

Моделі сталого розвитку

4.2.4. Прогнозування часових рядів за допомогою ARIMA-моделей

Організації і підприємства збирають і зберігають величезні обсяги даних в своїх базах даних. Цю колекцію даних можна ефективно використовувати для пошуку корисних шаблонів (закономірностей), які можуть поліпшити бізнес.

Data Mining — це процес виявлення раніше невиявлених закономірностей на основі величезних обсягів операційних даних. Після того, як шаблони знайдені, їх, крім того, можна використовувється для побудови певних рішень для зростання і поліпшення організацій. Збір даних — це процес прийняття рішень, який складається з різних методів і алгоритмів, включаючи класифікацію, асоціативні правила, класифікацію, часові ряди і т. д. для здобуття знань з баз даних.

Одним з найважливіших завдань для організацій є знаходження приблизного прогнозу на майбутнє. Перший етап в наближенні майбутнього складається з збору даних з минулого. У зв'язку з цим зазвичай мають справу зі статистичними даними, отриманими через регулярні проміжки часу, наприклад, дні, місяці, роки і т. д., які повинні бути однаковими для всіх вимірювань. Зазвичай такі дані називаються часовими рядами.

Часовий ряд — це упорядкований набір спостережень змінної, взятих через рівні проміжки часу. Прогнозування часових рядів — це використання моделі для прогнозування майбутніх подій на основі відомих минулих подій. В історичних даних може бути також відсутнє значення. Це значення має бути заповнено перед прогнозуванням будь-яким з методів, таких як просте середнє, ковзне середнє або зважене ковзне середнє. В часових рядах треба знайти закономірність в історичних даних і екстраполювати цю закономірність на майбутнє. Часовий ряд, що містить спостереження однієї змінної, називається одновимірним, тоді як спостереження кількох змінних називаються багатовимірними.

Методи моделювання часових рядів припускають, що історія повторюється якщо не точно, то досить близько, щоб, вивчаючи минуле, можна було приймати більш правильні рішення в майбутньому. Часові ряди можуть мати такі характеристики, як тренд, циклічність, сезонність і нерегулярність. Залежно від характеристик, представлених даними часового ряду, відповідна модель буде застосовуватися до даних часового ряду. Поступове зміщення рівня ряду вгору чи вниз, або тенденція значень ряду збільшуватися або зменшуватися з плином часу, називається віковим трендом або просто трендом. Сезонні коливання в часовому ряду — це коливання у році протягом сезону. Сере-

Розділ 4. Економіко-математичне моделювання

дньострокові зміни в ряду, викликані обставинами, повторюваними циклами, називаються циклічними варіаціями. Непередбачувані впливи, які не є регулярними, а також не повторюються в певному зразку, називаються нерегулярними варіаціями.

У традиційному аналізі часових рядів для опису цих чотирьох компонентів використовуються два підходи: мультиплікативний і адитивний.

Мультиплікативна модель:

$$Y = T \cdot S \cdot C \cdot I,$$

адитивна модель:

$$Y = T + S + C + I,$$

де Y позначає спостереження, а T , S , C і I — тренд, сезонна, циклічна і нерегулярна складові відповідно.

Мультиплікативна модель заснована на припущенні, що чотири компоненти часового ряду не обов'язково незалежні і можуть впливати один на одного, тоді як в адитивній моделі передбачається, що чотири компоненти часового ряду не залежать одне від одного.

Одним із факторів, який впливає на вибір методу прогнозування є встановлений для прогнозу часовий горизонт. Збільшення термінів прогнозу для досліджуваного процесу збільшує дисперсію випадкової складової. Сукупна похибка прогнозу може бути розкладена на три складові:

- похибки, які виникають при визначенні форми зв'язку, тобто ступеню співпадання певної функції із наявним трендом;
- похибки при визначенні параметрів вибраної в якості тренду функції (похибки, які залежать від методу аналітичного виразу залежності);
- випадкові відхилення від тренду, спричинені індивідуальними особливостями процесу.

Правильно вибрана форма аналітичного виразу та найбільш точний метод визначення параметрів прогнозної моделі значною мірою підвищують оцінку точності прогнозу. Точність прогнозної моделі визначається на основі похибок прогнозу: MAD (Mean Absolute Derivation), MSE (Mean Squared Error), MAPE (Mean Absolute Percentage Error), MPE (Mean Percentage Error). Очевидно, що зазначені показники можуть бути розраховані після того, як період упередження (часовий горизонт прогнозу) вже завершився та є в наявності фактичні дані про прогнозований показник, або значення показника на ретроспективній ділянці. В останньому випадку наявна інформація ділиться на дві частини: на основі першої — оцінюються параметри моделі, а дані другої частини розглядаються в якості фактичних. Похибки прогнозів, отримані ретроспективно (на дру-

Моделі сталого розвитку

гій ділянці) характеризують точність відповідної моделі.

Сьогодні доступні різні методи прогнозування, такі як методи ковзної середньої, експоненційне згладжування, регресійний підхід, інтегрована модель авторегресії — ковзного середнього ARIMA, нейронні мережі, і т.д.

Останнім часом популярними стали ARIMA-моделі (Autoregressive Integrated Moving Average), які пояснюють поведінку часового ряду, виходячи лише з його значень в попередні моменти часу, а також добре описують як стаціонарні, так і нестаціонарні часові ряди.

Як зазначає І. Лук'яненко, широкому застосуванню ARIMA-моделей сприяє те, що [1]:

1) не завжди попередня інформація про можливі взаємозв'язки між динамічними рядами може бути добре обґрунтована; у цьому випадку чисто статистична модель, що зв'яже поточні та попередні значення досліджуваного показника, може використовуватися для короткострокових прогнозів;

2) інколи з добре відомих структурних моделей економічної теорії можна отримати моделі типу авторегресійних або моделей ковзного середнього, особливо у процесі оцінювання наведеної форми симультивних систем рівнянь, тобто під час вираження ендогенних змінних структурної моделі через попередньо визначені та екзогенні змінні.

Значна частина методів аналізу зорієнтована на стаціонарні процеси, статистичні властивості яких не змінюються протягом часу (середнє та дисперсія постійні у випадку нормального розподілу). Однак багато часових рядів мають нестаціонарний характер. Оскільки, більшість часових рядів можна привести до стаціонарних після виділення тренду, сезонної компоненти чи взяття різниці, тому стаціонарним часовим рядам приділяється велика увага.

Розглянемо різні приклади нестаціонарних часових рядів.

Спочатку розглянемо тренд. Нехай часовий ряд можна представити

$$y_t = \alpha + \beta t + \varepsilon_t, \quad (4.2.1)$$

Тут ряд y_t представлено у вигляді композиції детермінованої складової $\alpha + \beta t$ (лінійний тренд) і випадкової складової ε_t , що є стаціонарним часовим рядом з нульовою середньою. Також зустрічаються інші приклади тренду: квадратичний $\alpha + \beta t + \gamma t^2$, експоненційний $\alpha e^{\beta t}$ і т. п.

Для того щоб виділити тренд в моделі (і їй подібних), можна застосувати звичайну техніку оцінювання параметрів регресійних рівнянь, вважаючи t незалежною змінною. Після цього ми отримуємо ряд

Розділ 4. Економіко-математичне моделювання

залишків, для опису якого можна буде застосувати моделі стаціонарних часових рядів.

Розглянемо сезонність. В економічних даних часто зустрічається сезонна компонента. Наприклад, в кварталних даних може спостерігатися сезонна компонента з періодом 4:

$$y_t = S(t) + \varepsilon_t, S(t+4) = S(t). \quad (4.2.2)$$

Тут ряд y_t представлений в вигляді композиції періодичної детермінованої складової $S(t)$ (сезонна компонента) і випадкової складової ε_t , що є стаціонарним часовим рядом з нульовою середньою. Сезонну компоненту $S(t)$ можна представити в вигляді

$$S(t) = \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t},$$

де d_i — фіктивні (бінарні) змінні для кварталів. Для виділення сезонної компоненти можна застосувати методи оцінювання параметрів регресії до рівняння:

$$y_t = \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \varepsilon_t. \quad (4.2.3)$$

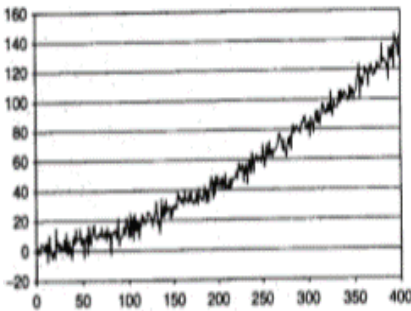
Часто модель (4.2.3) представляють в вигляді регресії з обмеженням, включаючи в неї константу:

$$y_t = \alpha + \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \varepsilon_t, \quad \sum \beta_i = 0, \quad (4.2.4)$$

В (4.2.4) коефіцієнти β_i представляють відхилення від середнього за рік рівня в кварталі i .

Як і в випадку виділення тренду, методи моделювання стаціонарних часових рядів застосовуються далі до ряду залишків регресії.

Типові приклади графіків нестационарних часових рядів приведені на рисунку 4.2.1.



а)



б)

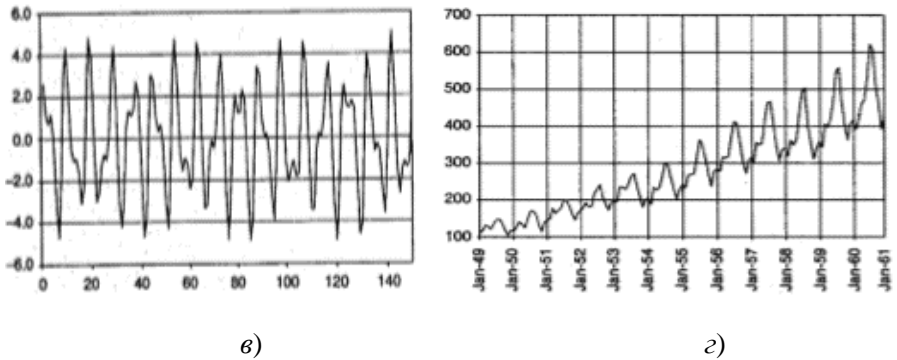


Рис. 4.2.1. Типові приклади графіків нестационарних часових рядів:

а) тренд, б) випадкове блукання, в) сезонність, г) тренд і сезонність

Розглянемо взяття послідовної різниці. Випадковий процес $y_t = y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$, $t = 1, \dots, n$ є прикладом нестационарного часового ряду. Проте, якщо до нього застосувати операцію взяття послідовної різниці, отримаємо стаціонарний часовий ряд:

$$z_t = \Delta y_t = (1 - L)y_t = y_t - y_{t-1}, \quad z_t = \varepsilon_t.$$

Для більш загального процесу $A(L)y_t = \varepsilon_t$, де $A(L)$ має один одиничний корінь, тобто, $A(L) = B(L)(1 - L)$ а всі корені $B(L)$ лежать зовні одиничного кола, це ж перетворення $z_t = \Delta y_t$ приводить до стаціонарного процесу $B(L)z_t = \varepsilon_t$.

Взяття різниці також приводить до стаціонарного процесу ряд (4.2.1) з лінійним трендом:

$$\Delta y_t = \beta + u_t, \quad u_t = \Delta \varepsilon_t = \varepsilon_t - \varepsilon_{t-1}. \quad (4.2.5)$$

У випадку квадратичного тренду $\alpha + \beta t + \gamma t^2$ взяття першої різниці не приводить до стаціонарного ряду, але якщо взяти другу різницю $\Delta^2 y_t = \Delta(\Delta y_t) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$, то $\Delta y_t = \beta + \gamma(2t - 1) + \Delta \varepsilon_t$, $\Delta^2 y_t = 2\gamma + \Delta^2 \varepsilon_t$ і $\Delta^2 y_t$ вже є стаціонарним часовим рядом.

У випадку наявності сезонної компоненти (4.2.2) усунути останню можна з допомогою оператора взяття сезонної послідовної різниці $\Delta_4 y_t = (1 - L^4)y_t = y_t - y_{t-4}$. Якщо період сезонної компоненти дорівнює 12 (це може бути для місячних даних), то треба застосувати оператор Δ_{12} .

Слід відмітити, що застосування оператора послідовної різниці не обов'язково приводить нестационарний ряд до стаціонарного.

Таким чином, застосовуючи виділення тренду, сезонності і / або

Розділ 4. Економіко-математичне моделювання

оператор послідовної (і сезонної) різниці, часто можна отримати з вихідного часового ряду стаціонарний.

Яким чином визначають за наявними спостереженнями, чи є ряд стаціонарним.

Перше — подивитись на графік отриманих спостережень. Можливо, він містить очевидний на око тренд або періодичну компоненту (сезонність). Також можливо, що розкид спостережень зростає чи спадає з часом. Це може служити вказівкою на залежність середньої або відповідно дисперсії від часу. В обох випадках ряд буде, найшвидше, нестаціонарний.

Друге — побудувати графік вибіркової автокореляційної функції (ACF) або корелограми (correlogram):

$$r_k = \hat{\rho}_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad k = 1, 2, \dots \quad (4.2.6)$$

Корелограма стаціонарного часового ряду «швидко спадає» з ростом k після декількох перших значень. Якщо ж графік спадає дуже повільно, то є основи припустити нестаціонарність ряду. Крім ACF, можна також побудувати графік часткової автокореляційної функції PACF, яка теж повинна швидко спадати для стаціонарного процесу.

У загальному випадку, коефіцієнт часткової кореляції між y_t та x_1 при виключенні впливу x_2 визначається за формулою:

$$r(y, x_1 | x_2) = \frac{r(y, x_1) - r(y, x_2)r(x_1, x_2)}{\sqrt{1 - r^2(y, x_2)}\sqrt{1 - r^2(x_1, x_2)}}. \quad (4.2.7)$$

Значення його лежить в інтервалі $[-1; 1]$, як у звичайного коефіцієнта кореляції.

Часткова автокореляційна функція $PACF(k)$ є «чиста кореляція» між y_t і y_{t-k} при виключенні впливу проміжних значень $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$.

Якщо застосувати процедуру обчислення вибіркового часткового коефіцієнта кореляції, то виявляється, що у випадку стаціонарного ряду y_t значення вибіркової часткової автокореляційної функції $PACF(k)$ обчислюється як МНК-оцінка останнього коефіцієнта β_k в $AR(k)$ регресійному рівнянні:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_k y_{t-k} + \varepsilon_t, \quad (4.2.8)$$

Третє — можна використовувати формальні тести на наявність одиничного кореня (тест Дікі – Фуллера DF, розширений тест Дікі – Фуллера ADF, тест МакКінлі і ін.).

Моделі сталого розвитку

Розглянемо частковий випадок моделі Бокса-Дженкінса — ARMA моделі для стаціонарних часових рядів.

Нехай є наступний клас моделей стаціонарних рядів:

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim iid(0, \sigma^2), \quad (4.2.9)$$

або в короткій формі

$$\Phi(L)y_t = \delta + \Theta(L)\varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2), \quad (4.2.10)$$

де $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ і $\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ — поліноми від оператора зсуву. Така модель називається моделлю авторегресії і ковзної середньої — ARMA(p, q).

Розглянемо прості ARMA моделі.

Процес AR(1) або ARMA(1,0) представляється формулою:

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2). \quad (4.2.11)$$

Його числові характеристики

$$E(y_t) = \frac{\delta}{1 - \phi_1}, \quad V(y_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}, \quad \gamma_k = \phi_1^k \gamma_0, \\ \rho_k = \gamma_k / \gamma_0 = \phi_1^k \quad (4.2.12)$$

Необхідною умовою стаціонарності процесу y_t є умова $|\phi_1| < 1$.

Часткова автокореляційна функція AR(1) процесу дорівнює нулю для значень $k > 1$.

Авторегресійний процес AR(2) або ARMA(2,0):

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2). \quad (4.2.13)$$

Нехай вільний член рівняння дорівнює нулю:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2). \quad (4.2.14)$$

$PACF(k) = 0$ для $k > 2$. Для AR(p) процесу часткова автокореляційна функція $PACF(k) = 0$, починаючи з $k = p + 1$. Необхідно наголосити, що цей результат вірний для теоретичної часткової автокореляційної функції і не завжди вірний для вибіркової часткової автокореляційної функції. Проте, на практиці відбувається різке спадання $PACF$ до значень близьких до нуля після k , що дорівнює порядку авторегресійного процесу.

Розглянемо процеси ковзної середньої. Моделлю ковзної середньої (MA) порядку q називається модель ARMA(0, q)

$$y_t = \delta + \Theta(L)\varepsilon_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim iid(0, \sigma^2), \quad (4.2.15)$$

яка позначається MA(q). З (4.2.15) видно, що процес MA(q) стаціонарний при будь-якому q і будь-яких θ_i .

Розділ 4. Економіко-математичне моделювання

Розглянемо $MA(1)$

$$y_t = \delta + \Theta(L)\varepsilon_t = \delta + \varepsilon_t - \theta_1\varepsilon_{t-1}, \varepsilon_t \sim iid(0, \sigma^2). \quad (4.2.16)$$

$MA(1)$ в вигляді авторегресійного процесу:

$$\Theta(L)^{-1}y_t = \Theta(L)^{-1}\delta + \varepsilon_t \quad (4.2.17)$$

або

$$y_t = \frac{\delta}{1-\theta_1} - \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \dots + \varepsilon_t. \quad (4.2.18)$$

Таке $AR(\infty)$ представлення $MA(1)$ процесу (4.2.16) можливе тільки у випадку оборотності оператора $\Theta(L) = 1 - \theta L$, тобто, коли виконується умова оборотності $|\theta| < 1$.

Середня і дисперсія процесу $MA(1)$:

$$E(y_t) = \delta, \quad V(y_t) = \sigma^2(1 + \theta_1^2). \quad (4.2.19)$$

Автокореляційна функція процесу $MA(1)$:

$$\gamma_1 = Cov(y_t, y_{t-1}) = -\theta_1\sigma^2. \quad (4.2.20)$$

Автокореляційна функція $ACF(k) = 0$ для $k > 1$, тобто її вигляд аналогічний вигляду $PACF$ для $AR(q)$ процесу.

Часткова автокореляційна функція $PACF(k)$ для $MA(q)$ процесу, аналогічно $ACF(k)$ для $AR(q)$ процесу, експоненційно спадає. Таким чином, має місце деяка симетрія: пара графіків ($ACF, PACF$) для $MA(q)$ процесу мають такий же вигляд, як пара графіків ($PACF, ACF$) для $AR(q)$.

Слід відмітити, що подібно $AR(\infty)$ представленню (4.2.18) для $MA(1)$ процесу (4.2.16) існує $MA(\infty)$ представлення для $AR(1)$ процесу (4.2.11):

$$y_t = (1 - \phi_1 L)^{-1}(\delta + \varepsilon_t) = \frac{\delta}{1 - \phi_1} + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \dots \quad (4.2.21)$$

Розглянемо найпростіший змішаний $ARMA(1,1)$ процес (4.2.10) з $\Phi(L) = 1 - \phi_1 L$, і $\Theta(L) = 1 - \theta_1 L$:

$$y_t - \phi_1 y_{t-1} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \varepsilon_t \sim iid(0, \sigma^2). \quad (4.2.22)$$

Будем вважати, що $|\phi_1| < 1$ і $|\theta_1| < 1$. Процес $ARMA(1,1)$ є стаціонарним і зворотним.

Середня, дисперсія і коваріація $ARMA(1,1)$ процесу:

$$E(y_t) = \frac{\delta}{1 - \phi_1} \quad (4.2.23)$$

Моделі сталого розвитку

$$\gamma_0 = V(y_t) = \sigma^2 \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2} \quad (4.2.24)$$

$$\gamma_1 = Cov(y_t, y_{t-1}) = \phi_1\gamma_0 - \theta_1\sigma^2 \quad (4.2.25)$$

Для автокореляцій порядку більше 1 отримуємо рекурентне співвідношення

$$\begin{aligned} \gamma_k = Cov(y_t, y_{t-k}) &= \phi_1\gamma_{k-1}, \rho_k = \phi_1^{k-1}\rho_1, k > 1, \\ \rho_1 &= \frac{(1 - \phi_1\theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1\theta_1}. \end{aligned} \quad (4.2.26)$$

З (4.2.26) видно, що ACF для $ARMA(1,1)$ процесу веде себе так як ACF для $AR(1)$ процесу. Хоча значення ρ_1 інше, але співвідношення між ρ_1 і наступними значеннями ACF точно такі ж.

Цей висновок можна узагальнити на випадок $ARMA(p, q)$ процесу. Перші q значень ACF визначаються взаємодією AR і MA компонент, а подальша поведінка автокореляційної функції така ж, як у $AR(p)$ процесі.

Аналогічний висновок справедливий для часткової автокореляційної функції $ARMA(p, q)$ процесу. Вона спадає подібно $PACF$ для $MA(q)$ процесу.

Узагальнимо вище приведені результати у вигляді таблиці 4.2.1.

Табл. 4.2.1. Поведінка автокореляційної і часткової автокореляційної функцій для різних процесів $ARMA$

Процеси	Функції	
	Автокореляційна функція (корелограма)	Часткова автокореляційна функція
$AR(1)$	Експоненційно спадає	$PACF(k) = 0$ при $k > 1$
$AR(2)$	Експоненційно спадає	$PACF(k) = 0$ при $k > 2$
$AR(p)$	Експоненційно спадає	$PACF(k) = 0$ при $k > p$
$MA(1)$	$ACF(k) = 0$ при $k > 1$	Експоненційно спадає
$MA(2)$	$ACF(k) = 0$ при $k > 2$	Експоненційно спадає
$MA(q)$	$ACF(k) = 0$ при $k > q$	Експоненційно спадає
$ARMA(p, q)$	Експоненційно спадає при $k > q$	Експоненційно спадає при $k > p$

Розділ 4. Економіко-математичне моделювання

Для вибору найкращої ARIMA-моделі переважно застосовують алгоритм за методом Бокса-Дженкінса. Він передбачає побудову ARIMA-моделі на основі наявного динамічного ряду. В деяких випадках навіть акцентують увагу на тому, що чим більша довжина динамічного ряду, тим буде кращою модель.

Деякі нестационарні часові ряди можуть бути приведені до стаціонарних з допомогою оператора послідовної різниці. Нехай, часовий ряд y_t після застосування d раз оператора послідовної різниці, став стаціонарним рядом $\Delta^d y_t$, що задовольняє ARMA(p, q) моделі (4.2.10). Тоді процес y_t називається інтегрованим процесом авторегресії і ковзної середньої ARIMA(p, d, q). Слід відмітити, що з моделі, наприклад, для ряду Δy_t легко отримати модель для вихідного ряду y_t , використовуючи співвідношення $y_t = y_{t-1} + \Delta y_t$.

Методологія Бокса-Дженкінса підбору ARIMA моделі для даного ряду спостережень складається з трьох етапів.

1. Ідентифікація моделі

1.1. Перший крок — отримання стаціонарного ряду. Тестується ряд на стаціонарність, використовуючи описані вище методи: візуальний аналіз графіку, візуальний аналіз ACF і PACF, тести на одиничні корені. Якщо отримується стаціонарний ряд, то переходимо до наступного пункту, якщо ні — застосовуємо оператор взяття послідовної різниці і повторяєм тестування. На практиці послідовна різниця береться, як правило, не більше двох разів.

1.2. Після того як отримано стаціонарний часовий ряд, будуються його вибіркові ACF і PACF, які дозволяють сформулювати декілька гіпотез про можливі порядки авторегресії p і ковзної середньої q . Зазвичай, використовуються моделі більш низьких порядків, як правило, з $p + q \leq 3$ (якщо немає сезонної компоненти).

Вибіркові ACF і PACF, зазвичай, не співпадають з теоретичними аналогами, але достатньо близькі до них.

2. Оцінювання моделі і перевірка адекватності моделі

2.1. Для кожної з вибраних на першому етапі моделей оцінюються їх параметри і обчислюються залишки.

2.2. Кожна з моделей перевіряється, наскільки вона відповідає даним. З моделей, адекватних даним, вибирається найпростіша (з меншою кількістю параметрів).

3. Прогнозування

Моделі сталого розвитку

Після того, як на другому етапі вибрана модель, можна будувати прогноз на один або декілька кроків за часом і оцінювати довірчі межі прогнозних значень.

У сучасних комп'ютерних пакетах є різні методи оцінювання ARMA моделей, такі, як лінійний або нелінійний МНК, повний або умовний метод максимальної правдоподібності.

Прогнозування обсягів портфельних інвестицій в Україну

Визначимо:

- найбільш оптимальну модель ARIMA для прогнозування;
- прогнозний обсяг портфельних інвестицій.

В якості інформаційної бази застосовувалися дані про щоквартальні обсяги портфельних інвестицій за період з I кварталу 1998 року по I квартал 2014 року (65 даних), які були опубліковані Національним банком України.

Вихідний часовий ряд портфельних інвестицій в Україну має наступний вигляд (рис. 4.2.2).

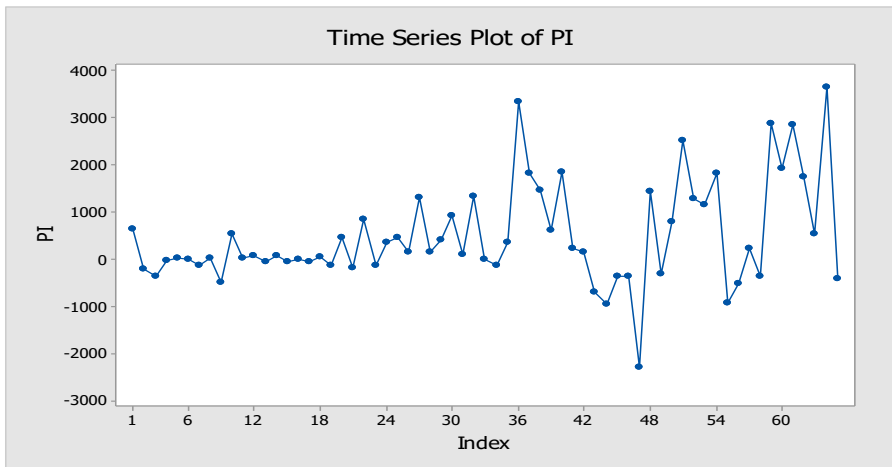


Рис. 4.2.2. Обсяг портфельних інвестицій в Україну (млн. дол. США)

Аналіз функції вибіркової автокореляції вихідних даних свідчить: коефіцієнти автокореляції швидко згасають, це означає, що вихідний ряд портфельних інвестицій є стаціонарним. Функція вибіркової часткової автокореляції, свідчить що перший, другий і четвертий коефіцієнти часткової автокореляції є значущі, всі інші — незначущі і вони плавно наближаються до нуля. В сукупності структура функцій вибір-

Розділ 4. Економіко-математичне моделювання

кової автокореляції і вибіркової часткової автокореляції відповідає процесу ARIMA(2,0,4).

Перевірка моделі на адекватність через аналіз залишків свідчить про те, що коефіцієнти автокореляції залишків є незначними (рис.4.2.3), знаходяться в межах нуля в діапазоні $\pm 2/\sqrt{n}$, а отже можемо зробити висновок — вибрана модель є адекватною. Прогноз для 66 і 67 періоду становить $\hat{Y}_{66} = 221,05$ млн. дол. США і $\hat{Y}_{67} = 576,34$ млн. дол. США відповідно.

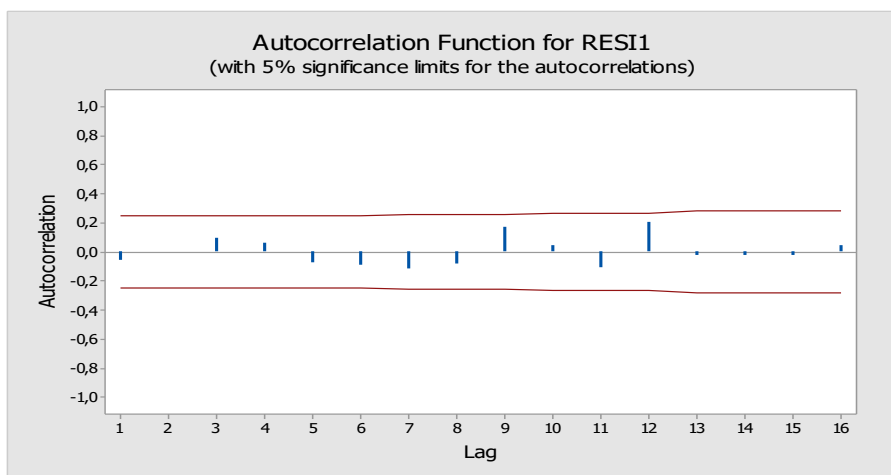


Рис. 4.2.3. Автокореляційна функція залишків для моделі ARIMA (2,0,4)

Оскільки набір автокореляційних функцій може відрізнятися від теоретичних функцій внаслідок варіацій вибірки, то ми вирішили також побудувати моделі з авторегресією і ковзним середнім нижчих порядків задля принципу економії, який полягає в переважанні простої моделі більш складній [2]. Показники їх точності наведені в таблиці 4.2.2.

Табл. 4.2.2. Показники точності моделей

Модель	MAPE	MAD	MSD
ARIMA (1, 0, 0)	2,250921	706,0783	1152435
ARIMA (2, 0, 0)	1,590415	670,3394	1012981
ARIMA (4, 0, 0)	1,997346	657,9827	940211,5
ARIMA (2, 0, 4)	3,471811	653,9083	913834,8
ARIMA (4, 0, 4)	1,49423	619,2669	835757,9

Моделі сталого розвитку

Як видно з таблиці 4.2.2 за показниками точності найкращою є модель високих порядків ARIMA (4, 0, 4). Якщо вибрати модель нижчих порядків, то за показником MAPE кращою є модель ARIMA (2, 0, 0). Проте прогнозні значення за цією моделлю для 66 і 67 періодів суттєво відрізняються від вище отриманих: вони становлять $\hat{Y}_{66} = 1328,48$ млн. дол. США і $\hat{Y}_{67} = 151,08$ млн. дол. США.

Прогнозування обсягів іноземних інвестицій на основі довгих динамічних рядів дає більш точний результат, оскільки прогнозування за допомогою моделей ARIMA здійснюється шляхом аналізу інформації, яка міститься в передісторії часового ряду. Відповідно чим довший часовий ряд, тим більша кількість інформації є доступною для аналізу та побудови якісного прогнозу.

Моделі сталого розвитку

33. Бутко М. Методологія оцінки ролі експортного потенціалу в економічному розвитку регіону [Електронний ресурс] / М. Бутко, М. Козік // Економіст. 2015. Режим доступу: http://nbuv.gov.ua/jpdf/econ_2015_10_4.pdf
34. Михайлов В.С. Деякі методологічні питання побудови індексів зовнішньої торгівлі в Україні / В. Михайлов // Статистика України, 2004. № 2. С. 21–23.

4.2.4

1. Лук'яненко І. Прогнозування податкових надходжень за допомогою моделей корегування помилки / І. Лук'яненко, Ю. Городніченко // Фінанси України. 2001. № 7. С. 89–9]
2. Березька К. М. Деякі аспекти прогнозування обсягів портфельних інвестицій в Україну / К. М. Березька, В.В. Маслій // Матеріали VIII-ої міжнародної конференції «Актуальні проблеми економіки 2014», м. Київ, 12 грудня 2014 р. К.: Національна академія управління, 2015. С. 7 – 12.

4.2.5–4.2.7

1. Братусь А. С., Новожилов А.С., Платонов А. П. Динамические системы и модели биологии. Москва : Физматлит, 2010. 400 с.
2. Алілуйко А.М. Дослідження динаміки взаємодії підприємств з використанням конкурентної моделі Лоткі–Вольтерра. Східно–Європейський журнал передових технологій. 2013. №1/3 (61). С. 25–29.
3. Алілуйко А.М. Дослідження конкурентної взаємодії на ринку послуг мобільного зв'язку. Інноваційна економіка. 2013. №2 (40). С. 221–226.
4. Vossara N. Modeling complex systems. New York : Springer–Verlag, 2003. 397 p.
5. Колесников А. А. Синергетические методы управления сложными системами: Теория системного синтеза. Москва : КомКнига, 2006. 240 с.