<u>Лабораторна робота №4</u>

Методичні вказівки до виконання

Тема: “Побудова математичних моделей з трендом та прогнозів на їх основі за допомогою пакету Eviews”

<u>Мета роботи:</u> *Навчитись враховувати тренд і сезонність при побудові математичних моделей часових рядів та будувати прогнози за допомогою пакету Eviews.*

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

ЗМІСТ

Список скорочень

АР – авторегресійне рівняння

КС – ковзне середнє

АРІКС – авторегресія з інтегрованим ковзним середнім

1 Стаціонарні та нестаціонарні процеси

Процеси, представлені часовими рядами, називають ***стаціонарними***, якщо три основні статистичні характеристики відповідного часового ряду не залежать від часу, а саме: *математичне сподівання, дисперсія* та *коваріація*.

Формально стаціонарність процесу формулюють наступним чином:

_ $\mu_x = E[x(k)] = const$

(постійне математичне сподівання);

_ $\mathrm{var}[x(k)] = E\left\{(x(k) - E[x(k)])^2\right\} = E\left\{(x(k) - \mu_x)^2\right\} = const$

(постійна дисперсія);

_ $\gamma_x = E\{[x(k) - \mu_x][x(k-s) - \mu_x]\} = const$, $s = 0,1,2,...$

(постійна коваріація).

Якщо хоча б одна із наведених статистичних характеристик процесу змінюється в часі, то процес називають ***нестаціонарним***. Існують деякі інші визначення стаціонарності, але наведене є самим поширеним і саме воно найчастіше використовується на практиці.

У випадку, коли $E[x(k)] \neq const$, тобто математичне сподівання змінюється в часі, то такий процес називають ***процесом з трендом*** або ***інтегрованим процесом*** (по аналогії із характером зміни сигналу на виході інтегратора) або ***процесом з одиничними коренями*** (відповідного характеристичного рівняння).

Тренд (поточне середнє) може бути ***зростаючим*** або ***спадаючим***, а за характером зміни в часі може бути ***детермінованим*** або ***стохастичним***.

Детермінований тренд описують вибраною функцією, наприклад, поліномом від часу, сплайном, експонентою, комбінацією тригонометричних функцій та інше. Часто використовують поліноми від часу вигляду:

$$y(k) = a_0 + a_1 \cdot k + a_2 \cdot k^2 + ... + a_m \cdot k^m + \varepsilon(k), \qquad (1.1)$$

де $k -$ дискретний час, який зв'язаний з неперервним реальним часом t через період реєстрації (дискретизації) даних: $t = kT_s$; $\varepsilon(k) -$ випадкова змінна, оцінку якої можна знайти після оцінювання рівняння: $\hat{\varepsilon}(k) = e(k)$, де $e(k) -$ похибка моделі. Очевидно, що після оцінювання моделі послідовність значень $\{e(k)\}$ буде містити всі коливання, що накладаються на тренд.

Випадкові тренди, тобто тренди, які не можна описати з необхідною точністю за допомогою детермінованих функцій, моделюють за допомогою випадкових процесів. В даній роботі цей підхід не розглядається.

<u>Таким чином, описуючи тренд рівнянням (1.1), ми фактично видаляємо його з процесу і повна модель процесу буде складатись щонайменше з двох рівнянь: рівняння (1.1) для тренду і рівняння АРКС(p,q), яке описує коливання, що накладаються на тренд.</u>

Тренд може бути видалений з процесу (даних) за допомогою різниць. Так, перші різниці видаляють тренд першого порядку (лінійний тренд), другі різниці видаляють квадратичний тренд і т.д. Наприклад, нехай $y(k) = a_0 + a_1 \cdot k$. Перші різниці цього процесу

$$\Delta y(k) = y(k) - y(k-1) = a_0 + a_1 \cdot k - [a_0 + a_1 \cdot (k-1)] = a_1$$

приводять до видалення лінійного тренду. Очевидно, що після видалення тренду ми вже не зможемо його спрогнозувати. Докладно задача моделювання процесів з трендом буде розглянута в подальшому.

<u>Якщо процес містить сезонний ефект, то він враховується шляхом введення в праву частину рівняння залежної змінної із затримкою, що дорівнює періоду сезонного ефекту, або введення складової ковзного середнього з тією ж затримкою. Для зменшення дисперсії процесу з сезонним ефектом застосовують так звані "сезонні" різниці, тобто різниці вигляду:</u>

$$\Delta_4 y(k) = y(k) - y(k-4),$$

де "4" означає періодичність сезонного ефекту.

Логарифмування даних

В процесі побудови математичних моделей зустрічаються статистичні дані, представлені досить великими числами, наприклад, десятки і сотні тисяч, мільйони і мільярди. Такі числа необхідно перетворювати у більш прийнятні для обчислювального процесу величини. Оскільки для алгоритмів оцінювання параметрів моделей характерним є накопичення похибок обчислень, а в деяких випадках необхідно обчислювати обернені матриці, то великі числа необхідно логарифмувати або нормувати у вибраному діапазоні значень.

2 Побудова моделі нестаціонарного процесу

Для побудови моделі нестаціонарного процесу скористайтесь даними, що зберігаються у файлі **US_M1.txt**. Цей файл містить дані щодо агрегату М1 з першого кварталу 1960 року по четвертий квартал 1991 року для США (**128** значень). Обробка даних виконується в наступній послідовності:

Етап 1. За допомогою пакета *Eviews* організуйте робочий файл (рис. 1) і введіть дані з диска (рис. 2).

File –> New –> WorkFile

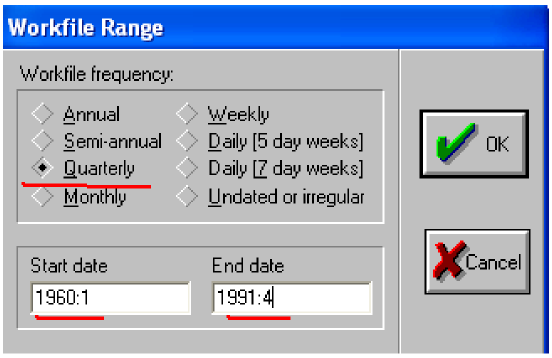


Рис. 1 Організація робочого файлу

Procs –> Import –> Read Text – Lotus – Excel **–>** файл з і'ям ***US_M1.txt***

Задати файлу ім'я ***"y"***, рис. 2.

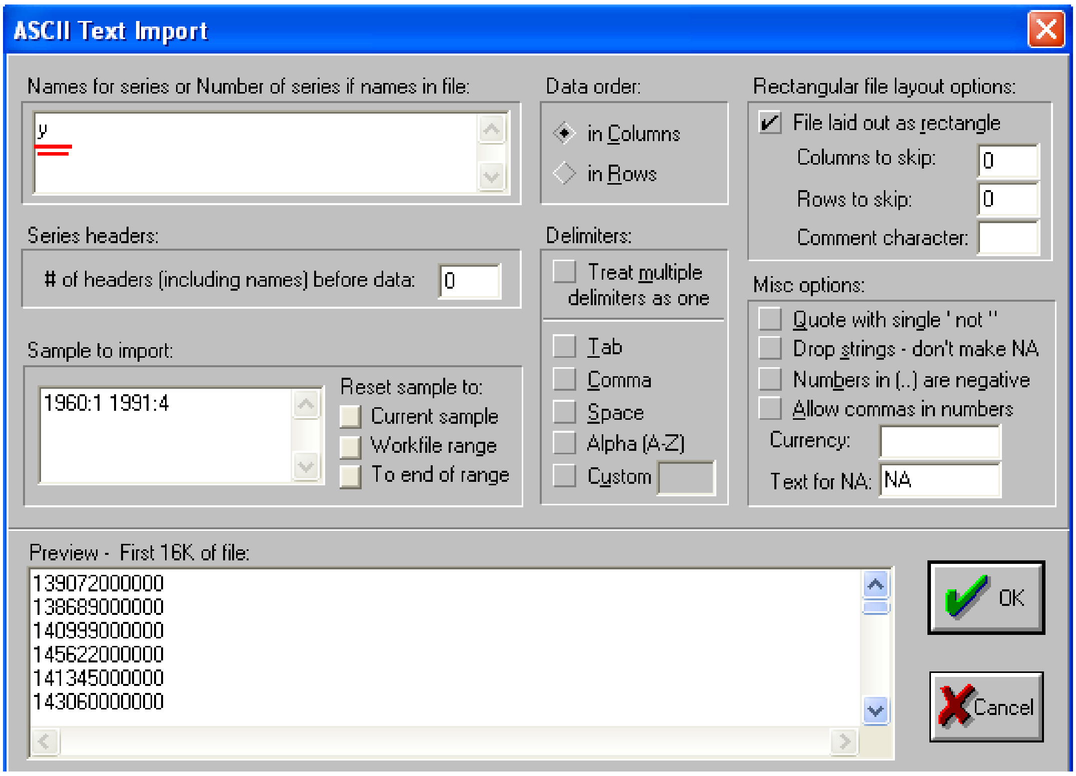


Рис. 2 Ввод даних з диску

Етап 2. Надрукуйте графік введеного ряду (рис. 3) і *візуально* визначте наявність нестаціонарності (тренду, рис. 3).

Графік часового ряду з рис. 3 схожий на частину параболи, отже з великою ймовірністю можна стверджувати, дані описуються трендом у вигляді полінома другого порядку.

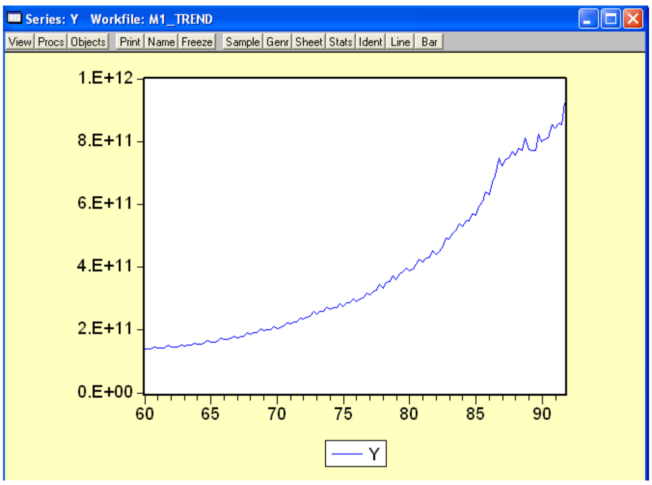


Рис .3 Графік агрегату М1

2.1 Побудова рівняння тренду

Рівняння тренду в даному випадку записується як

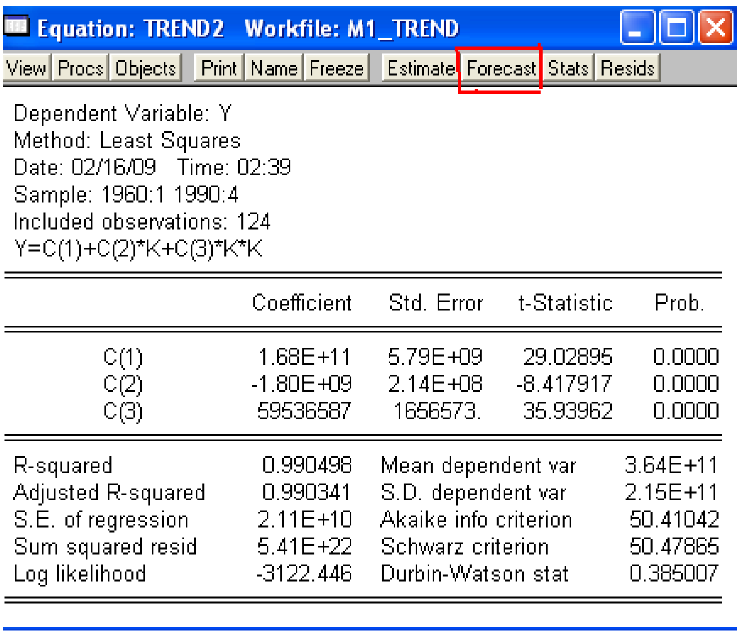
$$y(k) = a_0 + a_1 \cdot k + a_2 \cdot k^2 + \varepsilon(k),$$

де k – дискретний час (він знаходиться у файлі ***time.txt***) який представляє собою послідовність натуральних чисел від 1 до 128, а $\varepsilon(k) = resid$, тобто залишку, отриманому після оцінювання рівняння $y(k) = a_0 + a_1 \cdot k + a_2 \cdot k^2$.

Побудова моделі робиться на даних в часовому проміжку з 1960:1 по 1990:4. Останні чотири точки (з 1991:1 по 1991:4) будуть використані для оцінювання прогнозуючих якостей моделі.

smpl 1960:1 1990:4

equation trend2.ls y=c(1)+c(2)*k+c(3)*k*k


Equation: TREND2 Workfile: M1_TREND

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

Dependent Variable: Y
Method: Least Squares
Date: 02/16/09 Time: 02:39
Sample: 1960:1 1990:4
Included observations: 124
Y=C(1)+C(2)*K+C(3)*K*K

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.68E+11	5.79E+09	29.02895	0.0000
C(2)	-1.80E+09	2.14E+08	-8.417917	0.0000
C(3)	59536587	1656573.	35.93962	0.0000

R-squared	0.990498	Mean dependent var	3.64E+11
Adjusted R-squared	0.990341	S.D. dependent var	2.15E+11
S.E. of regression	2.11E+10	Akaike info criterion	50.41042
Sum squared resid	5.41E+22	Schwarz criterion	50.47865
Log likelihood	-3122.446	Durbin-Watson stat	0.385007

Рис. 4 Статистичні характеристики моделі

<u>**Побудуємо статичний однокроковий прогноз по історичним даним**</u>. Для цього введіть команду ***trend2.forecast*** або натисніть кнопку “forecast” (на рис. 4 обведено червоним) . Після чого з’явиться вікно діалогу (рис. 5) в якому

необхідно обрати тип прогнозування, задати ім’я ряду прогнозу та часовий інтервал для побудови прогнозу.

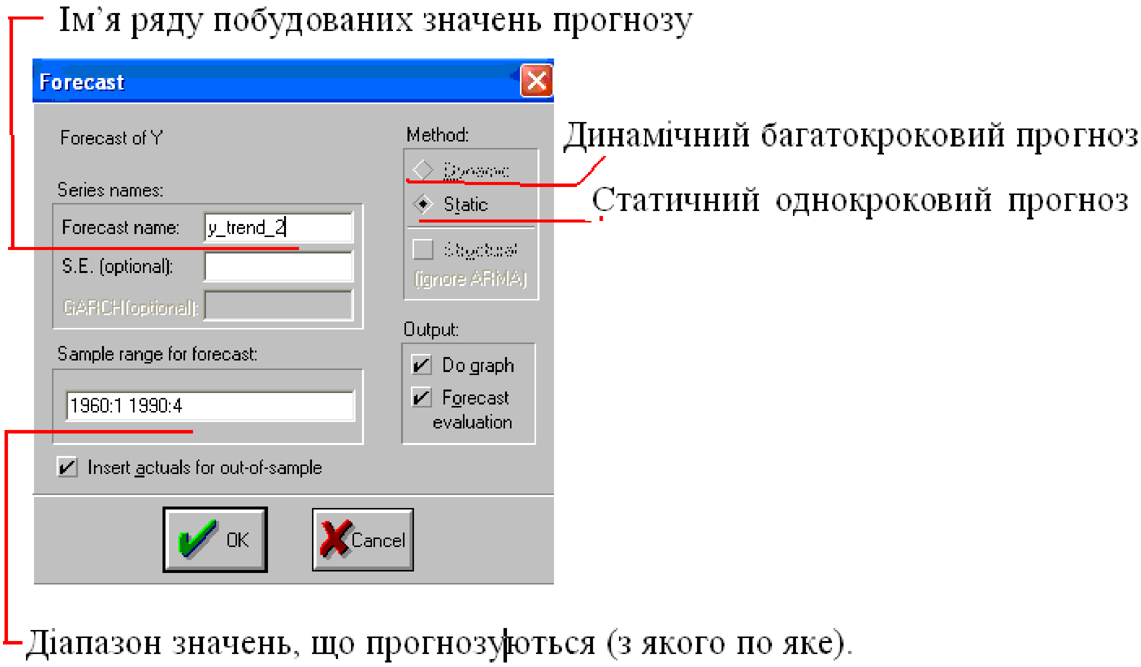


Рис. 5 Вікно побудови прогнозу

Статичний однокроковий прогноз – прогноз на наступний крок будується на основі попередніх реальних історичних даних

Динамічний багатокроковий прогноз – прогноз для 1-шого значення, що прогнозується, будується на основі попередніх реальних історичних даних; для 2-гого значення, що прогнозується, по спрогнозованому перед цим значенню та історичним даним; і так далі; для останнього значення, що прогнозується, по спрогнозованим перед цим значенням.

Після побудови прогнозу в робочому файлі буде створений часовий ряд зі значеннями прогнозу ***“y_trend_2”***. Побудувавши графіки ряду та тренду, рис. 6, можна побачити, що гіпотеза про наявність тренду у вигляді поліному другого порядку була вірна.

plot y y_trend_2

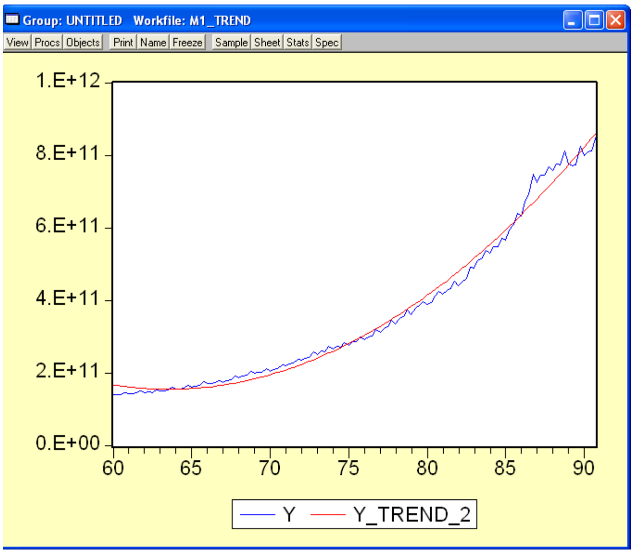


Рис. 6 Графіки часового ряду та тренду

<u>Побудуємо статичний однокроковий прогноз на 4-ри кроки вперед.</u>

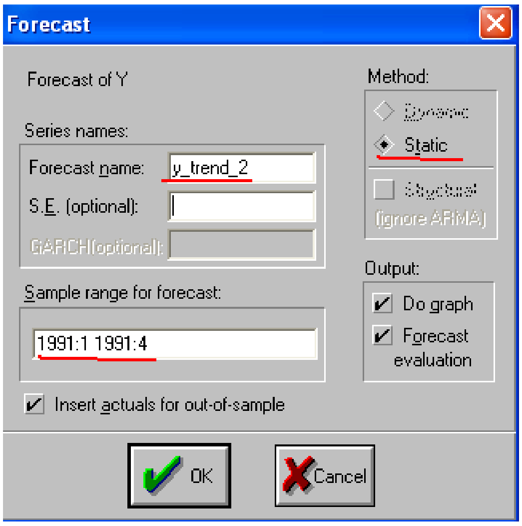


Рис. 7 Вікно побудови прогнозу

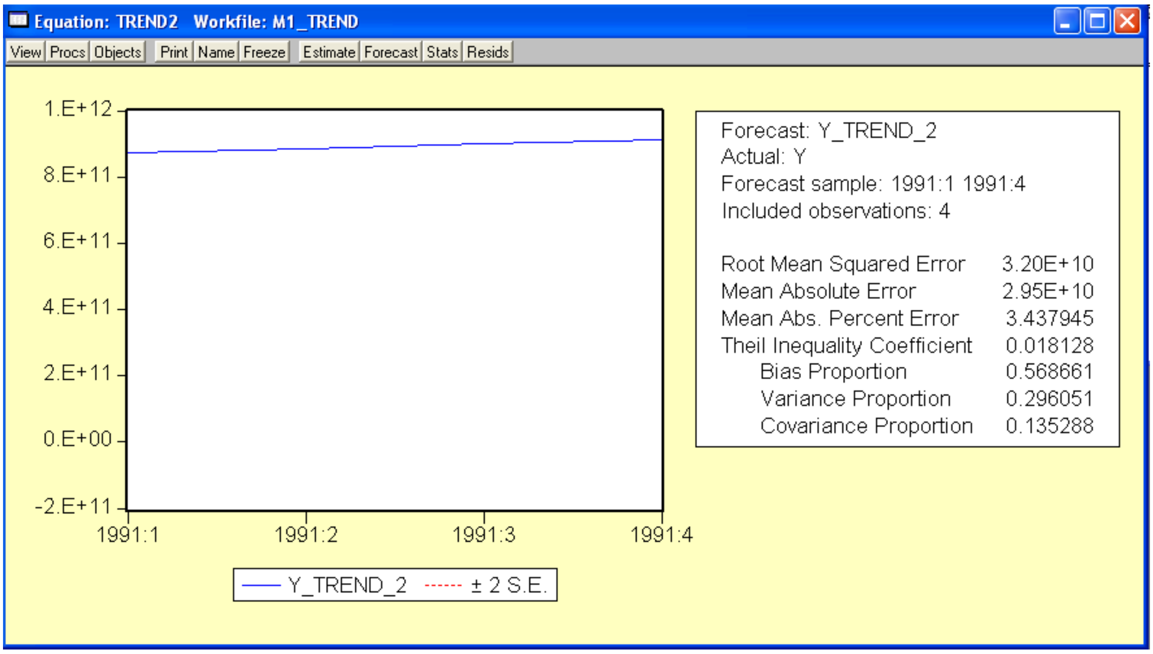


Рис. 9 Вікно побудованого прогнозу та статистичних характеристик отриманих результатів

Побудуємо графіки реальних та прогнозних значень на рис. 10.

smpl 1991:1 1991:4

plot y y_trend_2

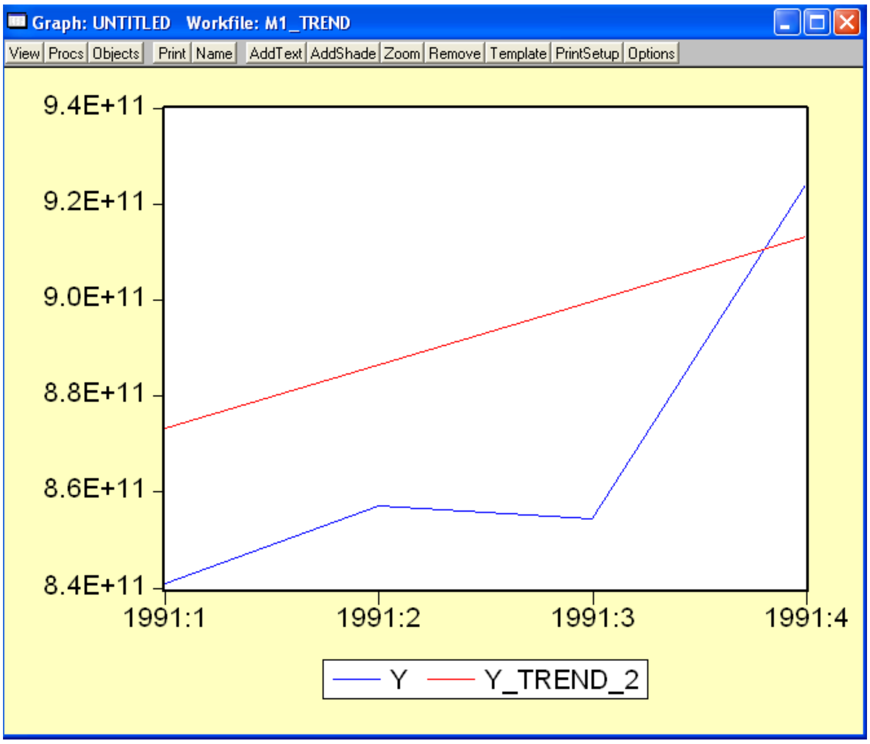


Рис. 10 Графіки реальних та прогнозних значень

Побудова математичної моделі для лагарифмованих значень.

По відношенню до економетричних даних, що складаються з великих додатних чисел, застосовують операцію логарифмування.

Прологарифмуємо часовий ряд ***“y”***, прологарифмовані дані збережемо в ряді ***“logy”*** та побудуємо графік.

smpl 1960:1 1991:4

series logy=log(y)

plot logy

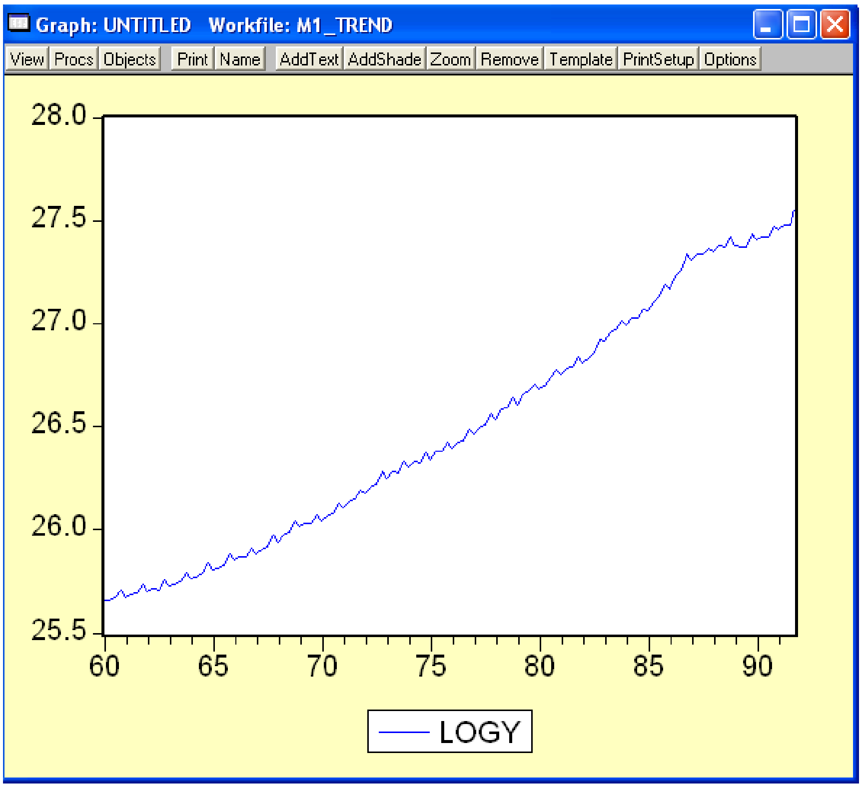


Рис. 11 Графік часового ряду ***“logy”***

Необхідно відмітити, що використання операції логарифмування в даному випадку, окрім приведення даних до меншого масштабу значень зробила ще одну важливу річ – зменшила порядок тренду з другого до першого.

smpl 1960:1 1990:4

equation trend1.ls logy=c(1)+c(2)*k

Sample: 1960:1 1990:4
Included observations: 124
LOGY=C(1)+C(2)*K

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	25.47495	0.012735	2000.320	0.0000
C(2)	0.015726	0.000177	88.93506	0.0000

R-squared	0.984810	Mean dependent var	26.45780
Adjusted R-squared	0.984685	S.D. dependent var	0.569517
S.E. of regression	0.070479	Akaike info criterion	-2.450993
Sum squared resid	0.606017	Schwarz criterion	-2.405505
Log likelihood	153.9616	Durbin-Watson stat	0.172794

Рис. 12 Статистичні характеристики моделі тренда 1-го порядку

$$\log y(k) = 25{,}4749 + 0{,}0157 \cdot k$$ тренд 1-го порядку

trend1.forecast

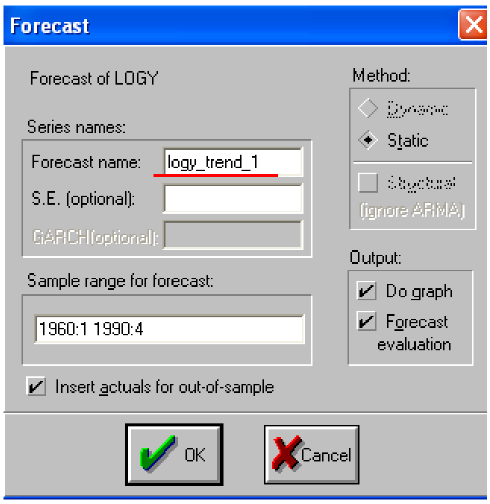


Рис. 13 Вікно побудови прогнозу

plot logy logy_trend_1

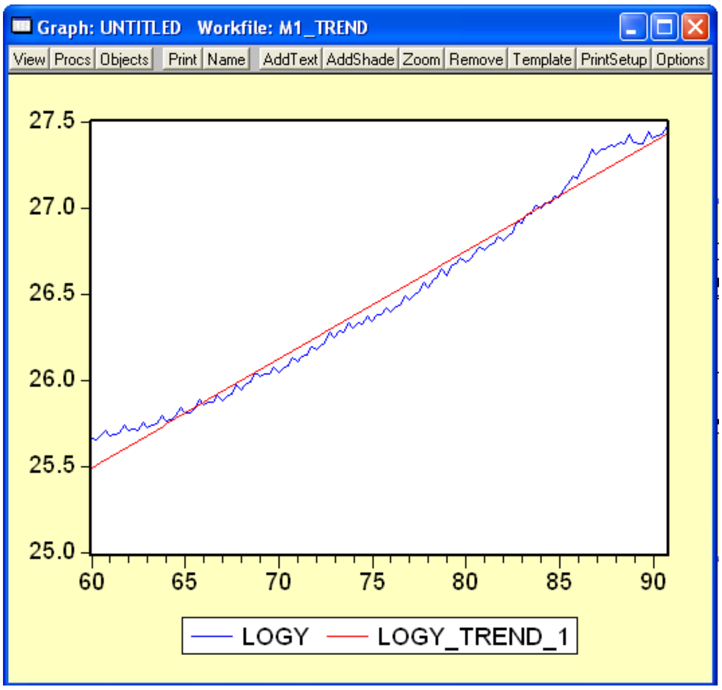


Рис. 14 Графіки часового ряду та тренду

Побудова прогнозу на 4-ри кроки вперед

smpl 1991:1 1991:4

trend1.forecast(g,e) logy_trend1_f

plot logy logy_trend1_f

Команда trend1.***forecast(g,e) logy_trend1_f*** складається з параметрів:
g – вивід на екран графіку прогнозу (+/- 2 математичних сподівання);
e – таблиця статистичних значень прогнозу;
logy_trend1_f – ім’я ряду прогнозу.

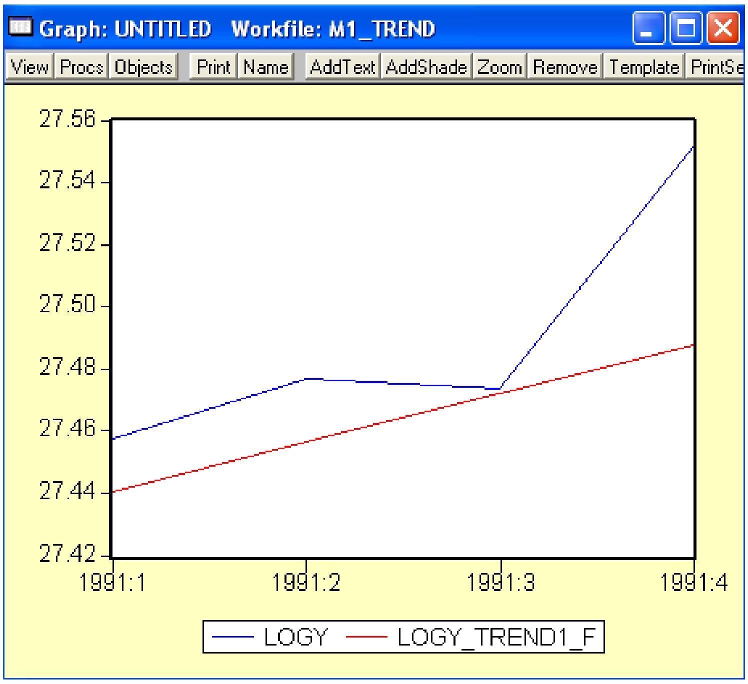


Рис. 15 Графіки реальних та прогнозних значень

<u>3. Побудова АРІКС(4,1,4) моделі для лагарифмованих значень ряду “у”</u>

АРІКС(p,d,q) – процес авторегресії з інтегрованим ковзним середнім, де $p-$ порядок авторегресії, $q-$ порядок ковзного середнього, а $d-$ кількість одиничних коренів характеристичного рівняння (порядок тренду).

Процеси цього класу *нестаціонарні* – вони мають тренд, порядок якого визначається числом одиничних коренів. Якщо $d=1$, то тренд лінійний; якщо $d=2$, то тренд квадратичний і т.д. Таким чином, *процеси з трендом*, *інтегровані процеси* і *процеси з одиничними коренями* – це різні назви процесів, що мають тренд.

Взагалі в назві АРІКС замість “інтегрований” більш коректно було б використовувати “сумований” (більш докладніше цей момент описаний в книзі Бокса і Дженкінса “Аналіз часових рядів” на сторінці 28).

Лагарифмовані дані часового ряду ***“у”*** мають лінійний тренд.

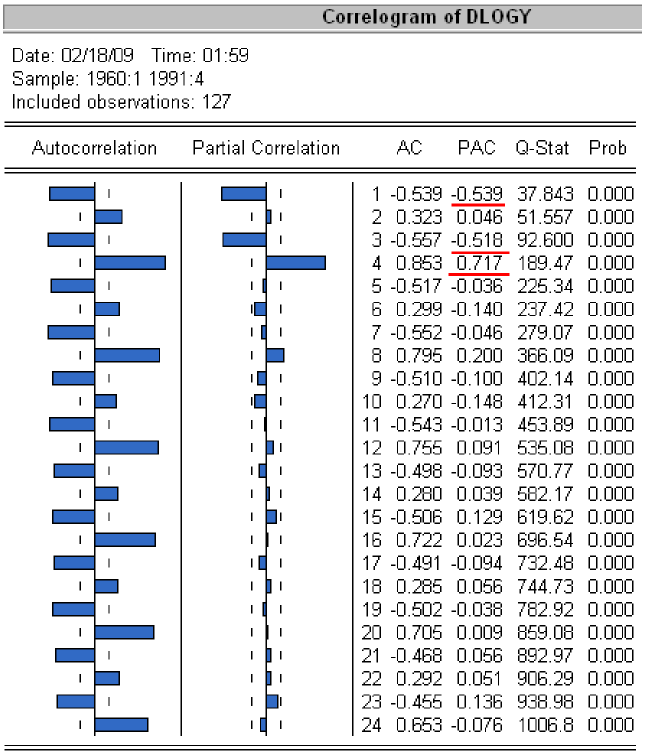
$$\log y(k) = 25{,}4749 + 0{,}0157 \cdot k$$ тренд 1-го порядку

Операція видалення перших різниць часового ряду призводить до видалення лінійного тренду (тренду першого порядку). Видалення других різниць – еквівалентно видаленню тренда, що описується поліномом другого порядку.

smpl 1960:1 1991:4

series dlogy=d(logy)

dlogy.correl



Correlogram of DLOGY

Date: 02/18/09 Time: 01:59
Sample: 1960:1 1991:4
Included observations: 127

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.539	-0.539	37.843	0.000
		2	0.323	0.046	51.557	0.000
		3	-0.557	-0.518	92.600	0.000
		4	0.853	0.717	189.47	0.000
		5	-0.517	-0.036	225.34	0.000
		6	0.299	-0.140	237.42	0.000
		7	-0.552	-0.046	279.07	0.000
		8	0.795	0.200	366.09	0.000
		9	-0.510	-0.100	402.14	0.000
		10	0.270	-0.148	412.31	0.000
		11	-0.543	-0.013	453.89	0.000
		12	0.755	0.091	535.08	0.000
		13	-0.498	-0.093	570.77	0.000
		14	0.280	0.039	582.17	0.000
		15	-0.506	0.129	619.62	0.000
		16	0.722	0.023	696.54	0.000
		17	-0.491	-0.094	732.48	0.000
		18	0.285	0.056	744.73	0.000
		19	-0.502	-0.038	782.92	0.000
		20	0.705	0.009	859.08	0.000
		21	-0.468	0.056	892.97	0.000
		22	0.292	0.051	906.29	0.000
		23	-0.455	0.136	938.98	0.000
		24	0.653	-0.076	1006.8	0.000

Рис. 16 Таблиця ЧАКФ (АКФ) часового ряду ***“dlogy”***

Як видно з аналізу АКФ часового ряду ***“dlogy”***, з того що $ACF_2, ACF_4, \ldots > 0{,}2$, $ACF_1, ACF_3, \ldots < -0{,}2$ можна зробити висновок, що в першому (перші місяці нового року майже завжди супроводжується зниженням економічних показників у порівнянні з останнім кварталом попереднього року) та третьому (кінець літніх відпусток та завершення фінансового року в США) кварталі в США спостерігається сезонне зниження об’єму грошової маси, а в другому (початок літніх відпусток) та четвертому (різдвяний бум подарунків) збільшення.

Взагалі сезонний ефект тому і називається сезонним, що спостерігається з відповідною сезонною періодичністю, наприклад раз в неділю (для щоденних даних), раз в рік (для квартальних та місячних даних)..Найчастіше сезонний ефект спостерігається один раз для заданого проміжку даних але з постійною періодичністю. В нашому прикладі трапилось так, що сезонність фактично спостерігається в кожному кварталі року, кожний рік спостерігається зниження в першому кварталі, незначний ріст в другому, в третьому знову зниження і значний (передноворічний) ріст в четвертому кварталі.

Але в даному випадку, як видно з аналізу ЧАКФ часового ряду ***“dlogy”***, в АР модель навіть без аналізу ряду на сезонність необхідно включити в модель авторегресійні змінні з 1, 3 та 4 лагами. Тобто сезонність включена вже автоматично

smpl 1960:1 1990:4

equation ar4.ls dlogy c ar(1) ar(3) ar(4)

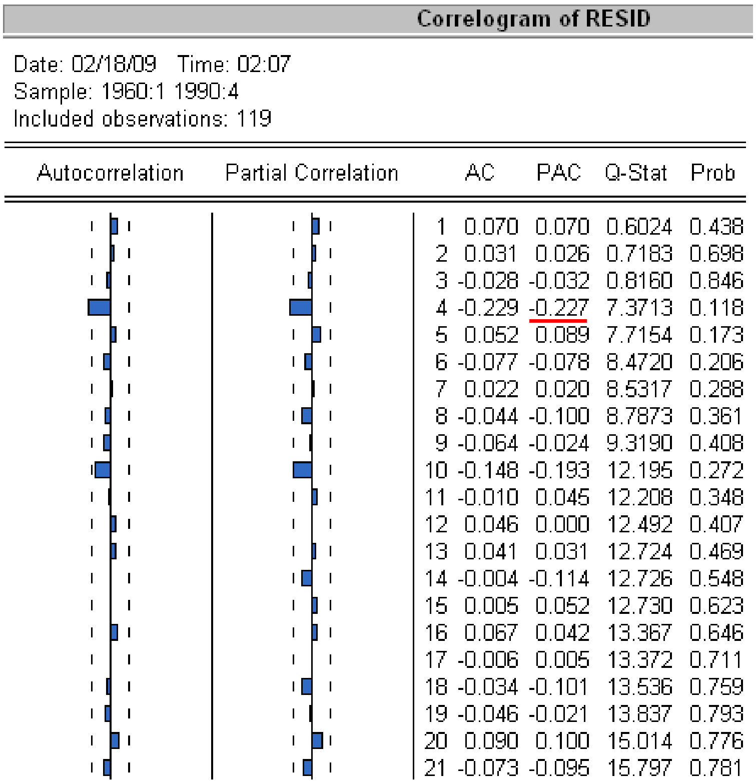
Sample(adjusted): 1961:2 1990:4
Included observations: 119 after adjusting endpoints
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.015171	0.002920	5.194977	0.0000
AR(1)	-0.071629	0.050341	-1.422887	0.1575
AR(3)	-0.126503	0.050068	-2.526633	0.0129
AR(4)	0.778280	0.057438	13.55002	0.0000

R-squared	0.798624	Mean dependent var	0.015113
Adjusted R-squared	0.793371	S.D. dependent var	0.029398
S.E. of regression	0.013363	Akaike info criterion	-5.759562
Sum squared resid	0.020537	Schwarz criterion	-5.666146
Log likelihood	346.6939	F-statistic	152.0240
Durbin-Watson stat	1.850656	Prob(F-statistic)	0.000000

Рис. 17 Статистичні характеристики побудованої АР(4) моделі ряду ***“dlogy”.***

В свою чергу аналіз залишків АР(4), рис. 18, моделі ряду ***“dlogy”*** показує що в побудовану модель АР(4) бажано включити ковзне середнє з 4-м лагом. Факт включення КС з 4-м лагом можна сприймати, як включення сезонності четвертого порядку. На рис. 16 аналіз АКФ чітко вказував на наявність в часових даних сезонної складової четвертого порядку (тому що на 4, 8, 12 і т.д лагах, тобто з періодичністю чотири, спостерігаються **значні сплески АКФ**).



Correlogram of RESID

Date: 02/18/09 Time: 02:07
Sample: 1960:1 1990:4
Included observations: 119

	AC	PAC	Q-Stat	Prob
1	0.070	0.070	0.6024	0.438
2	0.031	0.026	0.7183	0.698
3	-0.028	-0.032	0.8160	0.846
4	-0.229	-0.227	7.3713	0.118
5	0.052	0.089	7.7154	0.173
6	-0.077	-0.078	8.4720	0.206
7	0.022	0.020	8.5317	0.288
8	-0.044	-0.100	8.7873	0.361
9	-0.064	-0.024	9.3190	0.408
10	-0.148	-0.193	12.195	0.272
11	-0.010	0.045	12.208	0.348
12	0.046	0.000	12.492	0.407
13	0.041	0.031	12.724	0.469
14	-0.004	-0.114	12.726	0.548
15	0.005	0.052	12.730	0.623
16	0.067	0.042	13.367	0.646
17	-0.006	0.005	13.372	0.711
18	-0.034	-0.101	13.536	0.759
19	-0.046	-0.021	13.837	0.793
20	0.090	0.100	15.014	0.776
21	-0.073	-0.095	15.797	0.781

Рис. 18 ЧАКФ (АКФ) залишків АР(4) моделі ряду ***“dlogy”***

Побудуємо АРКС(4,4) модель для часового ряду ***“dlogy”***

smpl 1960:1 1990:4

equation ar4ma4.ls dlogy c ar(1) ar(3) ar(4) ma(4)

Sample(adjusted): 1961:2 1990:4
Included observations: 119 after adjusting endpoints
Convergence achieved after 18 iterations
Backcast: 1960:2 1961:1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.022841	0.004783	4.775908	0.0000
AR(1)	-0.015120	0.007449	-2.029927	0.0447
AR(3)	-0.007358	0.007298	-1.008158	0.3155
AR(4)	0.981706	0.009236	106.2903	0.0000
MA(4)	-0.938908	0.028787	-32.61587	0.0000

R-squared	0.848861	Mean dependent var	0.015113
Adjusted R-squared	0.843558	S.D. dependent var	0.029398
S.E. of regression	0.011628	Akaike info criterion	-6.029727
Sum squared resid	0.015413	Schwarz criterion	-5.912957
Log likelihood	363.7687	F-statistic	160.0680
Durbin-Watson stat	1.729863	Prob(F-statistic)	0.000000

Рис. 19 Статистичні характеристики побудованої АРКС(4, 4) моделі ряду ***"dlogy".***

В загальному вигляді на основі отриманнях рівнянь повна АРІКС модель для часового ряду ***"logy"*** (лагарифмованих значень ***"y"***) записується як:

:

<table>
<tr>
<td>

$$\log y(k) = d\log y(k) + \log y(k-1),$$

де

$$d\log y(k) = 0{,}0228 - 0{,}0151 \cdot d\log y(k-1) - 0{,}0073 \cdot d\log y(k-3) + \\ + 0{,}9817 \cdot d\log y(k-4) - 0{,}9389 \cdot ma(k-4)$$

</td>
<td>АРІКС(4,1,4)</td>
</tr>
</table>

Побудова статичного однокрокового і динамічного багатокрокового прогнозів на основі АРІКС(4,1,4) моделі

<u>**Однокроковий статичний прогноз на чотири кроки вперед.**</u>

smpl 1991:1 1991:4
ar4ma4.fit

Команда ***"fit"*** призначена для побудови статичного однокрокового прогнозу.

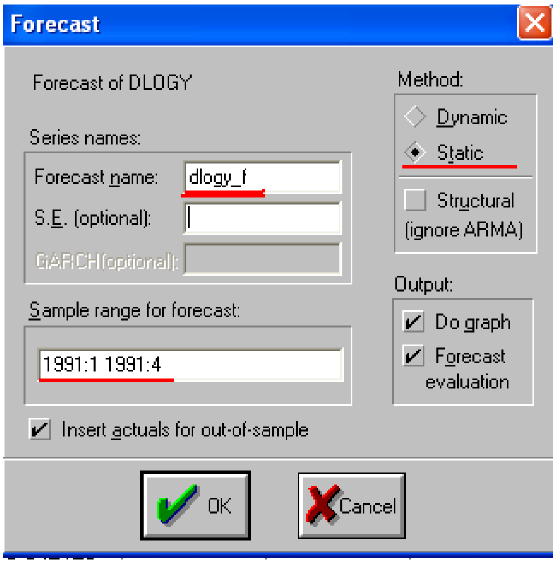


Рис. 20 Вікно побудови статичного прогнозу на чотири кроки вперед

Дані прогнозу зберігаються в часовому ряді ***“dlogy_f”*** в робочому файлі, рис. 21.

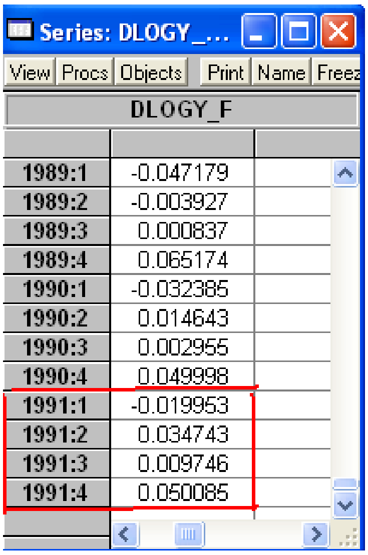


Рис. 21 Прогнозні значення для ряду ***“dlogy”***

Таблиця 1

Значення змінних АРІКС моделі при статичному однокроковому прогнозуванні

Час k	Прогнозне значення ряду ***“dlogy”*** dlogy_f	Значення ряду ***“logy”***	Прогнозне значення ряду “logy” ***logy(k)=dlogy(k)+logy(k-1)*** logy_f
1990/4	–	27.472889	–
1991/1	-0.01995	27.457827	27.45294
1991/2	0.034743	27.476812	27.49257
1991/3	0.009746	27.473663	27.48656
1991/4	0.050085	27.551847	27.52375

<u>Динамічний прогноз на чотири кроки вперед</u>

smpl 1991:1 1991:4

ar4ma4.forecast

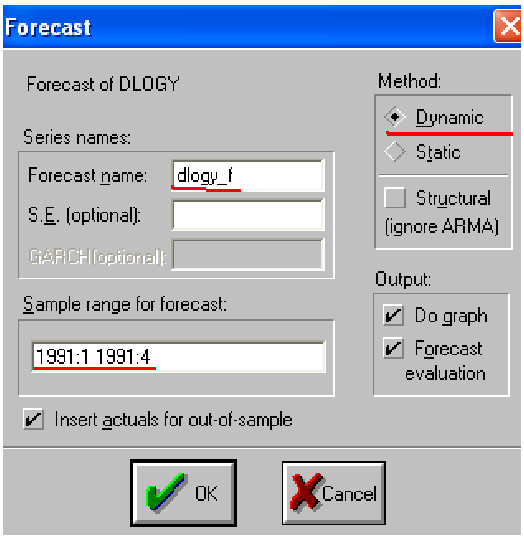


Рис. 21 Вікно побудови динамічного прогнозу на чотири кроки вперед

$$\log \hat{y}(k=1991/1) = d\log y(k=1991/1) + \log y(k=1990/4)$$

$$\log \hat{y}(k=1991/2) = d\log y(k=1991/2) + \log \hat{y}(k=1991/1)$$

$$\log \hat{y}(k=1991/3) = d\log y(k=1991/3) + \log \hat{y}(k=1991/2)$$

$$\log \hat{y}(k=1991/4) = d\log y(k=1991/4) + \log \hat{y}(k=1991/3)$$

Таблиця 2

Значення змінних АРІКС моделі при динамічному прогнозуванні на чотири кроки вперед

Час k	Прогнозне значення ряду ***“dlogy”*** dlogy_f	Значення ряду ***“logy”***	Прогнозне значення ряду “logy” ***logy(k)=dlogy(k)+logy(k-1)*** logy_f
1990/4	–	27.47288902	–
1991/1	-0.02033	–	27.45255612
1991/2	0.034753	–	27.48730916
1991/3	0.009635	–	27.49694384
1991/4	0.049368	–	27.54631223

4. Результати прогнозування

В попередніх підрозділах по лагарифмованим даним часового ряду були побудовані математичні моделі у вигляді тренду 1-го порядку та АРІКС

$$\log y(k) = 25{,}4749 + 0{,}0157 \cdot k$$ тренд 1-го порядку

$$\log y(k) = d\log y(k) + \log y(k-1),$$

де

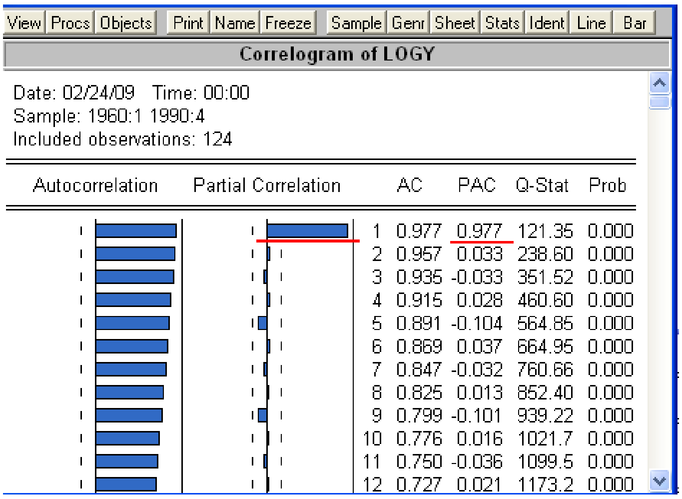
$$d\log y(k) = 0{,}0228 - 0{,}0151 \cdot d\log y(k-1) - 0{,}0073 \cdot d\log y(k-3) + \\ + 0{,}9817 \cdot d\log y(k-4) - 0{,}9389 \cdot ma(k-4)$$ АРІКС(4,1,4)

Для повноти аналізу побудуємо ще АР та АРКС моделі для ***“logy”***.

smpl 1960:1 1990:4

logy.correl

Аналіз ЧАКФ ряду “logy”, рис. 22, показує що необхідно в АР модель включити 1-й лаг.



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

Correlogram of LOGY

Date: 02/24/09 Time: 00:00
Sample: 1960:1 1990:4
Included observations: 124

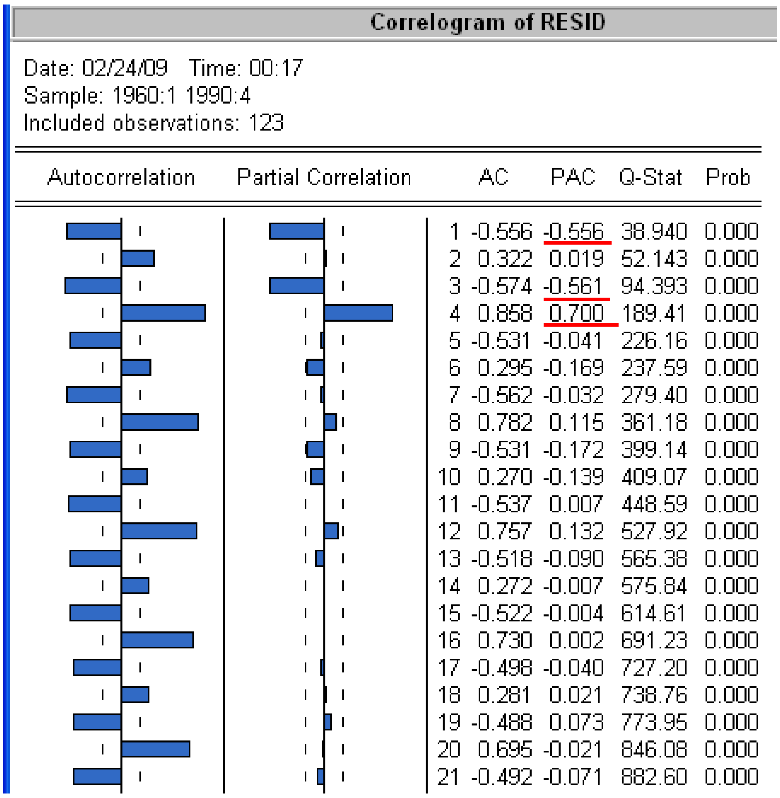
	AC	PAC	Q-Stat	Prob
1	0.977	0.977	121.35	0.000
2	0.957	0.033	238.60	0.000
3	0.935	-0.033	351.52	0.000
4	0.915	0.028	460.60	0.000
5	0.891	-0.104	564.85	0.000
6	0.869	0.037	664.95	0.000
7	0.847	-0.032	760.66	0.000
8	0.825	0.013	852.40	0.000
9	0.799	-0.101	939.22	0.000
10	0.776	0.016	1021.7	0.000
11	0.750	-0.036	1099.5	0.000
12	0.727	0.021	1173.2	0.000

Рис. 22 ЧАКФ ряду ***“logy”***

$$\log y(k) = 0{,}0838 + 1{,}0037 \cdot \log y(k-1) \qquad \text{АР(4)}$$

equation ar1.ls logy=c(1)+c(2)*logy(-1)

Аналіз залишків АР(1) моделі, рис. 23, показує що небхідно включити КС з 1, 3 та 4-м лагами.



Correlogram of RESID

Date: 02/24/09 Time: 00:17
Sample: 1960:1 1990:4
Included observations: 123

	AC	PAC	Q-Stat	Prob
1	-0.556	-0.556	38.940	0.000
2	0.322	0.019	52.143	0.000
3	-0.574	-0.561	94.393	0.000
4	0.858	0.700	189.41	0.000
5	-0.531	-0.041	226.16	0.000
6	0.295	-0.169	237.59	0.000
7	-0.562	-0.032	279.40	0.000
8	0.782	0.115	361.18	0.000
9	-0.531	-0.172	399.14	0.000
10	0.270	-0.139	409.07	0.000
11	-0.537	0.007	448.59	0.000
12	0.757	0.132	527.92	0.000
13	-0.518	-0.090	565.38	0.000
14	0.272	-0.007	575.84	0.000
15	-0.522	-0.004	614.61	0.000
16	0.730	0.002	691.23	0.000
17	-0.498	-0.040	727.20	0.000
18	0.281	0.021	738.76	0.000
19	-0.488	0.073	773.95	0.000
20	0.695	-0.021	846.08	0.000
21	-0.492	-0.071	882.60	0.000

Рис. 23 ЧАКФ залишків АР(1) моделі

smpl 1960:1 1990:4

equation ar1ma4.ls logy c ar(1) ma(1) ma(3) ma(4)

$$\log y(k) = 22{,}8322 + 1{,}004 \cdot \log y(k-1) + 0{,}2542 \cdot ma(k-1) - \\ - 0{,}2756 \cdot ma(k-3) + 0{,}6569 \cdot ma(k-4) \qquad \text{АРКС(4,4)}$$

Таблиця 3

Результати статичного прогнозування на чотири кроки вперед

Час	**Реальне значення** ряду ***"logy"***	Тренд 1-го порядку	**Математична модель** АРІКС(4,1,4)	АР(1)	АРКС(1,4)
1991/1	27.4578	27.4406	27.4529	27.4914	27.462
1991/2	27.4768	27.4563	27.4925	27.4763	27.4875
1991/3	27.4736	27.4721	27.4865	27.4953	27.4907
1991/4	27.5518	27.4878	27.5237	27.4922	27.5066
Середньоквадратична **похибка** прогнозу		0.0173	0.0087	0.0179	0.0124

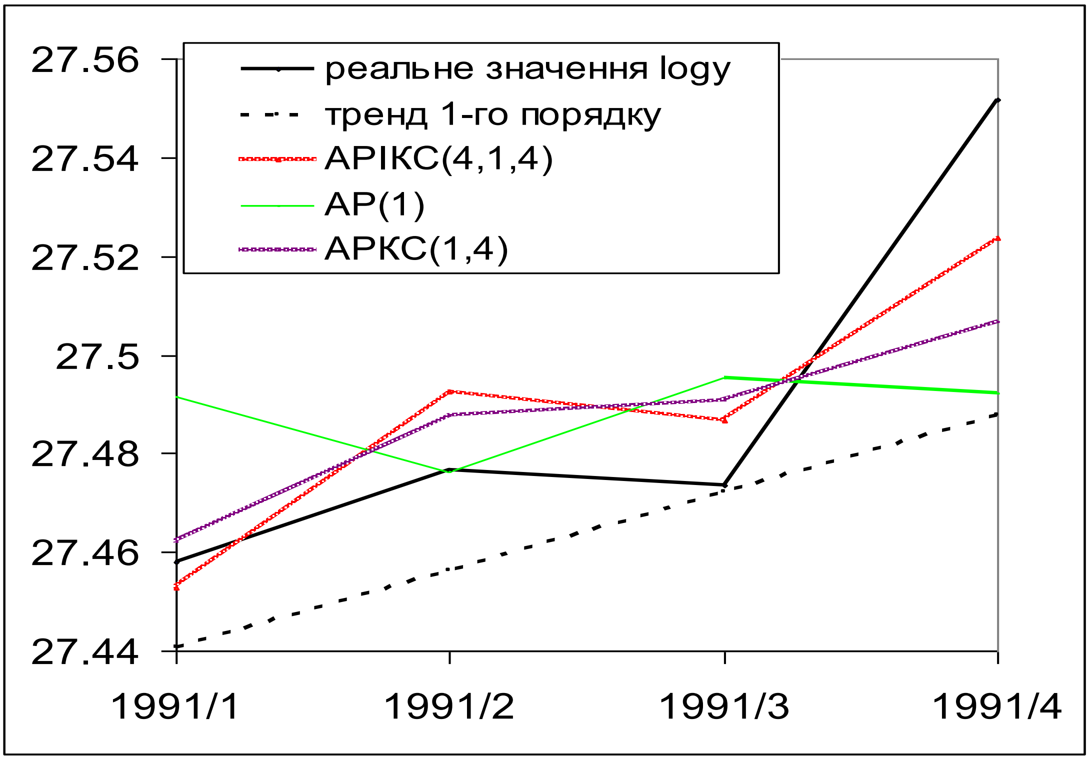


Рис. 24 Графіки реальних та прогнозних значень отриманих при статичному прогнозуванні

Таблиця 4

Результати динамічного прогнозування на чотири кроки вперед

Час	**Реальне значення** ряду *“logy”*	**Математична модель** АРІКС(4,1,4)	АР(1)	АРКС(1,4)
1991/1	27.4578	27.4525	27.4914	27.462
1991/2	27.4768	27.4873	27.51	27.4907
1991/3	27.4736	27.4969	27.5288	27.502
1991/4	27.5518	27.5463	27.5475	27.5295
Середньоквадратична **похибка** прогнозу		0.00666	0.01819	0.00971

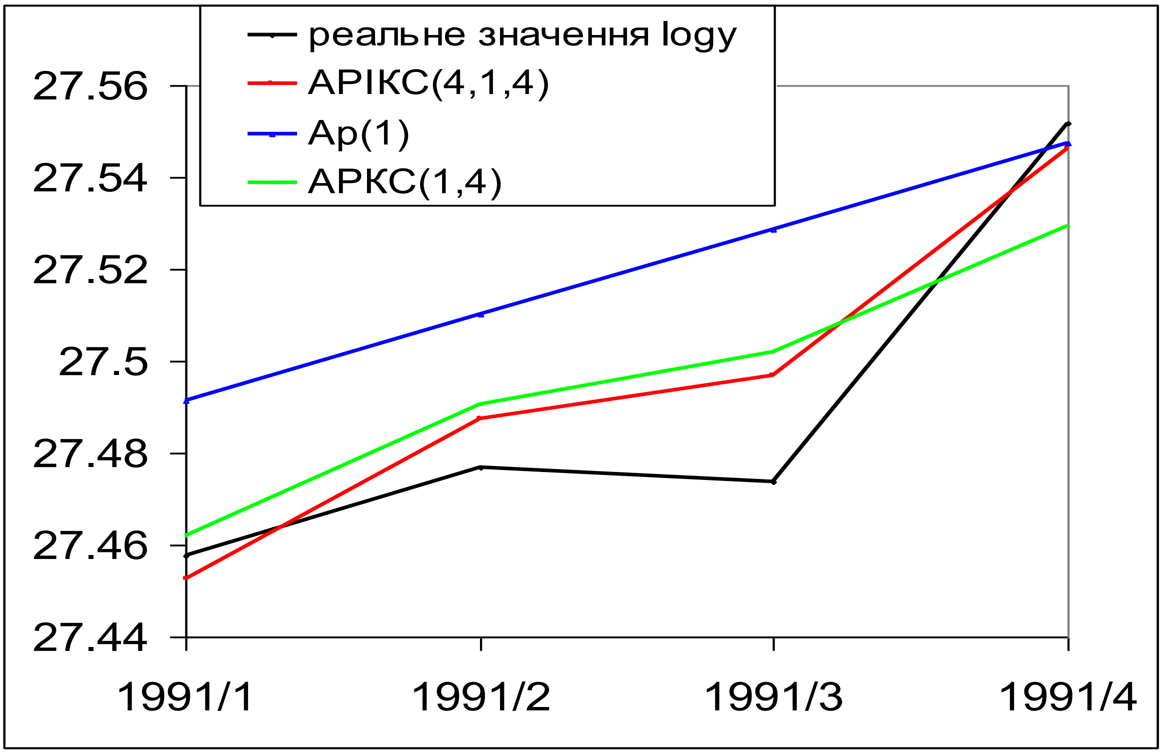


Рис. 25 Графіки реальних та прогнозних значень отриманих при динамічному прогнозуванні

Як видно при статичному та динамічному прогнозуванні за критерієм найменшої середньоквадратичної похибки показала АРІКС(4,1,4).

<u>Завдання на виконання лабораторної роботи</u>

Згідно з номером бригади визначте свої часові ряди в табл. 5.

Для кожного часового ряду побудуйте наступні математичні моделі:

1. у вигляді тренду (першого, другого або обох порядків).
2. АРІКС моделі
3. АР моделі
4. АРКС моделі

На основі кожної моделі побудуйте статичний та динамічний прогнози (для тренду не треба динамічний, тому що він співпадає зі статичним).

В файлі *02_report_example.doc* знаходиться приклад оформлення протоколу.

Важливе зауваження.

Для побудови використовуйте лише частину вибірки.

Приклад-1 вибірка складається з місячних даних в на часовому відрізку від 1947.1 по 2008.8, в цьому випадку для побудови використовуються дані з 1947.1 по 2007.12, а для перевірки прогнозуючих якостей моделі дані з 2008.1 по 2008.8.

Приклад-2 вибірка складається з місячних даних в на часовому відрізку від 1943.1 по 1999.04., в цьому випадку для побудови використовуються дані з 1943.01 по 1998.12, а для перевірки прогнозуючих якостей моделі дані з 1999.01 по 1999.04.

Таблиця 5

Набори даних для виконання лр№4 у відповідності до номера бригади

№ бригади	Файл даних
1	RSAFSNA.txt
	CURRNS.txt
2	Revolns.txt
	TXNAN.xls
3	EXPCA.txt
	RSAFSNA.txt
4	IMPGE.txt
	CURRNS.txt
5	NONREVNS.txt
	Revolns.txt
6	RSAFSNA.txt
	TOTALNS.txt
7	TXNAN.xls
	TOTALNS.txt
8	EXPCA.txt
	RSAFSNA.txt
9	IMPGE.txt
	CURRNS.txt
10	NONREVNS.txt
	Revolns.txt
11	RSAFSNA.txt
	TOTALNS.txt
12	TXNAN.xls
	TOTALNS.txt
13	RSAFSNA.txt
	IMPGE.txt

Запитання

1. В чому полягає різниця між стаціонарним і нестаціонарним процесами?

2. Дайте визначення стаціонарного процесу?

3. Чому є нестаціонарним процес, представлений даними у файлі US_M1.txt?

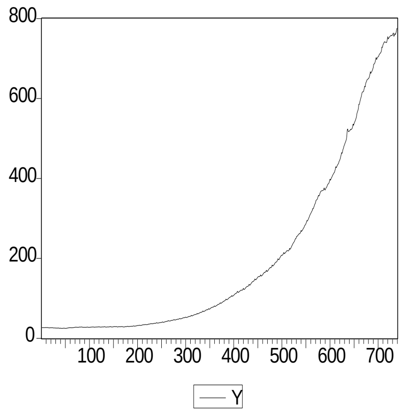
4. Розкажіть яка послідовність побудови моделі нестаціонарного процесу?

5. Для чого використовують перші та різниці вищих порядків?

6. Яка мета застосування "сезонних" різниць?

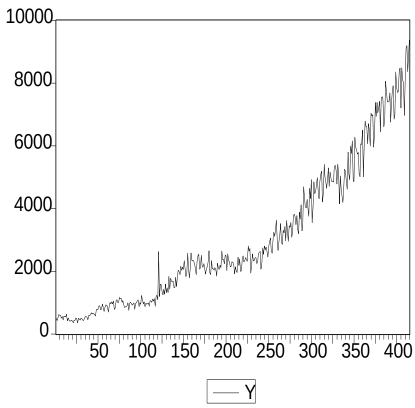
7. Опишіть методику побудови АРІКС моделі.

ДОДАТОК А. Опис даних



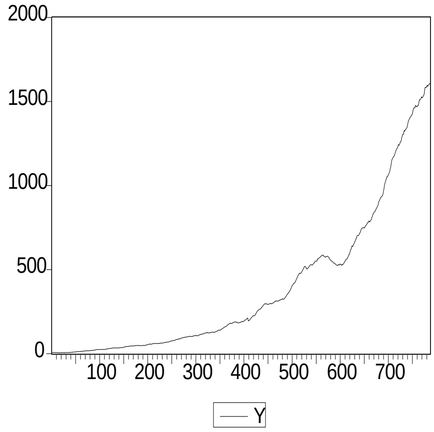
CURRNS.txt

Кількість доларів США в обігу (1 одиниця виміру = 1 мільярд долларів). Файл даних складається з 740 значень місячних даних на часовому проміжку з 1947.1 по 2008.8.



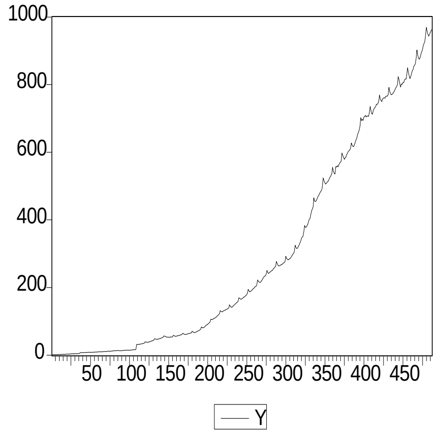
IMPGE.txt

Розмір імпорту з Німеччини до США (мільйонів доларів). Файл даних складається з 415 значень місячних даних на часовому проміжку з 1974.1 по 2008.7.



NONREVNS.txt

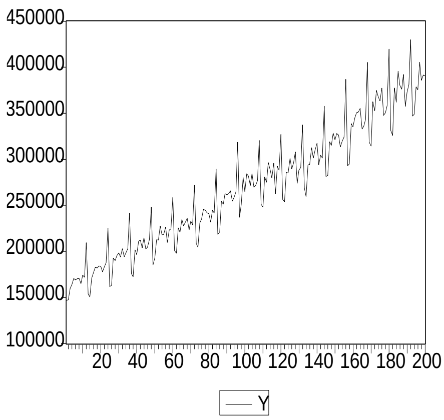
Загальний розмір не поновлюємих споживчих кредитів (в мільярдах доларів). Файл даних складається з 787 значень місячних даних на часовому проміжку з 1943.1 по 2008.7.



Revolns.txt

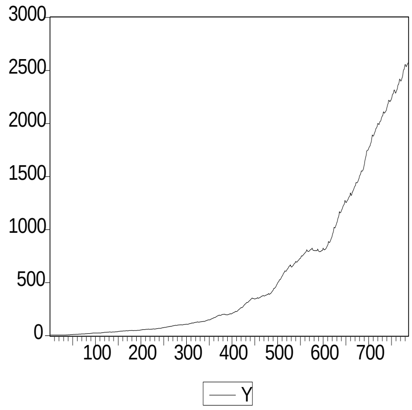
Загальний розмір кредитів, що автоматично поновлюються (в мільярдах доларів). Файл даних складається з 487 значень місячних даних на часовому проміжку з 1968.1 по 2008.7.

Поновлюваний кредит, автоматично поновлюваний кредит або револьверний кредит – надається у межах установленого ліміту заборгованості і строків погашення автоматично, без додаткових переговорів між сторонами кредитної угоди.



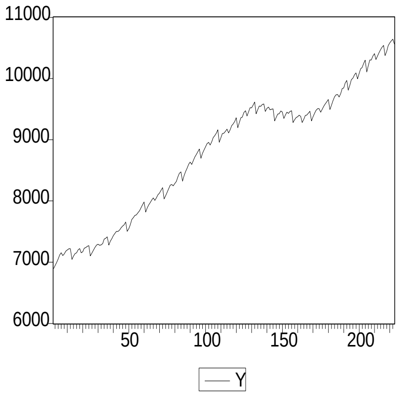
RSAFSNA.txt

Обсяг роздрібної торгівлі та послуг харчування, в мільйонах доларів. Файл даних складається з 200 значень місячних даних на часовому проміжку з 1992.1 по 2008.8.



TOTALNS.txt

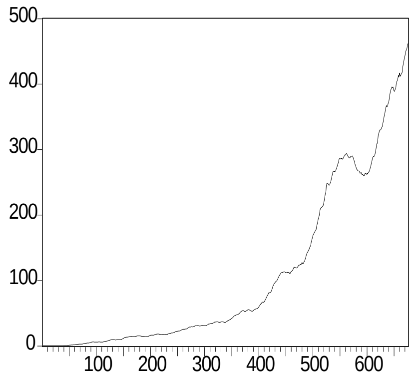
Загальний розмір споживчих кредитів (в мільярдах доларів). Файл даних складається з 787 значень місячних даних на часовому проміжку з 1943.1 по 2008.7.



TXNAN.xls

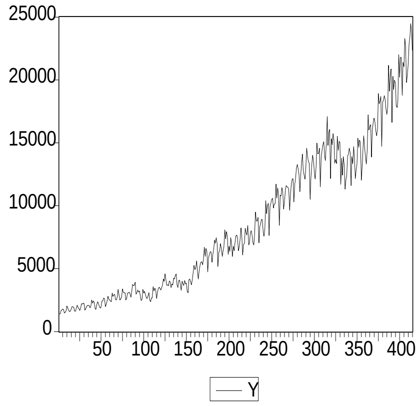
Кількість працівників, за винятком сільськогосподарських, в штаті Техас (1 одиниця виміру = 1000 чоловік). Файл даних складається з 223 значень місячних даних на часовому проміжку з 1990.1 по 2008.7.

Показник чисельності працівників, за винятком сільськогосподарських робітників, в США являється щотижневим індикатором безробіття, що дозволяє судити про стадії економічного циклу, що впливає на прийняття інвестиційних рішень, коливання цін акцій і валюти.



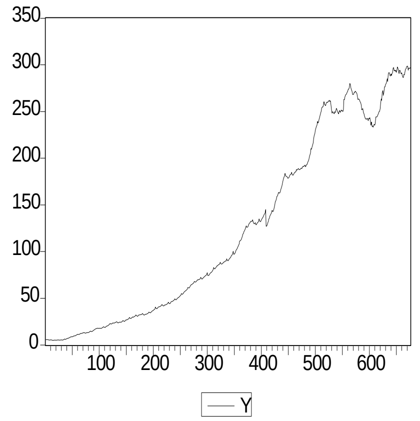
AUTONS.txt

Об'єм виданих кредитів для придбання автомобілів (в мільярдах доларів). Файл даних складається з 676 значень місячних даних на часовому проміжку з 1943.1 по 1999.04.



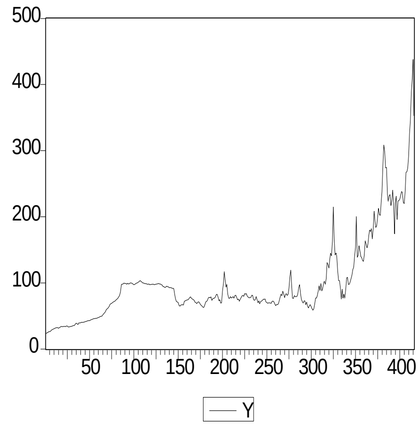
EXPCA.txt

Розмір експорту зі США до Канади (мільйони доларів). Файл даних складається з 415 значень місячних даних на часовому проміжку з 1974.1 по 2008.7.



OTHERNS.txt

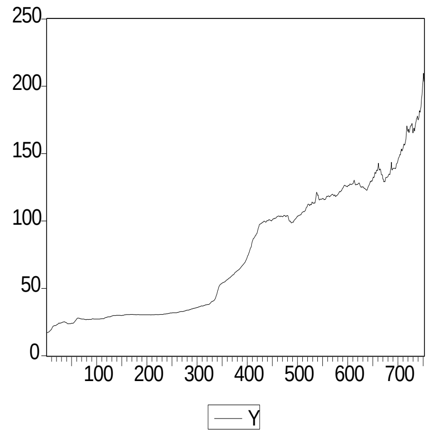
Обсяг виданих споживчих кредитів (в мільярдах доларів). Файл даних складається з 676 значень місячних даних на часовому проміжку з 1943.1 по 1999.4.



PPICEM.txt

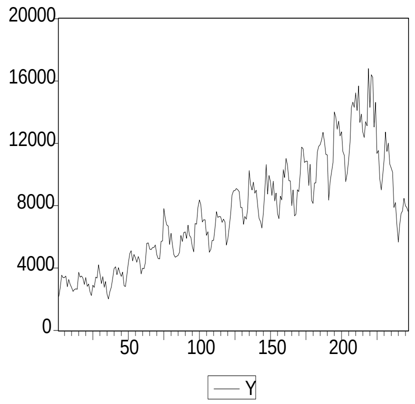
Індекс цін товаровиробників – первинні енергетичні матеріали. Одиниця виміру – індекс. В 1982 році була досягнута позначка в 100 одиниць. Файл даних складається з 416 значень місячних даних на часовому проміжку з 1974.1 по 2008.8.

Індекс цін виробників, індекс промислових цін або Producer Price Index (PPI) –сукупний індекс цін, за якими виробники продають свою продукцію; міра зміни оптових цін. Індекс оцінює середню зміну цін, обумовлених виробниками на всіх етапах виготовлення. Індекс не містить в собі імпортні товари, послуги та податки; відслідковується не окреме значення індексу, а його зміна протягом певного періоду часу; до 1978 року йменувався індексом оптових цін; один з показників інфляції; індикатор стану економіки в цілому; впливає на прийняття інвестиційних рішень.



PPIIDC.txt

Індекс цін товаровиробників – промислові товари. Одиниця виміру – індекс. В 1982 році була досягнута позначка в 100 одиниць. Файл даних складається з 752 значень місячних даних на часовому проміжку з 1946.1 по 2008.8.



TXBP1FH.txt

Кількість будинків в штаті Техас, що знаходяться в приватній власності (1 одиниця виміру = 1 приватний будинок). Файл даних складається з 247 значень місячних даних на часовому проміжку з 1988.1 по 2008.7.

ДОДАТОК Б. Статистичні показники для оцінювання якості побудови моделі та прогнозу

Середньо квадратична похибка RSME:

$$\text{RSME} = \sqrt{\frac{1}{N}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}.$$

Середня похибка ME :

$$\text{ME} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i).$$

Середня відсоткова похибка MPE:

$$\text{MPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{(y_i - \hat{y}_i)}{y_i} \times 100\%.$$

Середня відсоткова абсолютна похибка MAPE:

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\% .$$

Коефіцієнт нерівності Тейла U:

$$U = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{T}(y_i)^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{T}(\hat{y}_i)^2}}$$

Відношення предвзятості:

$$U^M = \frac{(y_i - \hat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2}$$

Відношення варіацій:

$$U^S = \frac{(\sigma_{actual} - \sigma_{fitted})^2}{\frac{1}{N}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2} .$$

Відношення коваріацій:

$$U^C = \frac{2(1-\rho)(\sigma_{actual} \times \sigma_{fitted})}{\frac{1}{N}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2} .$$

RSME як міра точності, є стандартним відхиленням залишків. ME вимірює упередженість в оцінюванні. За припущенням, середня помилка повинна дорівнювати нулю. Інакше, ми маємо упередженість у оцінюванні. Середня відсоткова похибка MPE забезпечує відносну оцінку зміщеності прогнозу. MAPE є подібною до RSME, але вона є відносною мірою точності моделі.

Коефіцієнт нерівності Тейла є дуже важливим індикатором точності моделі і її сумісності . За побудовою, його величина знаходиться між 0 і 1. Якщо U=1, модель не може бути використана для прогнозу. Прогнозовані і реальні ряди є некорельованими. У протилежному випадку, якщо U=0, прогнозовані ряди співпадають з реальними рядами і модель є ідеальною.

Цей коефіцієнт може бути розкладений на суму відношення U^M, відношення варіацій U^S і відношення коваріацій U^C. U^M використовується для перевірки наявності систематичних відхилень середніх для реальних і прогнозних рядів. Або, інакше кажучи, чи модель весь час завищує прогноз Чим менша величина U^M, тим краще. Якщо $U^M = 0$, то в прогнозованих значеннях відсутня упередженість і модель є якісною.

Відношення варіацій використовується для перевірки того, що модель має достатньо динамічних властивостей для поглинання варіації реальних рядів. Наприклад, модель може забезпечити систематично менші коливання ніж коливання реальних рядів. Аналогічно U^M, менші значення U^S є індикатором меншого зміщення.

Нарешті, відношення коваріацій вимірює, наскільки є корельованими прогнозовані та реальні ряди. Рівність U^C нулю є свідченням того, що прогнозовані і реальні ряди ідеально корелюють. Необхідно зазначити, що

$$U^C + U^M + U^S = 1.$$

Точки перегину є важливими, оскільки деякі моделі можуть мати більшу точність, але можуть погано спрацьовувати при прогнозуванні змін трендів (і, наприклад, циклів). Інші моделі можуть бути менш точними, але можуть мати більш багатий динамічний характер. Підсумовуючи, можна говорити про компроміс між точністю і динамічними властивостями. На жаль, не існує формального тесту цих властивостей. Проте, візуальна перевірка прогнозованих і реальних рядів швидко визначає, чи включає модель точки перегину.

Інший важливий тест якості моделі є аналіз чутливості до початкових (стартових) даних. Якщо модель дає результати, в цілому грубо незалежні від початкових даних, то така модель якісна. В протилежному випадку, це може здатись підозрілим, що ми вважаємо чимось невірним, якщо результати моделі залежать від початкової точки.