняються. Так трапилося, що фактично дві різні речі називають однією назвою.

Підхід №1.

Побудова АРКС(p,q) коли КС будується по залишкам АР(p) рівняння моделі.

Крок 1. Оцінюються коефіцієнти рівняння моделі АР(p)

$$y(k) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i) + e \quad , \qquad \text{АР(p)} \qquad (6.1)$$

де $e-$залишки моделі (***resid***).

Крок 2. По залишкам АР(p) моделі формують КС.

Введемо нові змінні ***r1=resid*** (справа в тому що змінна ***resid*** ще може знадобиться) та ***mv*** (вектор для зберігання КС).

```
series r1
r1=resid
series mv
mv=@movav(r1)
```

Крок 3. Визначення порядку КС(q).

Для цього будують **ЧАКФ(mv)** та визначають q.

Крок 4. Оцінювання коефіцієнтів КС(q).

Фактично треба визначити коефіцієнти рівняння

$$resid(k) = mv(k) + \sum_{j=1}^{q} b_j \cdot mv(k-j), \qquad \text{КС(q)} \qquad (6.2)$$

Використовуючи рівняння 6.1 та 6.2 здійснюється побудова АРКС(p,q) по частинам, тобто спочатку коефіцієнти авто регресійної частини а після коефіцієнти при ковзних середніх. Можна одразу оцінити всі коефіцієнти $a_0,...,a_p,b_1,...,b_q$ рівняння (6.3). За теорією вважається що оцінювання коефіцієнтів окремо (спочатку авто регресійної частини, а потім КС) дає більш якісний результат.

$$y(k) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i) + mv(k) + \sum_{j=1}^{q} b_j \cdot mv(k-j), \qquad \text{АРКС(p,q)} \qquad (6.3)$$

<u>Підхід №2.</u>

<u>Побудова АРКС(p,q) коли КС будується по вихідному сигналу *y*</u>

Крок 1. Побудова КС по вихідному сигналу *y*.

Введемо нову ***mv*** (вектор для зберігання КС).

```
series mv
mv=@movav(y)
```

Крок 2. Визначення порядку КС(q).

Для цього будують **ЧАКФ(mv)** та визначають q.

Треба зауважити, що на практиці в даному випадку (коли КС будують по вихідному сигналу *y*) дуже часто $p = q$.

Крок 3. Оцінювання коефіцієнтів $a_0,...,a_p,b_1,...,b_q$ АРКС(p,q), дивись рівняння 6.3.

При цьому бажано здійснити проміжне перетворення, перенести *mv(k)* в ліву частину рівняння 6.3. Інакше це відобразиться на якості отриманих коефіцієнтів.

$$y1(k) = y(k) - mv(k) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i) + \sum_{j=1}^{q} b_j \cdot mv(k-j)$$

Обчисленні коефіцієнти за теорією повинні бути $\sum_{i=1}^{p} a_i < 1$, а $\sum_{j=1}^{q} b_j \to 1$. Але на практиці ці умови не завжди виконуються. Погано це чи ні в кожному окремому випадку можна визначити проаналізувавши результати прогнозування за допомогою отриманої моделі на декілька кроків вперед. Якщо результати прогнозування прийнятні, то на невиконання умов, щодо коефіцієнтів моделі, можна закрити очі, якби мовити – для досягнення мети всі моделі гарні, за умови що вони дають гарні результати.

Зауваження.

Виконавши крок 2 можна самостійно задати коефіцієнти $b_1,...,b_q$. В цьому випадку вони повинні задовольняти умові $\sum_{j=1}^{q} b_j = 1$. Наприклад при $q = 3$ можна встановити $b_1 = 0{,}5;\ b_2 = 0{,}375;\ b_3 = 0{,}125$. Після чого по відношенню до рівняння 6.3 необхідно здійснити проміжну операцію перетворення

$$y1(k) = y(k) - mv(k) - \sum_{j=1}^{q} b_j \cdot mv(k-j) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i)$$.

7. Побудова ARMAX моделі

Загальний вигляд ARMAX|(p,q,d) моделі

$$y(k) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i) + mv(k) + \sum_{j=1}^{q} b_j \cdot mv(k-j) + \sum_{s=1}^{d} c_s \cdot x_s, \qquad (6.4)$$

де $p-$порядок авто регресійної частини, $q-$порядок КС, $d-$кількість включених пояснюючих змінних.

Назва ARMAX походить від **A**uto**R**egressive **M**oving **A**verage model with e**X**ogenous inputs model. З англійської exogenous перекладається як зовнішній або екзогенний фактор.

В рівнянні 6.4 сума $\sum_{s=1}^{d} c_s \cdot x_s$ представляє собою лінійну комбінацію зовнішніх пояснюючих змінних $x_1,...,x_d$.

Рішення о включенні в ARMAX рівняння моделі відповідної пояснюючої змінної x_s приймається на основі аналізу сумісної кореляції вихідного сигналу y та x_s $correl(y, x_s) = r_{y,x_s}$ Якщо $r_{y,x_s} > 0{,}5$ то змінну x_s необхідно включати.

В загальному випадку до складу ARMAX рівняння окрім регрессорів $x_1,...,x_d$ також можуть включатися лагові змінні $x_s(k-m)$. Для того щоб включити відповідну авто регресійну частину регрессора x_s необхідно виконати аналіз ЧКФ(y, x_s). На рис. 36 наведений приклад можливої діаграми ЧКФ між y та x_s. Так як ЧКФ на 3, 4 та 5 лагах перевищують порогів рівень то в рівняння 6.4 необхідно включити наступні авто регресійні члени регрессора: $c_3 \cdot x_s(k-3) + c_4 \cdot x_s(k-4) + c_5 \cdot x_s(k-5)$. При цьому на практиці умова $\sum c_j < 1$ може не виконуватися. Найчастіше це пов'язано з різним масштабом значень y та x_s, у випадку коли значення y та x_s спочатку нормують (наприклад всі змінні процесу приводять до діапазону $[-1;+1]$ або $[0;1]$), то умова $\sum c_j < 1$ скоріш за все буде виконуватися.

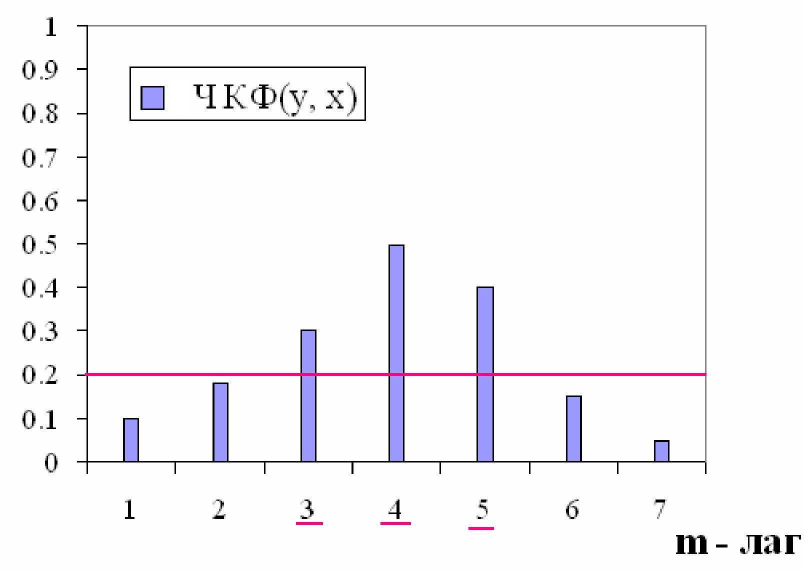


Рис. 36 Приклад ЧКФ(y, x_s).

8. Завдання на виконання ЛР№3

Частина перша: статистика Дурбина-Уотсона

Зробить програму в системі Eviews, яка повинна обчислювати статистику Дурбина-Уотсона. Формула для обчислення статистики наведена в розділі 3.3.

На вхід програми в якості параметра програми буде подаватися вектор часових даних. В папці *ATS_lab_03* знаходиться файл *example_for_DW.txt* з тестовими даними який ви можете використовувати для перевірки роботоспроможності вашої програми. Вірний результат роботи програми $DW = 1,2978$.

Частина друга: побудова адекватного рівняння для опису процесу.

Для кожного часового ряду з таблиці 1 лабораторної роботи №3 (та сама таблиця 1 з ЛР №2), згідно з номером бригади (1) виведіть описову статистику (*View–>Descriptive Statistics–>Histogram and Stats*) та (2) побудуйте рівняння АРКС яке найбільш адекватно описує відповідний набір даних (дивиться алгоритм підбора рівняння).

Алгоритм підбора рівняння.

Всі отримані в результаті аналізу рівняння (з чисельними коефіцієнтами, а не в загальному вигляді) необхідно включити до протоколу у вигляді таблиці 3.

1. Побудова АРКС(p,q) коли КС будується по залишкам АР(p) рівняння моделі.

1.1. Визначення $p-$ порядку авто регресійної складової.

1.1.1. Обчисліть АКФ та ЧАКФ для 12 лагів часового ряду

1.1.2. По отриманим значенням АКФ та ЧАКФ визначте $p-$ порядок авто регресійної складової.

1.2. Визначення $q-$порядку КС.

1.2.1. Обчисліть коефіцієнти моделі АР(p) : $y(k) = a_0 + \sum_{i=1}^{p} y(k-i)$.

Використовуючи команду LS. Отримані результати занесіть в таблицю 2..

1.2.2..Обчисліть коефіцієнти АКФ та ЧАКФ для 12 лагів залишків моделі АР(p).

1.2.3. По отриманим значенням АКФ та ЧАКФ визначте $q-$порядок КС.

Підхід №1. Застосування стандартного КС системи Eviews.

Використовуючи команду LS та відповідні специфікатори для авто регресійної частини та КС побудуйте АРКС(p,q).

Отримані результати занесіть в таблиці 2 та 3. Буде отримано чотири рівняння (1) для випадку коли просте КС з $N=5$; (2) просте КС з $N=10$; (3) експоненційне КС з $N=5$; (4) експоненційне КС з $N=10$.

Підхід №2. Застосування власного КС.

Побудуйте просте КС по залишкам АР(p) за допомогою команди @movav, з розміром вікна КС $N=5$ та $N=10$. Побудуйте АРКС(p,q) та занесіть в таблицю 2 отримані результати.

Побудуйте експоненційне КС по залишкам АР(p), для цього необхідно застосувати програму написану в попередній ЛР№2, з розміром вікна КС $N=5$ та $N=10$. Побудуйте АРКС(p,q) та занесіть в таблиці 2 та 3 отримані результати. Буде отримано чотири рівняння (1) для випадку коли просте КС з $N=5$; (2) просте КС з $N=10$; (3) експоненційне КС з $N=5$; (4) експоненційне КС з $N=10$.

<u>2. Побудова АРКС(p,q) коли КС будується по вихідному сигналу *y*</u>

2.1. Визначення $p-$ порядку авто регресійної складової.

2.1.1. Обчисліть АКФ та ЧАКФ для 12 лагів часового ряду

2.1.2. По отриманим значенням АКФ та ЧАКФ визначте $p-$ порядок авто регресійної складової.

2.2. Побудова КС по y.

Побудуйте просте КС по y за допомогою команди @movav, з розміром вікна КС $N=5$ та $N=10$.

Побудуйте експоненційне КС по y, для цього необхідно застосувати програму написану в попередній ЛР№2, з розміром вікна КС $N=5$ та $N=10$.

2.3. Визначення $q-$порядку КС.

Для цього побудуйте ЧАКФ(КС) та визначте q.

2.4. Побудуйте АРКС(p,q).

Підхід №1. Застосування власних коефіцієнтів при КС.

Обчислить коефіцієнти $b_1,...,b_q$ за формулою

$$b_i = \frac{(1-\alpha)^i}{\sum_{j=1}^{q}(1-\alpha)^j},$$

де $\alpha = \dfrac{2}{q+1}$.

Після чого обчислить коефіцієнти рівняння

$$y1(k) = y(k) - mv(k) - \sum_{j=1}^{q} b_j \cdot mv(k-j) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i),$$

де $mv-$КС. Отримані результати занесіть в таблиці 2 та 3. Буде отримано чотири рівняння (1) для випадку коли просте КС з $N=5$; (2) просте КС з $N=10$; (3) експоненційне КС з $N=5$; (4) експоненційне КС з $N=10$.

Підхід №2. Обчислення коефіцієнтів АРКС(p,q).

В даному випадку коефіцієнти $b_1,...,b_q$ вважаються невідомими і їх необхідно обчислити разом з коефіцієнтами $a_0, a_1,..., a_p$.

$$y(k) = a_0 + \sum_{i=1}^{p} a_i \cdot y(k-i) + mv(k) + \sum_{j=1}^{q} b_j \cdot mv(k-j)$$

.

Отримані результати занесіть в таблиці 2 та 3. Буде отримано чотири рівняння (1) для випадку коли просте КС з $N=5$; (2) просте КС з $N=10$; (3) експоненційне КС з $N=5$; (4) експоненційне КС з $N=10$.

Наприклад для випадку АРКС(3,3)

ls y=c(1)+c(2)*y(-1)+c(3)*y(-2)+c(4)*y(-3)+mv+c(5)*mv(-1)+c(6)*mv(-2)+c(7)*mv(-3)

або

series y1=y-mv

ls y1 y(-1) y(-2) y(-3) mv(-1) mv(-2) mv(-3)

3. Проаналізуйте отримані результати. Яка з отриманих моделей найбільш адекватна? Для цього серед отриманих моделей АРКС по статистичним характеристикам наведеним в таблиці 2 оберіть найкраще рівняння яке найбільш адекватно описує процес. Обґрунтуйте відповідь.

Таблиця 2

	R - squared	Sum squared resid	Akaike	Durbin – Watson stat
АРКС(p,q) де КС побудоване по залишкам АР(p) рівняння				
АРКС(p)				
АРКС(p,q) із застосуванням КС системи Eviews				
АРКС(p,q) із застосуванням власного простого КС, при N=5.				
АРКС(p,q) із застосуванням власного простого КС, при N=10.				
АРКС(p,q) із застосуванням власного експоненційного КС, при N=5.				
АРКС(p,q) із застосуванням власного експоненційного КС, при N=10.				
Побудова АРКС(p,q) де КС побудоване по вихідному сигналу у				
Підхід №1. Застосування власних коефіцієнтів при КС.				

АРКС(p,q) із застосуванням власного простого КС по y, при N=5.				
АРКС(p,q) із застосуванням власного простого КС по y, при N=10.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=5.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=10.				
Підхід №2. Обчислення коефіцієнтів АРКС(p,q).				
АРКС(p,q) із застосуванням власного простого КС по y, при N=5.				
АРКС(p,q) із застосуванням власного простого КС по y, при N=10.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=5.				
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=10.				

Таблиця 3

Рівняння з чисельними коефіцієнтами	
АРКС(p,q) де КС побудоване по залишкам АР(p) рівняння	
АРКС(p)	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}4833 \cdot y(k-2)$
АРКС(p,q) із застосуванням КС системи Eviews	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) +$ $+ ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного простого КС, при N=5.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) +$ $+ ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного простого КС, при N=10.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) +$ $+ ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного експоненційного КС, при N=5.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) +$ $+ ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного експоненційного КС, при N=10.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) +$ $+ ma(k) - 0{,}71 \cdot ma(k-1)$
Побудова АРКС(p,q) де КС побудоване по вихідному сигналу y	
Підхід №1. Застосування власних коефіцієнтів при КС.	
АРКС(p,q) із застосуванням власного простого КС по y, при N=5.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) +$ $+ ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного простого КС по y, при N=10.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) +$ $+ ma(k) - 0{,}71 \cdot ma(k-1)$

АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=5.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) + ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=10.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) + ma(k) - 0{,}71 \cdot ma(k-1)$
Підхід №2. Обчислення коефіцієнтів АРКС(p,q).	
АРКС(p,q) із застосуванням власного простого КС по y, при N=5.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) + ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного простого КС по y, при N=10.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) + ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=5.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) + ma(k) - 0{,}71 \cdot ma(k-1)$
АРКС(p,q) із застосуванням власного експоненційного КС по y, при N=10.	$y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}567 \cdot y(k-2) + ma(k) - 0{,}71 \cdot ma(k-1)$

Частина третя: побудова рівняння множинної регресії.

Побудуйте рівняння множинної регресії для індексу rts1, в якості регресорів необхідно розглянути галузеві індекси RTScr, RTSeu, RTSfn, RTSin, RTSmm, RTSog та RTStl.

В загальному вигляді рівняння множинної регресії має вигляд:

$$\mathrm{rts1} = a_0 + a_1 \cdot \mathrm{RTScr} + \mathrm{a}_2 \cdot \mathrm{RTSeu} + \mathrm{a}_3 \cdot \mathrm{RTSfn} + \mathrm{a}_4 \cdot \mathrm{RTSin} + \\ + \mathrm{a}_5 \cdot \mathrm{RTSmm} + \mathrm{a}_6 \cdot \mathrm{RTSog} + a_7 \cdot \mathrm{RTStl}$$

Але спочатку необхідно побудувати кореляційну матрицю індексів і на основі аналізу її значень включити в рівняння моделі саме ті індекси які дійсно впливають на процес.

Частина четверта: <u>творча, не обов’язкова</u>.

4.1. Написати програму для обчислення ЧКФ часткової кореляційної функції.

4.2. Побудувати ARMAX модель для індексу rts1, в якості регресорів необхідно розглянути галузеві індекси RTScr, RTSeu, RTSfn, RTSin, RTSmm, RTSog та RTStl, з включенням авто регресійних складових пояснюючих змінних регресорів.

Пункт 4.2 неможливо виконати без пункту 4.1.

Таблиця 1

№ бригади	Ім’я файлу	Рекомендоване ім’я ряду при створенні програми	Розмір вибірки даних (параметр **End date**)
1	rts1.txt	rts1	178
	1996rts1.txt	rts1996	249
2	rts2.txt	rts2	178
	1997rts1.txt	rts1997	251
3	RTScr.txt	RTScr	178
	1998rts1.txt	rts1998	251
4	RTSeu.txt	RTSeu	178
	1999rts1.txt	rts1999	251
5	RTSfn.txt	RTSfn	178
	2000rts1.txt	rts2000	250
6	RTSin.txt		178
	2001rts1.txt	rts2001	251
7	RTSmm.txt	RTSmm	178
	2002rts1.txt	rts2002	250
8	RTSog.txt	RTSog	178
	2003rts1.txt	rts2003	250
9	RTStl.txt	RTStl	178
	2004rts1.txt	rts2004	251
10	rts1.txt	rts1	178
	2005rts1.txt	rts2005	248
11	rts2.txt	rts2	178
	2006rts1.txt	rts2006	248
12	RTSfn.txt	RTSfn	178
	2007rts1.txt	rts2007	248
13	RTSmm.txt	RTSmm	178

	2007rts1.txt	rts2007	248

Всі файли знаходяться в папці …\ATS_lab3_new\Data

Розмір вибірки даних – це кількість даних в файлі. Це значення використовується при створенні робочого файлу програми, як параметр **End date**.

ДОДАТОК А:

Приклад побудови рівняння АРКС по економетричним даним часового ряду

Крок 1. Створіть робочий файл програми **Eviews** (**WorkFIle**) з нерегулярним типом даних (**Undated or irregular**) зі 100 значень (**Start date: 1** та **End date: 100**) та завантажте файл **File–>Import–>Read Text-Lotus-Excel** з тестовими даними ***example_data.txt*** (файл знаходиться в папці *ATS_lab_03*). Задайте ім’я ряду ***“y”***.

Крок 2. Обчислення коефіцієнтів АКФ та ЧАКФ ряду ***“y”***.

В робочому файлі (**Workfile**) подвійним кликом лівої кнопки миші відкрийте вікно ряду ***“y”***. У вікні, що з’явилося оберіть в меню закладку **View**, а в ній пункт **Correlogram**, як показано на рис. А.1 **(View–>Correlogram)**. Після цього з’явиться вікно, рис. А.2.

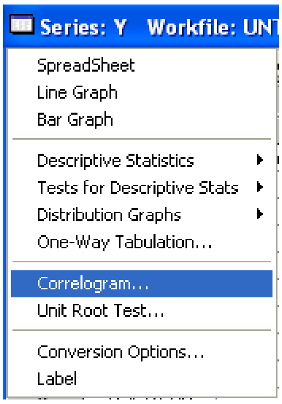


Рис. А.1

Опис опцій вікна **Correlogram Speciffication**, рис. А.2:

Level – побудова АКФ та ЧАКФ для ряду даних без змін.

1-st difference – побудова АКФ та ЧАКФ для ряду даних перших різниць. Від кожного *к*-го значення ряду буде відняте попереднє *к–1* значення, тобто спочатку буде виконана операція ***y(k)=y(k)-y(k-1)***.

2-nd difference – побудова АКФ та ЧАКФ для ряду даних перших різниць. Від кожного *к*-го значення ряду буде відняте попереднє *к–2* значення, тобто спочатку буде виконана операція ***y(k)=y(k)-y(k-2)***.

Lags to include – кількість значень ряду, для яких буде виведений результат.

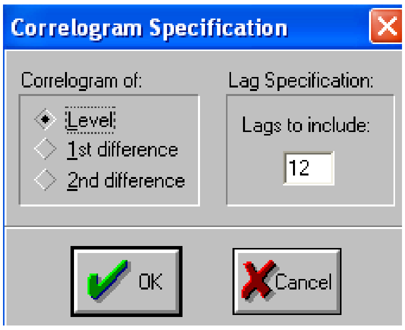
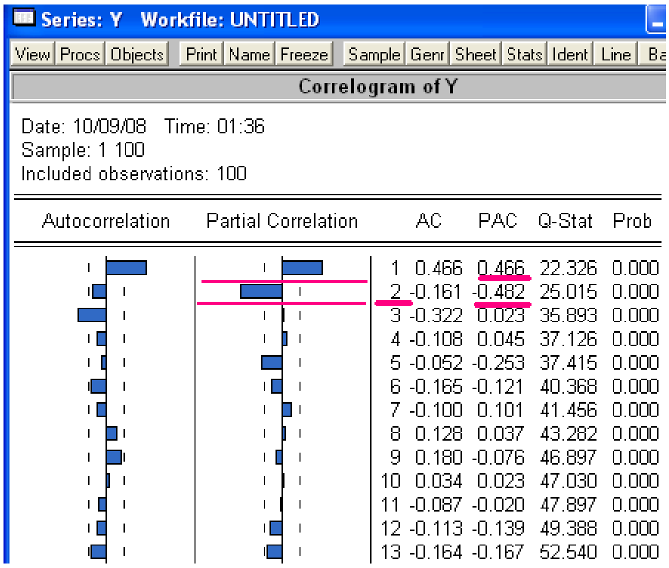


Рис. А.2



Series: Y Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Sample | Genr | Sheet | Stats | Ident | Line | Ba

Correlogram of Y

Date: 10/09/08 Time: 01:36
Sample: 1 100
Included observations: 100

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.466	0.466	22.326	0.000
		2	-0.161	-0.482	25.015	0.000
		3	-0.322	0.023	35.893	0.000
		4	-0.108	0.045	37.126	0.000
		5	-0.052	-0.253	37.415	0.000
		6	-0.165	-0.121	40.368	0.000
		7	-0.100	0.101	41.456	0.000
		8	0.128	0.037	43.282	0.000
		9	0.180	-0.076	46.897	0.000
		10	0.034	0.023	47.030	0.000
		11	-0.087	-0.020	47.897	0.000
		12	-0.113	-0.139	49.388	0.000
		13	-0.164	-0.167	52.540	0.000

Рис. А.3

На рис. А.3 наведена таблиця отриманих значень АКФ та ЧАКФ для ряду ***"у"***. Аналізуючи значення ЧАКФ видно що значимими являються 1 та 2 лаги, тому що $PAC(1) \geq 0{,}3$ ($PAC(1) = 0{,}466$) і $|PAC(2)| \geq 0{,}3$ ($PAC(2) = -0{,}482$). Знак чисельного значення ЧАКФ носить той же сенс що і звичайне значення кореляції двох змінних. Значення 0,3 було взяте в якості граничного (порогового значення), тобто якщо $|PAC(i)| \geq 0{,}3$ то $i-$й лаг являється значимим і його необхідно враховувати при побудові рівняння моделі. Відмітимо, що граничне значення не обов'язково повинно бути рівним 0,3, це суб'єктивне рішення аналітика, наприклад в MatLab граничне значення дорівнює 0,2. Але в цьому прикладі було взяте значення саме 0,3 щоб спростити аналіз, в противному випадку необхідно було б включити лаги включно до п'ятого.

Крок 3. Побудова АР(1) процесу.

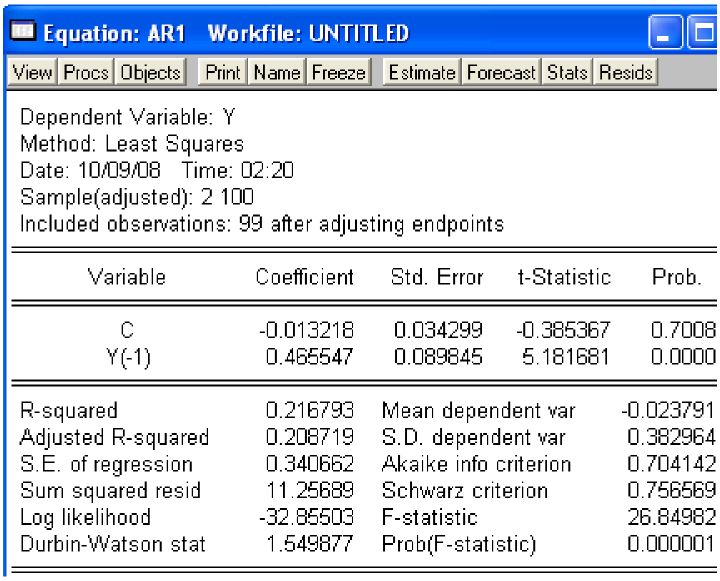
На попередньому кроці 2 аналізуючи ЧАКФ ряду було визначено що необхідно розглядати 1 та 2 лаги при побудові моделі, тобто порядок авто регресійного рівняння $p = 2$.

Але спочатку побудуємо АР(1).

Для обчислення коефіцієнтів АР(1) $y(k) = a_0 + a_1 \cdot y(k-1)$ виконайте в командній строчці Eviews команду

equation ar1 y c y(-1)

що створить об'єкт "equation" з ім'ям "***ar1***". Подвійним кліком лівої кнопки миші по цьому об'єкту "***ar1***" з'явиться вікно результатів оцінювання, рис. А.4.


Equation: AR1 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/09/08 Time: 02:20
Sample(adjusted): 2 100
Included observations: 99 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.013218	0.034299	-0.385367	0.7008
Y(-1)	0.465547	0.089845	5.181681	0.0000

R-squared	0.216793	Mean dependent var	-0.023791
Adjusted R-squared	0.208719	S.D. dependent var	0.382964
S.E. of regression	0.340662	Akaike info criterion	0.704142
Sum squared resid	11.25689	Schwarz criterion	0.756569
Log likelihood	-32.85503	F-statistic	26.84982
Durbin-Watson stat	1.549877	Prob(F-statistic)	0.000001

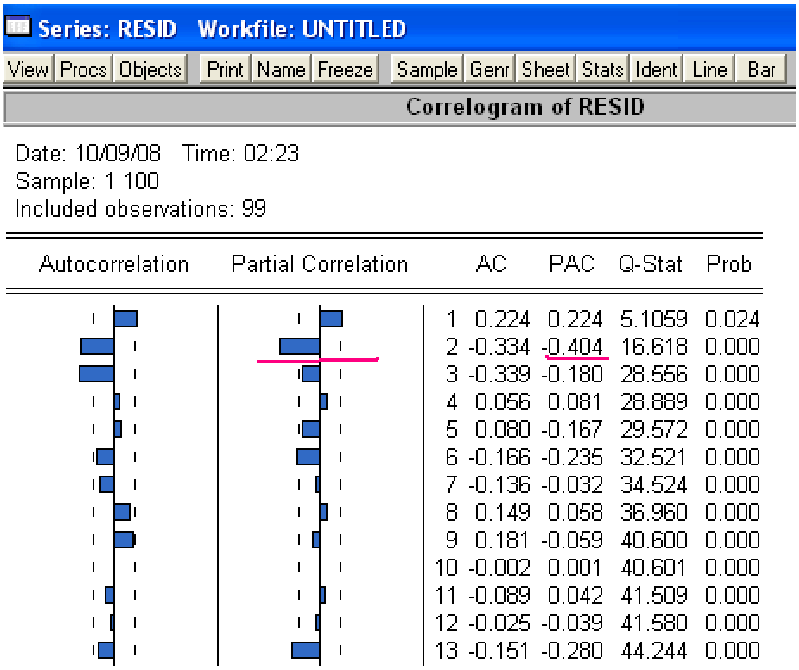
Рис. А.4

Тобто АР(1) процесу описується наступним рівнянням $y(k) = -0{,}013 + 0{,}465 \cdot y(k-1)$. Відмітимо, що $PAC(1) \approx a_1$ і це не збіжність обставин, так повинно бути за теорією.

Крок 4. Аналіз остатків моделі АР(1).

При побудові АР(1) процесу в системі Eviews автоматично були обчисленні остатки моделі, які зберігаються в векторі-змінній ***resid***.

Побудуємо АКФ та ЧАКФ остатків моделі АР(1), рис. А.5. Видно що остатки між собою корелюють $PAC(2) > 0{,}3$. Це означає що з часового ряду ***"y"*** при побудові АР(1) ми не задіяли всю інформацію, що в ньому міститься.


Series: RESID Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Sample | Genr | Sheet | Stats | Ident | Line | Bar

Correlogram of RESID

Date: 10/09/08 Time: 02:23
Sample: 1 100
Included observations: 99

	AC	PAC	Q-Stat	Prob
1	0.224	0.224	5.1059	0.024
2	-0.334	-0.404	16.618	0.000
3	-0.339	-0.180	28.556	0.000
4	0.056	0.081	28.889	0.000
5	0.080	-0.167	29.572	0.000
6	-0.166	-0.235	32.521	0.000
7	-0.136	-0.032	34.524	0.000
8	0.149	0.058	36.960	0.000
9	0.181	-0.059	40.600	0.000
10	-0.002	0.001	40.601	0.000
11	-0.089	0.042	41.509	0.000
12	-0.025	-0.039	41.580	0.000
13	-0.151	-0.280	44.244	0.000

Рис. А.5.

Крок 5. Побудова АР(2) процесу.

Для обчислення коефіцієнтів АР(2) $y(k) = a_0 + a_1 \cdot y(k-1) + a_2 \cdot y(k-1)$ виконайте в командній строчці Eviews команду

equation ar2 y c y(-1) y(-2)

що створить об’єкт “equation” з ім’ям ***“ar2”***. Подвійним кліком лівої кнопки миші по цьому об’єкту ***“ar2”*** з’явиться вікно результатів оцінювання, рис. А.6.

АР(2) : $y(k) = -0{,}0211 + 0{,}6901 \cdot y(k-1) - 0{,}4833 \cdot y(k-2)$.

Equation: AR2 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/09/08 Time: 02:04
Sample(adjusted): 3 100
Included observations: 98 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.021125	0.030504	-0.692555	0.4903
Y(-1)	0.690188	0.089816	7.684487	0.0000
Y(-2)	-0.483355	0.089892	-5.377076	0.0000

R-squared	0.399329	Mean dependent var	-0.025483
Adjusted R-squared	0.386683	S.D. dependent var	0.384561
S.E. of regression	0.301167	Akaike info criterion	0.467832
Sum squared resid	8.616667	Schwarz criterion	0.546964
Log likelihood	-19.92379	F-statistic	31.57818
Durbin-Watson stat	1.974741	Prob(F-statistic)	0.000000

Рис. А.6.

<u>Побудова АРСК(p,q) коли КС будується по залишкам АР(p) рівняння моделі</u>.

Крок 6. Аналіз остатків моделі АР(2).

При побудові АР(2) процесу в системі Eviews автоматично були обчисленні остатки моделі, які зберігаються в векторі-змінній ***resid***.

Побудуємо АКФ та ЧАКФ остатків моделі АР(2), рис. А.7. Видно що при граничному значенні 0,3 остатки між собою не корелюють.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.010	0.010	0.0102	0.919
		2	0.027	0.027	0.0860	0.958
		3	-0.127	-0.127	1.7445	0.627
		4	0.067	0.070	2.2124	0.697
		5	-0.017	-0.012	2.2417	0.815
		6	-0.194	-0.218	6.2631	0.394
		7	-0.127	-0.107	7.9955	0.333
		8	0.112	0.129	9.3548	0.313
		9	0.077	0.032	10.011	0.350
		10	-0.050	-0.077	10.294	0.415
		11	-0.097	-0.064	11.349	0.414
		12	0.087	0.062	12.209	0.429
		13	-0.146	-0.231	14.654	0.329

Рис. А.7.

Але якщо зменшити граничне значення до 0,2 то кореляція буде присутня. Це припущення робиться для того, щоб на наступних кроках показати послідовність побудови АРКС.

Крок 7. Побудова АРКС(2,1).

Варіант 1.

На кроку 5 при побудова АР(2) процесу в системі Eviews автоматично були обчисленні остатки моделі, які зберігаються в векторі-змінній ***resid***.

Створить новий часовий ряд з ім'ям ***"ar2_resid"*** і збережіть в ньому отримані остатки ***resid***.

```
series ar2_resid=resid
```

Згенеруйте ряд простого КС з ім'ям ***"mv"***, з розміром вікна п'ять, для остатків АР(2) моделі.

```
series mv=@movav(ar2_resid,5)
```

Після чого створить новий вектор з ім'ям ***"y1"*** в який запишіть значення часового ряду ***"y".***

```
series y1=y
```

З вектора ***"y1"*** відніміть вектор ***"mv".*** Сенс даної операції полягає в тому що АРКС(2,1) записується як

$$y(k) = a_0 + a_1 \cdot y(k-1) + a_2 \cdot y(k-1) + mv(k) + b_1 \cdot mv(k-1),$$

тобто апріорі відомо що при $ma(k)$ коефіцієнт одиниця. Для того щоб команда **ls** обчислила саме коефіцієнти АРКС(2,1) (без повторного обчислення коефіцієнта при $ma(k)$) виконується видалення $ma(k)$ з $y(k)$

$$y(k) - mv(k) = a_0 + a_1 \cdot y(k-1) + a_2 \cdot y(k-1) + b_1 \cdot mv(k-1).$$

```
y1=y1-mv
```

Після чого обчислить коефіцієнта АР(2,1)

equation arma_2_1.ls y1 c y(-1) y(-2) mv(-1)

На рис. А.8 наведені результати оцінювання.

АРКС(2,1) :

$$y(k) = -0{,}018 + 0{,}667 \cdot y(k-1) - 0{,}567 \cdot y(k-2) + mv(k) - 0{,}71 \cdot mv(k-1)$$

Equation: ARMA_2_1 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y1
Method: Least Squares
Date: 10/09/08 Time: 03:14
Sample(adjusted): 8 100
Included observations: 93 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.018153	0.025740	-0.705226	0.4825
Y(-1)	0.667454	0.076645	8.708354	0.0000
Y(-2)	-0.567041	0.080036	-7.084842	0.0000
MV(-1)	-0.708681	0.219589	-3.227311	0.0017

R-squared	0.548451	Mean dependent var	-0.022463
Adjusted R-squared	0.533230	S.D. dependent var	0.362307
S.E. of regression	0.247530	Akaike info criterion	0.087492
Sum squared resid	5.453145	Schwarz criterion	0.196421
Log likelihood	-0.068359	F-statistic	36.03304
Durbin-Watson stat	2.057499	Prob(F-statistic)	0.000000

Рис. А.8.

<u>Варіант 2.</u>

В системі Eviews є можливість об’являти специфікатори в команді **ls** для включення авторегресійної складової (**ar**) та ковзного середнього (**ma**).

equation arma_2_1_b.ls y c ar(1) ar(2) ma(1)

Результати виконання дивиться на рис. А.9.

Результати наведені на рис. А.8 та рис. А.9 відрізняються, хоча оцінювалися коефіцієнти для рівняння АРКС(2,1), тому що у другому випадку (при виконанні команди **equation arma_2_1_b.ls y c ar(1) ar(2) ma(1)**) КС обчислюється із використанням формули реалізованої в системі Eviews, яка відрізняється від формули при обчисленні ***mv***.

Equation: ARMA_2_1_B Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/09/08 Time: 03:20
Sample(adjusted): 3 100
Included observations: 98 after adjusting endpoints
Convergence achieved after 6 iterations
Backcast: 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.026556	0.039167	-0.678008	0.4994
AR(1)	0.664291	0.190244	3.491775	0.0007
AR(2)	-0.471170	0.121497	-3.878048	0.0002
MA(1)	0.033675	0.216088	0.155838	0.8765

R-squared	0.399535	Mean dependent var	-0.025483
Adjusted R-squared	0.380371	S.D. dependent var	0.384561
S.E. of regression	0.302713	Akaike info criterion	0.487898
Sum squared resid	8.613713	Schwarz criterion	0.593407
Log likelihood	-19.90699	F-statistic	20.84841
Durbin-Watson stat	1.992145	Prob(F-statistic)	0.000000

Рис. А.9.

Після того як обчислені коефіцієнти та характеристики рівняння АРКС(2,1) обчислюють коефіцієнти АРКС(2,2) і так далі до АРКС(2,q). Порядок КС $q-$визначається на основі аналізу ЧАКФ залишків рівняння АР(2) моделі наведених на рис. А.7 для даного прикладу достатньо буде взяти $q=6$.

<u>Побудова АРКС(p,q) коли КС будується по вихідному сигналу *y*</u>

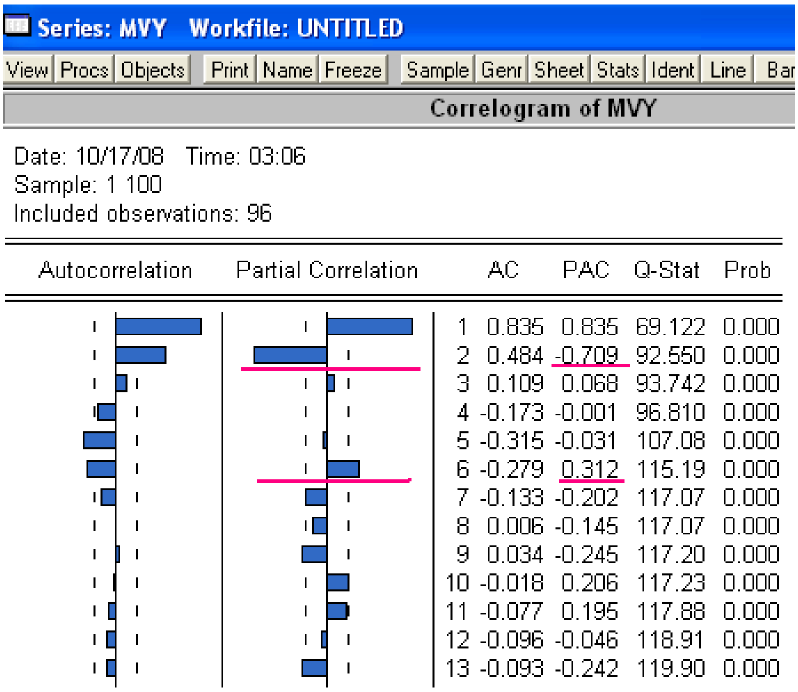
Крок 6. Побудова КС по вихідному сигналу *y*.

Введемо новий вектор ***mvy*** для зберігання КС.

```
series mvy=@movav(y,5)
```

Крок 7. Визначення порядку КС(q).

Побудуємо для цього **ЧАКФ(*mvy*)**, дивиться рис. А.10. В даному випадку можна розглядати 6-й порядок КС $q=6$ (за бажанням і 13-й, $q=13$) але щонайменш $q=2$.



Series: MVY Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Sample | Genr | Sheet | Stats | Ident | Line | Bar

Correlogram of MVY

Date: 10/17/08 Time: 03:06
Sample: 1 100
Included observations: 96

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.835	0.835	69.122	0.000
		2	0.484	-0.709	92.550	0.000
		3	0.109	0.068	93.742	0.000
		4	-0.173	-0.001	96.810	0.000
		5	-0.315	-0.031	107.08	0.000
		6	-0.279	0.312	115.19	0.000
		7	-0.133	-0.202	117.07	0.000
		8	0.006	-0.145	117.07	0.000
		9	0.034	-0.245	117.20	0.000
		10	-0.018	0.206	117.23	0.000
		11	-0.077	0.195	117.88	0.000
		12	-0.096	-0.046	118.91	0.000
		13	-0.093	-0.242	119.90	0.000

Рис. А.10

Крок 8. Побудова АРКС(2,1).

series y1=y-mvy

equation armay_2_1.ls y1 ar(1) ar(2) mvy(-1)

Equation: ARMAY_2_1 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y1
Method: Least Squares
Date: 10/17/08 Time: 03:08
Sample(adjusted): 8 100
Included observations: 93 after adjusting endpoints
Convergence achieved after 5 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MVY(-1)	-0.782354	0.175880	-4.448224	0.0000
AR(1)	0.624671	0.088749	7.038602	0.0000
AR(2)	-0.546118	0.091614	-5.961096	0.0000

R-squared	0.560639	Mean dependent var	0.003370
Adjusted R-squared	0.550876	S.D. dependent var	0.367681
S.E. of regression	0.246408	Akaike info criterion	0.068067
Sum squared resid	5.464503	Schwarz criterion	0.149764
Log likelihood	-0.165110	Durbin-Watson stat	1.954237

Рис. А.11

На рис. А.11 наведені коефіцієнти та характеристики рівняння АРКС(2,1).

Після того як обчислені коефіцієнти та характеристики рівняння АРКС(2,1) обчислюють коефіцієнти АРКС(2,2) і так далі до АРКС(2,q).

ДОДАТОК Б

Приклад побудови моделі множинної регресії

Для побудови моделі множинної регресії скористайтесь даними, наведеними в таблиці Б.1. Ці дані представлені у вигляді файлів *example_y.txt* (валовий продукт в млн. у.о.), *example_x1.txt* (об’єм робочої сили на 1000 осіб населення) та *example_x2.txt* (капіталовкладення в сільськогосподарському секторі, у.о.) що знаходиться в папці ATS_lab_03 на диску

Таблиця Б.1.

Дані для задачі побудови множинної регресії

Рік	Валовий продукт в млн. у.о.	Об’ем робочої сили на 1000 осіб населення	Капіталовкладення в сільськогосподарському секторі, у.о.
1980	9127.5	265.6	121753400
1981	11795.7	268.2	123251300
1982	12655.3	274.7	124998500
1983	13698.4	284.9	125670600
1984	14900.2	295.5	127460900
1985	16357.3	345.7	131740200
1986	19489.6	390.3	134530400
1987	22591.7	467.4	139880300
1988	24139.6	558.9	147158500
1989	27067.9	610.6	154890400
1990	31201.2	660.5	165743200
1991	35373.5	712.8	175494700
1992	39890.6	750.3	187665300
1993	44193.7	790.1	203472200
1994	49839.3	830.8	222565700
1995	55385.9	850.4	241337300

1996	61217.8	880.3	263955600
1997	66976.1	899.6	287554300
1998	72837.4	915.8	330498100
1999	79028.2	925.3	359684200
2000	85479.3	931.7	385674800

Крок 1. Створить робочий файл програми з щорічним типом даних (**Annual**) зі **Start date: 1980** та **End date: 2000**).

Крок 3.

Завантажте файл **File–>Import–>Read Text-Lotus-Excel** з даними ***example_y.txt*** (файл знаходиться в папці *ATS_lab_03*). Задайте ім’я ряду ***“y”***.

Завантажте файл ***example_x1.txt.***. Задайте ім’я ряду ***“x1”***.

Завантажте файл ***example_x2.txt.***. Задайте ім’я ряду ***“x2”***.

Крок 4.

Виділіть ***“y”, “x1”*** та ***“x2”*** в робочому файлі (утримуючи **shift**), натисніть праву кнопку миші, виберіть опцію **Open–>as Group**, дивися рис. Б.1.

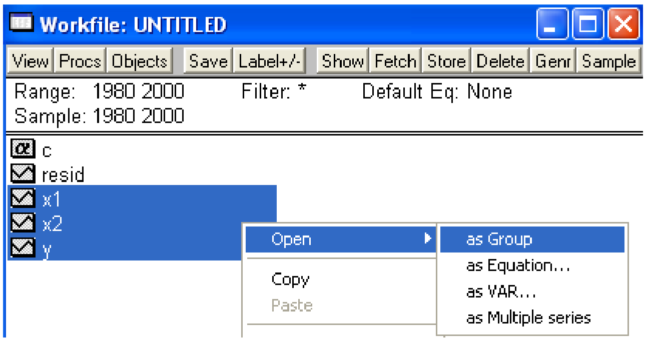
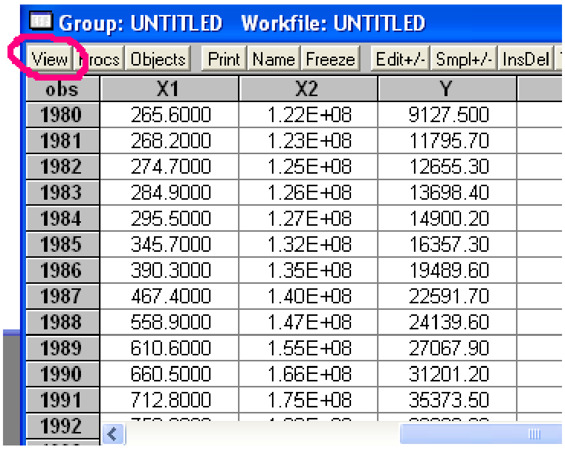


Рис. Б.1.

Крок 5.

У вікні **Group** що з’явилося натисніть кнопку **View**.



obs	X1	X2	Y
1980	265.6000	1.22E+08	9127.500
1981	268.2000	1.23E+08	11795.70
1982	274.7000	1.25E+08	12655.30
1983	284.9000	1.26E+08	13698.40
1984	295.5000	1.27E+08	14900.20
1985	345.7000	1.32E+08	16357.30
1986	390.3000	1.35E+08	19489.60
1987	467.4000	1.40E+08	22591.70
1988	558.9000	1.47E+08	24139.60
1989	610.6000	1.55E+08	27067.90
1990	660.5000	1.66E+08	31201.20
1991	712.8000	1.75E+08	35373.50
1992	[illegible]	[illegible]	[illegible]

Рис. Б.2.

Крок 6.

У вікні, що з’явилося виберіть опцію **Correlation**.

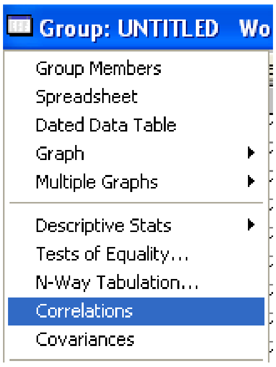


Рис. Б.3.

В результаті цієї операції з’явиться вікно **Correlation Matrix**, що містить кореляційну матрицю вимірності для трьох змінних ***“y”, “x1”*** та ***“x2”.*** Дивись рис. Б.4.

Correlation Matrix			
	X1	X2	Y
X1	1.000000	0.868769	0.936462
X2	0.868769	1.000000	0.985210
Y	0.936462	0.985210	1.000000

Рис. Б.4.

Аналізуючи значення кореляційної матриці наведеної на рис. Б.4 експерт приймає рішення щодо включення ***“x1”*** та ***“x2”*** в рівняння множинної регресії в якості пояснюючих змінних (регресорів). Якщо спостерігається сильний кореляційний зв'язок, як в нашому випадку де $r_{y,x_1} = 0{,}936$ та $r_{y,x_2} = 0{,}985$, то відповідні регресори необхідно включити в рівняння моделі. В таблиці Б.2 наведена класифікація видів кореляційного зв'язку.

Таблиця Б.2.

Класифікація видів кореляційного зв'язку

Кореляція	**Позитивна**	**Від'ємна**
слабка	[0,1 ; 0,3)	(-0,3 ; -0,1]
середня	[0,3 ; 0,5)	(-0,5 ; -0,3]
сильна	[0,5 ; 1]	[-1 ; -0,1]

Крок 7. Побудова рівняння множинної регресії.

<u>Підхід №1.</u>

Побудова рівняння по вхідним даним.

$$y(k) = c_0 + c_1 x_1(k) + c_2 x_2(k) + \varepsilon(k) \qquad \text{EQ1}$$

equation eq1.ls y=c(1)+c(2)*x1+c(3)*x2

Equation: EQ1 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/18/08 Time: 02:19
Sample: 1980 2000
Included observations: 21
Y=C(1)+C(2)*X1+C(3)*X2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-21509.38	816.5062	-26.34320	0.0000
C(2)	30.95857	2.419838	12.79365	0.0000
C(3)	0.000203	7.46E-06	27.26404	0.0000

R-squared	0.997091	Mean dependent var	37773.63
Adjusted R-squared	0.996768	S.D. dependent var	24200.44
S.E. of regression	1375.856	Akaike info criterion	17.42310
Sum squared resid	34073631	Schwarz criterion	17.57232
Log likelihood	-179.9426	Durbin-Watson stat	0.922350

Рис. Б.5.

Тобто рівняння множинної регресії має вигляд:

$$y(k) = -21509{,}38 + 30{,}95857 \cdot x_1(k) + 0{,}0002 \cdot x_2(k) + \varepsilon(k)$$

Підхід №2.

При роботі з економетричними даними часто використовують операцію логарифмування по відношенню до цих даних. Операція логарифмування дозволяє зменшити чисельний масштаб даних, що аналізуються, а також позбавитися квадратичного тренду (від полінома другого порядку перейти до поліному першого порядку). В нашому випадку саме так і відбувається, на рис. Б.6 та Б.7 наведені графіки ***"y"*** та ***log(y)***.

```
plot y
series lgy=log(y)
plot lgy
```

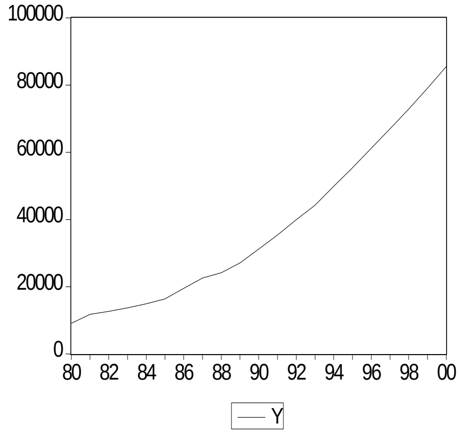


Рис. Б.6.

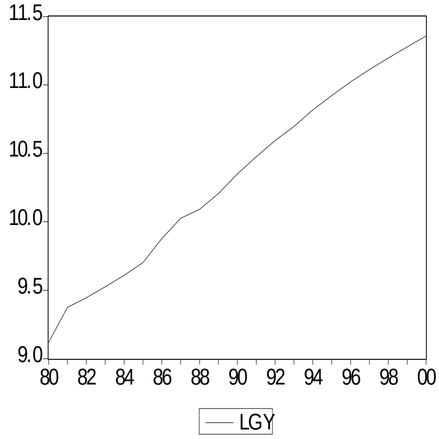


Рис. Б.7.

Побудова рівняння по лагарифмованим даним.

$$\ln[y(k)] = c_0 + c_1 \ln[x_1(k)] + c_2 \ln[x_2(k)] + \varepsilon(k) \qquad \text{EQ2}$$

equation eq2.ls log(y)=c(1)+c(2)*log(x1)+c(3)*log(x2)

Equation: EQ2 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: LOG(Y)
Method: Least Squares
Date: 10/18/08 Time: 02:20
Sample: 1980 2000
Included observations: 21
LOG(Y)=C(1)+C(2)*LOG(X1)+C(3)*LOG(X2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-10.44244	1.421790	-7.344571	0.0000
C(2)	0.830654	0.075926	10.94034	0.0000
C(3)	0.815394	0.095940	8.499039	0.0000

R-squared	0.988716	Mean dependent var	10.32344
Adjusted R-squared	0.987462	S.D. dependent var	0.696165
S.E. of regression	0.077952	Akaike info criterion	-2.133886
Sum squared resid	0.109377	Schwarz criterion	-1.984669
Log likelihood	25.40581	Durbin-Watson stat	0.821642

Рис. Б.8.

Рівняння множинної регресії для лагарифмованих даних має вигляд:

$$\log(y(k)) = -10{,}44 + 0{,}83 \cdot \log(x_1(k)) + 0{,}81 \cdot \log(x_2(k)) + \varepsilon(k)$$

Крок 8. Аналіз отриманих результатів.

В таблиці Б.3 наведені статистичні характеристики отриманих рівнянь множинної регресії Порівнювати суми квадратів похибок отримані при побудові EQ1 та EQ2 в даному випадку некоректно, тому що масштаб даних, що використовувалися для побудови рівнянь, різний – лагарифмовані дані на декілька порядків менші оригінальних. Коефіцієнт Акайке також некоректно порівнювати, так як він обчислюється на основі значення логарифмічної функції правдоподібності, яка в свою чергу обчислюється по сумі квадратів похибок. Аналіз коефіцієнта детермінації та коефіцієнта Дурбина-Уотсона показує, що рівняння EQ1 більше підходить для опису процесу.

Таблиця Б.3.

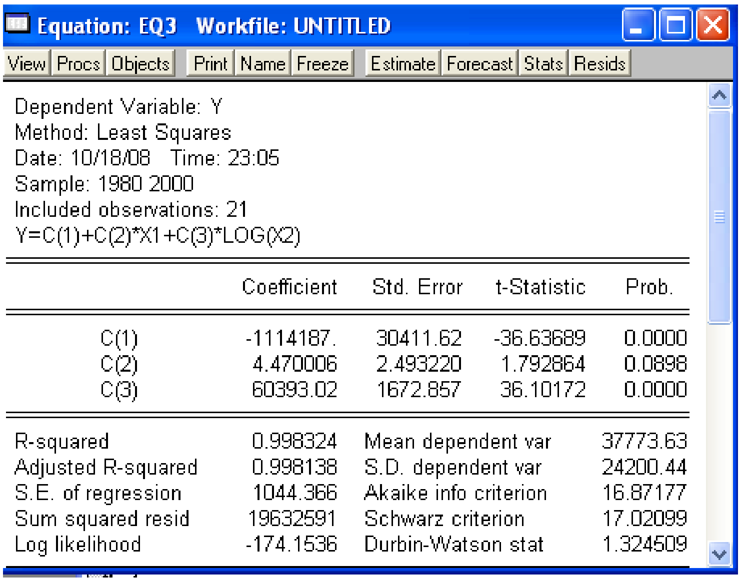
Таблиця статистичних характеристик рівнянь EQ1, EQ2 та EQ3

	EQ1	**EQ2**	**EQ3**	**EQ4**
R-squared	0.997091	0.988716	0,998324	0.999458

Sum squared resid	34073631	0.109377	19632591	5883709
Akaike	17.42310	-2.133886	16,87177	15.82984
Durbin-Watson stat	0.922350	0.821642	1,324509	1.938497

Але аналіз вхідних даних можна продовжити далі. Наприклад побувати рівняння

$$y(k) = c_0 + c_1 x_1(k) + c_2 \cdot \log(x_2(k)) + \varepsilon(k) \qquad \text{EQ3}$$


Equation: EQ3 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/18/08 Time: 23:05
Sample: 1980 2000
Included observations: 21
Y=C(1)+C(2)*X1+C(3)*LOG(X2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-1114187.	30411.62	-36.63689	0.0000
C(2)	4.470006	2.493220	1.792864	0.0898
C(3)	60393.02	1672.857	36.10172	0.0000

R-squared	0.998324	Mean dependent var	37773.63
Adjusted R-squared	0.998138	S.D. dependent var	24200.44
S.E. of regression	1044.366	Akaike info criterion	16.87177
Sum squared resid	19632591	Schwarz criterion	17.02099
Log likelihood	-174.1536	Durbin-Watson stat	1.324509

Рис. Б.9.

Як можна побачити з таблиці Б.3 рівняння EQ3 при порівнянні коефіцієнта детермінації та коефіцієнта Дурбина-Уотсона більш адекватне у порівнянні з EQ1 та EQ2.

Підбір моделей можна продовжити далі за рахунок включення авто регресійної частини для y.

Для цього необхідно побудувати ЧАКФ(y), рис.Б.10. Як можна побачити з рис.Б.10 $PAC(1) = 0{,}852 > 0{,}2$, тобто необхідно включити першу авто регресійну складову для y.

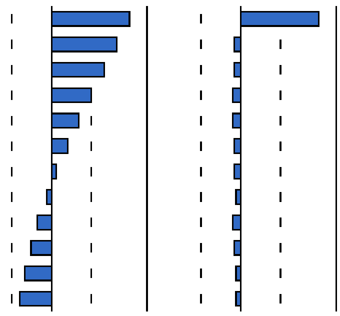
Series: Y Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Sample | Genr | Sheet | Stats | Ident | Line | Ba

Correlogram of Y

Date: 10/19/08 Time: 00:37
Sample: 1980 2000
Included observations: 21

Autocorrelation   Partial Correlation



	AC	PAC	Q-Stat	Prob
1	0.852	0.852	17.549	0.000
2	0.710	-0.062	30.349	0.000
3	0.568	-0.079	39.007	0.000
4	0.429	-0.082	44.226	0.000
5	0.293	-0.083	46.817	0.000
6	0.164	-0.080	47.683	0.000
7	0.047	-0.065	47.760	0.000
8	-0.055	-0.056	47.872	0.000
9	-0.150	-0.085	48.776	0.000
10	-0.232	-0.069	51.136	0.000
11	-0.297	-0.056	55.405	0.000
12	-0.344	-0.045	61.771	0.000

Рис.Б.10.

$$y(k) = c_0 + c_1 \cdot y(k-1) + c_2 x_1(k) + c_3 \cdot \log(x_2(k)) + \varepsilon(k) \qquad \text{EQ4}$$

equation eq4.ls y=c(1)+c(2)*y(-1)+c(3)*x1+c(4)*log(x2)

Equation: EQ4 Workfile: UNTITLED

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
Method: Least Squares
Date: 10/19/08 Time: 00:27
Sample(adjusted): 1981 2000
Included observations: 20 after adjusting endpoints
Y=C(1)+C(2)*Y(-1)+C(3)*X1+C(4)*LOG(X2)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-355611.8	167285.3	-2.125780	0.0494
C(2)	0.703726	0.152607	4.611353	0.0003
C(3)	4.181220	1.507515	2.773584	0.0136
C(4)	19279.83	9076.254	2.124206	0.0496

R-squared	0.999458	Mean dependent var	39205.93
Adjusted R-squared	0.999356	S.D. dependent var	23898.47
S.E. of regression	606.4090	Akaike info criterion	15.82984
Sum squared resid	5883709.	Schwarz criterion	16.02899
Log likelihood	-154.2984	Durbin-Watson stat	1.938497

Рис. Б.11.

$$y(k) = -355611 + 0{,}7 \cdot y(k-1) + 4{,}18 \cdot x_1(k) + 19279 \cdot \log(x_2(k)) + \varepsilon(k)$$

Модель EQ4 дає ще більш кращі результати у порівнянні з EQ3.

Але і це ще не кінець аналізу. Можна включити у рівняння авто регресійну частину для кожного з x_1 та x_2. Але для цього необхідно побудувати часткові кореляційні функції ЧКФ(y, x_1) та ЧКФ(y, x_2). Нажаль в системі Eviews ЧКФ не реалізована (в MatLab також). Але справжній дослідник-студент ІПСА повинен бути в змозі самостійно її запрограмувати та застосувати.

ЛІТЕРАТУРА

1. Довідкова система Eviews.
2. http://en.wikipedia.org
3. http://ru.wikipedia.org
4. http://masters.donntu.edu.ua/2007/fvti/karpunova/diss/index.html

Контрольні запитання

1. В чому полягає відмінність між автокореляційною та частковою АКФ?

2. Як обчислюється кореляційна матриця та яке її призначення? Поясніть значення, отримані вами при обчисленні кореляційної матриці.

3. В чому полягає різниця між кореляцією та коваріацією? Де можна використати коваріаційну функцію?

4. Поясніть в чому полягає відмінність вибіркової АКФ від теоретичної?

5. Для чого необхідно знаходити теоретичні статистичні характеристики процесів?

6. Які статистичні параметри відносяться до описової статистики?

7. Поясність, що характеризує коефіцієнт асиметрії, ексцес, статистика Жак-Бера?

8. Які типи рівнянь використовуються для описування часових рядів?

9. Для чого логарифмують значення часового ряду?

10. Поясніть значення статистичних величин, що використовуються для визначення ступеня адекватності моделі (які генерує пакет після обчислення коефіцієнтів різницевого рівняння): $R^2,\ \sum e^2(k),\ DW,\ AIC,\ BSC,\ t -$статистика.

11. Як отримати оцінки коефіцієнтів рівняння типу АР чи АРКС за допомогою пакету Eviews?

12. Яким чином будується рівняння АРКС?

13. Яким чином будується рівняння множинної регресії?