

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний аерокосмічний університет ім. М.Є. Жуковського
«Харківський авіаційний інститут»

Б.І. Душин

**МЕТОДИ ЕКОНОМІКО-СТАТИСТИЧНИХ
ДОСЛІДЖЕНЬ.
РЕГРЕСІЙНИЙ АНАЛІЗ СТАЦІОНАРНИХ
ЧАСОВИХ РЯДІВ**

Навчальний посібник

Харків «ХАІ» 2006

УДК 519.23/24

Методи економіко-статистичних досліджень. Регресійний аналіз стаціонарних часових рядів / Б.І. Душин. - Навч. посібник. - Харків: Нац. аерокосм. ун-т «Харк. авіац. ін-т», 2006. - 51 с.

Розглянуто сучасні методи економіко-статистичних досліджень даних, поданих у вигляді часових рядів, які враховують можливу наявність у розглянутих змінних стохастичного тренду.

Основна увага приділена роз'ясненню базових понять і основних процедур регресійного аналізу стаціонарних часових рядів із залученням достатньої кількості числових прикладів, що пояснюють теоретичний матеріал.

Для магістрів, що вивчають курс “Методи економіко-статистичних досліджень”. Може також використовуватись при підготовці й проведенні практичних занять, виконанні домашніх і контрольних завдань, підготовці дипломної роботи, у дослідницькій діяльності економіко-управлінських служб підприємства.

Іл. 15. Табл. 4. Бібліогр.: 13 назв.

Рецензенти: д-р техн. наук О.Я. Мовшович,
канд. екон. наук В.М. Писін

За редакцією акад. ІАН України, д-ра техн. наук, проф. А.І. Бабушкіна

ЗМІСТ

Передмова	4
1. Стаціонарні часові ряди	5
1.1. Поняття стаціонарного часового ряду.	5
1.2. Процес «білого шуму»	8
1.3. Процес авторегресії	10
1.4. Процес авторегресії порядку p	17
1.5. Процес ковзного середнього	23
1.6. Змішаний процес авторегресії - ковзного середнього (про- цес авторегресії із залишками у вигляді ковзного середньо- го)	27
1.7. Моделі ARMA, що враховують наявність сезонності	29
2. Підбір стаціонарної моделі ARMA.	31
2.1. Основні етапи підбору моделі	31
2.2. Ідентифікація стаціонарної моделі ARMA	32
2.3. Використання інформаційних критеріїв для підбору мо- делі	40
2.4. Оцінювання коефіцієнтів моделі	45
2.5. Діагностика оціненої моделі	47
Бібліографічний список	50

ПЕРЕДМОВА

Цей навчальний посібник призначено для магістрів за фахом «Економіка підприємства». Мета посібника – розвиток і поглиблення знань, пов'язаних з економіко-статистичними дослідженнями, набутими студентами на попередніх курсах при вивченні дисциплін «Теорія ймовірностей», «Математична статистика», «Економетрика», «Економічна статистика». Передбачається, що, вивчаючи посібник, студенти вже мають базові знання з перерахованих дисциплін, а також уявлення про похідну й інтеграл, матриці й операції над ними.

Посібник дає введення в сучасні методи економіко-статистичних досліджень даних, які являють собою часові ряди. Враховується можлива наявність у розглядуваних змінних стохастичного тренду. Найбільша увага приділяється статистичним часовим рядам типу AR, MA, ARMA та підбору ARMA-моделі досліджуваного ряду.

Роз'яснюються базові поняття і основні процедури регресійного аналізу стаціонарних часових рядів із залученням достатньої кількості числових прикладів.

Посібник написано на підставі лекцій і джерел [1,5,13]. Усі чисельні приклади перераховані заново.

1. СТАЦІОНАРНІ ЧАСОВІ РЯДИ

1.1. Поняття стаціонарного часового ряду

Під **часовим рядом (time series)** розуміється послідовність спостережень за значеннями деякої змінної, зроблених через однакові проміжки часу.

Якщо прийняти довжину такого проміжку за одиницю часу (рік, квартал, день та ін.), то можна вважати, що послідовні спостереження x_1, \dots, x_n зроблено в моменти $t = 1, \dots, n$.

Основна відмітна риса статистичного аналізу часових рядів полягає в тому, що послідовність спостережень x_1, \dots, x_n розглядається як реалізація послідовності, загалом кажучи, статистично залежних випадкових величин X_1, \dots, X_n , які мають спільний розподіл з функцією розподілу

$$F(v_1, \dots, v_n) = P(X_1 < v_1, \dots, X_n < v_n),$$

де $P(X_1 < v_1, \dots, X_n < v_n)$ – ймовірність того, що $X_1 < v_1, \dots, X_n < v_n$ одночасно.

У посібнику розглядаються часові ряди, в яких спільний розподіл випадкових величин X_1, \dots, X_n має спільну щільність розподілу

$$p(v_1, \dots, v_n) = \frac{d^n F(v_1, \dots, v_n)}{dv_1 \dots dv_n}.$$

Щоб зробити задачу статистичного аналізу часових рядів доступною для практичного розв'язання, доводиться так чи інакше обмежувати клас розглядуваних моделей часових рядів, вводячи ті чи інші припущення щодо структури ряду і структури його ймовірнісних характеристик. Одне з таких обмежень припускає стаціонарність часового ряду.

Ряд x_t , $t = 1, \dots, n$, називається **строго стаціонарним** (або **стаціонарним у вузькому розумінні**), якщо для кожного m ($m < n$) спільний розподіл імовірностей випадкових величин X_{t_1}, \dots, X_{t_m} такий же, як і для $X_{t_1+\tau}, \dots, X_{t_m+\tau}$, при будь-яких t_1, \dots, t_m і τ , таких, що $1 < t_1, \dots, t_m < n$ і $1 < t_1 + \tau, \dots, t_m + \tau < n$.

Інакше кажучи, властивості строго стаціонарного часового ряду не змінюються при зміні початку відліку часу. Зокрема, при $m = 1$ з припущення про строгу стаціонарність часового ряду x_t випливає, що закон розподілу ймовірностей випадкової величини X_t не залежить від t , а отже, не залежать від t і всі його основні числові характерис-

тики (якщо, звичайно, вони існують), у тому числі математичне сподівання $E(X_t) = \mu$ і дисперсія $D(X_t) = \sigma^2$ ($D(X_t) = E([x_t - E(X_t)]^2)$).

Значення μ визначає сталий рівень, відносно якого коливається аналізований часовий ряд x_t , а стала σ^2 характеризує розмах цих коливань.

Одна з головних відмінностей послідовності спостережень, що утворюють часовий ряд, полягає в тому, що члени часового ряду є, загалом кажучи, статистично взаємозалежними. Ступінь тісноти статистичного зв'язку між випадковими величинами X_t і $X_{t+\tau}$ може бути виміряна парним коефіцієнтом кореляції

$$\text{Corr}(X_t, X_{t+\tau}) = \frac{\text{Cov}(X_t, X_{t+\tau})}{\sqrt{D(X_t)}\sqrt{D(X_{t+\tau})}},$$

де $\text{Cov}(X_t, X_{t+\tau}) = E[(X_t - E(X_t))(X_{t+\tau} - E(X_{t+\tau}))]$ – коваріація випадкових величин X_t і $X_{t+\tau}$.

Якщо ряд x_t – стаціонарний, то значення $\text{Cov}(X_t, X_{t+\tau})$ не залежить від t і є функцією тільки від τ .

Будемо використовувати позначення

$$\gamma(\tau) = \text{Cov}(X_t, X_{t+\tau}).$$

Зокрема,

$$D(X_t) = \text{Cov}(X_t, X_t) \equiv \gamma(0).$$

Відповідно для стаціонарного ряду значення коефіцієнта кореляції $\text{Corr}(X_t, X_{t+\tau})$ також залежить тільки від τ .

Будемо використовувати для нього позначення

$$\rho(\tau) = \text{Corr}(X_t, X_{t+\tau}) = \frac{\gamma(\tau)}{\gamma(0)}.$$

Зокрема, $\rho(0) = 1$.

Практична перевірка строгої стаціонарності ряду x_t на підставі спостереження за значеннями x_1, \dots, x_n у загальному випадку утруднена.

У зв'язку з цим під стаціонарним рядом на практиці часто мають на увазі часовий ряд x_t , у якого:

- 1) $E(X_t) \equiv \mu$;
- 2) $D(X_t) \equiv \sigma^2$;
- 3) $\text{Cov}(X_t, X_{t+\tau}) = \gamma(\tau)$ для будь-яких t і τ .

Ряд, для якого виконуються зазначені три умови, називають **стаціонарним у широкому розумінні (слабо стаціонарним, стаціонарним другого порядку або коваріаційно стаціонарним)**.

Якщо ряд є стаціонарним у широкому розумінні, то він не обов'язково є строго стаціонарним. Водночас і строго стаціонарний ряд може не бути стаціонарним у широкому розумінні просто тому, що в нього можуть не існувати математичне сподівання і/або дисперсія (прикладом може бути випадкова вибірка з розподілу Коші).

Ряд x_t , $t=1, \dots, n$, називається **гауссівським**, якщо спільний розподіл випадкових величин X_1, \dots, X_n є n -мірним нормальним розподілом. Для гауссівського ряду поняття стаціонарності у вузькому і широкому розумінні збігаються.

У подальшому під стаціонарністю деякого ряду x_t (якщо не оговорюється протилежне) передбачатиметься стаціонарність у широкому розумінні.

Отже, нехай x_t – стаціонарний ряд з $E(X_t) \equiv \mu$, $D(X_t) \equiv \sigma^2$ і $\rho(\tau) = \text{Corr}(X_t, X_{t+\tau})$.

Оскільки в цьому випадку коефіцієнт $\rho(\tau)$ вимірює кореляцію між членами одного і того самого часового ряду, його прийнято називати **коефіцієнтом автокореляції** (або просто **автокореляцією**).

З тієї ж причини про коваріації $\text{Cov}(X_t, X_{t+\tau}) = \gamma(\tau)$ говорять як про **автоковаріації**. При аналізі зміни величини $\rho(\tau)$ залежно від значення τ прийнято говорити про **автокореляційну функцію** $\rho(\tau)$. Автокореляційна функція безмірна, тобто не залежить від масштабу виміру аналізованого часового ряду. Її значення можуть змінюватися в межах від -1 до +1, при цьому $\rho(0) = 1$. Крім того, зі стаціонарності ряду x_t випливає, що $\rho(\tau) = \rho(-\tau)$, отже, при аналізі поведінки автокореляційних функцій звичайно обмежуються розглядом тільки невід'ємних значень τ .

Графік залежності $\rho(\tau)$ від τ часто називають **корелограмою**. Він може використовуватися для характеристики деяких властивостей процесу, що породжує часовий ряд. Для подальшого зазначимо, що якщо x_t – стаціонарний часовий ряд і c – деяка стала, то часові ряди x_t і $x_t + c$ мають однакові корелограми.

Якщо припустити, що часовий ряд описується моделлю стаціонарного гауссівського процесу, то повний опис спільного розподілу випадкових величин X_1, \dots, X_n потребує задання $n+1$ параметрів: μ , $\gamma(0)$, ..., $\gamma(n-1)$ (або μ , $\rho(0)$, ..., $\rho(n-1)$). Це набагато менше, ніж без вимоги стаціонарності, але все ж більше, ніж кількість спостере-

жень. У зв'язку з цим навіть для стаціонарних гауссівських часових рядів доводиться робити подальше спрощення моделі для того, щоб обмежити кількість параметрів, які підлягають оцінюванню за наявними спостереженнями.

Далі розглядаються прості за структурою часові ряди, які водночас корисні для опису еволюції в часі багатьох реальних економічних показників.

1.2. Процес «білого шуму»

Процесом «білого шуму» («білим шумом», «чисто випадковим часовим рядом») називають стаціонарний часовий ряд x_t , для якого:

- математичне сподівання $E(X_t) \equiv 0$;
- дисперсія $D(X_t) \equiv \sigma^2 > 0$;
- $\rho(\tau) = 0$ при $\tau \neq 0$.

Остання умова означає, що при $t \neq s$ випадкові величини x_t і x_s , які відповідають спостереженням за процесом «білого шуму» в моменти t і s , – не корельовані.

У випадку, коли X_t має нормальний розподіл, випадкові величини X_1, \dots, X_n взаємно незалежні та мають однаковий нормальний розподіл $N(0, \sigma^2)$, який створює випадкову вибірку з цього розподілу, тобто $X_t \sim i.i.d. N(0, \sigma^2)$.

Такий ряд називають **гауссівським «білим шумом»**.

Водночас у загальному випадку, навіть якщо деякі випадкові величини X_1, \dots, X_n взаємно незалежні та мають однаковий розподіл, це ще не означає, що вони утворюють процес «білого шуму», тому що випадкова величина X_t може просто не мати математичного сподівання і/або дисперсії (як приклад знову можна вказати на розподіл Коші).

Часовий ряд, який відповідає процесу «білого шуму», поводить себе надто нерегулярним чином через некорельованості при $t \neq s$ випадкових величин x_t і x_s . Це ілюструє графік на рис. 1.1. Тому процес «білого шуму» непридатний для безпосереднього моделювання еволюції більшості часових рядів, які мають місце в економіці.

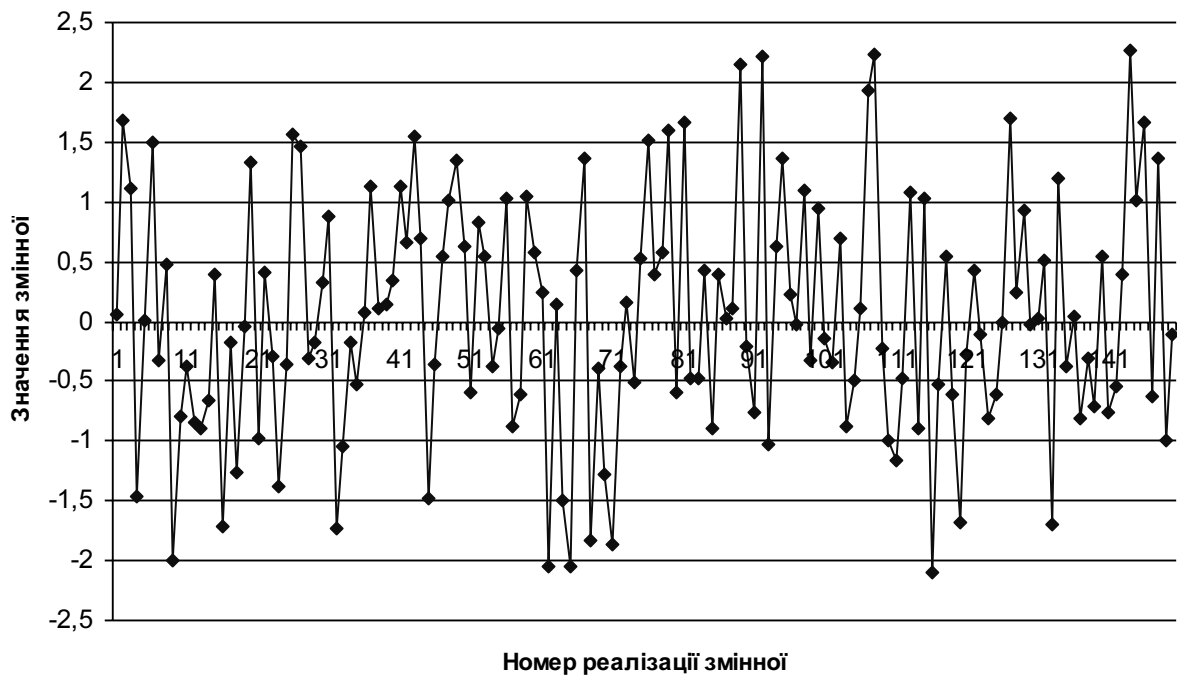


Рис. 1.1. Графік реалізації гауссівського процесу “білого шуму” при $\sigma^2 = 1$

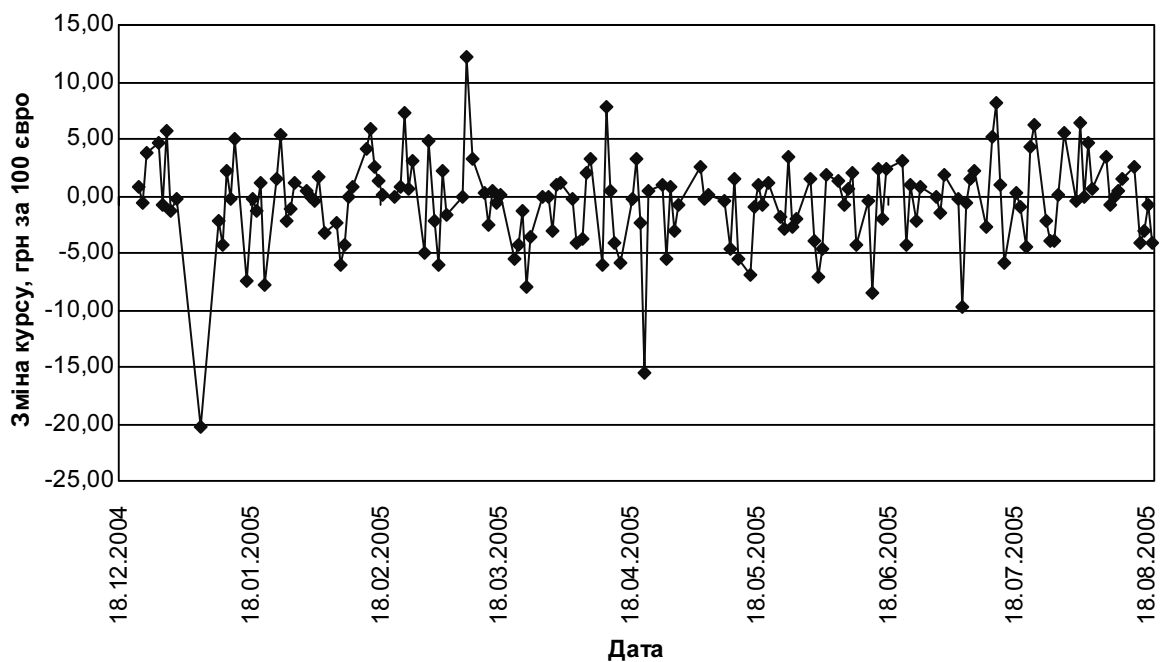


Рис. 1.2. Графік динаміки абсолютних змін курсу гривні відносно євро за перші 8 місяців 2005 року

Водночас такий процес є базою для побудови більш реалістичних моделей часових рядів, які породжують “більш гладкі” траєкторії ряду.

У зв'язку з частим використанням процесу «білого шуму» в подальшому будемо відрізняти цей процес від інших моделей часових рядів, використовуючи для нього позначення ε_t .

Як приклад ряду, траєкторія якого схожа на реалізацію процесу «білого шуму», можна вказати на ряд, утворений значеннями абсолютних змін курсу євро відносно гривні (рис. 1.2).

Неважко помітити, що в цьому ряді спостерігається асиметрія розподілу ймовірностей значень x_t (скісність цього розподілу вбік від'ємних значень). Це виключає опис моделі цього ряду як гауссівського «білого шуму».

1.3. Процес авторегресії

Однією з найвикористовуваніших моделей часових рядів є **процес авторегресії (модель авторегресії)**. У своїй найпростішій формі модель авторегресії (у короткому позначенні – AR(1)) описує механізм породження ряду в такий спосіб:

$$X_t = aX_{t-1} + \varepsilon_t, \quad t = 1, \dots, n,$$

де ε_t – процес «білого шуму», що має нульове математичне сподівання й дисперсію σ^2 ;

X_0 – деяка випадкова величина;

$a \neq 0$ – деякий сталий коефіцієнт.

При цьому $E(X_t) = aE(X_{t-1})$, отже, розглядуваний процес може бути стаціонарним, тільки якщо $E(X_t) = 0$ для всіх $t = 0, 1, \dots, n$.

Модель авторегресії дозволяє виразити X_t у такий спосіб:

$$X_1 = aX_0 + \varepsilon_1;$$

$$X_2 = aX_1 + \varepsilon_2 = a(aX_0 + \varepsilon_1) + \varepsilon_2 = a^2X_0 + a\varepsilon_1 + \varepsilon_2;$$

... ..;

$$X_t = aX_{t-1} + \varepsilon_t = \dots = a^tX_0 + a^{t-1}\varepsilon_1 + a^{t-2}\varepsilon_2 + \dots + \varepsilon_t.$$

Якщо випадкова величина X_0 не корельована з випадковими величинами $\varepsilon_1, \dots, \varepsilon_n$, то

$$Cov(X_0, \varepsilon_1) = 0, \quad Cov(X_1, \varepsilon_2) = 0, \quad \dots, \quad Cov(X_{t-1}, \varepsilon_t) = 0$$

і

$$D(X_t) = D(aX_{t-1} + \varepsilon_t) = a^2D(X_{t-1}) + D(\varepsilon_t), \quad t = 1, \dots, n.$$

Вважаючи, що

$$D(X_0) = D(X_t) = \sigma_X^2, \quad t = 1, \dots, n,$$

знаходимо $\sigma_X^2 = a^2 \sigma_X^2 + \sigma_\varepsilon^2$.

Остання рівність може виконуватися тільки при виконанні умови $a^2 < 1$, тобто $|a| < 1$.

$$\text{Отже, } \sigma_X^2 = \frac{\sigma_\varepsilon^2}{1 - a^2}.$$

Що стосується автоковаріацій і автокореляцій, то

$$\begin{aligned} \text{Cov}(X_t, X_{t+\tau}) &= \text{Cov}(a^t X_0 + a^{t-1} \varepsilon_1 + a^{t-2} \varepsilon_2 + \dots + \varepsilon_t, \\ &\quad a^{t+\tau} X_0 + a^{t+\tau-1} \varepsilon_1 + a^{t+\tau-2} \varepsilon_2 + \dots + \varepsilon_{t+\tau}) = \\ &= a^{2t+\tau} D(X_0) + a^\tau (1 + a^2 + \dots + a^{2(t-1)}) \sigma_\varepsilon^2 = \\ &= a^\tau \left[a^{2t} \frac{\sigma_\varepsilon^2}{1 - a^2} + (1 - a^{2t}) \frac{\sigma_\varepsilon^2}{1 - a^2} \right] = a^\tau \frac{\sigma_\varepsilon^2}{1 - a^2} \end{aligned}$$

(математичне сподівання добутку незалежних випадкових величин дорівнює добутку математичних очікувань), звідки

$$\text{Corr}(X_t, X_{t+\tau}) = a^\tau,$$

тобто при зроблених припущеннях автоковаріації й автокореляції залежать тільки від того, наскільки рознесені в часі відповідні спостереження.

Таким чином, механізм породження послідовних спостережень, заданий співвідношеннями $X_t = aX_{t-1} + \varepsilon_t$, $t = 1, \dots, n$, породжує стаціонарний часовий ряд, якщо:

- $|a| < 1$;
- випадкова величина X_0 не корельована з випадковими величинами $\varepsilon_1, \dots, \varepsilon_n$;
- $E(X_0) = 0$;
- $D(X_0) = \frac{\sigma_\varepsilon^2}{1 - a^2}$.

$$\text{Отже, } \text{Corr}(X_t, X_{t+\tau}) = \rho(\tau) = a^\tau.$$

Розглядувана модель породжує (при зазначених умовах) стаціонарний ряд, що має нульове математичне сподівання. Однак її можна легко поширити і на часові ряди y_t з ненульовим математичним сподіванням $E(Y_t) = \mu$, вважаючи, що зазначена модель належить до центрованого ряду $X_t = Y_t - \mu$:

$$Y_t - \mu = a(Y_{t-1} - \mu) + \varepsilon_t, \quad t = 1, \dots, n,$$

отже,

$$Y_t = aY_{t-1} + \delta + \varepsilon_t, \quad t = 1, \dots, n,$$

де $\delta = \frac{\mu}{1-a}$.

Тому без обмеження спільності можна обмежитися в поточному розгляді моделями авторегресії, які породжують стаціонарний процес з нульовим середнім.

Продовжуючи розгляд моделі для раніше визначеного процесу X_t (з нульовим математичним сподіванням), зазначимо, що для нього

$$\gamma(1) = E(X_t \cdot X_{t-1}) = E((aX_{t-1} + \varepsilon_t) \cdot X_{t-1}) = a\gamma(0),$$

тому

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)},$$

і при значеннях $a > 0$, близьких до 1, між сусідніми спостереженнями є сильна додатна кореляція, яка забезпечує “більш гладкий” характер поведіння траєкторій ряду порівняно з процесом «білого шуму». При $a < 0$ процес авторегресії, навпаки, має “менш гладкі” реалізації, оскільки в цьому випадку проявляється тенденція чергування знаків послідовних спостережень.

Графік на рис. 1.3 показує поведінку змодельованої реалізації часового ряду, породженого моделлю авторегресії $X_t = aX_{t-1} + \varepsilon_t$ при $X_0 = 0$, $a = 0,8$ і $\sigma_\varepsilon^2 = 1$, графік на рис. 1.4 – при $X_0 = 0$, $a = -0,8$ і $\sigma_\varepsilon^2 = 1$. При побудові цих графіків використано одні й ті ж значення «білого шуму», подані на рис. 1.1.

При розгляді авторегресії потрібно звернути увагу на таку важливу обставину. На практиці “стартове” значення $X_0 = x_0$, на основі якого відповідно до співвідношення $X_t = aX_{t-1} + \varepsilon_t$ будуються наступні значення ряду X_t , може належати кінцю попереднього періоду, на якому просто через інші економічні умови еволюція відповідного економічного показника потребує іншої моделі, наприклад моделі $X_t = aX_{t-1} + \varepsilon_t$ з іншими значеннями a та σ_ε^2 .

Більше того, статистичні дані про поведінку ряду до моменту $t = 0$ можуть зовсім бути відсутніми, так що значення x_0 є просто деякою спостережуваною числовою величиною. В обох випадках ряд X_t уже не буде стаціонарним навіть при $a < 1$.

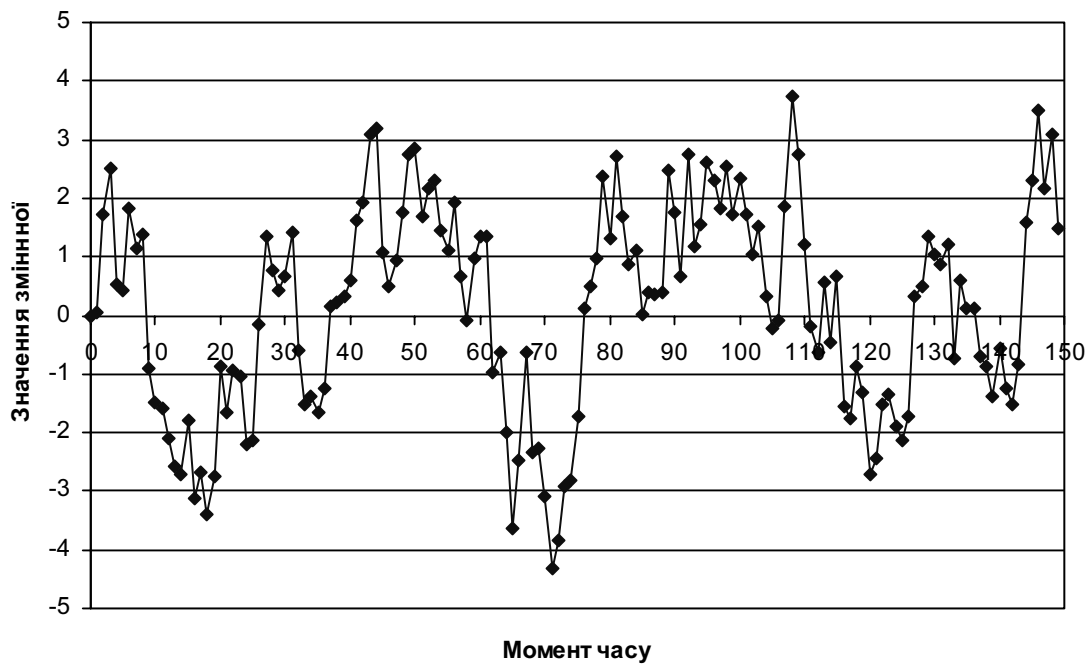


Рис. 1.3. Приклад поведження авторегресії при $x_0 = 0$, $a = 0,8$ і $\sigma_\varepsilon^2 = 1$

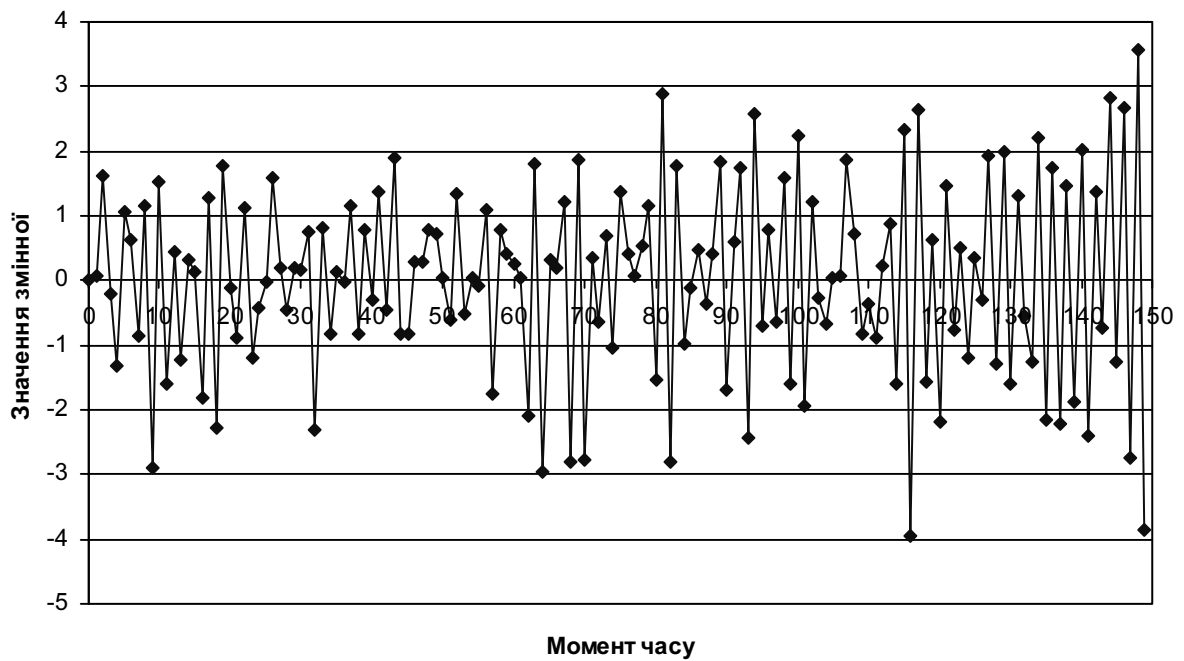


Рис. 1.4. Приклад поведження авторегресії при $x_0 = 0$, $a = -0,8$ і $\sigma_\varepsilon^2 = 1$

Розглянемо докладніше характеристики і поведження ряду в таких ситуаціях.

Якщо не конкретизувати модель, відповідно до якої породжувалися спостереження до моменту $t = 1$, то значення x_0 можна розглядати як фіксоване. При цьому

$$X_t = a^t x_0 + a^{t-1} \varepsilon_1 + a^{t-2} \varepsilon_2 + \dots + \varepsilon_t,$$

$$E(X_t) = a^t x_0 + a^{t-1} E(\varepsilon_1) + a^{t-2} E(\varepsilon_2) + \dots + E(\varepsilon_t) = a^t x_0,$$

$$D(X_t) = (a^{2(t-1)} + a^{2(t-2)} + \dots + 1) \sigma_\varepsilon^2 = \frac{(1 - a^{2t})}{(1 - a^2)} \sigma_\varepsilon^2 =$$

$$= \frac{\sigma_\varepsilon^2}{(1 - a^2)} - \frac{a^{2t} \sigma_\varepsilon^2}{(1 - a^2)},$$

$$\text{Cov}(X_t, X_{t+\tau}) = \text{Cov}(X_t - a^t x_0, X_{t+\tau} - a^{t+\tau} x_0) =$$

$$= a^\tau (1 + a^2 + \dots + a^{2(t-1)}) \sigma_\varepsilon^2 =$$

$$= a^\tau (1 - a^{2t}) \frac{\sigma_\varepsilon^2}{1 - a^2},$$

тому і математичне сподівання, і дисперсія випадкової величини X_t , а також коваріації $\text{Cov}(X_t, X_{t+\tau})$ залежать від t .

Водночас, якщо $a < 1$, то при $t \rightarrow \infty$ одержимо

$$E(X_t) \rightarrow 0,$$

$$D(X_t) \rightarrow \frac{\sigma_\varepsilon^2}{(1 - a^2)},$$

$$\text{Cov}(X_t, X_{t+\tau}) \rightarrow a^\tau \frac{\sigma_\varepsilon^2}{1 - a^2},$$

тобто при $t \rightarrow \infty$ значення математичного сподівання і дисперсії випадкової величини X_t , а також автоковаріації $\text{Cov}(X_t, X_{t+\tau})$ стабілізуються, наближаючись до своїх граничних значень.

Таким чином, умову $a < 1$ можна трактувати як умову стабільності ряду, що породжується моделлю $X_t = aX_{t-1} + \varepsilon_t$ при фіксованому значенні $X_0 = x_0$.

Розглянемо в цій ситуації поряд з тільки що дослідженим рядом

$$X_t = a^t x_0 + \sum_{k=0}^{t-1} a^k \varepsilon_{t-k}, \quad |a| < 1,$$

ряд, що породжується моделлю

$$\tilde{X}_t = \sum_{k=0}^{\infty} a^k \varepsilon_{t-k}, \quad |a| < 1.$$

Звідси маємо

$$\tilde{X}_t - X_t = -a^t x_0 + \sum_{k=t}^{\infty} a^k \varepsilon_{t-k}, \quad |a| < 1;$$

при $t \rightarrow \infty$

$$a^t x_0 \rightarrow 0$$

і

$$E \left| \sum_{k=t}^{\infty} a^k \varepsilon_{t-k} \right|^2 = \sigma_{\varepsilon}^2 \sum_{k=t}^{\infty} a^{2k} \rightarrow 0.$$

Таким чином, ряд \tilde{X}_t є граничним для X_t ; ряд X_t “виходить на режим” \tilde{X}_t при $t \rightarrow \infty$. При цьому вихід ряду X_t на режим \tilde{X}_t відбувається тим швидше, чим ближче X_0 і a до нуля.

Для ряду \tilde{X}_t

$$E(\tilde{X}_t) = E\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k}\right) = \sum_{k=0}^{\infty} a^k E(\varepsilon_{t-k}) = 0,$$

$$D(\tilde{X}_t) = D\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k}\right) = \sum_{k=0}^{\infty} a^{2k} D(\varepsilon_{t-k}) = \sigma_{\varepsilon}^2 \sum_{k=0}^{\infty} a^{2k} = \frac{\sigma_{\varepsilon}^2}{1-a^2},$$

$$\begin{aligned} \text{Cov}(\tilde{X}_t, \tilde{X}_{t+\tau}) &= E\left[\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t-k}\right)\left(\sum_{k=0}^{\infty} a^k \varepsilon_{t+\tau-k}\right)\right] = \\ &= a^{\tau} \left(\sum_{k=0}^{\infty} a^{2k} E(\varepsilon_{t-k}^2)\right) = a^{\tau} \frac{\sigma_{\varepsilon}^2}{1-a^2}, \end{aligned}$$

отже, \tilde{X}_t – стаціонарний ряд (у широкому розумінні). Крім того,

$$\tilde{X}_{t-1} = \frac{1}{a} \sum_{k=0}^{\infty} a^k \varepsilon_{t-k},$$

звідки маємо

$$a\tilde{X}_{t-1} + \varepsilon_t = \sum_{k=0}^{\infty} a^k \varepsilon_{t-k} = \tilde{X}_t,$$

тобто \tilde{X}_t задовольняє співвідношення

$$\tilde{X}_t = a\tilde{X}_{t-1} + \varepsilon_t.$$

Оскільки ε_t не входить до правої частини виразів для $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots$, то випадкова величина ε_t не корельована з $\tilde{X}_{t-1}, \tilde{X}_{t-2}, \dots$, тобто ε_t є інновацією (оновленням).

У підсумку одержимо, що \tilde{X}_t – стаціонарний процес авторегресії першого порядку, і фактично саме цей процес мається на увазі, коли говорять про стаціонарний процес AR(1).

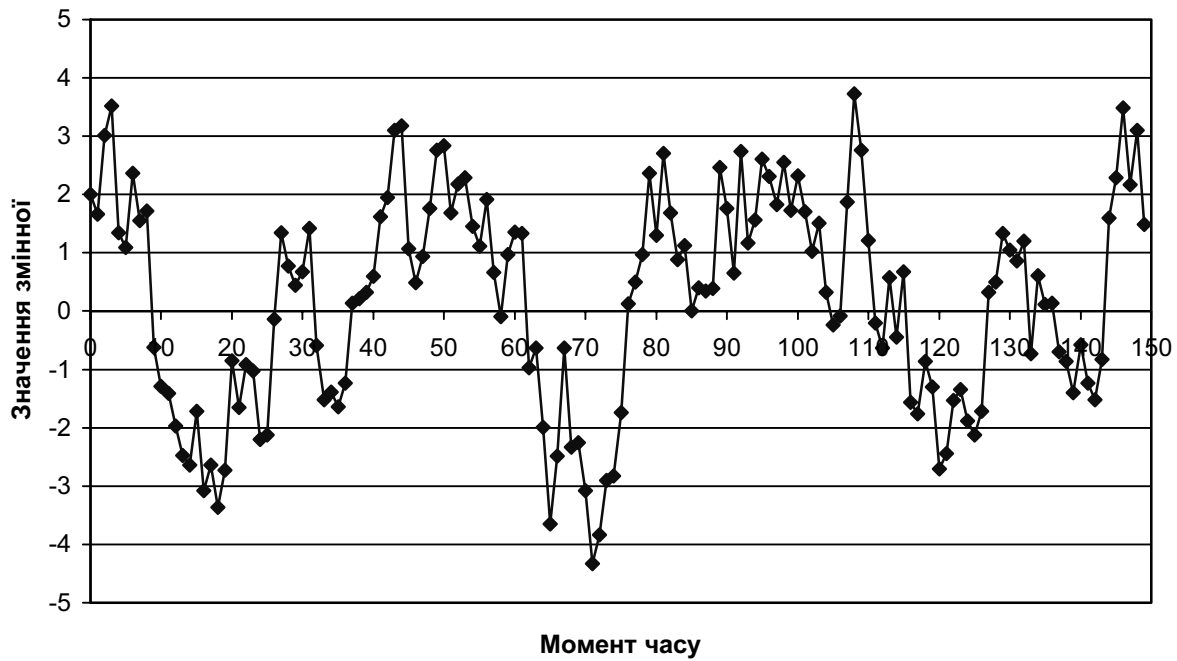


Рис. 1.5. Приклад поведження авторегресії при $x_0 = 2$, $a = 0,8$ і $\sigma_\varepsilon^2 = 1$

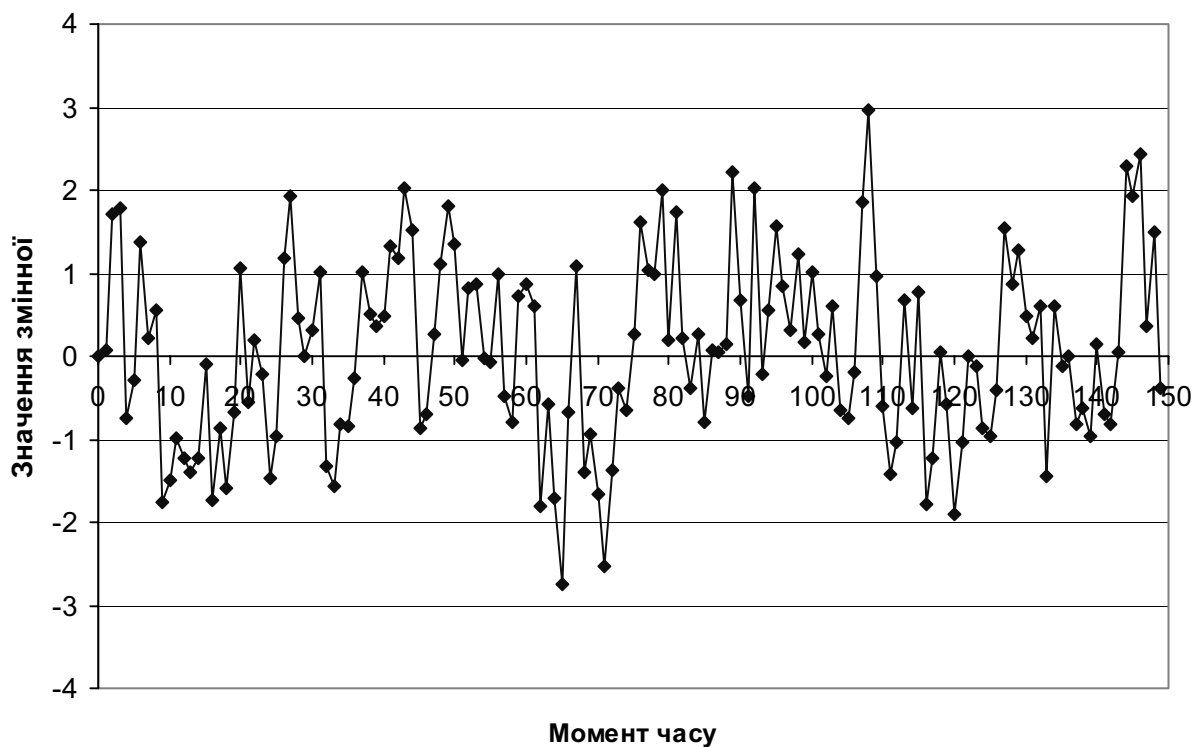


Рис. 1.6. Приклад поведження авторегресії при $x_0 = 0$, $a = 0,4$ і $\sigma_\varepsilon^2 = 1$

Ілюстраціями цього є змодельовані реалізації ряду x_t , породжені моделлю $X_t = aX_{t-1} + \varepsilon_t$ з $\sigma_\varepsilon = 1$ і різними значеннями коефіцієнта a і “стартового” значення x_0 (рис. 1.5, 1.6). При побудові цих графіків використано одні й ті ж значення «білого шуму», які подані на рис. 1.1.

Графік на рис. 1.5 показує, що зміна початкового значення $x_0 = 0$ на $x_0 = 2$ перестає позначатися, починаючи приблизно з 10-го моменту часу (порівняно з графіком на рис. 1.3). Графік на рис. 1.6 показує, як зменшення вдвічі параметра кореляції a знижує вплив попередніх значень і підсилює вплив «білого шуму».

1.4. Процес авторегресії порядку p

Розглянуту в попередньому підрозділі модель $X_t = aX_{t-1} + \varepsilon_t$ називають процесом авторегресії першого порядку.

Процес авторегресії порядку p (у короткому позначенні – AR(p)) визначається співвідношеннями

$$X_t = a_1X_{t-1} + a_2X_{t-2} + \dots + a_pX_{t-p} + \varepsilon_t, \quad a_p \neq 0,$$

де ε_t – процес «білого шуму» з $D(\varepsilon_t) = \sigma_\varepsilon^2$. Для простоти будемо тепер відразу вважати, що $Cov(X_{t-s}, \varepsilon_t) = 0$ для всіх $s > 0$.

Вважається, що випадкові величини ε_t утворюють **інноваційну (оновлюючу) послідовність**. Випадкова величина ε_t називається **інновацією** для спостереження в момент t .

Така термінологія пояснюється тим, що спостережуване значення ряду в момент t визначається як лінійна комбінація p попередніх значень цього ряду плюс не корельована з цими попередніми значеннями випадкова складова ε_t , що відбиває оновлену інформацію, наприклад, про стан економіки, на момент t , що впливає на спостережуване значення X_t .

При розгляді процесів авторегресії та деяких інших моделей зручно використовувати **оператор запізнювання L (lag operator)**, що впливає на часовий ряд і визначається співвідношенням

$$LX_t = X_{t-1}.$$

У деяких посібниках його називають оператором зворотного зсуву і використовують для нього позначення B (*backshift operator*).

Якщо оператор запізнювання застосовується k раз, що позначається як L^k , то в результаті це дає

$$L^k X_t = X_{t-k}.$$

Вираз

$$a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p}$$

можна записати тепер у вигляді

$$(a_1 L + a_2 L^2 + \dots + a_p L^p) X_t,$$

а співвідношення, що визначає процес авторегресії p -го порядку, – у вигляді

$$a(L) X_t = \varepsilon_t,$$

де

$$a(L) = 1 - (a_1 L + a_2 L^2 + \dots + a_p L^p).$$

Для того, щоб такий процес був стаціонарним, усі корені алгебричного рівняння

$$a(z) = 0$$

(дійсні та комплексні) повинні знаходитись поза одиничним колом $|z| \leq 1$ (зокрема, для процесу AR(1) маємо $a(z) = 1 - az$; рівняння

$a(z) = 0$ має корінь $z = \frac{1}{a}$, умова стаціонарності $z > 1$ рівносильна вже знайомій умові $|a| < 1$).

При цьому розв'язок рівняння $a(L) X_t = \varepsilon_t$ можна подати у вигляді

$$X_t = \frac{1}{a(L)} \varepsilon_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j}, \text{ де } \sum_{j=0}^{\infty} |b_j| < \infty,$$

звідки одержимо, що

$$E(X_t) = E\left(\sum_{j=0}^{\infty} b_j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} b_j E(\varepsilon_{t-j}) = 0.$$

Стаціонарний процес AR(p) з ненульовим математичним сподіванням μ задовольняє співвідношення

$$a(L)(X_t - \mu) = \varepsilon_t,$$

або

$$a(L) X_t = \delta + \varepsilon_t,$$

де

$$\delta = a(L)\mu = \mu(1 - a_1 - a_2 \dots - a_p) = \mu a(1).$$

Отже, розв'язок рівняння $a(L)(X_t - \mu) = \varepsilon_t$ має вигляд

$$X_t = \mu + \frac{1}{a(L)} \varepsilon_t.$$

Таким чином, якщо стаціонарний процес AR(p) задано у вигляді $a(L)X_t = \delta + \varepsilon_t$, то варто пам'ятати про те, що математичне сподівання цього процесу дорівнює не δ , а

$$\mu = \frac{1}{(1 - a_1 - a_2 \dots - a_p)}.$$

Для процесу AR(1) маємо $a(L) = 1 - aL$, отже (незалежно від того, дорівнює μ нулю чи ні),

$$X_t - \mu = \frac{1}{1 - aL} \varepsilon_t = (1 + aL + a^2L^2 + \dots)\varepsilon_t = \varepsilon_t + a\varepsilon_{t-1} + a^2\varepsilon_{t-2} + \dots$$

З останнього виразу відразу видно, що

$$\rho(k) = \text{Corr}(X_t, X_{t+k}) = a^k.$$

При $0 < a < 1$ корелограма (графік функції $\rho(k)$ для $k = 0, 1, 2, \dots$) відбиває показове убуття кореляцій зі зростанням інтервалу між спостереженнями; при $-1 < a < 0$ корелограма має характер згасаючої косінусоїди.

Приклади корелограм для простої авторегресії наведено на рис. 1.7 і 1.8.

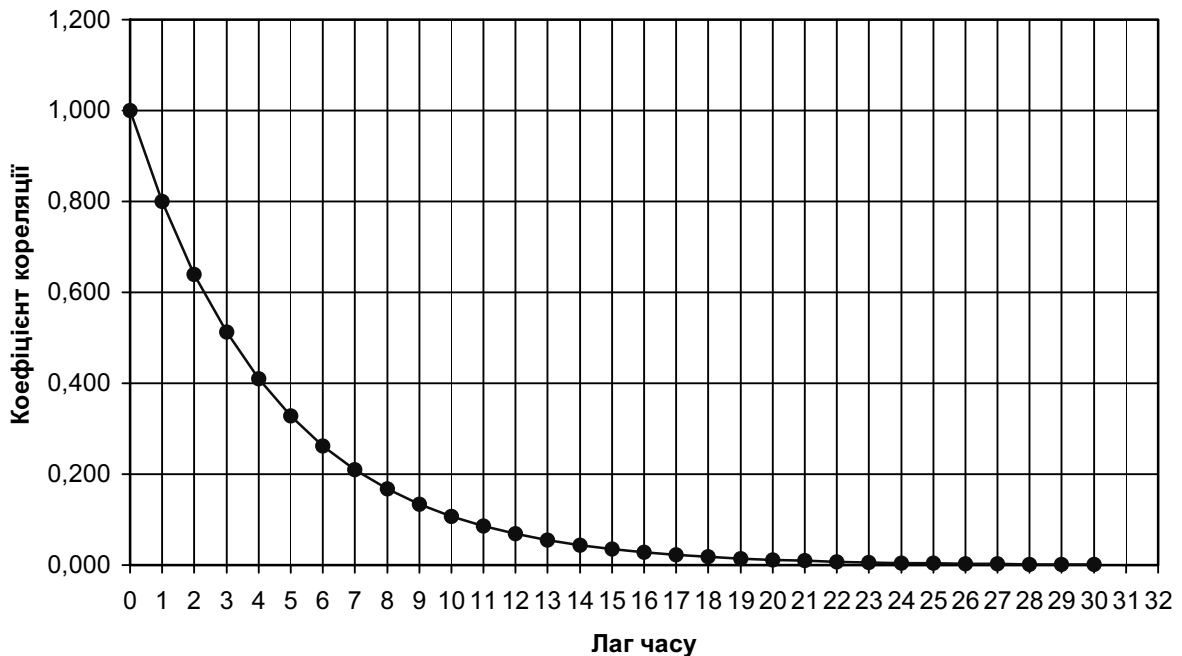


Рис. 1.7. Приклад корелограми простої авторегресії при $a = 0,8$

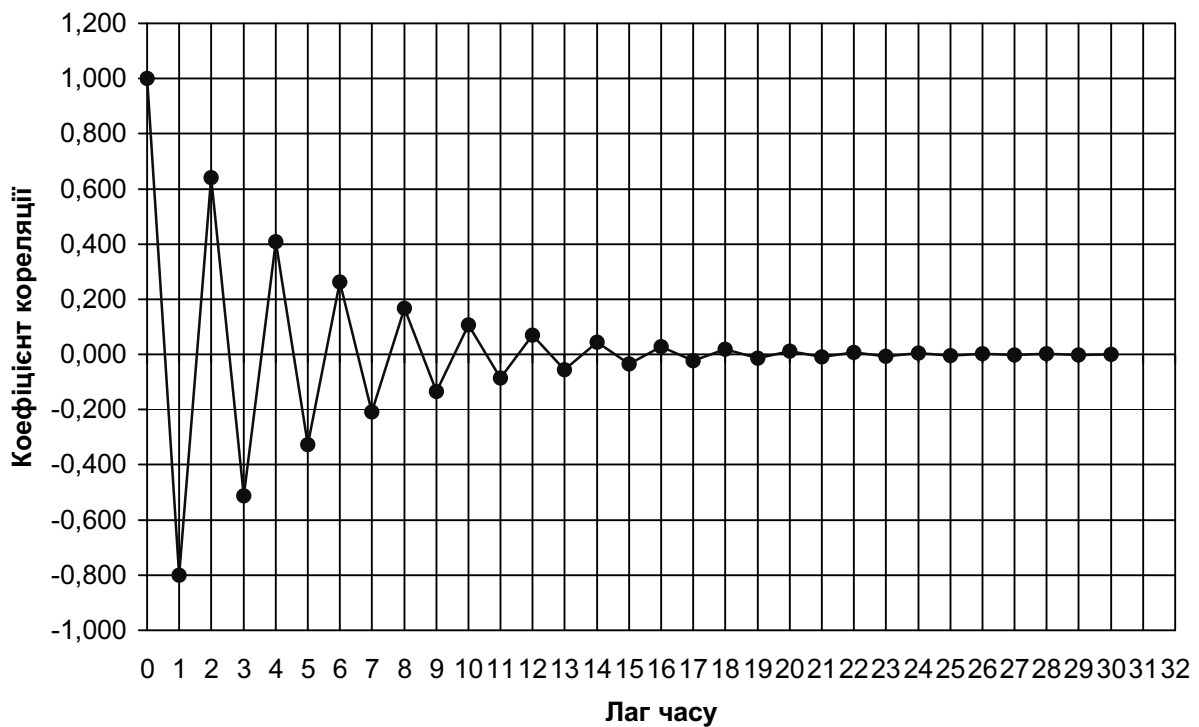


Рис. 1.8. Приклад корелограми простої авторегресії при $a = -0,8$

Корелограма процесу AR(p) при $p > 1$ має більш складну форму, що залежить від розміщення (на комплексній площині) коренів рівняння $a(z) = 0$.

Однак для великих значень k автокореляція $\rho(k)$ добре апроксимується значенням $A\theta^k$, де $\theta = \frac{1}{z_{\min}}$, z_{\min} – найменший за абсолютною величиною корінь рівняння $a(z) = 0$, якщо цей корінь є дійсним і додатним, або знаходиться в інтервалі $\pm A|\theta^k|$ в протилежному випадку, де $A > 0$ – деяка стала, зумовлена коефіцієнтами a_1, \dots, a_p .

Якщо помножити на X_{t-k} ($k > 0$) обидві частини співвідношення, що визначає процес AR(p), і після цього визначити для обох частин математичне сподівання, то одержимо співвідношення

$$\gamma(k) = a_1\gamma(k-1) + a_2\gamma(k-2) + \dots + a_p\gamma(k-p), \quad k > 0.$$

Розділивши обидві частини останнього на $\gamma(0)$, отримаємо **систему рівнянь Юла – Уокера**

$$\rho(k) = a_1\rho(k-1) + a_2\rho(k-2) + \dots + a_p\rho(k-p), \quad k > 0.$$

Ця система дозволяє послідовно знаходити значення автокореляцій і дає можливість, використовуючи перші p рівнянь, виразити коефіцієнти a_j через значення перших p автокореляцій, що можна безпосередньо використовувати при підборі моделі авторегресії до реальних статистичних даних.

Приклад 1.1. Розглянемо процес авторегресії AR(2)

$$X_t = 4,375 + 0,25 \cdot X_{t-1} - 0,125 \cdot X_{t-2} + \varepsilon_t .$$

У цьому випадку рівняння $a(z) = 0$ набирає вигляду

$$1 - 0,25 \cdot z + 0,125 \cdot z^2 = 0 \quad \text{або} \quad z^2 - 2z + 8 = 0 .$$

Корені цього рівняння $z_{1,2} = 1 \pm i\sqrt{7}$. Обидва корені за абсолютною величиною більше одиниці, отже, процес стаціонарний. Математичне сподівання цього процесу дорівнює

$$\mu = \frac{\delta}{1 - \alpha_1 - \alpha_2} = \frac{4,375}{1 - 0,25 - 0,125} = 5 ,$$

отже, траєкторії флюктуують навколо рівня 5.

Для побудови корелограми скористаємося рівняннями Юла – Уокера. У цьому випадку $p = 2$, отже,

$$\rho(k) = 0,25\rho(k-1) - 0,125\rho(k-2), \quad k > 0 .$$

За визначенням $\rho(0) = 1$. Для $\rho(1)$ маємо

$$\rho(1) = 0,25\rho(0) - 0,125\rho(-1) = 0,25 - 0,125\rho(1) ,$$

звідки

$$\rho(1) = \frac{0,25}{1 - 0,125} = 0,222 .$$

Далі послідовно знаходимо:

$$\rho(2) = 0,25\rho(1) - 0,125\rho(0) = 0,25 \cdot 0,222 - 0,125 = -0,069 ;$$

$$\rho(3) = 0,25\rho(2) - 0,125\rho(1) = 0,25 \cdot (-0,069) - 0,125 \cdot 0,222 = -0,045 ;$$

$$\rho(4) = -0,003 ; \quad \rho(5) = 0,005 .$$

Кореляції, навіть між сусідніми спостереженнями, малі, і можна сподіватися, що поведження траєкторій такого ряду не дуже істотно відрізняється від поведження реалізацій процесу «білого шуму». Корелограму розглядуваного процесу і змодельовану його реалізацію наведено на рис. 1.8 і 1.9.

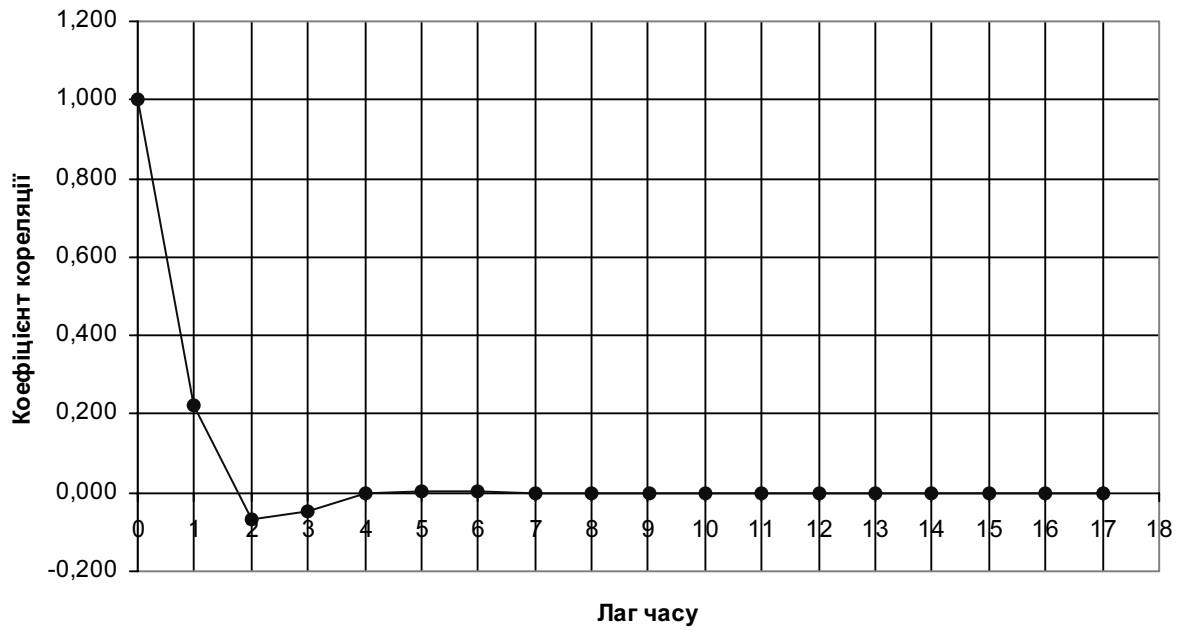


Рис. 1.9. Корелограма для процесу
 $X_t = 4,375 + 0,25 \cdot X_{t-1} - 0,125 \cdot X_{t-2} + \varepsilon_t$

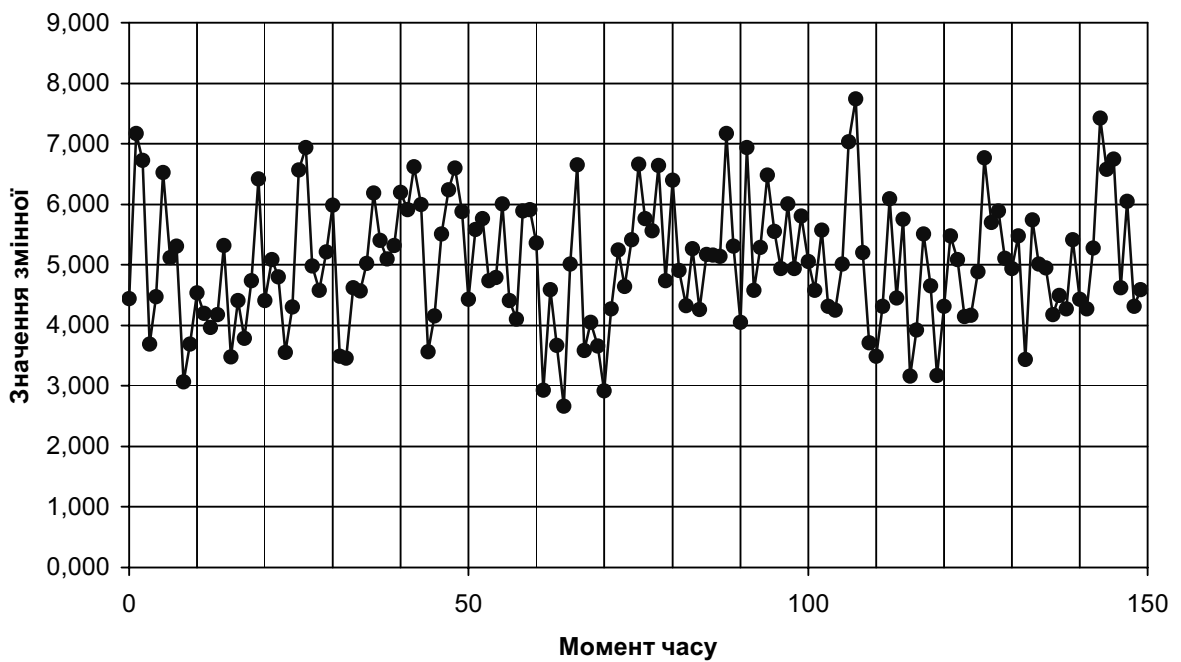


Рис. 1.10. Реалізація процесу $X_t = 4,375 + 0,25 \cdot X_{t-1} - 0,125 \cdot X_{t-2} + \varepsilon_t$

1.5. Процес ковзного середнього

Ще однією простою моделлю породження часового ряду є процес **ковзного середнього порядку q** .

Відповідно до цієї моделі

$$X_t = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q}, \quad b_q \neq 0,$$

де ε_t – процес «білого шуму».

Такий процес має нульове математичне сподівання. Модель можна узагальнити до процесу, що має ненульове математичне сподівання μ , покладаючи

$$X_t - \mu = \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q}, \quad b_q \neq 0,$$

тобто

$$X_t = \mu + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q}, \quad b_q \neq 0.$$

Для процесу ковзного середнього порядку q використовується позначення $MA(q)$ (ковзне середнє – **moving average**).

При $q = 0$ і $\mu = 0$ одержимо процес «білого шуму». Якщо $q = 1$, то $X_t - \mu = \varepsilon_t + b_1\varepsilon_{t-1}$ – **ковзне середнє першого порядку**. В останньому випадку

$$D(X_t) = (1 + b^2)\sigma_\varepsilon^2,$$

$$E[(X_t - \mu)(X_{t-1} - \mu)] = b\sigma_\varepsilon^2,$$

$$E[(X_t - \mu)(X_{t-k} - \mu)] = 0, \quad k > 1,$$

отже, процес X_t є стаціонарним з $E(X_t) = 0$,

$$D(X_t) = (1 + b^2)\sigma_\varepsilon^2,$$

$$\gamma(k) = \begin{cases} (1 + b^2)\sigma_\varepsilon^2, & k = 0, \\ b\sigma_\varepsilon^2, & k = 1, \\ 0, & k > 1. \end{cases}$$

Автокореляції цього процесу дорівнюють

$$\rho(k) = \begin{cases} 1, & k = 0, \\ \frac{b}{(1 + b^2)}, & k = 1, \\ 0, & k > 1, \end{cases}$$

тобто корелограма процесу має досить специфічний вигляд.

Корельованими виявляються тільки сусідні спостереження. Кореляція між ними додатна, якщо $b > 0$, і від'ємна при $b < 0$. Відповідно процес $MA(1)$ з $b > 0$ має «більш гладкі», порівняно з «білим шумом»,

реалізації, а процес MA(1) з $b < 0$ має “менш гладкі”, порівняно з “білим шумом”, реалізації. Зазначимо, що для будь-якого процесу MA(1)

$$|\rho(1)| \leq 0,5,$$

тобто кореляційний зв'язок між сусідніми спостереженнями невеликий, тоді як у процесу AR(1) такий зв'язок може бути як завгодно сильним (при значеннях $|a|$, близьких до 1).

Модель MA(q) коротко можна записати у вигляді

$$X_t - \mu = b(L)\varepsilon_t,$$

де

$$b(L) = 1 + b_1L + b_2L^2 + \dots + b_qL^q).$$

Для неї

$$\gamma(k) = E[(X_t - \mu)(X_{t-k} - \mu)] = \begin{cases} \left(\sum_{j=0}^{q-k} b_j b_{j+k} \right) \sigma_\varepsilon^2, & 0 \leq k \leq q, \\ 0, & k > q, \end{cases}$$

отже, MA(q) є стаціонарним процесом з нульовим математичним сподіванням, дисперсією

$$\sigma_X^2 = (1 + b_1^2 + \dots + b_q^2) \sigma_\varepsilon^2$$

і автокореляціями

$$\rho(k) = \begin{cases} \frac{\left(\sum_{j=0}^{q-k} b_j b_{j+k} \right)}{\left(\sum_{j=0}^{q-k} b_j^2 \right)}, & k = 0, 1, \dots, q, \\ \frac{b}{(1 + b^2)}, & k = q + 1, q + 2, \dots \end{cases}$$

Тут статистичний зв'язок між спостереженнями зберігається протягом q одиниць часу (тобто “тривалість пам'яті” процесу дорівнює q).

Такого роду часові ряди відповідають ситуації, коли деякий економічний показник перебуває в рівновазі, але відхиляється від положення рівноваги через послідовно виникаючі непередбачені події, причому система така, що вплив цих подій відзначається протягом деякого періоду часу.

Якщо вплив минулих подій слабшає з часом показовим чином, і тому $b_j = a^j$, $0 < a < 1$, то штучне припущення про те, що ряд ε_t починається в “нескінченному минулому”, приводить до моделі нескінченного ковзного середнього MA(∞)

$$X_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j} = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j},$$

де $\sum_{j=0}^{\infty} |b_j| < \infty$.

Раніше було показано, що таке ж подання допускає стаціонарний процес авторегресії першого порядку AR(1)

$$X_t = aX_{t-1} + \varepsilon_t, \quad |a| < 1,$$

тобто у розглядуваному випадку процес MA(∞) еквівалентний процесу AR(1).

Взагалі, всякий стаціонарний процес AR(p) можна записати у формі процесу MA(∞):

$$X_t = \mu + \frac{1}{1-aL} \varepsilon_t = \mu + \sum_{j=0}^{\infty} b_j \varepsilon_{t-j} = \mu + b(L)\varepsilon_t,$$

де $b(L) = \sum_{j=0}^{\infty} b_j L^j = \frac{1}{a(L)}$ і $\sum_{j=0}^{\infty} |b_j| < \infty$.

Приклад 1.2. Розглянемо процес MA(1) з $b = 0,8$ і $E(X_t) = 6$, тобто

$$X_t = 6 + \varepsilon_t + 0,8\varepsilon_{t-1}.$$

Для цього процесу $\rho(1) = \frac{0,8}{1-0,8^2} = 0,488$.

Змодельовану реалізацію цього процесу показано на рис. 1.11.

Приклад 1.3. Для MA(2) процесу

$$X_t = 5 + \varepsilon_t - 0,75\varepsilon_{t-1} + 0,125\varepsilon_{t-2}$$

маємо $\rho(1) = \frac{b_0 b_1 + b_1 b_2}{b_0^2 + b_1^2 + b_2^2} = \frac{-0,75 - 0,75 \cdot 0,125}{1 + 0,75^2 + 0,125^2} = -0,535$,

$$\rho(2) = \frac{0,125}{1,578} = 0,079.$$

Змодельовану реалізацію цього процесу показано на рис. 1.12.

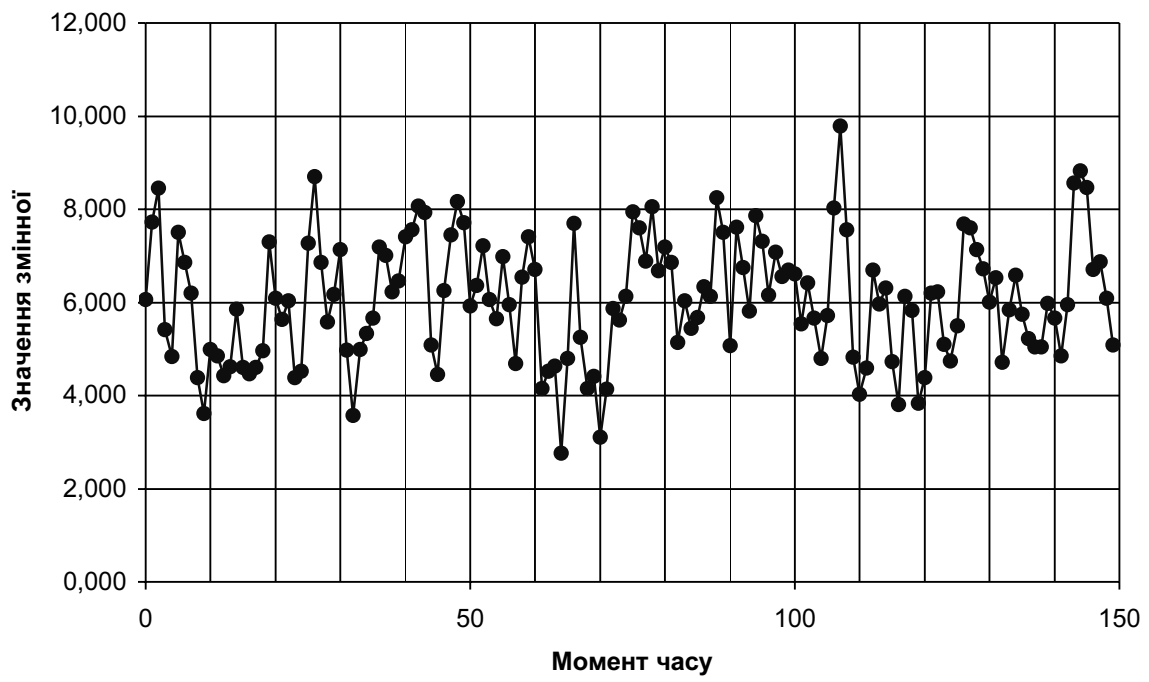


Рис. 1.11. Реалізація процесу $X_t = 6 + \varepsilon_t + 0,8\varepsilon_{t-1}$

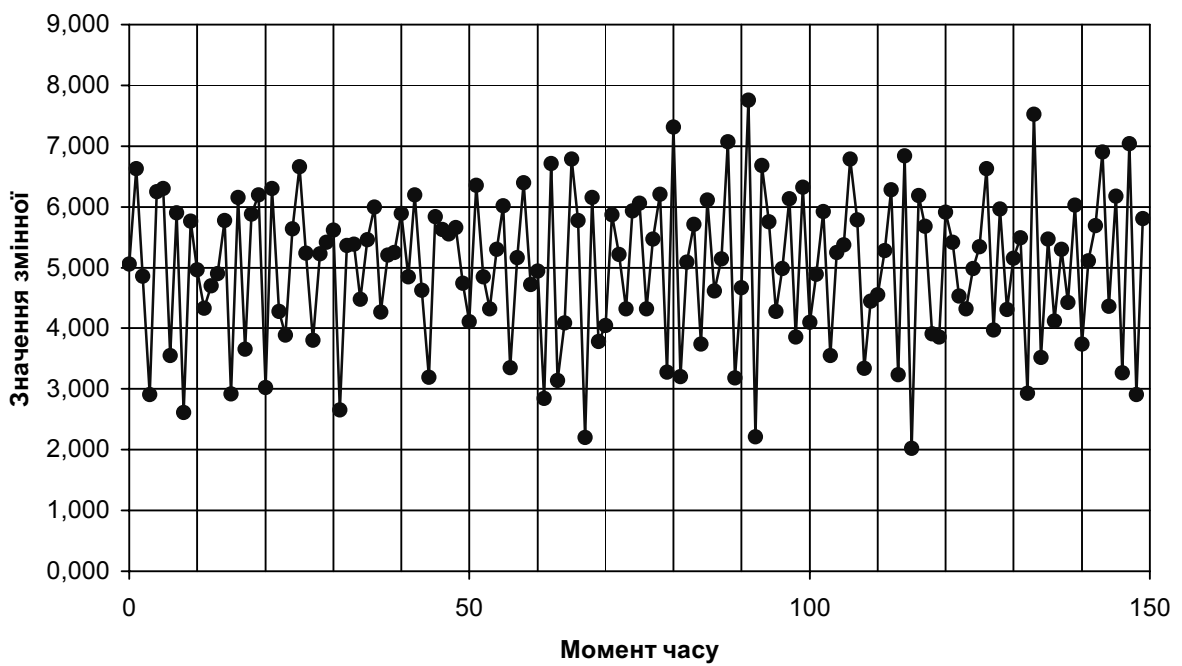


Рис. 1.12. Реалізація процесу $X_t = 5 + \varepsilon_t - 0,75\varepsilon_{t-1} + 0,125\varepsilon_{t-2}$

1.6. Змішаний процес авторегресії – ковзного середнього (процес авторегресії із залишками у вигляді ковзного середнього)

Процес X_t з нульовим математичним сподіванням, що належить класу змішаних процесів авторегресії – ковзного середнього, характеризується порядками p і q його AR і MA складових і позначається як процес ARMA(p, q) (***autoregressive moving average, mixed autoregressive moving average***). Точніше, процес X_t з нульовим математичним сподіванням належить класу ARMA(p, q), якщо

$$X_t = \sum_{j=1}^p a_j X_{t-j} + \sum_{j=0}^q b_j \varepsilon_{t-j}, \quad a_p \neq 0, \quad b_q \neq 0,$$

де ε_t – процес «білого шуму», $b_0 = 1$. В операторній формі останнє співвідношення можна записати як

$$a(L)X_t = b(L)\varepsilon_t,$$

де $a(L)$ і $b(L)$ мають той же вигляд, що й у визначених раніше моделях AR(p) і MA(q). Якщо процес має стале математичне сподівання μ , то у такому випадку його модель набуває вигляду

$$X_t - \mu = \sum_{j=1}^p a_j (X_{t-j} - \mu) + \sum_{j=0}^q b_j \varepsilon_{t-j}.$$

Процесу ARMA(p, q) з $E(X_t) = \mu$ притаманні такі властивості:

- 1) процес стаціонарний, якщо всі корені рівняння $a(z) = 0$ знаходяться поза одиничним колом $|z| \leq 1$;
- 2) якщо процес стаціонарний, то існує еквівалентний йому процес MA(∞)

$$X_t - \mu = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}, \quad c_0 = 1, \quad \sum_{j=0}^{\infty} |c_j| < \infty,$$

або

$$X_t - \mu = c(L)\varepsilon_t,$$

де

$$c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)};$$

- 3) якщо всі корені рівняння $b(z) = 0$ знаходяться поза одиничним колом $|z| \leq 1$ (***умова оборотності***), то існує еквівалентне подання процесу X_t у вигляді процесу авторегресії нескінченного порядку AR(∞):

$$X_t - \mu = \sum_{j=1}^{\infty} d_j (X_{t-j} - \mu) + \varepsilon_t,$$

або $d(L)(X_t - \mu) = \varepsilon_t$,

де $d(z) = 1 - \sum_{j=0}^{\infty} d_j z^j = \frac{a(z)}{b(z)}$,

Звідси випливає, що стаціонарний процес ARMA(p,q) завжди можна апроксимувати процесом ковзного середнього досить високого порядку, а при виконанні умови оборотності його також можна апроксимувати процесом авторегресії досить високого порядку.

Передумовою для обґрунтування використання моделей ARMA є такий факт. Якщо ARMA(p₁,q₁)-ряд X_t і ARMA(p₂,q₂)-ряд Y_t статистично незалежні між собою і $Z_t = X_t + Y_t$, то типовим є положення, коли Z_t є ARMA(p,q)-рядом, у якого:

- 1) $p = p_1 + p_2$;
- 2) $q = p_1 + q_2$, якщо $p_1 + q_2 > p_2 + q_1$;
- 3) $q = p_2 + q_1$, якщо $p_2 + q_1 > p_1 + q_2$.

Можливі також ситуації, коли значення p і q виявляються меншими від вказаних значень. Такі ситуації виникають у випадках, коли багаточлени $a_X(z)$ і $a_Y(z)$, що відповідають авторегресійним частинам процесів X_t і Y_t , мають загальні корені.

В економіці багато часових рядів є агрегованими. З зазначеного вище факту випливає, що якщо кожна з компонент відповідає простій моделі AR, то при незалежності цих компонент їхня сума буде ARMA-процесом. Процес такого ж роду буде отримано, якщо частина компонент має тип AR, а інші – тип MA. Єдиний виняток – це випадок, коли всі компоненти є MA-процесами, в результаті чого одержимо MA-процес.

Якщо «істинний» економічний ряд відповідає AR(p)-моделі, але значення цього ряду вимірюються з випадковими помилками, що утворюють процес «білого шуму» (тобто MA(0)), то спостережуваний ряд має тип ARMA(p,q).

Зауваження. Раніше вже зазначалося, що якщо ARMA(p,q)-процес X_t задовольняє умову оборотності, то його можна зобразити у вигляді стаціонарного процесу AR(∞). Останній, у свою чергу, можна апроксимувати стаціонарним процесом AR(p), який може бути досить високого порядку. Таким чином, у практичних задачах можна було б і зовсім обійтися без використання моделей ARMA, обмежуючись або AR, або MA-моделями. Однак при цьому кількість коефіцієнтів, що підлягають оцінюванню, може виявитися занадто великою (що знижує то-

чність оцінювання) і навіть може перевершувати кількість наявних спостережень. У цьому розумінні моделі ARMA можуть бути “більш економними”.

1.7. Моделі ARMA, що враховують наявність сезонності

Якщо спостережуваний часовий ряд має виражену сезонність, то модель ARMA, що відповідає цьому ряду, повинна містити складові, які забезпечують прояв відповідної сезонності в послідовності спостережень, що породжується цією моделлю.

Для квартальних даних чисто сезонними є стаціонарні моделі **сезонної авторегресії першого порядку (SAR(1))**

$$X_t = a_4 X_{t-4} + \varepsilon_t, \quad a_4 < 1,$$

і **сезонного ковзного середнього першого порядку (SMA(1))**

$$X_t = \varepsilon_t + b_4 \varepsilon_{t-4}.$$

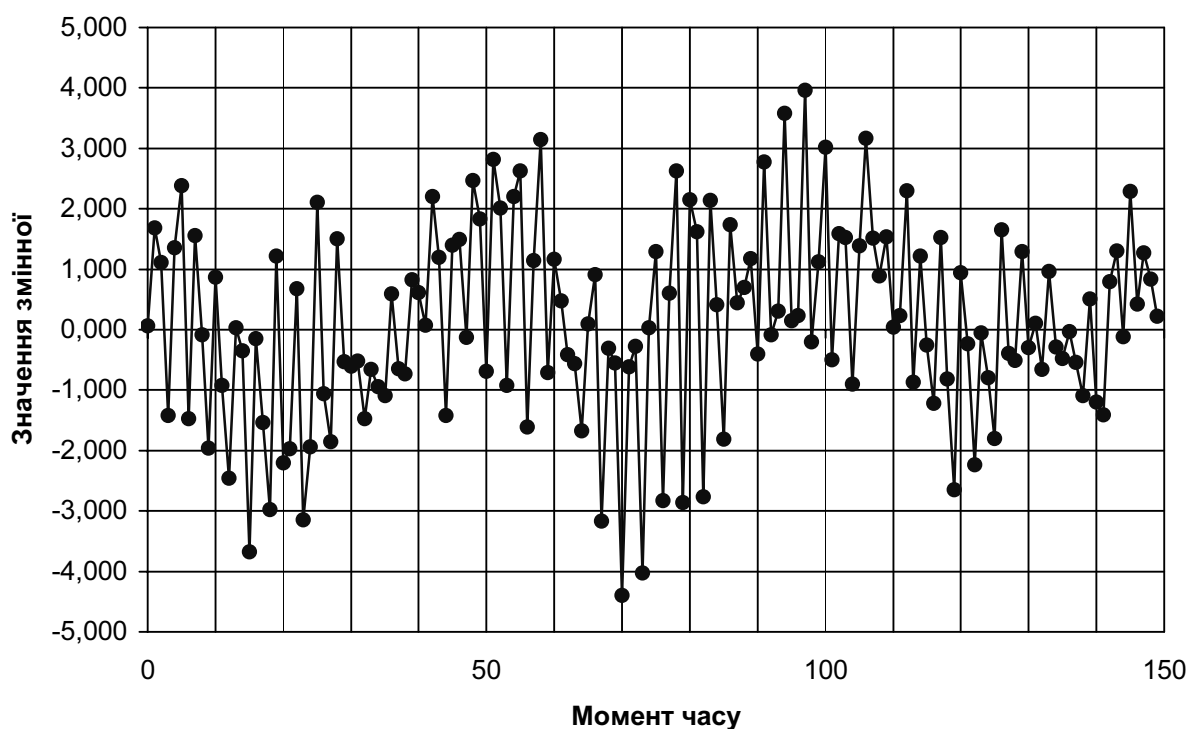


Рис. 1.13. Реалізація процесу $X_t = a_4 X_{t-4} + \varepsilon_t$

У першій моделі

$$\rho(k) = a_4^{k/4} \quad \text{для } k = 4m, \quad m = 0, 1, 2, \dots, \quad \rho(k) = 0 \quad \text{для інших } k > 0.$$

У другій моделі

$$\rho(0) = 1, \quad \rho(4) = b_4, \quad \rho(k) = 0 \text{ для інших } k > 0.$$

Змодельовані реалізації моделі SAR(1) з $a_4 = 0,8$ і моделі SMA(1) з $b_4 = 0,8$ наведено на рис. 1.13, 1.14. При моделюванні, як і раніше, використовувався “білий шум”, зображений на рис. 1.1.

Комбінації несезонних і сезонних змін реалізуються, наприклад, у моделях ARMA((1, 4), 1):

$$X_t = a_1 X_{t-1} + a_4 X_{t-4} + \varepsilon_t + b_1 \varepsilon_{t-1},$$

а також ARMA(1, (1, 4)):

$$X_t = a_1 X_{t-1} + \varepsilon_t + b_1 \varepsilon_{t-1} + b_4 \varepsilon_{t-4}.$$

Для першої моделі рівняння $a(z) = 0$ набирає вигляду $1 - \frac{2}{3}z + \frac{1}{48}z^4 = 0$, або $z^4 - 32z + 48 = 0$; корені цього рівняння $z_1 = 2$, $z_2 = 2$, $z_3 = -2 + i\sqrt{8}$, $z_4 = -2 - i\sqrt{8}$ знаходяться поза одиничним колом, що забезпечує стаціонарність розглядуваного процесу. Для другої моделі рівняння $a(z) = 0$ набирає вигляду $1 - 0,4z = 0$. Корінь цього рівняння $z = 2,5 > 1$, тому і ця модель – стаціонарна.

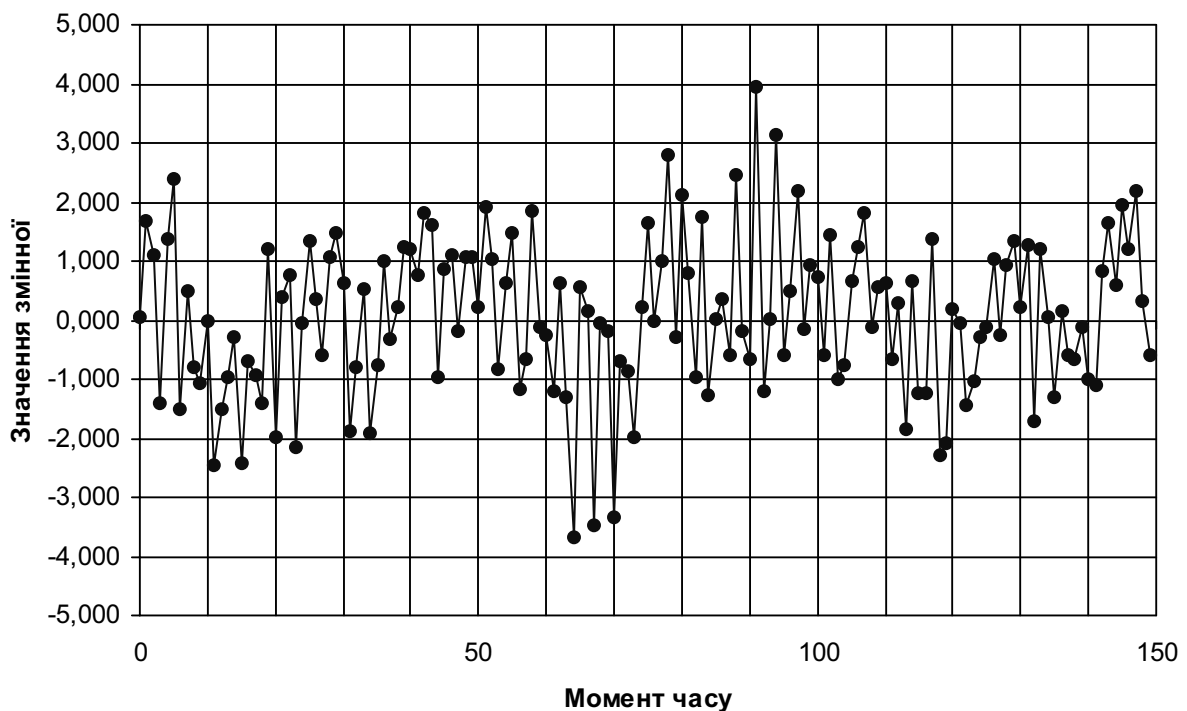


Рис. 1.14. Реалізація процесу $X_t = \varepsilon_t + b_4 \varepsilon_{t-4}$

Крім розглянутих прикладів адитивних сезонних моделей використовуються також і мультиплікативні варіанти, наприклад:

$$(1 - a_1L)X_t = (1 - b_1L)(1 - b_4L^4)\varepsilon_t;$$

$$(1 - a_1L)(1 - a_4L^4)X_t = (1 - b_1L)\varepsilon_t.$$

Перший варіант дає

$$X_t = a_1X_{t-1} + \varepsilon_t + b_1\varepsilon_{t-1} + b_4\varepsilon_{t-4} + b_1b_4\varepsilon_{t-5},$$

а другий –

$$X_t = a_1X_{t-1} + a_4X_{t-4} + a_1a_4X_{t-5} + \varepsilon_t + b_1\varepsilon_{t-1}.$$

У першій моделі допускається взаємодія складових ковзного середнього на лагах 1 і 4 (тобто значень ε_{t-1} і ε_{t-4}), а в другій – взаємодія авторегресійних складових на лагах 1 і 4 (тобто значень X_{t-1} і X_{t-4}).

Звичайно, ці дві моделі є окремими випадками адитивних моделей

$$X_t = a_1X_{t-1} + \varepsilon_t + b_1\varepsilon_{t-1} + b_4\varepsilon_{t-4} + b_5\varepsilon_{t-5},$$

$$X_t = a_1X_{t-1} + a_4X_{t-4} + a_5X_{t-5} + \varepsilon_t + b_1\varepsilon_{t-1} \text{ с } b_5 = b_1b_4, a_5 = -a_1a_4.$$

При наближеному виконанні останніх співвідношень (принаймні, якщо гіпотези про наявність таких співвідношень не відкидаються), природно перейти від оцінювання адитивної моделі до оцінювання мультиплікативної моделі, знову дотримуючись **принципу “економності”** (**“parsimony model”**). Проте якихось теоретичних підстав, що ведуть до переваги однієї форми сезонності перед іншою (мультиплікативної або адитивної), не існує.

Докладніше із сезонними ARMA моделями можна ознайомитися, наприклад, у книзі [2].

2. ПІДБІР СТАЦІОНАРНОЇ МОДЕЛІ ARMA

2.1. Основні етапи підбору моделі

Якщо припускається, що деякий спостережуваний часовий ряд x_1, \dots, x_T породжується моделлю ARMA, то при цьому виникає проблема підбору конкретної моделі із цього класу, вирішення якої передбачає три етапи:

- 1) **ідентифікація** моделі;
- 2) **оцінювання** моделі;
- 3) **діагностика** моделі.

На першому етапі здійснюється вибір деякої частинної моделі з усього класу ARMA, тобто вибір значень p і q . Використовувані при цьому процедури є не цілком точними, що може при наступному аналі-

зі призвести до висновку про непридатність ідентифікованої моделі та необхідність заміни її альтернативною. На цьому ж етапі робляться попередні грубі оцінки коефіцієнтів $a_1, \dots, a_p, b_1, \dots, b_q$ ідентифікованої моделі.

На другому етапі здійснюється уточнення оцінок коефіцієнтів моделі з використанням ефективних статистичних методів. Для оцінених коефіцієнтів обчислюються наближені стандартні помилки, що дають можливість, при додаткових припущеннях про розподіли випадкових величин X_1, \dots, X_T , будувати довірчі інтервали для цих коефіцієнтів і перевіряти гіпотези про їхні "істинні" значення з метою уточнення специфікації моделі.

На третьому етапі застосовуються різні діагностичні процедури перевірки адекватності вибраної моделі наявним даним (*misspecification tests*). Неадекватності, виявлені в процесі такої перевірки, вказують на необхідне коректування моделі, після чого здійснюється новий цикл підбору і т.д. доти, доки не буде отримана задовільна модель.

Зрозуміло, якщо ми маємо справу з ситуацією, коли вже є досить відпрацьована і розумно інтерпретована модель еволюції того чи іншого показника, можна обійтися без етапу ідентифікації.

Якщо ряд породжується моделлю $ARMA(p, q)$, то у подальшому для стислості це буде позначатися як $X_t \sim ARMA(p, q)$. Відповідно, якщо ряд породжується моделлю $AR(p)$, то $X_t \sim AR(p)$, а якщо – моделлю $MA(q)$, то $X_t \sim MA(q)$.

2.2. Ідентифікація стаціонарної моделі ARMA

Основною відправною точкою для ідентифікації стаціонарної моделі ARMA є різниця у поведженні **автокореляційних (ACF)** і **частинних автокореляційних (PACF)** функцій (*ACF – autocorrelation function, PACF – partial autocorrelation function*) рядів, що відповідають різним моделям ARMA.

Про поведження автокореляційних функцій для різних моделей ARMA вже говорилося. Однак за поведженням тільки автокореляційної функції важко ідентифікувати навіть порядок чистого (без MA складової) процесу авторегресії. Вирішенню цього питання допомагає розгляд поведження **частинної автокореляційної функції (PACF)** стаціонарного процесу X_t .

Значення цієї функції $\rho_{part}(k)$ на лозі k визначається як значення коефіцієнта кореляції між випадковими величинами X_t і X_{t+k} , звільненого від впливу випадкових величин $X_{t+1}, \dots, X_{t+k-1} \dots$.

Це відповідає тому, що $\rho_{part}(k)$ є коефіцієнтом при X_{t-k} у лінійній комбінації випадкових величин X_{t-1}, \dots, X_{t-k} , яка найкраще наближає випадкову величину X_t . Виходячи з останнього встановлено [3], що $\rho_{part}(k)$ визначається як розв'язання відносно a_k системи перших k рівнянь Юла – Уокера

$$\rho(s) = a_1\rho(s-1) + a_2\rho(s-2) + \dots + a_k\rho(s-k), \quad s = 1, 2, \dots, k,$$

яку в цьому випадку зручніше записати у вигляді

$$a_1\rho(s-1) + a_2\rho(s-2) + \dots + a_k\rho(s-k) = \rho(s), \quad s = 1, 2, \dots, k,$$

підкреслюючи, що невідомими тут є a_1, a_2, \dots, a_k , а $\rho(s-1), \dots, \rho(s-k)$ – відомі коефіцієнти.

Виходячи з цього і застосовуючи відоме з алгебри правило Крамера розв'язання системи k лінійних рівнянь з k невідомими, знаходимо, що PACF можна обчислити за формулами:

$$\rho_{part}(0) = 1;$$

$$\rho_{part}(1) = \rho(1);$$

$$\rho_{part}(2) = \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)};$$

$$\rho_{part}(3) = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}};$$

... .. ;

$$\rho_{part}(k) = \frac{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(3) \\ \dots & \dots & \dots & \dots & \dots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & \rho(k) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(k-3) \\ \dots & \dots & \dots & \dots & \dots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{vmatrix}}.$$

Тут визначник у чисельнику виразу для $\rho_{part}(k)$ відрізняється від визначника в знаменнику цього виразу тільки заміною останнього стовпця стовпцем, що складається зі значень $\rho(1), \rho(2), \dots, \rho(k)$.

Визначальним є той факт, що якщо X_t – процес типу AR(p), то

$$\begin{aligned} \rho_{part}(p) &\neq 0, \\ \rho_{part}(k) &\neq 0 \text{ для } k > p. \end{aligned}$$

Це дозволяє за графіком PACF визначити порядок процесу авторегресії та відрізнити процес авторегресії від процесів ковзного середнього і ARMA(p,q) з $q > 0$.

Нагадаємо, що занулення ACF після лага q відповідає процесу MA(q).

Тепер же очевидно, що занулення PACF після лага p відповідає процесу AR(p).

Тому ідентифікація цих моделей за ACF і PACF більш визначена порівняно з ідентифікацією моделей ARMA(p,q) з $p \neq 0, q \neq 0$.

Водночас замість невідомих “істинних” послідовностей автокореляцій $\rho(k)$ і частинних автокореляцій $\rho_{part}(k)$ можна задовольнятися тільки їхніми обґрунтованими оцінками – **вибірковою ACF**, утвореною **вибірковими автокореляціями**

$$r(k) = \frac{\frac{1}{T-k} \sum_{t=1}^{T-k} (x_t - \hat{\mu})(x_{t+k} - \hat{\mu})}{\frac{1}{T} \sum_{t=1}^T (x_t - \hat{\mu})^2} = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \quad k=1, \dots, T-1,$$

де $\hat{\mu} = \bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$ – оцінка для $\hat{\mu} = E(X_t)$,

$$\hat{\gamma}(k) = \frac{1}{T-k} \sum_{t=1}^{T-k} (x_t - \hat{\mu})(x_{t+k} - \hat{\mu}) \text{ – оцінка для } \gamma(k),$$

і **вибірковою PACF**, утвореною **вибірковими частинними автокореляціями** $\rho_{part}(k)$. Останні можна одержати, замінюючи автокореляції $\rho(s)$, що входять до виразу для $\rho_{part}(k)$, їхніми оцінками $r(s)$. Однак простіше зробити інакше, виходячи з того, що $\rho_{part}(k)$ є коефіцієнтом при X_{t-k} у лінійній комбінації випадкових величин X_{t-1}, \dots, X_{t-k} , яка найкраще наближає випадкову величину X_t .

Коефіцієнти в моделі $X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k} + \mu_t$ можна просто оцінити методом найменших квадратів (на складову $\mu_t = X_t - (a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k})$ не накладаються якісь попередні обмеження).

Отримана в результаті оцінка коефіцієнта a_k і є $r_{part}(k)$.

Якщо X_t є стаціонарним процесом типу ARMA(p,q) і $E(X_t^4) < \infty$, то зазначені оцінки $\hat{\mu}$, $\hat{\gamma}(k)$, $r(k)$ і $r_{part}(k)$ є обґрунтованими оцінками для μ , $\gamma(k)$, $\rho(k)$ і $\rho_{part}(k)$ відповідно [3].

Оскільки $r(k)$ і $r_{part}(k)$ – лише оцінки для $\rho(k)$ і $\rho_{part}(k)$, то їхні спостережувані значення можуть значно відрізнятись від $\rho(k)$ і $\rho_{part}(k)$.

Зокрема, якщо при деяких $k = k_1$ і $k = k_2$ в моделі, яка породжує спостереження, $\rho(k_1) = 0$ і $\rho_{part}(k_2) = 0$, то, як правило, $r(k_1) \neq 0$ і $r_{part}(k_2) \neq 0$, що вносить додаткову невизначеність у задачу ідентифікації.

Більше того, характер зміни теоретичної автокореляційної функції зовсім не обов'язково буде відтворюватися в її вибірковому аналізі – вибірковій автокореляційній функції.

Проте у багатьох випадках поведінка теоретичних ACF і PACF якоюсь мірою відбивається і на поведінці їхніх вибіркових аналогів.

Тому уявлення про поведінку теоретичних ACF і PACF може допомогти в роз'язанні задачі ідентифікації відповідних моделей у рамках загального класу моделей ARMA.

У табл. 2.1 наведено властивості ACF і PACF для деяких популярних моделей стаціонарних часових рядів.

Таблиця 2.1

Властивості функцій ACF і PACF для різних моделей

Модель	ACF	PACF
“Білий шум”, MA(0)	$\rho(k) = 0$ для $k \neq 0$	$\rho_{part}(k) = 0$ для $k \neq 0$
AR(1), $a_1 > 0$	Експоненціальне убуння $\rho(k) = a_1^k$	$\rho_{part}(1) = a_1$; $\rho_{part}(k) = 0$ для $k \geq 2$
AR(1), $a_1 < 0$	Коливне убуння $\rho(k) = a_1^k$	$\rho_{part}(k) = 0$ для $k \geq 2$
AR(p)	Убуння до нуля з можливою осциляцією	Занулення при $k \geq p$
MA(1), $b_1 > 0$	Додатний пік при $k = 1$; занулення при $k > 1$	Коливне убуння; $\rho_{part}(1) > 0$
MA(1), $b_1 < 0$	Від'ємний пік при $k = 1$; занулення при $k > 1$	Убуння за абсолютною величиною; $\rho_{part}(k) < 0$ при $k \geq 1$
MA(q)	Занулення при $k > q$	
ARMA(1,1), $a_1 > 0$	Експоненціальне убуння з лага 1; знак $\rho(1)$ збігається зі знаком $(a_1 + b_1)$	Коливне убуння з лага 1; $\rho_{part}(1) = \rho(1)$
ARMA(1,1), $a_1 < 0$	Коливне убуння з лага 1; знак $\rho(1)$ збігається зі знаком $(a_1 + b_1)$	Експоненціальне убуння з лага 1; $\rho_{part}(1) = \rho(1)$; знак $\rho_{part}(k)$ збігається зі знаком $\rho(1)$, $k > 1$
ARMA(p,q)	Коливне або пряме убуння, що починається з лага q	Коливне або пряме убуння, що починається з лага p

Модель	ACF	PACF
SAR(1)	Загасання на лагах, кратних періоду сезонності; занулення на інших лагах	Пік на лазі, кратному періоду сезонності; занулення на інших лагах
SMA(1)	Пік на лазі, кратному періоду сезонності; занулення на інших лагах	Загасання на лагах, кратних періоду сезонності; занулення на інших лагах

Для ідентифікації моделей $AR(p)$ і $MA(q)$ за графіками функцій $r(k)$ і $r_{part}(k)$ використовують статистичні критерії перевірки гіпотез про рівність нулю тих чи інших значень $\rho(k)$ і $\rho_{part}(k)$ на підставі спостережуваних значень $r(k)$ і $r_{part}(k)$.

Питання це складне, і часто обмежуються такими двома наближеними рецептами, які передбачають гауссівість інновацій (тобто що ε_t – гауссівський “білий шум”):

1. Якщо X_t – процес типу $MA(q)$, то при великих T

$$D(r(k)) \approx \frac{1}{T} \left(1 + \sum_{j=1}^q r^2(j) \right) \text{ для } k > q,$$

отже, чим довший ряд спостережень, тим надійніше виявляються нульові значення $\rho(k)$, $k > q$.

При цьому $\lim_{T \rightarrow \infty} E(r(k)) = \rho(k)$.

Більше того, при великих T і $k > q$ розподіл випадкової величини $r(k)$ близький до нормального.

Звідси випливає, що природний наближений критерій перевірки гіпотези H_0 (“ X_t – процес типу $MA(q)$ ”) полягає в тому, щоб нехтувати цією гіпотезою, якщо

$$|r(k)| > \frac{2}{\sqrt{T}} \sqrt{1 + \sum_{j=1}^q r^2(j)} \text{ для } k > q.$$

Рівень значущості такого критерію приблизно дорівнює 0,05.

Цей критерій базується на тому, що для $r(k) \sim N(0, D(r(k)))$ 95%-й довірчий інтервал є

$$\left[r(k) - 1,96\sqrt{D(r(k))}; r(k) + 1,96\sqrt{D(r(k))} \right],$$

де 1,96 – квантиль $z_{0,975}$.

Зокрема, якщо $q = 0$, то $X_t \sim \text{MA}(0)$ – “білий шум”, і гіпотеза H_0 (“ X_t – “білий шум””) нехтується зазначеним наближеним критерієм при

$$|r(k)| > \frac{2}{\sqrt{T}}, \quad k > 0.$$

2. Якщо X_t – процес типу $\text{AR}(p)$, то при великих T і $k > p$ розподіл $r_{part}(k)$ можна апроксимувати нормальним розподілом

$$r_{part}(k) \approx N(0, T^{-1}) \quad (\text{так що } D(r_{part}(k)) \approx T^{-1}).$$

Отже, якщо гіпотезою H_0 ($X_t \sim \text{AR}(p)$) нехтувати при

$$|r_{part}(k)| > \frac{2}{\sqrt{T}}, \quad k > p,$$

то одержимо критерій, рівень значущості якого приблизно дорівнює 0,05.

Маючи на увазі два зазначених наближених критерії, у процедурах аналізу часових рядів звичайно передбачена роздруківка графіків вибірових ACF і PACF, на які нанесено межі смуги $\pm \frac{2}{\sqrt{T}}$. У цих межах з імовірністю, близькою до 0,95, повинні міститися значення $r(k)$, якщо X_t – “білий шум”, і значення $r_{part}(k)$, якщо $X_t \sim \text{AR}(p)$.

Тут варто зробити одне важливе попередження. Саме обидва побудованих критерії мають рівень значущості, близький до 0,05, тільки коли перевіряється гіпотеза H_0 при деякому фіксованому k .

Розглянемо застосування цих критеріїв на прикладі змодельованого «білого шуму» (див. рис. 1.1). Значення ACF і PACF, розраховані для цього часового ряду, наведено в табл. 2.2; значення $\frac{2}{\sqrt{T}}$ дорівнює 0,16329932.

У табл. 2.2 показано, що жодне з розрахованих значень ACF не виходить за межі інтервалу $\pm \frac{2}{\sqrt{T}} = 0,16329932$, що підтверджує гіпотезу H_0 для розглядуваного часового ряду (див. рис. 1.1).

У цілому до інтерпретації значень вибірових ACF і PACF слід підходити обережно і застосовувати інші додаткові критерії перевірки можливих гіпотез.

Так, у роздруківках аналізу часових рядів разом зі значеннями (графіками) вибірових ACF і PACF звичайно друкуються значення *Q-статистики*, що відносяться до критерію перевірки гіпотези про те, що спостережувані дані є реалізацією процесу «білого шуму».

Таблиця 2.2

Розрахунок значень функцій ACF і PACF

Лag часу	Вибіркова ко-варіація	Вибіркова автокореляція (ACF)	Вибіркова частинна автокореляція (PACF)	Q-статистика Бокса - Пірса	P-значення Q-статистики Бокса - Пірса
0	1	1	1	-	-
1	0,051441266	0,05200936	0,05200936	0,4057460	0,524136612
2	-0,000810705	-0,00081966	-0,00353419	0,4058468	0,816340788
3	0,080549586	0,08143913	0,08188825	1,4006967	0,705371441
4	-0,03352767	-0,03389793	-0,04284033	1,5730571	0,813626247
5	0,043997748	0,04448364	0,04990185	1,8698762	0,866841242
6	0,07484274	0,07566926	0,06402345	2,7287518	0,842039588
7	-0,038520559	-0,03894596	-0,04035672	2,9562700	0,889017918
8	0,023285591	0,02354275	0,02031922	3,0394091	0,931860100
9	-0,051545698	-0,05211494	-0,06406305	3,4468042	0,943924120
10	-0,06103918	-0,06171327	-0,04606824	4,0180833	0,946527526

Існує кілька варіантів **Q-статистик**. Одна з таких статистик (**статистика Бокса – Пірса**) була запропонована Боксом і Пірсом [4]:

$$Q = T \sum_{k=1}^M r^2(k).$$

Оскільки величини $r(1), \dots, r(M)$ асимптотично незалежні у випадку, коли X_t – “білий шум”, то при великих T маємо

$$\sqrt{T} \cdot r(k) \approx N(0,1),$$

так що

$$T \cdot r^2(k) \approx [N(0,1)]^2 = \chi^2(1)$$

(зазначимо, що в цій ситуації не потрібна гауссівість X_t [5]).

Звідси випливає, що при великих T приблизно маємо

$$Q = \chi^2(M).$$

Всупереч гіпотезі H_0 свідчать більші значення цієї статистики. Тому якщо вибрати рівень значущості таким, що дорівнює 0,05, то цією гіпотезою нехтують (з рівнем довіри 95%) при виконанні нерівності

$$Q > t_{0,95}(M).$$

У роздруківках корелограм звичайно наводяться **P-значення** статистики Q для послідовних значень $M = 1, 2, \dots$

P-значення статистики Q – це ймовірність того, що випадкова величина з розподілом $\chi^2(M)$ перевищить значення статистики Q .

При конкретному значенні M гіпотезою H_0 нехтують, коли відповідне P -значення менше 0,05. У табл. 2.2 наведено P -значення статистики Q для «білого шуму». Всі ці значення більші, ніж 0,05. Тому з рівнем довіри 95% може бути прийнята гіпотеза, що вихідний часовий ряд є «білим шумом».

Більш пізні дослідження показали, що статистика Бокса – Пірса погано наближається розподілом $\chi^2(M)$ при помірних значеннях T . Замість її в таких випадках переважніше використовувати **статистику Льюнга – Бокса** [6]

$$Q = T(T + 2) \sum_{k=1}^M \frac{r^2(k)}{(T - k)},$$

яка (при $T \rightarrow \infty$) також має асимптотичний розподіл $\chi^2(M)$, але ближче до цього розподілу при помірних значеннях T , ніж статистика Бокса - Пірса. У пакеті EVIEWS (*Econometric Views*) значення статистики Льюнга - Бокса роздруковуються разом з наближеними P -значеннями, що відповідають розподілам $\chi^2(M)$.

З розглянутого видно, що на етапі вибору придатної моделі серед усієї множини ARMA-моделей використовувані процедури є не цілком точними і часто призводять до досить невизначених висновків. У підсумку цього етапу можливий вибір для подальшого дослідження не однієї, а декількох потенціальних моделей.

2.3. Використання інформаційних критеріїв для підбору моделі

Більш означені висновки при виборі моделі на першому етапі можна одержати, застосовуючи інформаційні критерії відбору моделей.

Якщо заздалегідь обмежуватися розглядом тільки AR-моделей, тобто вважати, що процес X_t потребує моделі AR(k)

$$a(L)(X_t - \mu) = \varepsilon_t, \quad \text{або} \quad (X_t - \mu) - \sum_{j=1}^k a_{kj}(X_{t-j} - \mu) = \varepsilon_t,$$

з невідомим “істинним” порядком k , то для визначення k в таких ситуаціях довгий час використовувався **інформаційний критерій Акаїке** [7].

Відповідно до цього критерію серед альтернативних значень k вибирають значення, що мінімізує величину

$$AIC(k) = \ln \hat{\sigma}_k^2 + \frac{2k}{T} \rightarrow \min,$$

де T – кількість спостережень, а $\hat{\sigma}_k^2$ – оцінка дисперсії інновацій ε_t в AR-моделі k -го порядку. Для обчислення $\hat{\sigma}_k^2$ здійснюють підбір коефіцієнтів моделі k -го порядку з використанням рівнянь Юла – Уокера

$$r(s) = \sum_{j=1}^k \hat{a}_j r(s-j), \quad s = 1, \dots, k,$$

отримані оцінки коефіцієнтів \hat{a}_j , $j = 1, \dots, k$ (або, що те ж саме, $\rho_{part}(j)$, див. підрозд. 2.2.) підставляють замість a_j у рівняння моделі, μ заміняють на \bar{x} так, що одержують оцінки для ε_t :

$$\hat{\varepsilon}_{kt} = (x_t - \bar{x}) - \sum_{j=1}^k \hat{a}_j (x_{t-j} - \bar{x}),$$

після чого

$$\hat{\sigma}_k^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_{kt}^2.$$

Згодом було з'ясовано, що оцінка Акаїке необгрунтована і асимптотично переоцінює (завищує) справжнє значення k_0 з ненульовою ймовірністю. У зв'язку з цим були запропоновані обгрунтовані критерії, що базуються на мінімізації суми

$$\ln \hat{\sigma}_k^2 + k \cdot c_T,$$

де $c_T = O\left(\frac{\ln T}{T}\right)$ (тобто c_T при $T \rightarrow \infty$ має той самий порядок малості, що і $\frac{\ln T}{T}$).

Одним з таких критеріїв є часто використовуваний у цей час **інформаційний критерій Шварца** [8]

$$SIC = \ln \hat{\sigma}_k^2 + k \frac{\ln T}{T}.$$

Трохи пізніше було запропоновано **критерій Хеннана – Куїнна** [9], у якому $c_T = 2c \cdot k \left(\frac{\ln \ln T}{T}\right)$, $c > 1$,

$$HQ = \ln \hat{\sigma}_k^2 + k \frac{2c \ln \ln T}{T},$$

що має більш швидку збіжність з “істинним” значенням k_0 при $T \rightarrow \infty$.

Однак при невеликих значеннях T цей критерій недооцінює порядок авторегресії.

Приклад 2.1. Розглянемо модель процесу AR(2)

$$X_t = 1,2X_{t-1} - 0,36X_{t-2} + \varepsilon_t.$$

У цьому випадку рівняння $a(z) = 0$ набирає вигляду

$$1 - 1,2z - 0,36z^2 = 0$$

і має подвійний корінь $z = \frac{5}{3} > 1$, так що процес, який породжується такою моделлю, – стаціонарний.

Змодельована реалізація цього процесу для $t = 1, 2, \dots, 150$ має вигляд, показаний на рис. 2.1.

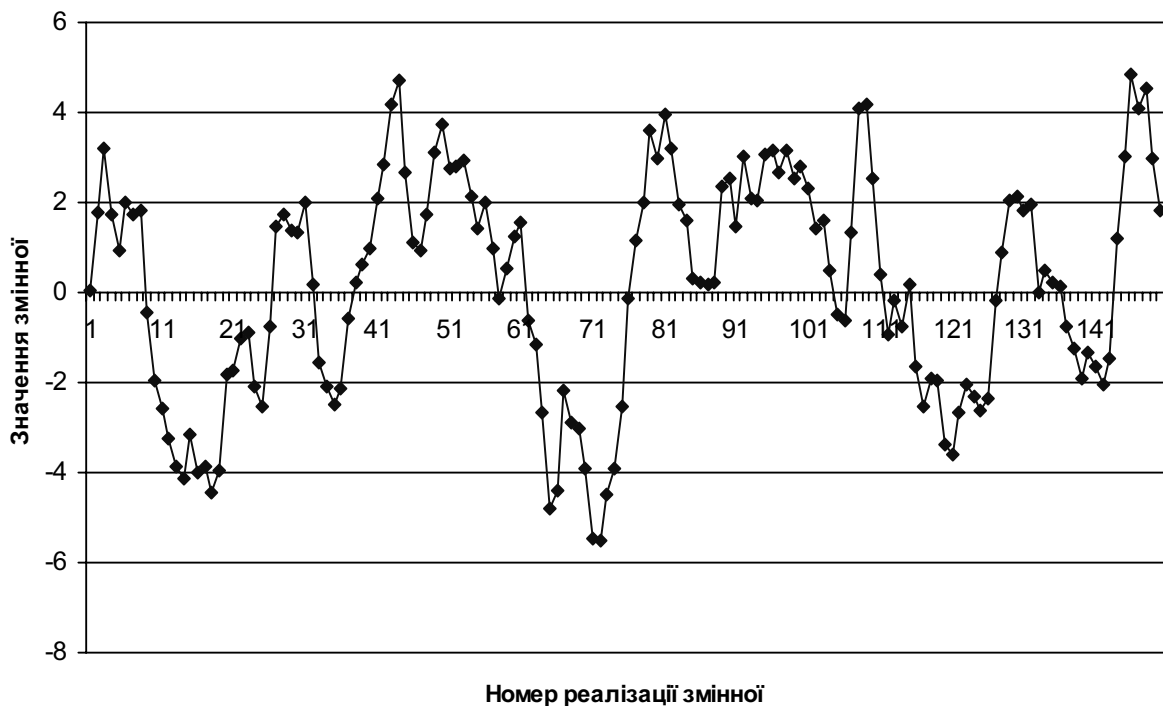


Рис. 2.1. Реалізація процесу $X_t = 1,2X_{t-1} - 0,36X_{t-2} + \varepsilon_t$

Значення коефіцієнтів кореляції, Q-статистики і P-значення Q-статистик наведено в табл. 2.3.

Дані табл. 2.3 показують, що за межі смуги $\pm \frac{2}{\sqrt{T}} = \pm 0,16329932$ виходять тільки значення вибіркової PACF, що відповідають лагам $k = 1, 2$. Відповідно до наближеного критерію, згаданого раніше, гіпотезою H_0 ($X_t \sim \text{AR}(2)$) нехтувати не можна.

Для підтвердження цієї гіпотези порівняємо значення інформаційних критеріїв Акаїке і Шварца, одержані при оцінюванні AR-моделей 1 - 5-го порядків, що припускають ненульове математичне сподівання відповідних AR-процесів (табл. 2.4).

Таблиця 2.3

Значення функцій ACF, PACF, Q-статистики і P-значення
Q-статистик для реалізації процесу AR(2)

Лаг часу	Вибіркова автокореляція (ACF)	Вибіркова приватна автокореляція (PACF)	Q-статистика	P-значення Q-статистики
1	0,90450099	0,90450099	122,71831	1,6072 E-28
2	0,74391183	-0,40802198	205,72903	2,1208 E-45
3	0,58238748	0,0265797	256,60531	2,4385 E-55
4	0,4348822	-0,05587555	284,97368	1,8861 E-60
5	0,31563132	0,03655456	299,91715	1,0435 E-62
6	0,21324292	-0,08316592	306,73804	2,9430 E-63
7	0,11696174	-0,07046871	308,79004	8,0195 E-63
8	0,03944498	0,03077929	309,02343	4,9386 E-62
9	-0,02763917	-0,07957421	309,13802	3,0040 E-61
10	-0,07523611	0,04405340	309,98709	1,2008 E-60

Таблиця 2.4

Значення інформаційних критеріїв Акаїке і Шварца
для реалізації процесу AR(2)

Порядок моделі p	Критерій Акаїке	Критерій Шварца
1	3,054	3,074
2	2,690	2,730
3	3,403	3,463
4	2,704	2,784
5	2,731	2,831

Обидва ці критерії вибирають модель AR(2).

Якщо не обмежувати себе моделями AR і припускати, що модель, яка породжує дані, має вигляд $ARMA(p_0, q_0)$ (з невідомими p_0, q_0)

$$a(L)(X_t - \mu) = b(L)\varepsilon_t,$$

то в цьому випадку є кілька процедур оцінювання пари (p_0, q_0) , одна з яких полягає в такому [10].

На першому кроці цієї процедури вже відомими методами здійснюється підбір моделі авторегресії AR(p)

$$(X_t - \mu) - \sum_{j=1}^p a_{pj}(X_{t-j} - \mu) = \varepsilon_t,$$

обчислюються оцінки коефіцієнтів \hat{a}_{pj} , $j = 1, \dots, p$, і на їхній основі одержуються оцінки інновацій

$$\hat{\varepsilon}_{p,t} = (x_t - \bar{x}) - \sum_{j=1}^p \hat{a}_{pj} (x_{t-j} - \bar{x}).$$

Порядок p авторегресійної моделі на цьому кроці має бути досить високим. Його можна вибрати, спираючись на порівняння значень критерію Акаїке для оцінених моделей авторегресії різних порядків (згадаємо, що критерій Акаїке схильний завищувати порядок моделі, що в цьому випадку й потрібно).

На другому кроці беруть регресії X_t на $\hat{\varepsilon}_{p,t-j}$, $j = 1, \dots, q$:

$$X_t = \mu + \hat{\varepsilon}_{p,t} + b_1 \hat{\varepsilon}_{p,t-1} + b_2 \hat{\varepsilon}_{p,t-2} + \dots + b_q \hat{\varepsilon}_{p,t-q}, \quad b_q \neq 0.$$

За допомогою методу найменших квадратів одержують оцінки \hat{b}_j для b_j , $j = 1, \dots, q$.

Відповідно оцінками поліномів $a(z)$, $b(z)$ служать

$$a(z) = \sum_{j=0}^p \hat{a}_j z^j, \quad \hat{a}_0 = 1,$$

$$b(z) = \sum_{j=0}^q \hat{b}_j z^j, \quad \hat{b}_0 = 1,$$

і за допомогою цих поліномів одержують уточнену оцінку для інновацій

$$\hat{\varepsilon}_t = \frac{\hat{a}(L)}{\hat{b}(L)} (x_t - \bar{x}),$$

на підставі якої – уточнену оцінку для дисперсії інновацій

$$\hat{\sigma}_{p,q}^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2.$$

Отже, передбачається, що самі інновації, знаючи точно коефіцієнти ARMA-моделі, можна знайти за формулою

$$\varepsilon_t = \frac{a(L)}{b(L)} x_t,$$

що відповідає оборотності цієї моделі.

Як оцінки для p_0, q_0 береться пара значень (\tilde{p}, \tilde{q}) , при якій мінімізується величина

$$SIC(p, q) = \ln \tilde{\sigma}_{p,q}^2 + (p + q) \frac{\ln T}{T}.$$

Істотно, що $SIC(p, q)$ – зростаюча функція від p і q , коли $p \geq p_0$, $q \geq q_0$, що веде до обґрунтованості оцінок (\tilde{p}, \tilde{q}) .

2.4. Оцінювання коефіцієнтів моделі

Після того, як зроблено ідентифікацію моделі ARMA (наприклад стаціонарної), тобто на підставі наявних спостережень прийнято рішення про значення p, q у моделі ARMA(p, q), яка породжує дані, переходять до етапу оцінювання коефіцієнтів моделі. На цьому етапі звичайно використовується метод максимальної правдоподібності, який в остаточному підсумку зводиться до методу найменших квадратів. За винятком деяких найпростіших випадків (наприклад моделі AR(1)), цю задачу розв'язують ітераційними методами, що потребують задання деяких "початкових" ("стартових") значень параметрів, які потім послідовно уточнюються.

Як такі початкові значення можна використовувати попередні оцінки, отримані на першому етапі. Ці значення можна знайти, прирівнюючи невідомі "істинні" значення автокореляцій $\rho(k)$ до значень $r(k)$ вибіркової автокореляційної функції та використовуючи функціональний зв'язок між значеннями $\rho(k)$ і значеннями коефіцієнтів моделі. Наприклад, якщо оцінюється модель AR(p), то коефіцієнти a_1, \dots, a_p визначаються із системи перших p рівнянь Юла – Уокера

$$\rho(s) = a_1\rho(s-1) + a_2\rho(s-2) + \dots + a_p\rho(s-p), \quad s = 1, 2, \dots, p,$$

в які замість невідомих значень $\rho(1), \dots, \rho(p)$ автокореляцій підставляють спостережувані (що обчислюються за реалізацією ряду) значення $r(1), \dots, r(p)$ вибірових автокореляцій.

При оцінюванні моделей з MA(q) складовою ($q > 0$) істотною виявляється умова оборотності, сформульована в підрозд. 1.6. Покажемо це на прикладі MA(1)-моделі

$$X_t - \mu = \varepsilon_t + b\varepsilon_{t-1}, \quad t = 1, \dots, T.$$

Маючи спостережувані значення x_1, \dots, x_T , послідовно виражаємо $\varepsilon_1, \dots, \varepsilon_T$ через ці значення та (неспостережуване) значення ε_0 :

$$\begin{aligned} \varepsilon_1 &= X_1 - \mu - b\varepsilon_0; \\ \varepsilon_2 &= X_2 - \mu - b\varepsilon_1 = X_2 - \mu - b(X_1 - \mu - b\varepsilon_0) = \\ &= X_2 - \mu - b(X_1 - \mu) + b^2\varepsilon_0; \\ &\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots; \\ \varepsilon_T &= X_T - \mu - b\varepsilon_{T-1} = (X_T - \mu) - b(X_{T-1} - \mu) + b^2(X_{T-2} - \mu) + \\ &+ (-1)^{T-1}b^{T-1}(X_1 - \mu) + (-1)^T b^T \varepsilon_0. \end{aligned}$$

Максимізація (за b) умовної функції правдоподібності, що відповідає спостережуваним значенням x_1, \dots, x_T при фіксованому значенні ε_0 , рівносильна мінімізації суми квадратів

$$Q(b) = \varepsilon_1^2 + \dots + \varepsilon_T^2,$$

яка є нелінійною функцією від b . Для пошуку мінімуму цієї суми квадратів доводиться використовувати чисельні ітераційні методи оптимізації, які, у свою чергу, потребують задання початкового (“стартового”) значення параметра b . Таке “стартове” значення може бути отримане на етапі ідентифікації моделі. Однак отримане в підсумку ітерацій “оптимальне” значення b залежить від невідомого значення ε_0 , що утруднює інтерпретацію результатів. Задача інтерпретації полегшується, якщо виконується умова оборотності $b < 1$ і значення b істотно менше 1.

Дійсно, при виконанні цієї умови можна просто покласти $\varepsilon_0 = 0$. Ефект від такої заміни “істинного” значення ε_0 на нульове швидко збуває, так що сума квадратів, яка одержується в припущенні $\varepsilon_0 = 0$, може служити хорошою апроксимацією для суми, отриманої при “істинному” значенні ε_0 , при досить великій кількості спостережень. Ті ж аргументи придатні й для моделі MA(q) з $q > 1$: у цьому випадку можна покласти $\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0$. Для одержання більш точної апроксимації у пакетах статистичних програм (у тому числі й в EVIEWS) передбачено процедуру **backcasting**, в якій процес ітерацій містить також і оцінювання значень $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-q+1}$ шляхом побудови для них “зворотного прогнозу”.

Якщо в результаті оцінювання отримано модель, у якій умова оборотності не виконується, рекомендується повторити процедуру оцінювання з використанням іншого набору початкових значень.

Більш докладний виклад процедур оцінювання стаціонарних ARMA моделей методом максимальної правдоподібності можна знайти, наприклад, у книзі [3]. Там же можна прочитати про те, як обчислюються наближення для стандартних помилок оцінок коефіцієнтів цих моделей, які можна використовувати при великій кількості спостережень звичайним чином.

На завершення необхідно тільки зробити одне важливе зауваження.

Нехай маємо стаціонарну AR(p)-модель

$$a(L)X_t = \delta + \varepsilon_t.$$

У цьому випадку математичне сподівання μ процесу X_t пов'язане з константою δ співвідношенням

$$\mu = \frac{\delta}{(1 - a_1 - \dots - a_p)}.$$

При цьому можна спочатку оцінити коефіцієнти a_1, \dots, a_p і δ , застосовуючи звичайний метод найменших квадратів до моделі

$$X_t = \delta + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t,$$

а потім, використовуючи отримані оцінки $\hat{a}_1, \dots, \hat{a}_p$ і $\hat{\delta}$, одержати оцінку для μ у вигляді

$$\mu = \frac{\hat{\delta}}{(1 - \hat{a}_1 - \dots - \hat{a}_p)}.$$

Можна зробити інакше, як це передбачено, наприклад, у пакеті EVIEWS (*Econometric Views*): записати ту ж модель у вигляді

$$X_t = \mu(1 - a_1 - \dots - a_p) + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t$$

і одночасно оцінювати a_1, \dots, a_p і μ . Така процедура теоретично більш ефективна. Однак у такій формі модель виявляється нелінійною за параметрами, і ця обставина, як і при оцінюванні МА-моделей, потребує застосування нелінійного методу найменших квадратів (NLLS – *nonlinear least squares*) і чисельних ітераційних методів оптимізації.

2.5. Діагностика оціненої моделі

Після вибору типу і оцінювання коефіцієнтів моделі здійснюється **діагностика** оціненої моделі, тобто з'ясування того, наскільки добре модель відповідає даним спостережень (адекватна даним спостережень). Це є третім етапом процедури підбору моделі.

Для цілей діагностики можна використовувати цілий ряд різних статистичних процедур, які спрямовані в основному на перевірку гіпотези H_0 про те, що в моделі, яка породжує спостереження, послідовність ε_t дійсно утворить процес «білого шуму».

Нехай на етапі ідентифікації моделі ARMA(p, q) вибір зупинено на моделі ARMA(p, q)

$$a(L)X_t = b(L)\varepsilon_t,$$

$$\text{тобто } X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q},$$

і на другому етапі оцінили її як

$$\hat{a}(L)X_t = \hat{b}(L)\varepsilon_t,$$

$$\text{де } \hat{a}(L) = 1 - \hat{a}_1 L - \dots - \hat{a}_p L^p,$$

$$\hat{b}(L) = 1 + \hat{b}_1 L + \dots + \hat{b}_q L^q.$$

Якщо МА складова моделі ARMA(p, q) оборотна, то

$$\varepsilon_t = \frac{a(L)}{b(L)} X_t,$$

і оцінки для ε_t теоретично можна одержати заміною $a(L)$ і $b(L)$ на $\hat{a}(L)$ і $\hat{b}(L)$ відповідно:

$$\hat{\varepsilon}_t = \frac{\hat{a}(L)}{\hat{b}(L)} X_t.$$

На практиці можна використовувати цю формулу лише частково, оскільки нескінченний ряд у правій частині доводиться обривати через наявність тільки скінченної кількості спостережень.

При великій кількості спостережень поведження $\hat{\varepsilon}_t$ повинно імітувати поведження самих ε_t . Отже, якщо помилки ε_t утворюють процес «білого шуму», то залишки мусять імітувати процес «білого шуму».

Ґрунтуючись на цьому міркуванні, Бартлетт [11], Бокс і Пірс [4] запропонували досліджувати статистичну значущість вибірових автокореляцій для ряду інновацій ε_t

$$r_\varepsilon(k) = \frac{\sum_{t=1}^{T-k} \hat{\varepsilon}_t \hat{\varepsilon}_{t+k}}{\sum_{t=1}^T \hat{\varepsilon}_t^2} X_t$$

сумою їхніх квадратів

$$Q_{BP} = T \sum_{k=1}^M r_\varepsilon^2(k).$$

Якщо модель правильно специфікована, то Q_{BP} має розподіл, наближений до розподілу $\chi^2(M - p - q)$ за умови, що T і M великі, а відношення $\frac{M}{T}$ мале. Гіпотезою адекватності підібраної моделі нехтують, якщо $Q_{BP} > \chi_{0,95}^2(M - p - q)$.

Однак згодом було помічено, що при скінченних T розподіл статистики Q_{BP} може істотно відрізнятись від розподілу $\chi^2(M - p - q)$. Використовуючи результати Люнга й Бокса [6], можна показати, що

$$E(Q_{BP}) = (M - p - q) - \frac{M(M + 5)}{2T + 2}.$$

Отже, якщо відношення $\frac{M(M + 5)}{2T + 2}$ істотне, то використання $\chi^2(M - p - q)$ -наближення не є обґрунтованим.

Люнг і Бокс запропонували два способи подолання проблеми зсуву. Перший – прямий метод – полягає у використанні наближення

$$Q_{BP} = \chi^2(E(Q_{BP})) ,$$

де для $E(Q_{BP})$ використовується зазначений вище вираз (модифікований критерій Бокса – Кокса).

Другий спосіб ураховує більш точний вираз для $D(r_\varepsilon(k))$: замість $\frac{1}{T}$ беруть $\frac{T-k}{T^2+2T}$. Це приводить до Q-статистики Люнга – Бокса

$$Q_{LB} = T(T+2) \sum_{k=1}^M \frac{r_\varepsilon^2(k)}{(T-k)},$$

яка має той самий асимптотичний розподіл $\chi^2(M-p-q)$, що й Q_{BP} , але при скінченних T розподіл статистики Q_{LB} набагато ближче до $\chi^2(M-p-q)$, ніж розподіл статистики Q_{BP} . Отже, якість наближення погіршується, якщо значення параметрів знаходиться поблизу межі стаціонарності або оборотності моделі; особливо це помітно при малих M . Хоча спочатку висновок асимптотичного розподілу статистики Люнга – Бокса робився за припущенням, що ε_t – гауссівський “білий шум”, у подальшому було встановлено, що цей критерій досить стійкий до відхилень розподілу ε_t від нормального. Важливо тільки, щоб дисперсія $D(\varepsilon_t)$ була скінченною. Остання умова часто порушується для рядів, що описують еволюцію фінансових показників, які швидко змінюються, – цін на акції, біржових індексів, обмінних курсів. Для таких рядів розподіл ε_t звичайно має “**важкі хвости**”, тобто досить часто спостерігаються більші за абсолютною величиною значення ε_t . І це потребує залучення для опису таких рядів більш складних моделей.

БІБЛІОГРАФІЧНИЙ СПИСОК

1. Носко В.П. Эконометрика. Введение в регрессионный анализ временных рядов. – М.: Ин-т экономики переходного периода, 2002. – 182 с. (www.ict.ru/mipt/2/text/curs_econometrics.htm)
2. Enders W., *Applied Econometric Time Series*, Wiley, New York, 1995, 416 p.
3. Hamilton, James D. () *Time Series Analysis*, Princeton University Press, Prinseton, 1994, 820p.
4. Box G.E.P., Pierce D.A. “Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models”,

- Journal of the American Statistical Association*, 1970, 65, p. 1509 – 1526.
5. Хеннан Э. *Многомерные временные ряды*: Пер с англ. – М.: Мир, 1974.
 6. Ljung G., G.E.P. Box “On a Measure of Lack of Fit in Time Series Models”, *Biometrika*, 66, 1979, p. 255 - 270.
 7. Akaike “Information Theory and an Extension of the Maximum Likelihood Principle”, in Petrov B.N. and Csöki F. (Eds), *Proceedings, 2 nd International Symposium on Information Theory*, 1973, p. 267 - 281 . Akademia Kiado, Budapest.
 8. Schwarz G. “Estimating the Dimension of a Model”, *The Annals of Statistics*, 16, 1978, p. 461 - 464.
 9. Hannan E.J., Quinn B.G. “The Determination of the Order of an Autoregression”, *Journal of the Royal Statistical Society, Series B*, 41, 1979, p. 190 - 195.
 10. Kavalieris “A Note on Estimating Autoregressive-Moving average Order”, *Biometrika*, 78, 1991, p. 920 - 922.
 11. Bartlett M.S. “On the Theoretical Specification of sampling properties of Autocorrelated Time Series”, *Journal of the Royal Statistical Society, Series B*, 8, 1946, p. 27 - 41.
 12. Ljung G., G.E.P. Box (1 979) “On a Measure of Lack of Fit in Time Series Models”, *Biometrika*, 66, p. 255 - 270.
 13. Носко В.П. *Эконометрика для начинающих*. – М.: Ин-т экономики переходного периода, 2002. – 182 с. (www.ict.ru/mipt/2/text/curs_econometrics.htm)

Душин Борис Ігорович

**МЕТОДИ ЕКОНОМІКО-СТАТИСТИЧНИХ ДОСЛІДЖЕНЬ.
РЕГРЕСІЙНИЙ АНАЛІЗ СТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ**

Редактори: А.М. Ємленінова, Л.О. Кузьменко

Зв. план, 2006

Підписано до друку 17.07.2006.

Формат 60x84 1/16. Папір. офс. № 2. Офс. друк.

Ум. друк. арк. 2,8. Обл.-вид. арк. 3,18. Наклад 100 прим.

Замовлення 404. Ціна вільна

Національний аерокосмічний університет ім. М.Є. Жуковського

«Харківський авіаційний інститут»

61070, Харків-70, вул. Чкалова, 17

<http://www.khai.edu>

Видавничий центр «Х А І»

61070, Харків-70, вул. Чкалова, 17

izdat@khai.edu