Лекція 6

Особливості дослідження часових рядів. Різновиди моделей часових рядів.

Аналіз часового ряду (time series)

Нагадаємо приклади часових рядів з попередньої лекції:

- ***Моментальний часовий ряд***

Дата надання позички	01.10.	05.10.	12.10.	23.10.	03.11.	07.11.
Розмір наданої позички, тис. гр. од.	3747	3710	3839	3783	3747	3710

- ***Інтервальний часовий ряд***

Місяць	Січень	Лютий	Березень	Квітень	Травень
Валовий внутрішній продукт, млн грн	6578	7016	7353	7353	7941

- ***Часовий ряд, утворений із середніх значень***

Місяць	Січень	Лютий	Березень
Середня зарплата загалом, грн/міс.	152,2	153,7	165,8

Аналіз часового ряду (time series)

Умови коректності застосування математичного апарату для аналізу часових рядів:

- порівнянність;
- однорідність;
- сталість;
- достатня сукупність спостережень (репрезентативність).

Аналіз часового ряду (time series)

- ***Мета***: побудова математичної моделі ряду, за допомогою якої можна пояснити поведінку ряду і здійснити прогноз на майбутні періоди.
- Основні ***задачі*** дослідження часових рядів:
  - Статистичний аналіз (визначення природи й закону розподілу ряду, вивчення статистичних параметрів ЧР);
  - Прогнозування на підставі адекватних моделей.
- ***Природа ряду:***

1. ***Стаціонарні*** (білий шум, випадкове блукання, марківський процес)

2. ***Нестаціонарні*** (типів TS і DS; тренд-сезонні; нелінійні)

Аналіз часового ряду (time series)

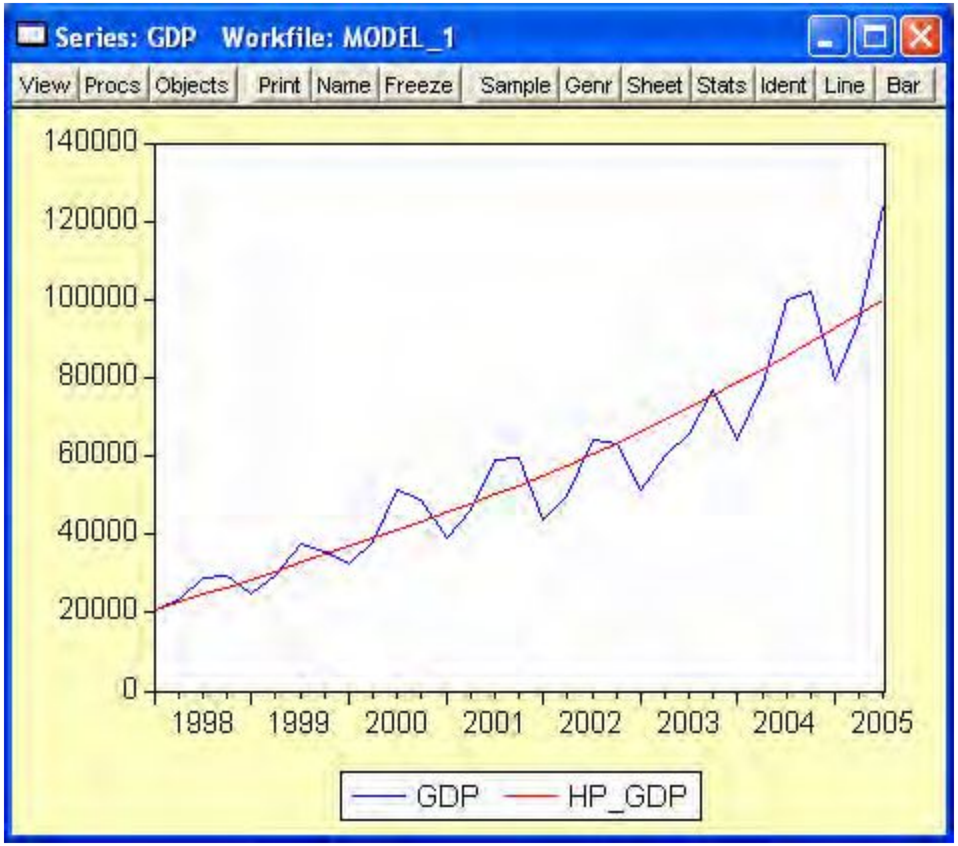
- ***Мета***: побудова математичної моделі ряду, за допомогою якої можна пояснити поведінку ряду і здійснити прогноз на майбутні періоди.
- Основні етапи:

(1) Побудова і вивчення ***графіка часового ряду*** стосовно наявності тренду, сезонності, циклічності та аномальних даних;

(2) Визначення основних числових характеристик – ***описової статистики*** (середнє, медіана, дисперсія, стандартне відхилення, коефіцієнти асиметрії й ексцесу);

(3) побудова ***гістограми*** для перевірки динамічного ряду на відповідність закону нормального розподілу;

Графічний аналіз ЧР



Графік 1 – Приклад ЧР із зростаючим трендом (виділений червоним) та сезонною складовою (у вигляді сезонних коливань)

Виявлення сезонності

- У тому випадку, коли коливання показника не мають чіткої тенденції, індекси сезонності визначаються за формулою:

$$I_{\text{сез}} = \frac{\overline{y_i}}{\bar{y}_{\text{заг}}} \times 100\%,$$

- де $\bar{y}_i$ — середнє значення показника за і-й період року (зазвичай береться період не менше 3х років – якщо це місячні дані, то розраховується середнє за 3 роки по кожному місяцю)
- $\bar{y}_{\text{заг}}$ — загальне середнє значення за всі роки.

Значні відхиленні (> 10%) від усередненого показника (100%) та їх повторення у певні періоди часу вказують на наявність сезонності.

Зазвичай, сезонну компоненту розраховують і виключають із ряду (тобто ряд коректують на сезонність), потім виявляють тенденцію, або трендову компоненту.

Виявлення циклічних (регулярних) коливань

Для пошуку циклічних закономірностей у динаміці певних показників використовується ***<u>спектральний аналіз</u>*** (аналіз Фур'є, перетворення Фур'є, гармонійний аналіз; Spectral (Fourier) analysis), що передбачає перетворення динамічного ряду на послідовність синусоїдальних і косинусоїдальних функцій різних частот.

Основними характеристиками при спектральному аналізі є частота (*f*, довжина хвилі функції, що виражається кількістю циклів за одиницю часу) та період (*T=1/f*, тривалість одного повного циклу).

Методи згладжування часового ряду

5. ***Мультиплікативна модель Холта-Вінтерса***

Дана модель використовує індекс сезонності:

$$S_{t+p} = (a_t + b_t p) c_{t+p},$$

де

$$a_t = \alpha \left(\frac{y_t}{c_{t-s}} \right) + (1-\alpha)(a_{t-1} + b_{t-1}),$$

$$b_t = \beta (a_t - a_{t-1}) + (1-\beta) b_{t-1},$$

$$c_t = \gamma \left(\frac{y_t}{a_t} \right) + (1-\gamma) c_{t-s}, \; 0 \le \alpha, \beta, \gamma \le 1, \quad t = \overline{2s+1, \; T},$$

де s - кількість циклів сезонності. Цей метод вимагає щонайменше 2 повних цикли сезонності по минулих спостереженнях для побудови розрахунків.

Прогноз на період $T + p$ будується наступним чином:

$$\hat{y}_{T+p} = (y_T + p b_T) c_{T-s+p}, \qquad p = 1, \; 2, \; \ldots, \; s,$$

$$\hat{y}_{T+p} = (y_T + p b_T) c_{T-2s+p}, \qquad p = s+1, \; s+2, \; \ldots, \; 2s.$$

Виявлення аномальних рівнів (при нормальному розподілі даних) метод Ірвіна (модифікований)

ґрунтується на порівнянні сусідніх значень ряду

Послідовний розрахунок за трьома спостереженнями величин:

$$\hat{\sigma}_y = \sqrt{\frac{\sum_{t=1}^{n} (y_{t-1} - \overline{y_t})^2 + (y_{t+1} - \overline{y}_t)^2}{2}}$$

$$\overline{y}_t = \frac{(y_{t-1} + y_{t+1})}{2}, \; t = 2,\, 3, \ldots n-1,$$

Метод Ірвіна (продовження)

2. Обчислити величину:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\hat{\sigma}_y}, \quad t = 2, 3, \ldots, n.$$

3. Розраховані ковзні значення λ_t порівнюють із критичним значенням λ для *n=3* (λ_3=2,3).

Якщо вони не перевищують критичне, то відповідні рівні вважаються нормальними.

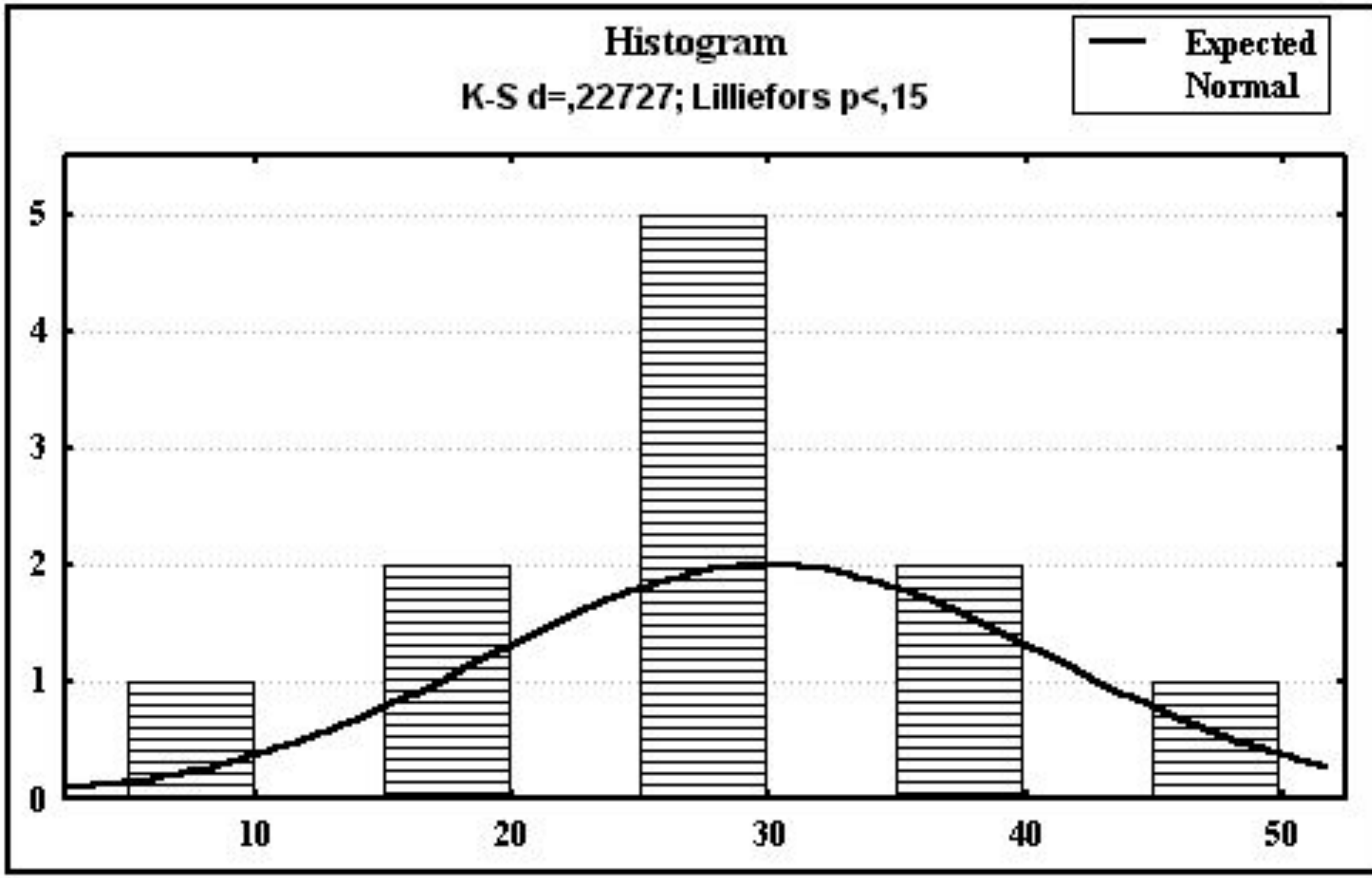
Перевірка динамічного ряду на відповідність закону нормального розподілу

Точна форма нормального розподілу – “дзвоноподібна крива” – визначається лише двома параметрами: середнім і стандартним відхиленням. Для нормального розподілу характерним є те, що 68% всіх значень знаходяться в межах ±1 стандартне відхилення (standard deviation) від середнього, а діапазон ±2 стандартних відхилення включає 95% значень. Про відхилення від нормального розподілу свідчать коефіцієнти асиметрії (skewness) та ексцесу (kurtosis).

Перевірка динамічного ряду на відповідність закону нормального розподілу

Найпростішим способом перевірки динамічного ряду на відповідність нормальному розподілу є побудова *гістограми* (histograms) – діаграми, яка показує, скільки значень попадає в кожний з інтервалів, на які можна розбити весь діапазон зміни певного показника, тобто частоту розподілу значень по інтервалах. На гістограму зазвичай накладається крива нормального розподілу, що дозволяє “на око” оцінити різні характеристики емпіричного розподілу.

Перевірка динамічного ряду на відповідність закону нормального розподілу



Перевірка динамічного ряду на відповідність закону нормального розподілу

Точнішу інформацію про відповідність динамічного ряду нормальному розподілу можна отримати за допомогою критеріїв нормальності, наприклад, **критерію Пірсона (χ^2),** критерію Колмогорова-Смірнова (K-S test for normality) чи критерію Ліллефорса (Lilliefors test for normality).

Обрахунок стандартних статистичних параметрів, що характеризують динамічний ряд

Для <u>стаціонарних динамічних рядів</u> обраховуються *стандартне відхилення* (standard deviation, застаріла назва – середньоквадратичне відхилення) та *стандартна помилка* (standard error). Стандартне відхилення розраховується за формулою:

$$SD = \sqrt{\sigma^2} = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{\sum\left(y_t - \overline{y}\right)^2}{n-1}}$$

де σ^2 – дисперсія, SS – сума квадратів (sum of squares), n – величина динамічного ряду.

Стандартна помилка розраховується за формулою:

$$SE = \sqrt{\frac{\sigma^2}{n}}$$

Обрахунок стандартних статистичних параметрів, що характеризують динамічний ряд

Для **нестаціонарних динамічних рядів** основним статистичним параметром є відхилення реальних значень від тренду – *середньоквадратична помилка* (інколи – стандартне відхилення), що розраховується за формулою:

$$S_y = \sqrt{\frac{\sum\left(y_t - \hat{y}_t\right)^2}{n}}$$

де y_t – реальні значення, $\hat{y}_t$ – трендові значення, n – кількість неспівпадінь реальних значень із трендовими

Етапи аналізу ЧР на стаціонарність

(3) ***Перевірка часового ряду на стаціонарність***:

3.1. Побудова автокореляційної (корелограми) і часткової автокореляційної функції для рівнів часових рядів, їх перших та других різниць.

3.2. Тест на наявність одиничних корнів – розширений тест Дікі-Фуллєра (augmented Dickey – Fuller test, ADF, або unit root test).

(4) ***Приведення до стаціонарного виду*** шляхом:

4.1. Виділення тренда – логарифмування – позначається літерою ***l*** ; знаходження перших або других різниць ряду – позначається літерою ***d*** (“difference”);

4.2. Сезонного згладжування –***sa*** (“seasonal ajustment”);

4.3. Усунення аномальних значень, нормалізація ряду.

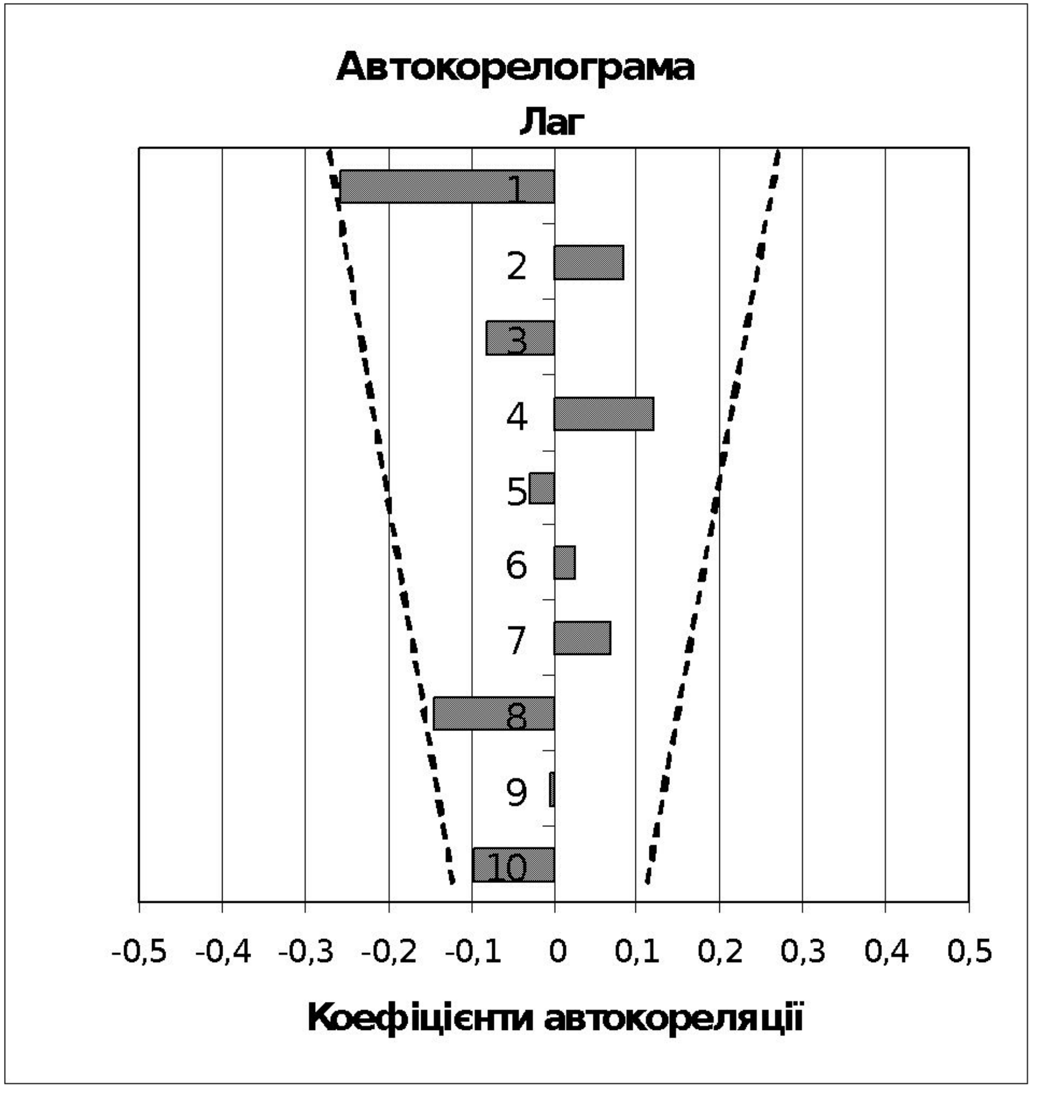
Дослідження стаціонарності часового ряду в EViews

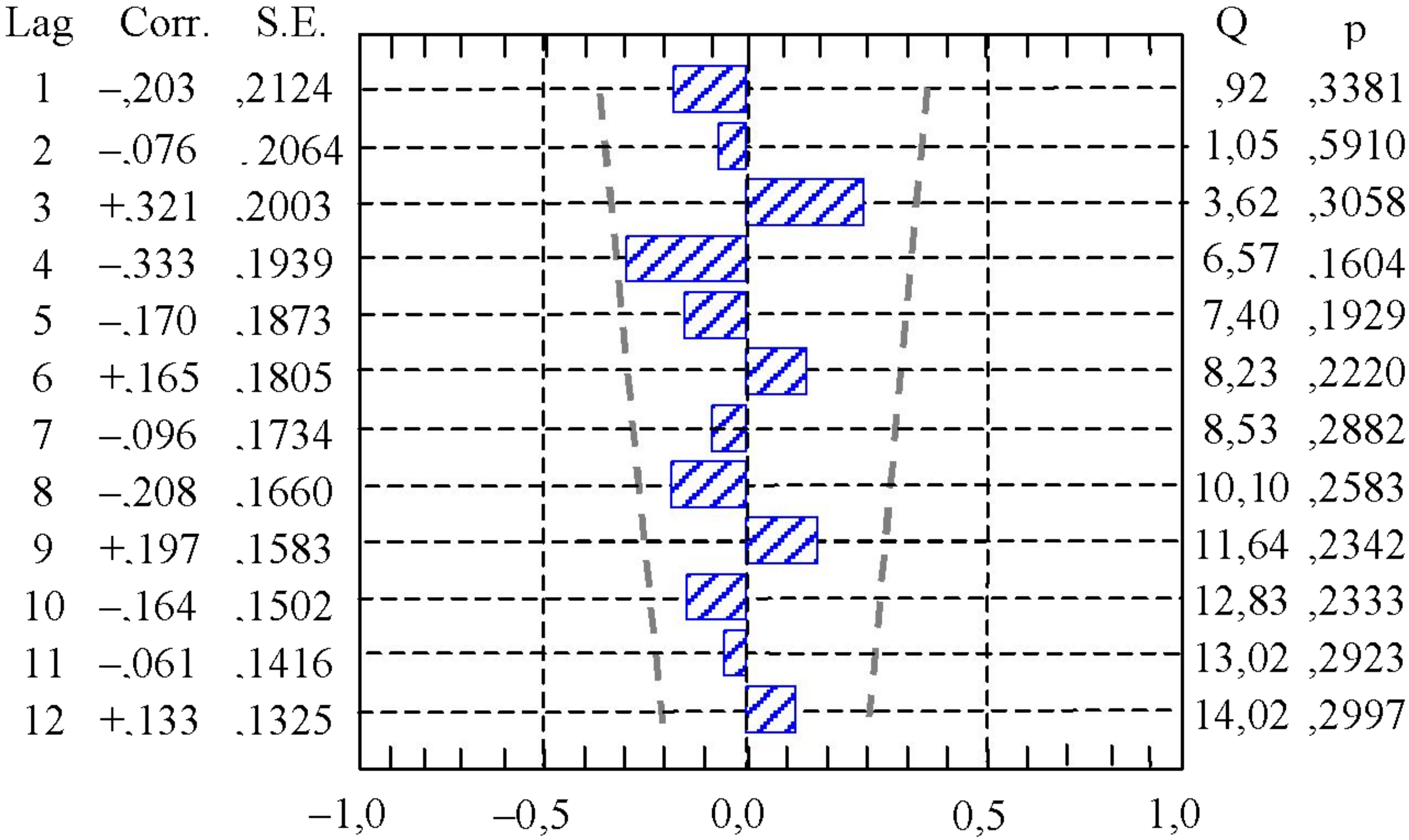
3.1. Побудова автокореляційної (корелограми) і часткової автокореляційної функції з різними часовими лагами для рівнів часових рядів, їх перших та других різниць .

$$r_a = \frac{\sum y_t y_{t-k} - n\overline{y}_t^{\,2}}{\sum y_t^{\,2} - n\overline{y}_t^{\,2}}$$

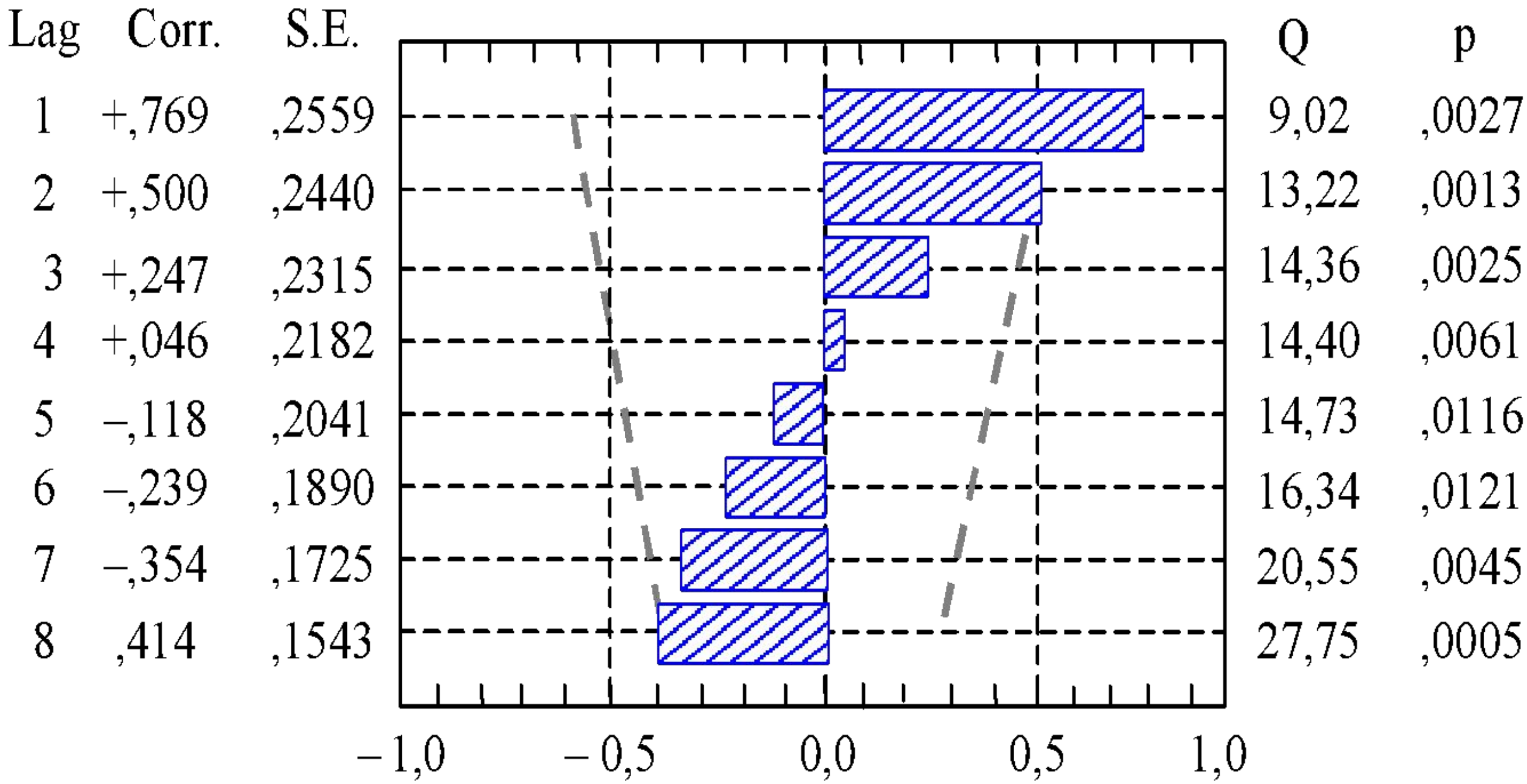
Автокореляційна функція показує ступінь тісноти зв’язку між t спостереженнями часового ряду. Часткова автокореляційна функція розраховується за формулою часткового коефіцієнту кореляції. Корелограма і графік часткової автокореляційної функції мають швидко спадати зі зростанням t у разі стаціонарного ряду. Перевірка гіпотези про те, що автокореляція відсутня до *k*-го лагу включно, здійснюється за допомогою статистики Люнга-Бокса та її *p*-значення.

Автокорелограма



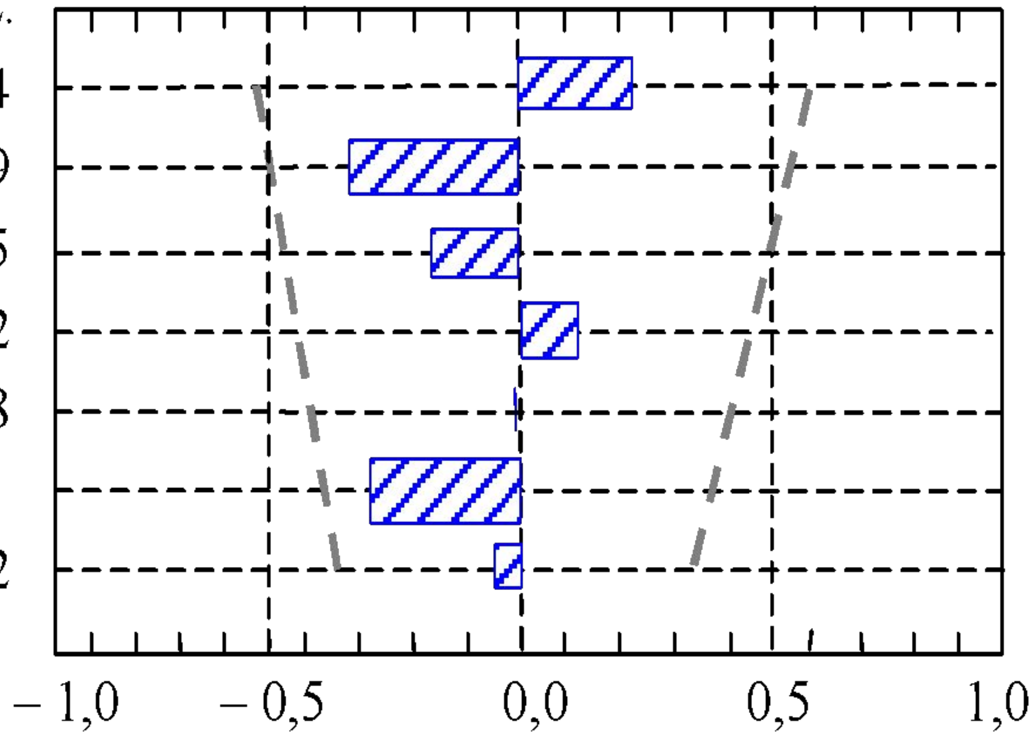


Корелограма для ряду, що не має тренду (стаціонарного ряду)



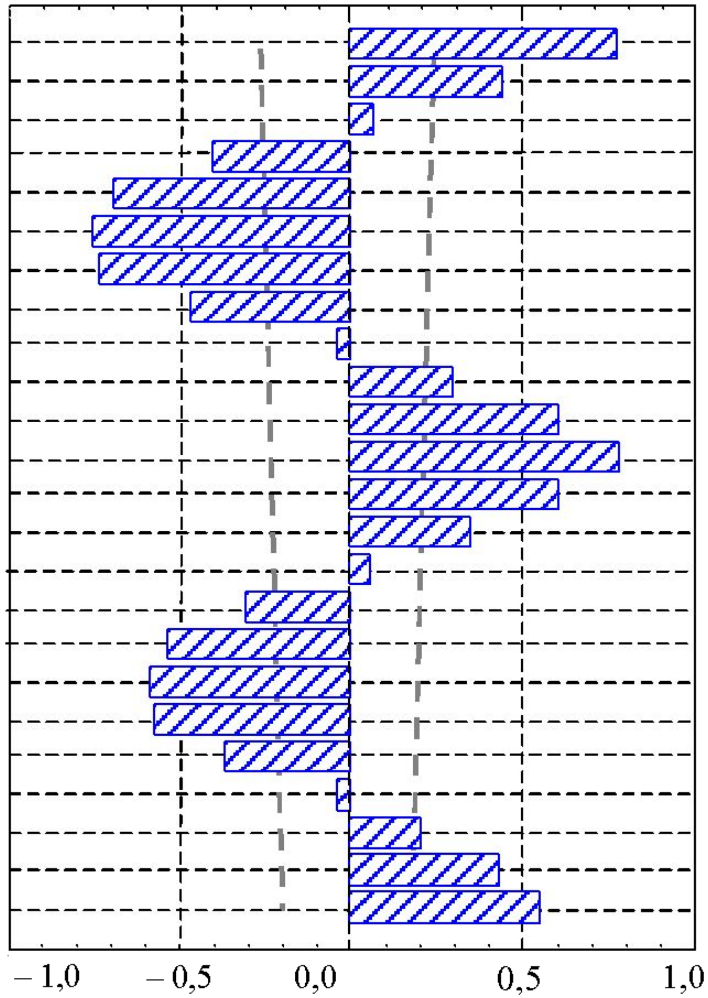
Корелограма для випадку лінійно-адитивного тренду (нульові різниці)

Lag	Corr.	S.E.	Q	p
1	+,162	,2644	,38	,5393
2	–,309	,2509	1,90	,3874
3	–,158	,2365	2,34	,5048
4	+,076	,2212	2,46	,6520
5	–,004	,2048	2,46	,7826
6	–,272	,1870	4,58	,5985
7	–,039	,1672	4,64	,7044



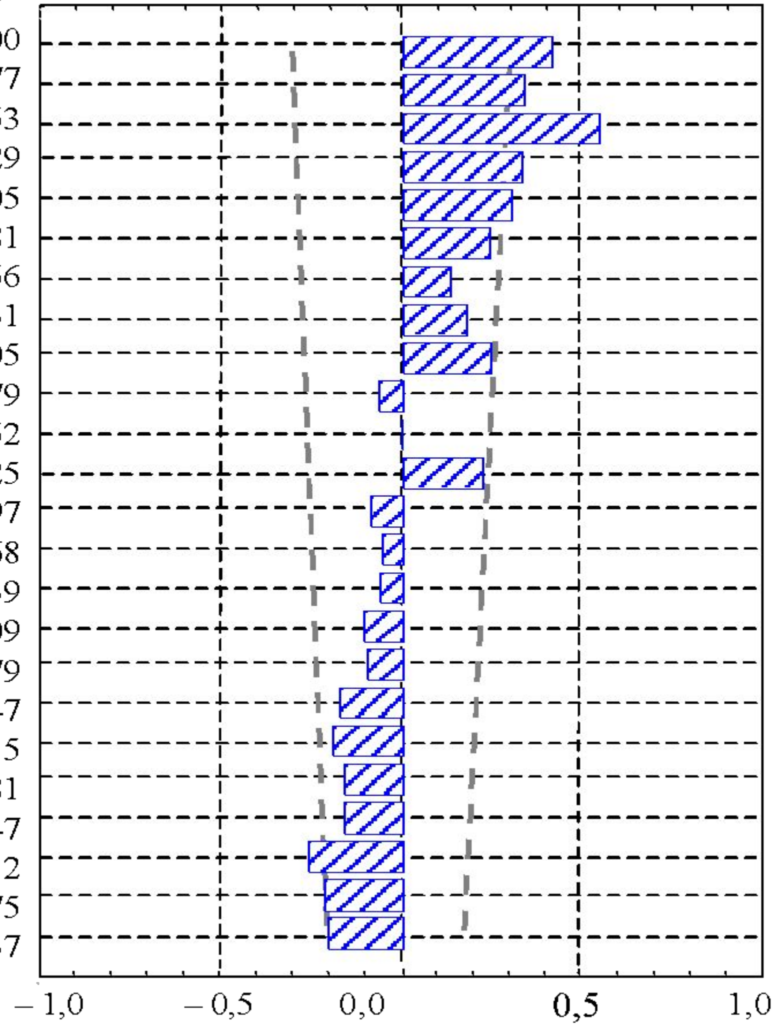
Корелограма для випадку лінійно-адитивного тренду (перші різниці)

Lag	Corr.	S.E.	Q	p
1	+,776	,1259	37,94	,0000
2	+,441	,1249	50,41	,0000
3	+,066	,1238	50,70	,0000
4	–,398	,1227	61,24	,0000
5	–,687	,1216	93,18	,0000
6	–,747	,1205	131,6	,0000
7	–,728	,1194	168,8	0,000
8	–,464	,1182	184,2	0,000
9	–,035	,1171	184,3	0,000
10	+,298	,1159	190,9	0,000
11	+,604	,1148	218,6	0,000
12	+,781	,1136	265,9	0,000
13	+,606	,1124	294,9	0,000
14	+,350	,1112	304,8	0,000
15	+,056	,1100	305,0	0,000
16	–,306	,1088	312,9	0,000
17	–,529	,1075	337,1	0,000
18	–,582	,1063	367,1	0,000
19	–,570	,1050	396,6	0,000
20	–,364	,1037	408,9	0,000
21	–,036	,1024	409,0	0,000
22	+,206	,1011	413,2	0,000
23	+,433	,0997	432,0	0,000
24	+,551	,0984	463,4	0,000



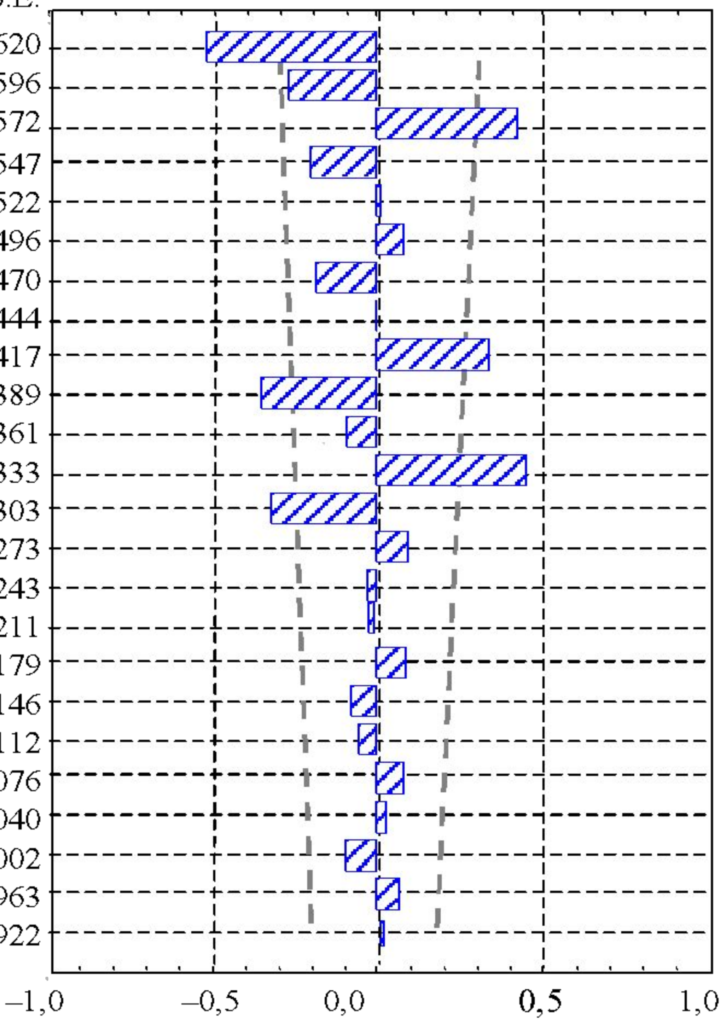
Корелограма ряду із сезонним коливанням (нульові різниці)

Lag	Corr.	S.E.	Q	p
1	+,419	,1600	6,85	,0088
2	+,343	,1577	11,58	,0031
3	+,554	,1553	24,29	,0000
4	+,333	,1529	29,02	,0000
5	+,306	,1505	33,15	,0000
6	+,245	,1481	35,89	,0000
7	+,131	,1456	36,70	,0000
8	+,177	,1431	38,24	,0000
9	+,247	,1405	41,34	,0000
10	–,069	,1379	41,59	,0000
11	–,003	,1352	41,59	,0000
12	+,221	,1325	44,39	,0000
13	–,094	,1297	44,91	,0000
14	–,058	,1268	45,12	,0000
15	–,065	,1239	45,39	,0001
16	–,113	,1209	46,27	,0001
17	–,104	,1179	47,05	,0001
18	–,179	,1147	49,49	,0001
19	–,201	,1115	52,75	,0001
20	–,166	,1081	55,10	,0000
21	–,168	,1047	57,68	,0000
22	–,270	,1012	64,78	,0000
23	–,224	,0975	70,08	,0000
24	–,212	,0937	75,19	,0000



Корелограма ряду з лінійним трендом та сезонно-адитивною складовою (нульові різниці)

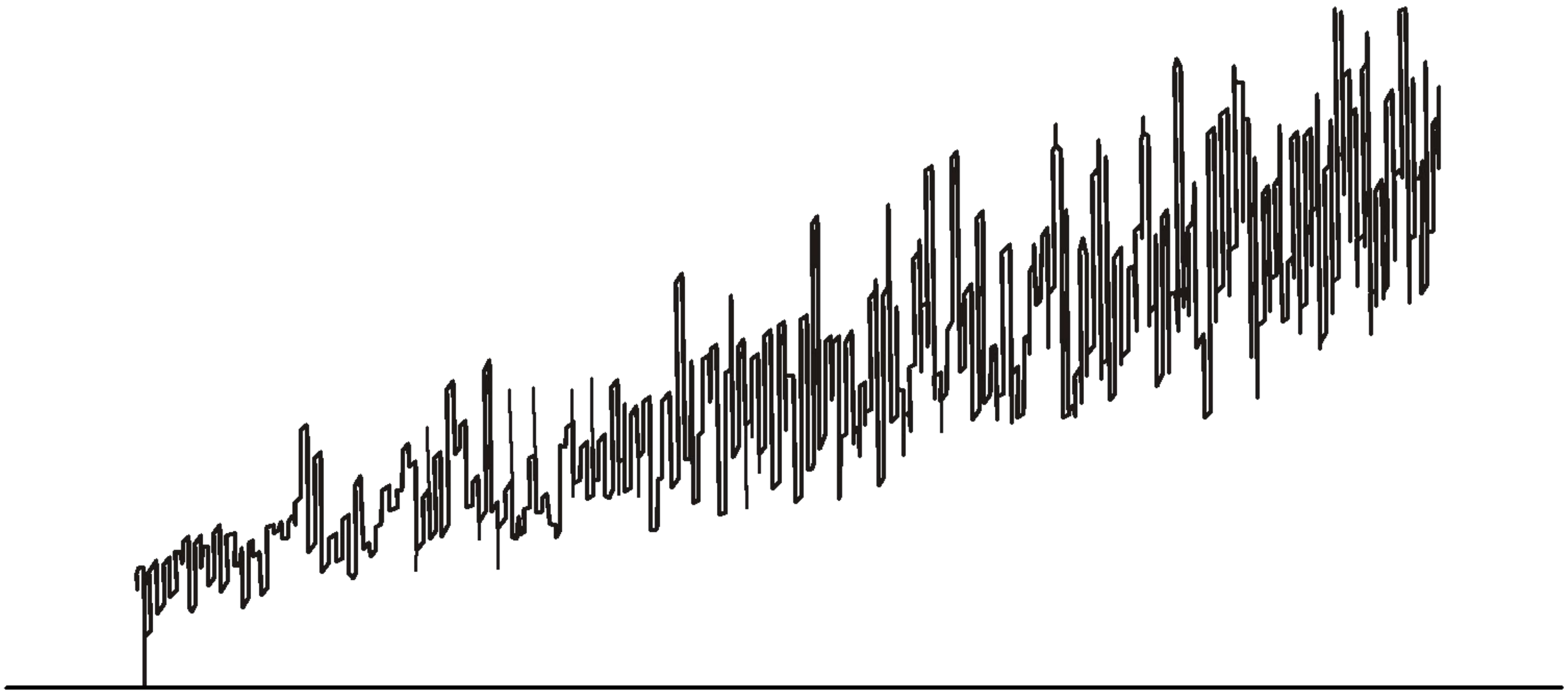
Lag	Corr.	S.E.	Q	p
1	–,465	,1620	8,25	,0041
2	–,240	,1596	10,50	,0052
3	+,382	,1572	16,40	,0009
4	–,181	,1547	17,77	,0014
5	+,007	,1522	17,77	,0033
6	+,071	,1496	17,99	,0063
7	–,166	,1470	19,26	,0074
8	+,001	,1444	19,26	,0135
9	+,305	,1417	23,90	,0045
10	–,315	,1389	29,04	,0012
11	–,082	,1361	29,40	,0020
12	+,406	,1333	38,70	,0001
13	–,286	,1303	43,50	,0000
14	+,084	,1273	43,94	,0001
15	–,024	,1243	43,98	,0001
16	–,016	,1211	43,99	,0002
17	+,077	,1179	44,42	,0003
18	–,068	,1146	44,77	,0004
19	–,050	,1112	44,98	,0007
20	+,072	,1076	45,43	,0010
21	+,025	,1040	45,49	,0015
22	–,085	,1002	46,21	,0019
23	+,062	,0963	46,63	,0025
24	+,008	,0922	46,64	,0037



Корелограма ряду з лінійним трендом та сезонно-адитивною складовою (перші різниці)

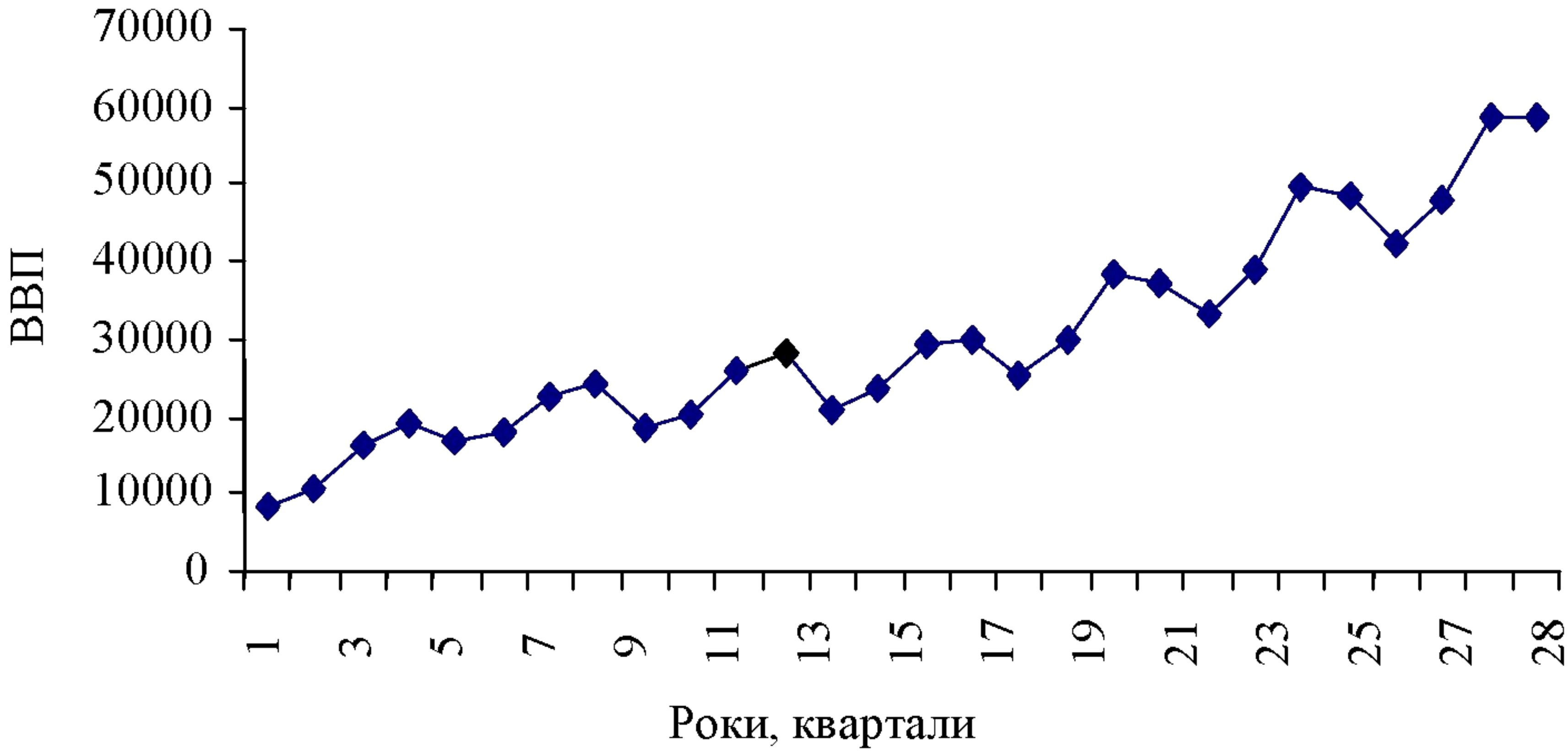
Ряд з лінійно-адитивним трендом

Сезонний ряд



Ряд з лінійно-мультиплікативним трендом

Динаміка попиту з лінійним трендом і сезонно-адитивною складовою



Динаміка ВВП із сезонно-мультиплікативним трендом

Дослідження стаціонарності часового ряду в EViews

3.13.2. Тест на наявність одиничних корнів – розширений тест Дікі-Фуллєра (augmented Dickey – Fuller test, ADF, або unit root test).

Часовий ряд має одиничний корінь, або порядок інтеграції один ($y(k) \sim I(1)$), якщо його перші різниці ($\Delta y(k)=y(k)-y(k-1)$) утворюють стаціонарний ряд ($\Delta y(k) \sim I(0)$).

ADF–тест дозволяє тестувати гіпотезу про наявність одиничних коренів у моделях, де кількість лагів більше одного.

ADF-статистики розраховується спочатку для ряду, потім для перших, других і т.д. різниць. Умова стаціонарності виконується, якщо значення ADF-статистики не перевищує відповідне критичне значення на встановленому рівні значущості. При цьому ряд, k-і різниці якого стаціонарні, називають інтегрованим рядом k-го порядку і позначають $I(k)$. Стаціонарний ряд позначають $I(0)$