

Зміст

Вступ	7
1 Часові ряди в економіці	12
1.1 Означення випадкового процесу	12
1.2 Поняття часового ряду	13
1.3 Математичне сподівання і дисперсія	14
1.4 Автоковаріації часового ряду	16
1.5 Стаціонарність	17
1.6 Ергодичність	19
2 Стаціонарні ARMA процеси	22
2.1 Процес білого шуму	22
2.2 Процес рухомого середнього першого порядку .	24
2.3 Процеси рухомого середнього вищих порядків .	28
2.4 Процеси рухомого середнього нескінченного по- ряду	30
2.5 Авторегресійний процес першого порядку	33
2.6 Авторегресійний процес другого порядку	38
2.7 Авторегресійні процеси p - го порядку	49
2.8 Мішані моделі авторегресії і рухомого середнього	53
2.9 Векторна авторегресійна модель	57
3 Властивості ARMA процесів	67
3.1 Умови стаціонарності ARMA процесів	67

3.2	Одиничні корені. <i>ARIMA</i> процеси	71
3.3	Тотожні корені	76
3.4	Автоковаріаційна генеруюча функція	79
3.5	Зворотність процесів часових рядів	85
3.5.1	Зворотність <i>MA(1)</i> процесу	85
3.5.2	Зворотність <i>MA(q)</i> процесу	91
3.6	Сумування <i>ARMA</i> процесів	93
3.6.1	Додавання <i>MA(1)</i> процесу і білого шуму	93
3.6.2	Сумування двох процесів рухомих середніх	98
3.6.3	Сумування двох авторегресійних процесів	101
3.7	Трикутна факторизація матриць	103
3.7.1	Трикутна факторизація додатно визначених симетричних матриць	103
3.7.2	Єдиність трикутної факторизації	110
3.7.3	Факторизація Колеського	110
3.7.4	Трикутна факторизація матриці автокореляцій для <i>MA(1)</i> процесу	112
4	Ідентифікація моделі	115
4.1	Тестування стаціонарності	116
4.1.1	Параметричні тести	117
4.1.2	Непараметричні тести	122
4.2	Автокореляційна функція <i>ACF</i>	127
4.3	Часткова автокореляційна функція <i>PACF</i>	131
4.4	Моделювання Бокса - Дженкінса	136
4.5	Діагностування моделі	139
4.6	Придатність моделі до прогнозування	141
4.7	Статистичні критерії для вибору моделі	142
5	Оцінювання <i>ARMA</i> моделей	145
5.1	Метод найменших квадратів	145
5.2	Оцінювання за допомогою коефіцієнтів автокореляції	147
5.2.1	Оцінювання параметрів <i>AR(p)</i> процесу	147

5.2.2	Оцінювання параметрів $MA(q)$ процесу	149
5.2.3	Оцінювання параметрів $ARMA(p, q)$ процесу	151
5.3	Оцінки максимальної правдоподібності	152
5.3.1	Функція правдоподібності Гаусівського $AR(1)$ процесу	152
5.3.2	Функція правдоподібності Гаусівського $AR(p)$ процесу	159
5.3.3	Функція правдоподібності Гаусівського $MA(1)$ процесу	164
5.3.4	Функція правдоподібності Гаусівського $MA(q)$ процесу	169
5.3.5	Функція правдоподібності Гаусівського $ARMA(p, q)$ процесу	172
5.4	Тестування гіпотез на підставі ОМП	173
5.4.1	Асимптотичні стандартні похибки оцінок максимальної правдоподібності	173
5.4.2	Тест відношення правдоподібності	174
5.4.3	Тест множників Лагранжа	175
6	Прогнозування	177
6.1	Принципи прогнозування	177
6.1.1	Прогнози, які ґрунтуються на умовних математичних сподіваннях	177
6.1.2	Прогнози, які ґрунтуються на лінійній проекції	180
6.1.3	Властивості лінійної проекції	182
6.1.4	Прогнозовані вектори	183
6.2	Прогнози, які ґрунтуються на нескінченній кількості спостережень	184
6.2.1	Прогнози, побудовані на запізненнях ε	184
6.2.2	Прогнози, побудовані на запізненнях y . Формула прогнозування Вінера - Колмогорова	187

6.3	Застосування формули Вінера - Колмогорова	189
6.3.1	Прогнозування $AR(1)$ процесу	189
6.3.2	Прогнозування $AR(p)$ процесу	191
6.3.3	Прогнозування $MA(1)$ процесу	193
6.3.4	Прогнозування $MA(q)$ процесу	194
6.3.5	Прогнозування $ARMA(1, 1)$ процесу	195
6.3.6	Прогнозування $ARMA(p, q)$ процесу	197
6.4	Прогнози, побудовані на скінченій кількості спостережень	198
6.4.1	Наближення оптимальних прогнозів	198
6.4.2	Точні скінченороджені прогнози	200
6.5	Уточнення лінійної проекції	203
6.5.1	Зв'язок між лінійною проекцією і трикутною факторизацією коваріаційної матриці	203
6.5.2	Уточнення скінченородженого прогнозу для $MA(1)$ процесу	209
6.5.3	Блочна трикутна факторизація	214
6.6	Оптимальні прогнози для Гауссівських процесів	217
7	Нестационарні часові ряди	221
7.1	Процеси з детермінованим часовим трендом і процеси з одиничним коренем	221
7.2	Тести на одиничний корінь	227
7.2.1	Тестування одиничного кореня в $AR(1)$ моделі	232
7.2.2	Тестування на одиничний корінь в авторегресійних моделях вищих порядків	237
7.3	Порівняння трендово-стаціонарних процесів і процесів з одиничним коренем	239
7.3.1	Порівняння прогнозованих значень	239
7.3.2	Порівняння похибок прогнозу	242
7.3.3	Порівняння динамічних множників	245

7.3.4	Різниця у двох специфікаціях стосовно перетворення для досягнення стаціонарності	246
7.4	Інші підходи до моделювання нестационарних часових рядів	248
7.4.1	Дробове інтегрування	248
7.4.2	Випадкові розриви в трендах	251
8	Процеси з детермінованим часовим трендом	253
8.1	Асимптотичний розподіл МНК оцінки простої моделі часового тренда	253
8.2	Тестування гіпотез для моделі часового тренда	264
8.3	Асимптотичні висновки для авторегресійного $AR(p)$ процесу з детермінованим часовим трендом	269
8.3.1	Перетворення регресії	269
8.3.2	Асимптотичний розподіл МНК оцінок параметрів перетвореної регресії	273
8.3.3	Асимптотичний розподіл МНК оцінки для початкової регресії	277
8.3.4	Тестування гіпотез	279
9	Одновимірні процеси з одиничним коренем	281
9.1	Асимптотичні властивості МНК оцінки параметра Гауссівського $AR(1)$ процесу	281
9.2	Броунівський рух	285
9.3	Асимптотичні властивості авторегресії першого порядку з одиничним коефіцієнтом	291
9.4	Тести Дікі - Фуллера	293
9.4.1	Регресія без сталої і часового тренда у випадку, коли істинний процес є випадковим блуканням	293
9.4.2	Регресія, яка містить сталу, але не містить часового тренда, у випадку, коли істинний процес є випадковим блуканням	297

9.4.3	Регресія, яка містить сталу, але не містить часового тренда у випадку, коли істинний процес є випадковим блуканням зі зміщенням	304
9.4.4	Регресія, яка містить сталий член і часовий тренд. Істинний процес є випадковим блуканням зі зміщенням або без нього	308
9.5	Тест Філіпса - Перрона	313
Список літератури		321

Вступ

Широке коло соціально - економічних, технічних і природничих процесів часто зображають набором послідовних значень y_1, y_2, \dots, y_T , що зафіксовані в рівновіддалені між собою моменти часу $t = 1, 2, \dots, T$. Тобто, маємо спостереження економічного показника впорядковані за часом, зокрема, це можуть бути щомісячні, щоквартальні або щорічні дані. Такий набір значень зазвичай називається часовим рядом.

У реальному житті зміни значень y_t в часі відбуваються під впливом певних причин і факторів. Проте у зв'язку з їх чисельністю, складністю вимірювання, нерозробленістю теоретичних передумов стосовно взаємозв'язків зі змінною y обґрунтувати і побудувати для опису процесу y_t ($t = 1, 2, \dots, T$) багатофакторну економетричну модель класичного типу не завжди можливо. В результаті щодо ряду y_t висувається припущення, що сукупний вплив цих факторів формує внутрішні закономірності у розвитку процесу y_t . Це дає змогу застосувати для його опису економетричну модель із специфічного класу моделей часових рядів.

Моделі часових рядів активно застосовують у дослідженнях динаміки значної кількості реальних процесів різноманітної природи. Їх часто використовують у дослідженні динаміки пасажиропотоків, попиту на різні види продукції, міграційних процесів, у радіотехніці, моделюванні природних подій: динаміки кількості сонячних плям, природних катастроф і багатьох інших процесів.

Найчастіше моделі часових рядів застосовують у дослідженні фінансових ринків, аналізі динаміки фінансових показників, прогнозуванні цін на різноманітні товари, курси акцій, співвідношення курсів валют і т.д.

Загальне припущення при моделюванні за допомогою часових рядів полягає в тому, що значення процесу y_t значною

мірою визначає його передісторія, тобто величина показника y_t залежить від його попередніх значень y_{t-1}, y_{t-2}, \dots відповідно до характерних для заданого ряду закономірностей. Математично це припущення можна виразити за допомогою такого загального рівняння

$$y_t = f(y_{t-1}, y_{t-2}, \dots, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \quad (1)$$

де ε_t – похибка моделі в момент часу t . У цій моделі поточне значення змінної є функцією лише від набору її минулих значень. Отже, маємо одновимірну модель, в якій інформація про інші змінні не враховується.

Функція f виражає характер взаємозв'язку, який склався в часовому ряді $y_t, t = 1, 2, \dots$. У випадку вдалого вибору цієї функції права частина виразу (1) буде в деякому сенсі близькою до реальних значень цього ряду. Для широкого кола процесів функція f має лінійний вигляд. В цьому посібнику ми розглянемо питання побудови лінійних моделей часових рядів і проаналізуємо їхні основні властивості.

Існує тісний зв'язок між економетричними моделями й одновимірними моделями часових рядів. Розглянемо приклад простої моделі економіки країни, яка задається такими рівняннями:

$$Y_t = C_t + I_t + G_t, \quad (2)$$

$$C_t = \alpha + \beta Y_t + \varepsilon_t, \quad (3)$$

$$I_t = \delta (Y_t - Y_{t-1}), \quad (4)$$

$$G_t = \rho Y_{t-1} + G', \quad (5)$$

де ендогенні змінні Y_t, C_t, I_t, G_t є, відповідно, реальним доходом, споживанням, інвестиціями та урядовими витратами у момент часу t , а $\alpha, \beta, \delta, \rho, G'$ – додатні сталі. Зауважимо, що в рівняння (4), (5) входить минуле значення однієї ендогенної змінної, а саме значення реального доходу Y_{t-1} у момент часу

$t - 1$. Збурення ε_t є білим шумом, тобто $E[\varepsilon_t] = 0$ для всіх t , $E[\varepsilon_t \varepsilon_s] = 0$ для $t \neq s$, а величина $E[\varepsilon_t^2] = \sigma^2$ є сталою.

У загальному випадку можна розв'язати систему рівнянь, одержавши рівняння зведеного вигляду, в яких кожна ендогенна змінна записана як функція від усіх екзогенних змінних і значень ендогенних змінних системи у попередні моменти часу. Отже, підставивши (3), (4) та (5) у (2), отримаємо рівняння для доходу у зведеному вигляді

$$Y_t = \lambda + \pi Y_{t-1} + u_t, \quad (6)$$

де

$$\lambda = \frac{\alpha + G'}{1 - \beta - \delta}, \quad \pi = \frac{\rho - \delta}{1 - \beta - \delta}, \quad u_t = \frac{\varepsilon_t}{1 - \beta - \delta}.$$

У (6) величина Y виражена через своє попереднє значення, стало та збурення. Отже, (6) є одновимірним зображенням для Y , яке відоме як авторегресія першого порядку або $AR(1)$ процес. Отже, ми показали, що економічну модель (2) - (5) можна звести до простої одновимірної моделі.

Якщо застосувати оператор зсуву L (у часі), визначений рівняннями

$$LY_t = Y_{t-1}, \quad L^n Y_t = Y_{t-n},$$

то можемо записати (6) як

$$Y_t = \lambda + \pi LY_t + u_t$$

або

$$Y_t = (1 - \pi L)^{-1} \lambda + (1 - \pi L)^{-1} u_t. \quad (7)$$

Якщо $-1 < \pi < 1$, то підстановка в (7) розвинення у геометричний ряд

$$(1 - \pi L)^{-1} = 1 + \pi L + \pi^2 L^2 + \pi^3 L^3 + \dots$$

дає рівняння

$$Y_t = (1 - \pi L)^{-1} \lambda + u_t + \pi u_{t-1} + \pi^2 u_{t-2} + \pi^3 u_{t-3} + \dots \quad (8)$$

У (8) значення Y виражається як функція від сталої та нескінченної кількості похибок, які є білим шумом. Якщо ряд можна записати таким способом, то кажуть, що це є процес рухомого середнього або *МА* процес.

Виразивши з (3) Y через C і підставивши замість Y_t та Y_{t-1} у (6), можемо в результаті перетворень переписати (6) у вигляді

$$C_t = (\alpha - \alpha\pi + \beta\lambda) + \pi C_{t-1} + \varepsilon_t - \pi \varepsilon_{t-1} + \beta u_t.$$

Тут C виражене через своє значення у попередній момент та значення залишків у поточний і попередній моменти часу. Процеси такого типу відомі як авторегресійні процеси рухомого середнього або *ARMA* процеси.

З цього простого прикладу видно, як дві з ендогенних змінних структурної моделі Y та C можна зобразити у вигляді одновимірних часових рядів і виразити лише у термінах їх власних минулих значень та випадкових похибок. Все це можна також застосувати до решти ендогенних змінних I та G . Звичайно, якщо змінити модель, включивши додаткові ендогенні змінні і значення інших змінних у попередні моменти часу, то зміняться також одновимірні зображення. Хоча адекватна модель економіки в цілому невідома, але якою б модель не була, при виконанні певних умов кожна з ендогенних змінних має одновимірне зображення. Тому одним з аргументів на користь використання одновимірної моделі є те, що її можна одержати зі структурної моделі.

У багатьох галузях економіки та підприємництва існує потреба робити прогнози. Звичайно, найкращий за інших рівних умов метод прогнозування значень заданої змінної величини

такий: використовуючи правильну економічну теорію, побудувати структурну економетричну модель, оцінити її параметри на підставі точних даних і застосувати цю модель для передбачення майбутніх значень змінної. Оскільки за побудовою така модель втілює конкретну економічну теорію, то вона повинна давати прогноз, який переважає прогнози, одержані іншими методами.

Однак на практиці рідко трапляється така ідеальна ситуація і не завжди можливо сконструювати економетричну модель. По-перше, дослідник може бути непевним стосовно вибору відповідної економічної теорії. По-друге, надійні дані про значення змінних, які відносяться до моделі, можуть не існувати. По-третє, вартість побудови та оцінювання економетричної моделі може бути більшою за очікуваний прибуток від неї. Тому шукають дешевші методи прогнозування. Однією з таких процедур одновимірного прогнозування є методи часових рядів.

Позаяк в одновимірних моделях інформація про інші змінні не враховується, то для прогнозу на наступний період дотатньо знати лише минулі значення цієї змінної. Тому ще одним аргументом на користь одновимірної моделі є те, що вони дають простий і недорогий метод побудови прогнозів. Особливо методи часових рядів корисні для прогнозування значень екзогенних змінних в економетричних моделях, на основі яких вже потім прогнозують ендогенну змінну.

Розділ 1

Часові ряди в економіці

Вихідною точкою у моделюванні рядів спостережень, впорядкованих у часі, є припущення, що ці величини породжені випадковим або стохастичним процесом.

1.1 Означення випадкового процесу

Будь-яка економічна або фінансова структура розвивається і змінюється протягом деякого часу. Якщо цю структуру можна охарактеризувати в кожний момент часу певними числовими параметрами, то одержимо набір випадкових величин, які змінюються з часом, тобто випадковий процес.

Теорія випадкових процесів вивчає випадкові величини, які залежать від параметра t з деякої множини Q .

Означення 1.1.1. Якщо кожному t з множини Q поставлено у відповідність випадкову величину Y_t , то говорять, що на Q задано випадковий процес $\{Y_t, t \in Q\}$.

Означення 1.1.2. Реалізацією випадкового процесу $\{Y_t, t \in Q\}$ є функція $Y_t : Q \rightarrow \Omega$, яка ставить у відповідність кожному $t \in Q$ одне з можливих значень $y_t \in \Omega$ (Ω –

множина можливих значень величин Y_t).

Ми розглядатимемо випадкові процеси, для яких $Q = \mathbb{Z}$ і інтерпретуватимемо t як час.

1.2 Поняття часового ряду

Часовий ряд – це послідовність числових даних (скалярних або векторних), які пов'язані з певними моментами часу і найчастіше корельовані між собою.

Більшість економічних даних можна подати у вигляді часових рядів. Такими будуть, наприклад ціни на певний товар, сподівана вартість грошей, банківські активи, доходи компанії, розмір дивідентів на акції, капітал інвестора тощо, якщо ці параметри простежуються на деякому проміжку часу.

Отже, можемо подати економічні дані у вигляді послідовності значень

$$\{y_1, y_2, \dots, y_T\}, \quad (1.2.1)$$

які розглядаємо як послідовність значень деяких випадкових величин Y_t , $t = 1, \dots, T$. Модель (1.2.1) презентує T окремих чисел, які є лише одним з можливих результатів стохастичного процесу, що породжує дані.

Якщо припускати, що маємо спостережуваний процес для нескінченного періоду часу, то приходимо до послідовності

$$\{y_t\}_{t=-\infty}^{\infty} = \{\dots, y_{-1}, y_0, y_1, \dots, y_T, y_{T+1}, y_{T+2}, \dots\}.$$

Нескінченна послідовність $\{y_t\}_{t=-\infty}^{\infty}$ є реалізацією стохастичного процесу $\{Y_t, t \in (-\infty, +\infty)\}$.

Приклад 1.2.1. Розглянемо вибірку обсягу T : $\{e_1, \dots, e_T\}$. Ці значення трактуватимемо як T реалізацій незалежних однаково розподілених (*i.i.d.*) випадкових величин ε_t ,

$t = 1, \dots, T$. Якщо припустити, що $\varepsilon_t \sim N[0, \sigma^2]$, то ця вибірка є прикладом процесу, який має назву Гаусівського білого шуму (Gaussian white noise) обсягу T .

Розглянемо приклад генерування часових рядів. За допомогою комп'ютера можемо згенерувати нескінченну послідовність випадкових значень з $N[0, \sigma^2]$, $\{y_t^{(1)}\}_{t=-\infty}^{\infty}$. За допомогою іншої спроби генеруємо послідовність $\{y_t^{(2)}\}_{t=-\infty}^{\infty}$. Ці дві послідовності можна розглядати як дві незалежні реалізації Гаусівського білого шуму. Розглянемо множину N таких спроб, що генерують послідовності $\{y_t^{(1)}\}_{t=-\infty}^{\infty}, \{y_t^{(2)}\}_{t=-\infty}^{\infty}, \dots, \{y_t^{(N)}\}_{t=-\infty}^{\infty}$ і виберемо спостереження асоційовані з часом t для кожної послідовності

$$\{y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(N)}\}, \quad t \in (-\infty, +\infty). \quad (1.2.2)$$

Побудовану послідовність можна розглядати як модель з N реалізацій випадкових змінних Y_t .

Випадкові величини Y_t мають деяку щільність $f_{Y_t}(y_t)$, яка називається безумовною щільністю часового ряду.

Наприклад, для Гаусівського білого шуму ця щільність задається такою функцією:

$$f_{Y_t}(y_t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{y_t^2}{2\sigma^2}\right).$$

1.3 Математичне сподівання і дисперсія часового ряду

Математичне сподівання часового ряду, якщо воно існує, дорівнює

$$E[Y_t] \equiv \int_{-\infty}^{\infty} y_t f_{Y_t}(y_t) dy_t. \quad (1.3.1)$$

Математичне сподівання $E[Y_t]$ можемо також визначити як ймовірнісну границю середніх

$$E[Y_t] = \text{plim} \frac{1}{N} \sum_{i=1}^N y_t^{(i)}. \quad (1.3.2)$$

Приклад 1.3.1. Нехай $\{y_t\}_{t=-\infty}^{\infty}$ зображається як сума сталої μ і Гаусівського білого шуму $\{\varepsilon_t\}_{t=-\infty}^{\infty}$, тобто

$$y_t = \mu + \varepsilon_t. \quad (1.3.3)$$

Математичне сподівання цього ряду таке

$$E[y_t] = \mu + E[\varepsilon_t] = \mu. \quad (1.3.4)$$

Приклад 1.3.2. Нехай y_t – часовий тренд δt плюс Гаусівський білий шум

$$y_t = \delta t + \varepsilon_t, \quad (1.3.5)$$

тоді математичне сподівання становитиме

$$E[y_t] = \delta t. \quad (1.3.6)$$

Значення $E[y_t]$ називають безумовним математичним сподіванням часового ряду y_t і позначають μ_t : $E[y_t] = \mu_t$. Зазначимо, що математичне сподівання може бути функцією від часу спостереження t . Зокрема, для процесу (1.3.5), який містить часовий тренд, математичне сподівання є функцією часу, тоді як математичне сподівання процесу (1.3.3) не залежить від t .

Дисперсія часового ряду – це дисперсія випадкової змінної Y_t , яка позначається γ_{0t} і визначається як

$$\gamma_{0t} \equiv E[(Y_t - \mu_t)^2] = \int_{-\infty}^{\infty} (y_t - \mu_t)^2 f_{Y_t}(y_t) dy_t. \quad (1.3.7)$$

Наприклад, для процесу (1.3.5) дисперсія дорівнює

$$\gamma_{0t} = E[(y_t - \delta t)^2] = E[\varepsilon_t^2] = \sigma^2.$$

1.4 Автоковаріації часового ряду

Для кожної окремої реалізації $\{y_t^{(i)}\}_{t=-\infty}^{\infty}$ часового ряду побудуємо вектор $x_t^{(i)}$, який визначений у момент часу t і містить $k + 1$ спостережень $y_t^{(i)}$

$$x_t^{(i)} \equiv \begin{pmatrix} y_t^{(i)} \\ y_{t-1}^{(i)} \\ \vdots \\ y_{t-k}^{(i)} \end{pmatrix}.$$

Кожна реалізація $\{y_t^{(i)}\}_{t=-\infty}^{\infty}$ задає певне значення вектора $x_t^{(i)}$ і ми можемо обчислити ймовірнісний розподіл векторів x_t . За допомогою цього розподілу можна обчислити k -ту автоковаріацію Y_t , яка позначається γ_{kt}

$$\begin{aligned} \gamma_{kt} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (y_t - \mu_t)(y_{t-k} - \mu_{t-k}) \times \\ &\quad \times f_{Y_t, Y_{t-1}, \dots, Y_{t-k}}(y_t, y_{t-1}, \dots, y_{t-k}) dy_t dy_{t-1} \dots dy_{t-k} = \\ &= E[(Y_t - \mu_t)(Y_{t-k} - \mu_{t-k})]. \end{aligned} \quad (1.4.1)$$

Вираз (1.4.1) описує коваріацію Y_t з його власним лаговим значенням Y_{t-k} , тому γ_{kt} називається автоковаріацією. Зауважимо, що в (1.4.1) нульова автоковаріація є дисперсією змінної Y_t , як і передбачалося виразом для γ_{0t} в (1.3.7).

Автоковаріація описується як другий момент процесу y_t . Зокрема, автоковаріацію γ_{kt} можна визначити як $(1, k+1)$ елемент коваріаційної матриці вектора x_t .

Автоковаріацію можна також розглядати як ймовірнісну границю

$$\gamma_{kt} = \text{plim} \frac{1}{N} \sum_{i=1}^N (y_t^{(i)} - \mu_t)(y_{t-k}^{(i)} - \mu_{t-k}). \quad (1.4.2)$$

Наприклад, обчислимо автоковаріації процесу з прикладу

1.3.1

$$\gamma_{kt} = E[(y_t - \mu)(y_{t-k} - \mu)] = E[\varepsilon_t \varepsilon_{t-k}] = 0 \quad \text{для } k \neq 0,$$

тобто всі автоковаріації процесу (1.3.3) є нульовими.

1.5 Стаціонарність

Означення 1.5.1. Якщо математичне сподівання μ_t і автоковаріації γ_{kt} для всіх k не залежать від часу, то часовий ряд y_t називається коваріаційно - стаціонарним або слабко стаціонарним.

Тобто, для коваріаційно - стаціонарного процесу

$$\begin{aligned} E[Y_t] &= \mu && \text{для всіх } t, \\ E[(Y_t - \mu)(Y_{t-k} - \mu)] &= \gamma_k && \text{для всіх } t \text{ і будь-якого } k. \end{aligned}$$

Приклад 1.5.1. Процес (1.3.3) є коваріаційно - стаціонарним, оскільки

$$E[y_t] = \mu,$$

$$E[(y_t - \mu)(y_{t-k} - \mu)] = \begin{cases} \sigma^2, & \text{для } k = 0, \\ 0, & \text{для } k \neq 0. \end{cases}$$

Процес (1.3.5) не є коваріаційно - стаціонарним, тому що математичне сподівання δt є функцією часу.

Зауважимо таке: коли процес коваріаційно - стаціонарний, то коваріація між Y_t і Y_{t-k} залежить тільки від k , проміжку часу між спостереженнями, і не залежить від t , часу спостереження.

Лема 1.5.1. Для коваріаційно - стаціонарного процесу γ_k і γ_{-k} збігаються, тобто

$$\gamma_k = \gamma_{-k} \quad \text{для всіх цілих } k. \quad (1.5.1)$$

Д о в е д е н н я . З означення коваріації маємо

$$\gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)]. \quad (1.5.2)$$

Якщо процес є коваріаційно - стаціонарним, тоді ця величина є однаковою для будь-якого t . Тому можемо замінити t на $t+k$

$$\begin{aligned} \gamma_k &= E[(Y_t - \mu)(Y_{t-k} - \mu)] = \\ &= E[(Y_{t+k} - \mu)(Y_{(t+k)-k} - \mu)] = E[(Y_{t+k} - \mu)(Y_t - \mu)] = \\ &= E[(Y_t - \mu)(Y_{t+k} - \mu)] = E[(Y_t - \mu)(Y_{t-(-k)} - \mu)]. \end{aligned}$$

Але за означенням коваріації (1.4.1) останній вираз – це γ_{-k} . Отже, для кожного коваріаційно - стаціонарного процесу виконується (1.5.1). \square

Крім слабкої стаціонарності, розрізняють сильну стаціонарність.

Означення 1.5.2. Процес називається сильно стаціонарним, якщо для різних j_1, j_2, \dots, j_n сукупний розподіл $(Y_t, Y_{t+j_1}, Y_{t+j_2}, \dots, Y_{t+j_n})$ залежить тільки від інтервалів часу (j_1, j_2, \dots, j_n) і не залежить від самого t . Зокрема, у цьому випадку розподіл Y_t не залежатиме від t .

Зауважимо, якщо процес сильно стаціонарний з обмеженими другими моментами, то він є коваріаційно - стаціонарним. Це випливає з того, що щільність, яка визначена в (1.3.1) і (1.4.1), не залежить від часу, тому моменти μ_t і γ_{kt} не будуть залежати від часу.

Обернене твердження не правильне. Можна зазначити коваріаційно - стаціонарний процес, який не є сильно стаціонарним. Зокрема, математичне сподівання і автоковаріації можуть не бути функціями часу, але вищі моменти, зокрема $E[Y_t^3]$, залежатимуть від t .

Означення 1.5.3. Процес $\{Y_t\}$ називається Гаусівським, якщо сукупна щільність

$$f_{Y_t, Y_{t+j_1}, \dots, Y_{t+j_n}}(y_t, y_{t+j_1}, \dots, y_{t+j_n})$$

є щільністю нормального стандартного багатовимірного розподілу для всіх j_1, \dots, j_n .

Оскільки Гаусівський розподіл повністю визначається математичним сподіванням і коваріаційною матрицею, то коваріаційно - стаціонарний Гаусівський процес є сильно стаціонарним.

Надалі термін "стаціонарність" будемо використовувати для коваріаційно - стаціонарних процесів.

1.6 Ергоди́чність

Ми зображали математичне сподівання і автоковаріації часових рядів у термінах вибірових середніх, таких як в (1.3.2) і (1.4.2). Розглянемо реалізації процесу розміру T : $\{y_1^{(i)}, y_2^{(i)}, \dots, y_T^{(i)}\}$, $i = 1, 2, \dots$. З цих спостережень для кожного i можна обчислити середнє $\bar{y}^{(i)}$. Це середнє є часовим середнім

$$\bar{y}^{(i)} \equiv \frac{1}{T} \sum_{t=1}^T y_t^{(i)}, \quad i = 1, 2, \dots \quad (1.6.1)$$

Виникає питання, чи для стаціонарного ряду часові середні (1.6.1) збігаються до $E[Y_t]$.

Означення 1.6.1. Коваріаційно - стаціонарний процес називається ергоди́чним для математичного сподівання, якщо послідовність (1.6.1) збігається за ймовірністю до $E[Y_t]$ при $T \rightarrow \infty$.

Процес буде ергоди́чним для математичного сподівання тоді, коли автоковаріації γ_k прямують до 0 зі зростанням k .

Можна довести таке: якщо для коваріаційно - стаціонарного процесу виконується умова

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty, \quad (1.6.2)$$

то часовий ряд y_t є ергодичним для математичного сподівання.

Означення 1.6.2. Коваріаційно - стаціонарний процес називається ергодичним для других моментів, якщо

$$\frac{1}{T-k} \sum_{t=k+1}^T (y_t - \mu)(y_{t-k} - \mu) \xrightarrow{p} \gamma_k$$

для кожного k .

В окремих випадках, зокрема, якщо y_t є стаціонарним Гаусівським процесом, то умова (1.6.2) є достатньою для забезпечення ергодичності для усіх моментів.

У багатьох застосуваннях стаціонарність і ергодичність вимагають виконання тих самих умов. Проте можна навести приклад стаціонарного, але не ергодичного процесу.

Приклад 1.6.1. Нехай середнє $\mu^{(i)}$ для i -тої реалізації $\{y_t^{(i)}\}_{t=-\infty}^{\infty}$ є нормально розподіленою випадковою величиною з $N(0, \lambda^2)$, тобто

$$y_t^{(i)} = \mu^{(i)} + \varepsilon_t, \quad (1.6.3)$$

де ε_t – Гаусівський білий шум з математичним сподіванням 0 і дисперсією σ^2 , що не залежить від $\mu^{(i)}$. Зауважимо, що

$$\mu_t = E[\mu^{(i)}] + E[\varepsilon_t] = 0.$$

Також

$$\gamma_{0t} = E[(\mu^{(i)} + \varepsilon_t)^2] = \lambda^2 + \sigma^2$$

і

$$\gamma_{kt} = E[(\mu^{(i)} + \varepsilon_t)(\mu^{(i)} + \varepsilon_{t-k})] = \lambda^2 \quad \text{для кожного } k \neq 0.$$

Отже, процес (1.6.3) коваріаційно - стаціонарний. Але для нього не виконується умова ергоди́чності для математичних сподівань. Справді, часові середні

$$\frac{1}{T} \sum_{t=1}^T y_t^{(i)} = \frac{1}{T} \sum_{t=1}^T (\mu^{(i)} + \varepsilon_t) = \mu^{(i)} + \frac{1}{T} \sum_{t=1}^T \varepsilon_t$$

збігаються до $\mu^{(i)}$, а не до математичного сподівання y_t , яке дорівнює нулю $E[y_t] = 0$.

Розділ 2

Стаціонарні ARMA процеси

2.1 Процес білого шуму

Розглянемо послідовність випадкових змінних $\{\varepsilon_t\}_{t=-\infty}^{\infty}$, елементи якої мають математичне сподівання 0, дисперсію σ^2 , тобто

$$E[\varepsilon_t] = 0, \quad (2.1.1)$$

$$\text{Var}[\varepsilon_t] = \sigma^2 \quad (2.1.2)$$

і не корелюють з часом

$$\text{Cov}[\varepsilon_t, \varepsilon_s] = 0 \quad \text{для } t \neq s. \quad (2.1.3)$$

Процес, який задовольняє умови (2.1.1) - (2.1.3), називають білим шумом (white noise).

Зауважимо таке: якщо для всіх t виконується умова (2.1.1), то

$$\begin{aligned} \text{Var}[\varepsilon_t] &= E[\varepsilon_t^2], \\ \text{Cov}[\varepsilon_t, \varepsilon_s] &= E[\varepsilon_t \varepsilon_s]. \end{aligned}$$

Іноді умову (2.1.3) заміняють сильнішою умовою

$$\varepsilon_t \text{ і } \varepsilon_s \text{ — незалежні при } t \neq s. \quad (2.1.4)$$

Зазначимо, що з (2.1.4) випливає (2.1.3), але не навпаки. Якщо, крім (2.1.1) - (2.1.3), виконується умова

$$\varepsilon_t \sim N[0, \sigma^2], \quad (2.1.5)$$

то процес називають Гаусівським білим шумом (Gaussian white noise).

На рис. 2.1.1. зображено динаміку значень часового ряду, який описується процесом білого шуму.

Рис. 2.1.1. Білий шум $y_t = \varepsilon_t$

2.2 Процес рухомого середнього першого порядку $MA(1)$

Нехай послідовність ε_t є білим шумом. Розглянемо процес

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \quad (2.2.1)$$

де μ і θ – деякі сталі. Цей часовий ряд називають процесом рухомого середнього першого порядку (a first-order moving average process) і позначають $MA(1)$.

Тоді

$$\begin{aligned} E[y_t] &= E[\mu + \varepsilon_t + \theta\varepsilon_{t-1}] = \\ &= \mu + E[\varepsilon_t] + \theta E[\varepsilon_{t-1}] = \mu, \end{aligned} \quad (2.2.2)$$

$$\begin{aligned} \gamma_0 = \text{Var}[y_t] &= E[(y_t - \mu)^2] = \\ &= E[\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2] = (1 + \theta^2)\sigma^2. \end{aligned} \quad (2.2.3)$$

Крім того, перша автоковаріація дорівнює

$$\begin{aligned} \gamma_1 &= E[(y_t - \mu)(y_{t-1} - \mu)] = \\ &= E[\varepsilon_t\varepsilon_{t-1} + \theta\varepsilon_{t-1}^2 + \theta\varepsilon_t\varepsilon_{t-2} + \theta^2\varepsilon_{t-1}\varepsilon_{t-2}] = \theta\sigma^2, \end{aligned} \quad (2.2.4)$$

а автоковаріації вищих порядків дорівнюють нулю

$$\begin{aligned} \gamma_k &= E[(y_t - \mu)(y_{t-k} - \mu)] = \\ &= E[(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-k} + \theta\varepsilon_{t-k-1})] = 0 \quad \text{для } k > 1. \end{aligned} \quad (2.2.5)$$

Оскільки математичне сподівання і автоковаріації $MA(1)$ процесу не є функціями часу, то цей процес коваріаційно - стаціонарний незалежно від значення θ .

Враховуючи, що

$$\sum_{k=0}^{\infty} |\gamma_k| = \gamma_0 + |\gamma_1| = (1 + \theta^2)\sigma^2 + |\theta|\sigma^2 < \infty,$$

матимемо таке: якщо ε_t – Гаусівський білий шум, то $MA(1)$ процес (2.2.1) є ергодичним для всіх моментів.

Для коваріаційно - стаціонарного процесу k -ту автокореляцію визначають формулою

$$\rho_k = \text{Corr}[y_t, y_{t-k}] = \frac{\text{Cov}[y_t, y_{t-k}]}{\sqrt{\text{Var}[y_t]}\sqrt{\text{Var}[y_{t-k}]}} = \frac{\gamma_k}{\gamma_0}. \quad (2.2.6)$$

Оскільки ρ_k є кореляцією, то згідно з нерівністю Коші-Шварца $|\rho_k| \leq 1$ для всіх k . Зазначимо, що нульова автокореляція $\rho_0 = 1$ для всіх коваріаційно - стаціонарних процесів.

З (2.2.3) і (2.2.4) випливає, що для процесу $MA(1)$

$$\rho_1 = \frac{\theta}{1 + \theta^2}, \quad \rho_k = 0 \quad \text{для } k = 2, 3, \dots \quad (2.2.7)$$

На рис. 2.2.2, 2.2.3 зображено динаміку часових рядів, які описуються $MA(1)$ процесами $y_t = \varepsilon_t + 0,4\varepsilon_{t-1}$ і $y_t = \varepsilon_t - 0,9\varepsilon_{t-1}$ відповідно.

Зауважимо, що зображення рядів для різних $\theta > 0$ подібні між собою. Також подібними є зображення рядів для різних $\theta < 0$. Якщо ж порівнювати зображення рядів з додатним і від'ємним значенням параметра θ відповідно, то з рис. 2.2.2 і 2.2.3 видно, що зображення ряду з $\theta < 0$ є густішим, ніж для $\theta > 0$.

Автокореляції ρ_k як функції від k зображені на рис. 2.2.4 (ліворуч для ряду з $\theta = 0,4$ і праворуч для ряду з $\theta = -0,9$).

Рис. 2.2.2. $MA(1)$: $y_t = \varepsilon_t + 0,4\varepsilon_{t-1}$

Рис. 2.2.3. $MA(1)$: $y_t = \varepsilon_t - 0,9\varepsilon_{t-1}$

Рис. 2.2.4. Автокореляції ρ_k , $k = 0, 1, 2, \dots, 5$

Розглянемо залежність першої автокореляції ρ_1 $MA(1)$ процесу від параметра процесу θ .

Згідно з (2.2.7) при різних θ отримуємо різні значення першої автокореляції ρ_1 . Додатні θ спричиняють додатню автокореляцію в рядах. У цьому випадку незвично великі значення y_t породжують більше за середнє значення y_{t+1} , а мале значення y_t породжує менше за середнє значення y_{t+1} . З іншого боку, від'ємне значення θ означає від'ємну автокореляцію - велике y_{t+1} може бути спричинене малим значенням y_t .

Залежність автокореляції ρ_1 від θ показано на рис. 2.2.5.

Рис. 2.2.5. Залежність ρ_1 від параметра $MA(1)$ процесу

Зауважимо, що найбільшим можливим значенням $\rho_1 \in 0,5$ при $\theta = 1$. Найменшим значенням $\rho_1 \in -0,5$ при $\theta = -1$. Для інших значень θ значення ρ_1 містяться між $-0,5$ і $0,5$, причому є два різних значення θ , що можуть спричинити ту саму автокореляцію. Це випливає з того, що значення $\theta/(1+\theta^2)$ не змінюється, якщо θ замінити на $1/\theta$

$$\rho_1 = \frac{(1/\theta)}{1 + (1/\theta)^2} = \frac{\theta^2(1/\theta)}{\theta^2[1 + (1/\theta)^2]} = \frac{\theta}{\theta^2 + 1}.$$

Наприклад, процеси

$$y_t = \varepsilon_t + 0,5\varepsilon_{t-1}$$

і

$$y_t = \varepsilon_t + 2\varepsilon_{t-1}$$

мають однакову функцію автокореляції

$$\rho_1 = \frac{2}{(1+2^2)} = \frac{0,5}{1+0,5^2} = 0,4.$$

2.3 Процеси рухомого середнього вищих порядків

Нехай знову ε_t є білим шумом. Розглянемо процес

$$y_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q}, \quad (2.3.1)$$

де $\mu, \theta_1, \dots, \theta_q$ – деякі сталі. Цей часовий ряд називають процесом рухомого середнього порядку q (a q th-order moving average process) і позначають $MA(q)$. Тоді

$$E[y_t] = E[\mu] + E[\varepsilon_t] + \sum_{j=1}^q \theta_j E[\varepsilon_{t-j}] = \mu,$$

$$\begin{aligned} \gamma_0 &= E[(y_t - \mu)^2] = & (2.3.2) \\ &= E[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q})^2] = \\ &= (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2. \end{aligned}$$

Для $k = 1, 2, \dots, q$

$$\begin{aligned} \gamma_k &= E[(\varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q}) \times & (2.3.3) \\ &\quad \times (\varepsilon_{t-k} + \theta_1\varepsilon_{t-k-1} + \theta_2\varepsilon_{t-k-2} + \dots + \theta_q\varepsilon_{t-k-q})] = \\ &= E[\theta_k\varepsilon_{t-k}^2 + \theta_{k+1}\theta_1\varepsilon_{t-k-1}^2 + \theta_{k+2}\theta_2\varepsilon_{t-k-2}^2 + \dots + \theta_q\theta_{q-k}\varepsilon_{t-q}^2]. \end{aligned}$$

Очевидно, що для $k > q$ всі γ_k дорівнюють нулю. Отже,

$$\gamma_k = \begin{cases} (\theta_k + \theta_{k+1}\theta_1 + \theta_{k+2}\theta_2 + \dots + \theta_q\theta_{q-k})\sigma^2, & k = 1, 2, \dots, q, \\ 0, & k > q. \end{cases} \quad (2.3.4)$$

Наприклад, для процесу $MA(2)$

$$\begin{aligned} \gamma_0 &= (1 + \theta_1^2 + \theta_2^2)\sigma^2, \\ \gamma_1 &= (\theta_1 + \theta_2\theta_1)\sigma^2, \\ \gamma_2 &= \theta_2\sigma^2, \\ \gamma_3 &= \gamma_4 = \dots = 0. \end{aligned}$$

Відповідно автокореляції $MA(2)$ процесу дорівнюють

$$\begin{aligned} \rho_0 &= 1, \\ \rho_1 &= \frac{\theta_1 + \theta_2\theta_1}{1 + \theta_1^2 + \theta_2^2}, \\ \rho_2 &= \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}, \\ \rho_3 &= \rho_4 = \dots = 0. \end{aligned} \quad (2.3.5)$$

Тобто, всі автокореляції порядку вищого за 2 дорівнюють нулю.

З (2.3.2) і (2.3.4) видно, що для довільних значень $(\theta_1, \theta_2, \dots, \theta_q)$ процес $MA(q)$ коваріаційно - стаціонарний. Крім того, виконується умова ергодичності

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty.$$

На рис. 2.3.6 зображено динаміку часового ряду, що описується $MA(2)$ процесом $y_t = \varepsilon_t - 0,5\varepsilon_{t-1} + 0,4\varepsilon_{t-2}$. Автокореляції цього процесу ρ_k , $k = 0, 1, 2, \dots, 6$ як функції від k зображено зліва на рис. 2.3.7. У правій частині рис. 2.3.7 зображено автокореляції процесу $MA(4)$, а саме $y_t = \varepsilon_t - 0,7\varepsilon_{t-1} + 0,3\varepsilon_{t-2} - 0,5\varepsilon_{t-3} + 0,5\varepsilon_{t-4}$.

Рис. 2.3.6. $MA(2)$: $y_t = \varepsilon_t - 0,5\varepsilon_{t-1} + 0,4\varepsilon_{t-2}$

Рис. 2.3.7. Автокореляції ρ_k , $k = 0, 1, 2, \dots, 8$

2.4 Процеси рухомого середнього нескінченного порядку

Процес $MA(q)$ можна записати у вигляді

$$y_t = \mu + \sum_{j=0}^q \theta_j \varepsilon_{t-j},$$

де $\theta_0 = 1$. Якщо спрямувати q до нескінченності, то одержимо процес

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}, \quad (2.4.1)$$

який називають процесом $MA(\infty)$ (an infinite-order moving average process). Будемо використовувати символи ψ для позначення коефіцієнтів процесів рухомого середнього нескінченного порядку і θ для позначення коефіцієнтів процесів рухомого середнього скінченного порядку.

Припускаємо, що коефіцієнти процесу є абсолютно сумовними, тобто виконується умова

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \quad (2.4.2)$$

Тоді

$$\begin{aligned} E[y_t] &= \lim_{q \rightarrow \infty} E[\mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q}] = \\ &= \mu, \end{aligned} \quad (2.4.3)$$

$$\begin{aligned} \gamma_0 &= E[(y_t - \mu)^2] = \\ &= \lim_{q \rightarrow \infty} E[(\psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q})^2] = \\ &= \lim_{q \rightarrow \infty} (\psi_0^2 + \psi_1^2 + \psi_2^2 + \dots + \psi_q^2) \sigma^2 = \\ &= \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty, \end{aligned} \quad (2.4.4)$$

$$\begin{aligned} \gamma_k &= E[(y_t - \mu)(y_{t-k} - \mu)] = \\ &= \sigma^2 (\psi_k \psi_0 + \psi_{k+1} \psi_1 + \psi_{k+2} \psi_2 + \psi_{k+3} \psi_3 + \dots) = \\ &= \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k} < \infty. \end{aligned} \quad (2.4.5)$$

Зауважимо, що абсолютна сумовність $\{\psi_j\}_{j=0}^{\infty}$ (умова 2.4.2) і існування другого моменту $E[\varepsilon_t^2]$ є достатніми умовами для того, щоб могли переставити операції обчислення математичного сподівання і сумування в рівностях (2.4.3) – (2.4.5). Це впливає з такого твердження: що якщо послідовність випадкових величин $\{Y_t\}_{t=1}^{\infty}$ володіє властивістю

$$\sum_{t=1}^{\infty} E[|Y_t|] < \infty,$$

то

$$E \left\{ \sum_{t=1}^{\infty} Y_t \right\} = \sum_{t=1}^{\infty} E[Y_t].$$

Отже, ми довели лему.

Лема 2.4.1. *Якщо виконується умова (2.4.2), то $MA(\infty)$ процес коваріаційно - стаціонарний.*

Дослідимо ергодичність $MA(\infty)$ процесу.

Лема 2.4.2. *Якщо виконується умова (2.4.2), то процес $MA(\infty)$ ергодичний для всіх моментів.*

Д о в е д е н н я . Справді, оскільки

$$\gamma_k = \sigma^2 \sum_{j=0}^{\infty} \psi_{k+j} \psi_j,$$

то

$$\sum_{k=0}^{\infty} |\gamma_k| \leq \sigma^2 \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} |\psi_{k+j}| |\psi_j| = \sigma^2 \sum_{j=0}^{\infty} |\psi_j| \sum_{k=0}^{\infty} |\psi_{k+j}|.$$

На підставі умови (2.4.2) існує така стала M , що

$$\sum_{k=0}^{\infty} |\psi_{k+j}| \leq M$$

для всіх $j = 0, 1, 2, \dots$. Тому

$$\sum_{k=0}^{\infty} |\gamma_k| \leq \sigma^2 M \sum_{j=0}^{\infty} |\psi_j| \leq \sigma^2 M^2.$$

Тобто $MA(\infty)$ процеси з абсолютно сумовними коефіцієнтами мають абсолютно сумовні автоковаріації

$$\sum_{k=0}^{\infty} |\gamma_k| < \infty. \quad (2.4.6)$$

Отже, $MA(\infty)$ процеси, які задовольняють (2.4.2) ергодичні для математичного сподівання.

Якщо ж всі ε_t є Гаусівськими, тоді процес $MA(\infty)$ ергодичний для усіх моментів. \square

2.5 Авторегресійний процес першого порядку

Процес вигляду

$$y_t = c + \phi y_{t-1} + \varepsilon_t, \quad (2.5.1)$$

де c і ϕ – деякі сталі, а ε_t – білий шум, називають авторегресійним процесом першого порядку (a first-order autoregressive process) і позначають $AR(1)$.

Підставляючи послідовно значення y_t у праву частину (2.5.1), одержимо вираз

$$\begin{aligned} y_t &= (c + \varepsilon_t) + \phi(c + \varepsilon_{t-1}) + \phi^2(c + \varepsilon_{t-2}) + \dots = \\ &= c(1 + \phi + \phi^2 + \phi^3 + \dots) + \\ &\quad + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \dots \end{aligned} \quad (2.5.2)$$

Якщо $|\phi| < 1$, то задовольняється умова (2.4.2), оскільки ряд

$$\sum_{j=0}^{\infty} |\psi_j| = \sum_{j=0}^{\infty} |\phi|^j$$

є збіжним і його сума дорівнює $1/(1-|\phi|)$. Отже, процес (2.5.2) можна трактувати як процес $MA(\infty)$, де $\psi_j = \phi^j$.

Надалі припускатимемо, що умова $|\phi| < 1$ виконується. Тоді

$$\begin{aligned} E[y_t] &= c(1 + \phi + \phi^2 + \phi^3 + \dots) + 0 + 0 + \dots = \\ &= \frac{c}{1 - \phi}, \end{aligned}$$

тобто

$$\mu = \frac{c}{1 - \phi}. \quad (2.5.3)$$

Крім того,

$$\begin{aligned} \gamma_0 &= E[(y_t - \mu)^2] = E[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots)^2] = \\ &= \sigma^2(1 + \phi^2 + \phi^4 + \phi^6 + \dots) = \frac{\sigma^2}{1 - \phi^2}, \end{aligned} \quad (2.5.4)$$

$$\begin{aligned} \gamma_k &= E[(y_t - \mu)(y_{t-k} - \mu)] = \quad (2.5.5) \\ &= E[(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots + \phi^k\varepsilon_{t-k} + \phi^{k+1}\varepsilon_{t-k-1} + \dots) \times \\ &\quad \times (\varepsilon_{t-k} + \phi\varepsilon_{t-k-1} + \phi^2\varepsilon_{t-k-2} + \dots)] = \\ &= (\phi^k + \phi^{k+2} + \phi^{k+4} + \dots) \sigma^2 = \sigma^2 \frac{\phi^k}{1 - \phi^2}. \end{aligned}$$

З (2.5.4) і (2.5.5) випливає, що

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k. \quad (2.5.6)$$

Додатне значення ϕ , подібно як додатне значення θ для $MA(1)$ процесу, означає додатну кореляцію між y_t і y_{t+1} . Від'ємне значення ϕ означає від'ємну автокореляцію першого порядку, але додатну другого порядку.

У виразах (2.5.3) - (2.5.5) моменти стаціонарного процесу $AR(1)$ одержано шляхом зведення його до процесу $MA(\infty)$. Проте ті самі результати можна одержати на підставі (2.5.1), якщо при цьому наперед припустити коваріаційну стаціонарність процесу. Тобто, припускаємо, що

$$E[y_t] = E[y_{t-1}] = \mu$$

і обчислюємо моменти безпосередньо з (2.5.1). Тоді з

$$E[y_t] = c + \phi E[y_{t-1}] + E[\varepsilon_t]$$

одержимо

$$\mu = c + \phi\mu + 0,$$

або

$$\mu = \frac{c}{1 - \phi}. \quad (2.5.7)$$

Для того щоб знайти другий момент, запишемо (2.5.1) у вигляді

$$y_t = \mu(1 - \phi) + \phi y_{t-1} + \varepsilon_t$$

або

$$(y_t - \mu) = \phi(y_{t-1} - \mu) + \varepsilon_t. \quad (2.5.8)$$

Звідси

$$E[(y_t - \mu)^2] = \phi^2 E[(y_{t-1} - \mu)^2] + 2\phi E[(y_{t-1} - \mu)\varepsilon_t] + E[\varepsilon_t^2]. \quad (2.5.9)$$

Згідно з (2.5.2) $y_{t-1} - \mu$ є лінійною функцією змінних $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$, тобто

$$y_{t-1} - \mu = \varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \dots$$

Але ε_t не корелює з $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$, тому ε_t не корелює з $y_{t-1} - \mu$. Отже,

$$E[(y_{t-1} - \mu)\varepsilon_t] = 0. \quad (2.5.10)$$

З припущення коваріаційної стаціонарності одержуємо, що

$$E[(y_t - \mu)^2] = E[(y_{t-1} - \mu)^2] = \gamma_0. \quad (2.5.11)$$

Підставляючи (2.5.10) і (2.5.11) у (2.5.9), маємо

$$\gamma_0 = \phi^2 \gamma_0 + 0 + \sigma^2,$$

або

$$\gamma_0 = \frac{\sigma^2}{1 - \phi^2},$$

тобто формулу (2.5.4).

Аналогічно помноживши (2.5.8) на $y_{t-k} - \mu$ і обчислюючи математичне сподівання, одержимо

$$\begin{aligned} E[(y_t - \mu)(y_{t-k} - \mu)] &= \\ &= \phi E[(y_{t-1} - \mu)(y_{t-k} - \mu)] + E[\varepsilon_t(y_{t-k} - \mu)]. \end{aligned} \quad (2.5.12)$$

Оскільки член $y_{t-k} - \mu$ є лінійною функцією змінних $\varepsilon_{t-k}, \varepsilon_{t-k-1}, \dots$, то він не корелює з ε_t . Тому для $k > 0$ останній доданок у правій частині (2.5.12) дорівнює нулю. Зазначимо, що

$$E[(y_{t-1} - \mu)(y_{t-k} - \mu)] = E[(y_{t-1} - \mu)(y_{(t-1)-(k-1)} - \mu)]$$

є автоковаріацією спостережень y_t порядку $(k-1)$. Отже, для $k > 0$ з (2.5.12) одержуємо

$$\gamma_k = \phi \gamma_{k-1}, \quad (2.5.13)$$

звідки

$$\gamma_k = \phi^k \gamma_0,$$

Рис. 2.5.8. $AR(1) : y_t = 0,1 y_{t-1} + \varepsilon_t$

Рис. 2.5.9. $AR(1) : y_t = 0,8 y_{t-1} + \varepsilon_t$

що дає формулу (2.5.6).

На рис. 2.5.8 – 2.5.11 зображено динаміку часових рядів, що описуються $AR(1)$ процесами для різних ϕ , а також їхніх автокореляцій ρ_k як функцій від k .

Зазначимо, що для рядів з додатним значенням ϕ ряд виглядає тим плавнішим, чим ближче ϕ до 1. Для від'ємних ϕ подібно як і для від'ємних θ процесу $MA(1)$ зображення ряду є густішим, ніж зображення для будь-якого ряду з додатним ϕ . Для ϕ близького до -1 поведінка ряду має характерне особливе зображення, як на рис. 2.5.11.

Рис. 2.5.10. $AR(1)$: $y_t = -0,4 y_{t-1} + \varepsilon_t$

Рис. 2.5.11. $AR(1)$: $y_t = -0,8 y_{t-1} + \varepsilon_t$

2.6 Авторегресійний процес другого порядку $AR(2)$

Авторегресійний процес другого порядку (a second-order autoregressive process), який позначають $AR(2)$, має вигляд

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t. \quad (2.6.1)$$

Введемо лаговий оператор (оператор зсуву) L , дія якого на y_t дає його попереднє значення, тобто

$$\begin{aligned} L y_t &= y_{t-1}, \\ L^2 y_t &= L(L y_t) = L y_{t-1} = y_{t-2}, \\ &\dots \\ L^n y_t &= y_{t-n}. \end{aligned}$$

Оператор L також називають оператором зсуву, або лаговим оператором.

За допомогою лагового оператора L модель (2.6.1) можна записати у вигляді

$$(1 - \phi_1 L - \phi_2 L^2)y_t = c + \varepsilon_t. \quad (2.6.2)$$

Лема 2.6.1. *Оператор $1 - \phi_1 L - \phi_2 L^2$ має обернений, якщо корені характеристичного рівняння*

$$1 - \phi_1 z - \phi_2 z^2 = 0 \quad (2.6.3)$$

лежать поза одиничним кругом.

Д о в е д е н н я . Розглянемо квадратний тричлен $1 - \phi_1 z - \phi_2 z^2$ і розкладемо його на множники

$$1 - \phi_1 z - \phi_2 z^2 = (1 - \lambda_1 z)(1 - \lambda_2 z),$$

де λ_1, λ_2 такі, що

$$\lambda_1 + \lambda_2 = \phi_1, \quad \lambda_1 \lambda_2 = -\phi_2.$$

Тоді значення

$$z_1 = \frac{1}{\lambda_1}, \quad z_2 = \frac{1}{\lambda_2}$$

є коренями характеристичного рівняння (2.6.3).

Розглянемо оператор $1 - \lambda_1 L$. Якщо $|\lambda_1| < 1$, то до цього оператора існує обернений

$$(1 - \lambda_1 L)^{-1} = 1 + \lambda_1 L + \lambda_1^2 L^2 + \lambda_1^3 L^3 \dots$$

Аналогічно, якщо $|\lambda_2| < 1$, то існує

$$(1 - \lambda_2 L)^{-1} = 1 + \lambda_2 L + \lambda_2^2 L^2 + \lambda_2^3 L^3 \dots$$

Отже, якщо обидва $|\lambda_i| < 1$, то існує оператор обернений до $1 - \phi_1 L - \phi_2 L^2$, який визначається як

$$\begin{aligned} (1 - \phi_1 L - \phi_2 L^2)^{-1} &= (1 - \lambda_1 L)^{-1} (1 - \lambda_2 L)^{-1} = \\ &= (1 + \lambda_1 L + \lambda_1^2 L^2 + \dots)(1 + \lambda_2 L + \lambda_2^2 L^2 + \dots) = \\ &= 1 + (\lambda_1 + \lambda_2) L + (\lambda_1^2 + \lambda_1 \lambda_2 + \lambda_2^2) L^2 + \dots \end{aligned}$$

Оскільки $z_i = 1/\lambda_i$, то умова $|\lambda_i| < 1$ означає, що $|z_i| > 1$. Отже, для існування оберненого оператора другого порядку необхідно, щоб обидва корені характеристичного рівняння по модулю були більшими за одиницю, тобто лежали зовні одиничного круга. \square

Лема 2.6.2. *Якщо корені характеристичного рівняння (2.6.3) лежать поза одиничним кругом, то AR(2) процес (2.6.1) можна подати у вигляді MA(∞) зображення*

$$y_t = \mu + \psi_0 \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots, \quad (2.6.4)$$

де

$$\psi_0 = 1, \quad \mu = \frac{c}{1 - \phi_1 - \phi_2}, \quad (2.6.5)$$

причому виконується умова

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \quad (2.6.6)$$

Д о в е д е н н я. Якщо умови леми виконуються, то з доведення леми 2.6.1 випливає, що існує оператор

$$\psi(L) = (1 - \phi_1 L - \phi_2 L^2)^{-1} = (1 - \lambda_1 L)^{-1} (1 - \lambda_2 L)^{-1},$$

де $|\lambda_1| < 1, |\lambda_2| < 1$.

При $\lambda_1 \neq \lambda_2$ дія цього оператора еквівалентна дії

$$(\lambda_1 - \lambda_2)^{-1} \left(\frac{\lambda_1}{1 - \lambda_1 L} - \frac{\lambda_2}{1 - \lambda_2 L} \right). \quad (2.6.7)$$

Справді,

$$\begin{aligned} (\lambda_1 - \lambda_2)^{-1} \left(\frac{\lambda_1}{1 - \lambda_1 L} - \frac{\lambda_2}{1 - \lambda_2 L} \right) &= \\ &= (\lambda_1 - \lambda_2)^{-1} \left(\frac{\lambda_1(1 - \lambda_2 L) - \lambda_2(1 - \lambda_1 L)}{(1 - \lambda_1 L)(1 - \lambda_2 L)} \right) = \\ &= \frac{1}{(1 - \lambda_1 L)(1 - \lambda_2 L)}. \end{aligned}$$

Помноживши рівність (2.6.2) на $\psi(L)$, одержимо

$$y_t = \psi(L)c + \psi(L)\varepsilon_t. \quad (2.6.8)$$

Покажемо, що

$$\psi(L)\varepsilon_t = (\psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \dots)\varepsilon_t,$$

причому

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

На основі (2.6.7) отримаємо

$$\begin{aligned} \psi(L)\varepsilon_t &= (\lambda_1 - \lambda_2)^{-1} [\lambda_1(1 - \lambda_1 L)^{-1} - \lambda_2(1 - \lambda_2 L)^{-1}] \varepsilon_t = \\ &= \left(\frac{\lambda_1}{\lambda_1 - \lambda_2} [1 + \lambda_1 L + \lambda_1^2 L^2 + \lambda_1^3 L^3 + \dots] - \right. \\ &\quad \left. - \frac{\lambda_2}{\lambda_1 - \lambda_2} [1 + \lambda_2 L + \lambda_2^2 L^2 + \lambda_2^3 L^3 + \dots] \right) \varepsilon_t. \end{aligned}$$

Позначивши

$$c_1 = \frac{\lambda_1}{\lambda_1 - \lambda_2}, \quad c_2 = -\frac{\lambda_2}{\lambda_1 - \lambda_2},$$

матимемо

$$\begin{aligned} \psi(L)\varepsilon_t = & [c_1 + c_2]\varepsilon_t + [c_1\lambda_1 + c_2\lambda_2]\varepsilon_{t-1} + [c_1\lambda_1^2 + c_2\lambda_2^2]\varepsilon_{t-2} + \\ & + [c_1\lambda_1^3 + c_2\lambda_2^3]\varepsilon_{t-3} + \dots \end{aligned}$$

Звідси

$$\psi_j = c_1\lambda_1^j + c_2\lambda_2^j$$

і оскільки $|\lambda_1| < 1, |\lambda_2| < 1$, то

$$\sum_{j=0}^{\infty} |\psi_j| \leq |c_1| \sum_{j=0}^{\infty} |\lambda_1|^j + |c_2| \sum_{j=0}^{\infty} |\lambda_2|^j < \infty.$$

Оскільки $c_1 + c_2 = 1$, то $\psi_0 = 1$.

Крім того, безпосередні обчислення дають

$$\psi(L)c = \frac{c}{1 - \phi_1 - \phi_2},$$

тобто

$$E[y_t] = \mu = \frac{c}{1 - \phi_1 - \phi_2}.$$

□

Зауважимо таке: оскільки виконується умова (2.6.6), то $MA(\infty)$ процес (2.6.3) коваріаційно - стаціонарний. Тому за виконання умов леми $AR(2)$ процес можна записати за допомогою стаціонарного $MA(\infty)$ зображення

$$y_t = \mu + \varepsilon_t + \psi_1\varepsilon_{t-1} + \dots$$

Наслідок 2.6.1. *Якщо корені характеристичного рівняння лежать поза одиничним кругом, то $AR(2)$ процес коваріаційно - стаціонарний.*

Моменти $AR(2)$ процесу обчислюватимемо за допомогою алгоритму, який ґрунтується на припущенні коваріаційної стаціонарності процесу. А саме, спочатку припустимо, що корені

характеристичного рівняння лежать поза одиничним кругом, тобто процес $AR(2)$ є коваріаційно - стаціонарний, а потім обчислимо його математичне сподівання використовуючи зображення (2.6.1)

$$E[y_t] = c + \phi_1 E[y_{t-1}] + \phi_2 E[y_{t-2}] + E[\varepsilon_t].$$

Звідси

$$\mu = c + \phi_1 \mu + \phi_2 \mu,$$

що веде до (2.6.5).

Для того щоб знайти другий момент, запишемо (2.6.1) у вигляді

$$y_t = \mu(1 - \phi_1 - \phi_2) + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

або

$$(y_t - \mu) = \phi_1 (y_{t-1} - \mu) + \phi_2 (y_{t-2} - \mu) + \varepsilon_t. \quad (2.6.9)$$

Помножимо (2.6.9) на $(y_{t-k} - \mu)$ і обчислимо математичне сподівання обох частин одержаної рівності

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2}, \quad k = 1, 2, \dots \quad (2.6.10)$$

Поділивши (2.6.10) на γ_0 , отримаємо різницеve рівняння другого порядку для знаходження автокореляцій, яке називається рівнянням Юла - Уолкера (Yule - Walker)

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k = 1, 2, \dots \quad (2.6.11)$$

Зокрема для $k = 1$, оскільки $\rho_{-1} = \rho_1$ і $\rho_0 = 1$, то матимемо

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

або

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}. \quad (2.6.12)$$

Знаючи ρ_1 , розглянемо рівняння (2.6.11) для $k = 2$

$$\rho_2 = \phi_1 \rho_1 + \phi_2,$$

звідки матимемо

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2. \quad (2.6.13)$$

Автокореляції вищих порядків знаходимо послідовно з рекурсії (2.6.11).

Для знаходження дисперсії процесу $AR(2)$ помножимо (2.6.9) на $(y_t - \mu)$ і обчислимо математичне сподівання

$$\begin{aligned} E[(y_t - \mu)^2] &= \phi_1 E[(y_{t-1} - \mu)(y_t - \mu)] + \\ &+ \phi_2 E[(y_{t-2} - \mu)(y_t - \mu)] + E[\varepsilon_t(y_t - \mu)]. \end{aligned} \quad (2.6.14)$$

Звідси

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2, \quad (2.6.15)$$

оскільки

$$E[\varepsilon_t(y_t - \mu)] = E[\varepsilon_t(\phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \varepsilon_t)] = \sigma^2.$$

Рівняння (2.6.15) можна записати у вигляді

$$\gamma_0 = \phi_1 \rho_1 \gamma_0 + \phi_2 \rho_2 \gamma_0 + \sigma^2. \quad (2.6.16)$$

Підставляючи (2.6.12) і (2.6.13) у (2.6.16), одержимо

$$\gamma_0 = \left[\frac{\phi_1^2}{1 - \phi_2} + \frac{\phi_2 \phi_1^2}{1 - \phi_2} + \phi_2^2 \right] \gamma_0 + \sigma^2$$

або

$$\gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}.$$

На рис. 2.6.12 – 2.6.15 зображено динаміку часових рядів, які описуються $AR(2)$ процесами для різних ϕ_1, ϕ_2 , а також їхніх автокореляцій ρ_k як функцій від k .

Рис. 2.6.12. $AR(2) : y_t = y_{t-1} - 0,16 y_{t-2} + \varepsilon_t,$
 $\lambda_1 = 0,2, \lambda_2 = 0,8$

Рис. 2.6.13. $AR(2) : y_t = -y_{t-1} - 0,16 y_{t-2} + \varepsilon_t,$
 $\lambda_1 = -0,2, \lambda_2 = -0,8$

Рис. 2.6.14. $AR(2) : y_t = 0,8 y_{t-1} - 0,5 y_{t-2} + \varepsilon_t,$
 $\lambda_{1,2} = 0,4 \pm 0,3i$

Рис. 2.6.15. $AR(2) : y_t = -0,8 y_{t-1} - 0,5 y_{t-2} + \varepsilon_t,$
 $\lambda_{1,2} = -0,4 \pm 0,3i$

Визначимо умови, які повинні задовольняти коефіцієнти ϕ_j $AR(2)$ процесу, щоб він був коваріаційно - стаціонарний. Згідно з наслідком леми 2.6.2 одержуємо таке: $AR(2)$ процес буде коваріаційно - стаціонарним, якщо корені характеристичного рівняння (2.6.3) лежать ззовні одиничного круга, що еквівалентно тому, що корені рівняння

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0$$

лежать всередині одиничного круга, тобто $|\lambda_i| < 1$ ($i = 1, 2$).

Зауважимо, що

$$\lambda_{1,2} = \frac{\phi_1}{2} \pm \sqrt{\frac{\phi_1^2 + 4\phi_2}{4}}.$$

Корені рівняння будуть комплексно спряжені, якщо

$$\phi_1^2 + 4\phi_2 < 0,$$

що відповідає області нижче параболи на рис 2.6.16. Тоді

$$\lambda_{1,2} = \frac{\phi_1}{2} \pm \sqrt{\frac{|\phi_1^2 + 4\phi_2|}{4}} i$$

і

$$|\lambda_{1,2}| = \sqrt{\left(\frac{\phi_1}{2}\right)^2 - \frac{\phi_1^2 + 4\phi_2}{4}} = \sqrt{-\phi_2} < 1$$

при

$$\phi_1^2 + 4\phi_2 < 0, \quad -1 < \phi_2 < 0.$$

У випадку дійсних власних значень більше власне значення є більше від одиниці, якщо

$$\sqrt{\phi_1^2 + 4\phi_2} > 2 - \phi_1.$$

За умови, що власні значення дійсні, ліва частина нерівності невід'ємна і нерівність виконується для всіх $\phi_1 > 2$. Якщо $\phi_1 \leq 2$, то піднісши обидві частини нерівності до квадрата, одержимо

$$\phi_1^2 + 4\phi_2 > 4 - 4\phi_1 + \phi_1^2$$

або

$$\phi_2 > 1 - \phi_1.$$

Отже, $\lambda_1 > 1$, якщо

$$(\phi_1, \phi_2) \in \{(\phi_1, \phi_2) : \phi_1 > 2, \phi_2 \in \mathbb{R}\} \cup \{(\phi_1, \phi_2) : \phi_1 + \phi_2 > 1\}.$$

Аналогічно менше дійсне значення $\lambda_2 < -1$, якщо

$$\sqrt{\phi_1^2 + 4\phi_2} > 2 + \phi_1.$$

Ця нерівність правильна для всіх $\phi < -2$. Якщо $\phi_1 \geq -2$, то, піднісши до квадрата обидві частини нерівності, матимемо

$$\phi_2 > 1 + \phi_1.$$

Отже, обидва власні значення за модулем менші від одиниці, якщо точка (ϕ_1, ϕ_2) лежить всередині трикутника, зображеного на рис. 2.6.16.

Рис. 2.6.16.

2.7 Авторегресійні процеси p -го порядку

Авторегресія p -го порядку (а p th-order autoregressive process), яку позначають $AR(p)$, має вигляд

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t. \quad (2.7.1)$$

Використовуючи лаговий оператор L , регресію (2.7.1) можна записати у вигляді

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = c + \varepsilon_t.$$

Аналогічно як і лему 2.6.1 можна довести таке: якщо корені рівняння

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0 \quad (2.7.2)$$

лежать поза одиничним кругом, то існує оператор обернений до оператора $(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$:

$$\begin{aligned} \psi(L) &= (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)^{-1} = \\ &= (1 - \lambda_1 L)^{-1} (1 - \lambda_2 L)^{-1} \dots (1 - \lambda_p L)^{-1}, \end{aligned}$$

де $|\lambda_i| < 1$ для всіх $i = 1, \dots, p$.

Лема 2.7.1. *Якщо корені характеристичного рівняння (2.7.2) лежать поза одиничним кругом, то $AR(p)$ процес (2.7.1) можна звести до $MA(\infty)$ процесу*

$$\begin{aligned} y_t &= \mu + \psi(L)\varepsilon_t = \\ &= \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots, \end{aligned} \quad (2.7.3)$$

причому

$$\sum_{j=0}^{\infty} |\psi_j| < \infty. \quad (2.7.4)$$

Д о в е д е н н я . Помноживши (2.7.1) на $\psi(L)$, матимемо

$$y_t = \psi(L)c + \psi(L)\varepsilon_t.$$

Знайдемо такі сталі c_1, c_2, \dots, c_p , для яких

$$\begin{aligned} \frac{1}{(1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z)} &= \\ &= \frac{c_1}{(1 - \lambda_1 z)} + \frac{c_2}{(1 - \lambda_2 z)} + \cdots + \frac{c_p}{(1 - \lambda_p z)}. \end{aligned}$$

Для цього помножимо обидві частини останньої рівності на $(1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z)$. В результаті отримаємо рівність

$$\begin{aligned} 1 &= c_1(1 - \lambda_2 z)(1 - \lambda_3 z) \cdots (1 - \lambda_p z) + \\ &+ c_2(1 - \lambda_1 z)(1 - \lambda_3 z) \cdots (1 - \lambda_p z) + \dots + \\ &+ c_p(1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_{p-1} z), \end{aligned}$$

яка виконується для всіх z . Підставивши $z = \lambda_1^{-1}$, одержимо

$$1 = c_1(1 - \lambda_2 \lambda_1^{-1})(1 - \lambda_3 \lambda_1^{-1}) \cdots (1 - \lambda_p \lambda_1^{-1}),$$

звідки

$$c_1 = \frac{\lambda_1^{p-1}}{(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \cdots (\lambda_1 - \lambda_p)}.$$

Аналогічно послідовно підставляючи $z = \lambda_2^{-1}$, $z = \lambda_3^{-1}$, ..., $z = \lambda_p^{-1}$, матимемо

$$\begin{aligned} c_2 &= \frac{\lambda_2^{p-1}}{(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3) \cdots (\lambda_2 - \lambda_p)}, \\ &\dots \\ c_p &= \frac{\lambda_p^{p-1}}{(\lambda_p - \lambda_1)(\lambda_p - \lambda_2) \cdots (\lambda_p - \lambda_{p-1})}. \end{aligned}$$

Тому

$$\begin{aligned}\psi(L)\varepsilon_t &= \frac{c_1}{(1 - \lambda_1 L)}\varepsilon_t + \frac{c_2}{(1 - \lambda_2 L)}\varepsilon_t + \dots + \frac{c_p}{(1 - \lambda_p L)}\varepsilon_t = \\ &= c_1(1 + \lambda_1 L + \lambda_1^2 L^2 + \lambda_1^3 L^3 + \dots)\varepsilon_t + \\ &\quad + c_2(1 + \lambda_2 L + \lambda_2^2 L^2 + \lambda_2^3 L^3 + \dots)\varepsilon_t + \\ &\quad + \dots + \\ &\quad + c_p(1 + \lambda_p L + \lambda_p^2 L^2 + \lambda_p^3 L^3 + \dots)\varepsilon_t\end{aligned}$$

або

$$\begin{aligned}\psi(L)\varepsilon_t &= (c_1 + c_2 + \dots + c_p)\varepsilon_t + \\ &\quad + (c_1\lambda_1 + c_2\lambda_2 + \dots + c_p\lambda_p)\varepsilon_{t-1} + \\ &\quad + (c_1\lambda_1^2 + c_2\lambda_2^2 + \dots + c_p\lambda_p^2)\varepsilon_{t-2} + \\ &\quad + (c_1\lambda_1^3 + c_2\lambda_2^3 + \dots + c_p\lambda_p^3)\varepsilon_{t-3} + \dots\end{aligned}$$

Отже,

$$\psi_j = c_1\lambda_1^j + c_2\lambda_2^j + \dots + c_p\lambda_p^j$$

і оскільки $c_1 + c_2 + \dots + c_p = 1$, то $\psi_0 = 1$.

Звідси видно, що оскільки за припущення леми корені характеристичного рівняння лежать поза одиничним кругом, то $|\lambda_i| < 1$ для всіх $i = 1, \dots, p$, тому ряд $\sum_{j=0}^{\infty} |\psi_j|$ є збіжним. \square

Оскільки виконується умова (2.7.4), то $MA(\infty)$ процес (2.7.3) коваріаційно - стаціонарний. Отже, $AR(p)$ - стаціонарний, якщо корені характеристичного рівняння (2.7.2) лежать ззовні одиничного круга.

Припускаючи стаціонарність процесу, з (2.7.1) одержуємо

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}. \quad (2.7.5)$$

На підставі (2.7.5) регресію (2.7.1) можемо записати у вигляді

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \dots + \phi_p(y_{t-p} - \mu) + \varepsilon_t. \quad (2.7.6)$$

Автоковаріації знайдемо, помноживши обидві частини рівності (2.7.6) на $(y_{t-k} - \mu)$ $k = 0, 1, 2, \dots$ і обчисливши математичне сподівання

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma^2, \quad (2.7.7)$$

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}, \quad k = 1, 2, \dots \quad (2.7.8)$$

Якщо поділити (2.7.8) на γ_0 , то одержимо систему рівнянь Юла - Уолкера

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p}, \quad k = 1, 2, \dots \quad (2.7.9)$$

Оскільки $\rho_{-i} = \rho_i$, то перші p рівнянь системи (2.7.9) є системою p лінійних рівнянь з невідомими $\rho_1, \rho_2, \dots, \rho_p$. Тому розв'язавши її, ми можемо знайти перші p автокореляцій. Щоб знайти дисперсію, підставимо їх у рівняння (2.7.7). Матимемо

$$\gamma_0 = \phi_1 \gamma_0 \rho_1 + \dots + \phi_p \gamma_0 \rho_p + \sigma^2,$$

звідки

$$\gamma_0 = \frac{\sigma^2}{1 - \phi_1 \rho_1 - \dots - \phi_p \rho_p}. \quad (2.7.10)$$

Наступні автокореляції $\rho_{p+1}, \rho_{p+2}, \dots$ знаходимо послідовно з (2.7.9).

Зокрема, для процесу $AR(3)$ система для знаходження перших трьох коефіцієнтів автокореляції матиме вигляд

$$\begin{cases} \rho_1 = \phi_1 + \phi_2 \rho_2 + \phi_3 \rho_2, \\ \rho_2 = \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1, \\ \rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1 + \phi_3, \end{cases}$$

звідки

$$\rho_1 = \frac{\phi_1 + \phi_2 \phi_3}{1 - \phi_2 - \phi_3(\phi_1 + \phi_3)},$$

$$\rho_2 = \frac{(1 - \phi_2)\phi_2 - \phi_1(\phi_1 + \phi_3)}{1 - \phi_2 - \phi_3(\phi_1 + \phi_3)}.$$

Підставляючи в останнє рівняння системи, знаходимо ρ_3 і за допомогою (2.7.10) дисперсію γ_0 .

На рис. 2.7.17 – 2.7.20 зображено динаміку часових рядів, які описуються $AR(3)$ процесами, а також поведінку їхніх автокореляцій ρ_k як функцій від k .

Рис. 2.7.17. $AR(3) : y_t = 0,8 y_{t-1} - 0,4 y_{t-2} + 0,5 y_{t-3} + \varepsilon_t$

Рис. 2.7.18. $AR(3) : y_t = -0,1 y_{t-1} - 0,1 y_{t-2} + 0,7 y_{t-3}$

2.8 Мішані моделі авторегресії і рухомого середнього

Мішана модель авторегресії порядку p і рухомого середнього порядку q (a mixed autoregressive moving average process), яку

Рис. 2.7.19. $AR(3) : y_t = 0,5 y_{t-1} + 1,4 y_{t-2} - 0,5 y_{t-3} + \varepsilon_t$

Рис. 2.7.20. $AR(3) : y_t = -0,2 y_{t-1} + 0,1 y_{t-2} + 0,7 y_{t-3} + \varepsilon_t$

позначають $ARMA(p, q)$, має вигляд

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}. \quad (2.8.1)$$

За допомогою лагового оператора (2.8.1) можна записати у вигляді

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) y_t = c + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t. \quad (2.8.2)$$

Якщо корені рівняння (2.7.2) лежать ззовні одиничного круга і виконується умова (2.7.4), то (2.8.2) можна записати так:

$$y_t = \mu + \psi(L) \varepsilon_t,$$

де

$$\psi(L) = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p}, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty,$$

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}.$$

Отже, стаціонарність $ARMA(p, q)$ процесу залежить лише від параметрів авторегресії $\phi_1, \phi_2, \dots, \phi_p$. Часто буває зручно записати процес (2.8.1) у формі відхилень від середнього

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \phi_2(y_{t-2} - \mu) + \dots + \phi_p(y_{t-p} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}. \quad (2.8.3)$$

Щоб знайти автоковаріацію треба помножити обидві сторони (2.8.3) на $(y_{t-k} - \mu)$ і обчислити математичне сподівання. Для $k > q$ результуюче рівняння набере вигляду

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p} \quad \text{для} \quad k = q + 1, q + 2, \dots. \quad (2.8.4)$$

Отож, для $k > q$ автоковаріаційна функція γ_k (і автокореляційна функція ρ_k) задаються різницеvими рівняннями p -го порядку з авторегресійними параметрами.

Зауважимо, що (2.8.4) не правильне при $k \leq q$, внаслідок кореляції між $\theta_k \varepsilon_{t-k}$ і y_{t-k} . Тому $ARMA(p, q)$ процес матиме складніші автоковаріації для порядків від 1 до q , ніж в $AR(p)$ процесі.

Зокрема, $ARMA(1, 1)$ процес

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

стаціонарний, якщо $|\phi_1| < 1$. У цьому випадку матимемо

$$\begin{aligned}\gamma_0 &= \phi_1 \gamma_1 + \sigma^2 (1 + \theta_1^2), \\ \gamma_1 &= \phi_1 \gamma_0 + \sigma^2 \theta_1, \\ \gamma_2 &= \phi_1 \gamma_1, \\ \gamma_3 &= \phi_1 \gamma_2 = \phi_1^2 \gamma_1, \\ &\dots \\ \gamma_k &= \phi_1^{k-1} \gamma_1.\end{aligned}$$

Звідси одержимо

$$\begin{aligned}\gamma_0 &= \frac{\sigma^2 (1 + \theta_1^2 + \theta_1 \phi_1)}{1 - \phi_1^2}, \\ \gamma_1 &= \sigma^2 \left[\frac{\phi_1 (1 + \theta_1^2 + \theta_1 \phi_1)}{1 - \phi_1^2} + \theta_1 \right].\end{aligned}$$

Тому

$$\begin{aligned}\rho_1 &= \frac{\phi_1 (1 + \theta_1^2 + \theta_1 \phi_1) + \theta_1 (1 - \phi_1^2)}{1 + \theta_1^2 + \theta_1 \phi_1}, \\ \rho_k &= \phi_1^{k-1} \rho_1, \quad k = 2, 3, \dots\end{aligned}$$

Отже, якщо $|\phi_1| < 1$, то автокореляційна функція є спадною після ρ_1 . В загальному випадку для стаціонарного $ARMA(p, q)$ процесу автокореляційна функція є спадною після ρ_p .

На рис. 2.8.21. зображено динаміку часового ряду, що описується $ARMA(1, 1)$ процесом.

Рис. 2.8.21. $ARMA(1, 1)$: $y_t = 0,6y_{t-1} + \varepsilon_t + 0,8\varepsilon_{t-1}$

2.9 Векторна авторегресійна модель

Векторна авторегресія першого порядку $VAR(1)$

Векторну авторегресію першого порядку можна подати у вигляді

$$Y_t = c + AY_{t-1} + e_t \quad (2.9.1)$$

де c, Y_t, e_t – стовпці розміру n ; A – матриця розміру $n \times n$. Щодо векторів e_t припустатимемо, що їхнім математичним сподіванням є нульовий вектор

$$E[e_t] = 0 \quad \text{для всіх } t,$$

вектори e_t мають ту саму коваріаційну матрицю

$$\text{Var}[e_t] = E[e_t e_t'] = \Sigma \quad \text{для всіх } t$$

і є незалежними, звідки випливає, що вони не корелюють за всіма компонентами, тобто

$$E[e_t e_s'] = 0, \quad t \neq s.$$

Часовий ряд (2.9.1) буде коваріаційно - стаціонарним або слабко стаціонарним, якщо перші два моменти розподілу не залежать від t , тобто

$$E[Y_t] = \mu, \quad E[(Y_t - \mu)(Y_{t-k} - \mu)'] = \Gamma_k$$

для всіх t . Матрицю Γ_k називають матрицею автоковаріації. Її (i, j) -ий елемент

$$\gamma_{ij,k} = E[(y_{it} - \mu_i)(y_{j,t-k} - \mu_j)]$$

виражає коваріацію між y_{it} і $y_{j,t-k}$. Зазначимо, що матриця Γ_0 є коваріаційною матрицею вектора Y_t .

Якщо припустити стаціонарність процесу й обчислити математичне сподівання (2.9.1), то отримаємо

$$\mu = c + A\mu.$$

Звідси за умови існування матриці $(I - A)^{-1}$,

$$\mu = (I - A)^{-1}c.$$

Отже, рівняння (2.9.1) можна записати у вигляді відхилень від середнього

$$Y_t - \mu = A(Y_{t-1} - \mu) + e_t. \quad (2.9.2)$$

Помножимо (2.9.2) справа на $(Y_{t-k} - \mu)'$, $k \geq 1$ й обчислимо математичне сподівання

$$\Gamma_k = A\Gamma_{k-1} = A^2\Gamma_{k-2} = A^k\Gamma_0. \quad (2.9.3)$$

Для визначення матриці Γ_0 помножимо рівняння (2.9.2) справа на вектор транспонований до нього і обчислимо математичне сподівання

$$E[(Y_t - \mu)(Y_t - \mu)'] = A E[(Y_{t-1} - \mu)(Y_{t-1} - \mu)'] A' + E[e_t e_t'].$$

Звідси матимемо

$$\Gamma_0 = A\Gamma_0 A' + \Sigma.$$

Рівняння (2.9.2) можна також подати у вигляді

$$Y_t - \mu = e_t + A e_{t-1} + A^2 e_{t-2} + \dots \quad (2.9.4)$$

Матрицю коваріації Γ_0 можна визначити, якщо помножити справа рівняння (2.9.4) на транспоноване до нього і обчислити математичне сподівання, використавши незалежність e_t і e_{t-k} для $k \geq 1$

$$\Gamma_0 = \Sigma + A\Sigma A' + A^2\Sigma A'^2 + \dots \quad (2.9.5)$$

Позначимо через B матрицю, складену зі стовпців, які є власними векторами матриці A , а через D – діагональну матрицю, яка на діагоналі містить її власні значення, що позначимо λ_i ($i = 1, \dots, n$). За умови, що матриця B^{-1} існує, маємо $AB = BD$, а також $A = BDB^{-1}$. Зазначимо, що $A^k = BD^k B^{-1}$. Рівняння (2.9.5) тепер можемо записати у вигляді

$$\begin{aligned} \Gamma_0 &= \Sigma + BDB^{-1}\Sigma(B')^{-1}DB' + BD^2B^{-1}\Sigma(B')^{-1}D^2B' + \dots = \\ &= BWB' + BDWDB' + BD^2WD^2B' + \dots = \\ &= B[W + DWD + D^2WD^2 + \dots]B', \end{aligned}$$

де $W = B^{-1}\Sigma(B')^{-1}$. Нехай w_{ij} – (i, j) -ий елемент матриці W , тоді

$$\Gamma_0 = B[w_{ij} + w_{ij}\lambda_i\lambda_j + w_{ij}\lambda_i^2\lambda_j^2 + \dots]B'.$$

Ряд у квадратних дужках збіжний тоді і тільки тоді, коли модулі всіх власних значень матриці A менші від одиниці. Якщо

ця умова виконується, то

$$\Gamma_0 = B \left[\frac{w_{ij}}{1 - \lambda_i \lambda_j} \right] B'. \quad (2.9.6)$$

Звідси випливає, що часовий ряд, генерований рівнянням (2.9.1), коваріаційно - стаціонарний тоді і тільки тоді, коли модулі всіх власних значень матриці A менші від одиниці. У цьому випадку існує матриця $(I - A)^{-1}$, також можна визначити матрицю Γ_0 , а згідно з (2.9.3) – матриці $\Gamma_k = A^k \Gamma_0$. Формули (2.9.5) і (2.9.6) альтернативні для обчислення Γ_0 . Зазначимо таке: якщо модулі всіх власних значень матриці A менші від одиниці, то матриця $A^k = BD^k B^{-1}$ збігається до нульової матриці при зростанні k .

Зведення AR(p) моделі до векторної авторегресії

Використовуючи результат (2.9.3), дослідимо поведінку функції автоковаріації для авторегресійного процесу $AR(p)$ з нульовим математичним сподіванням

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t.$$

Запишемо цей процес у вигляді векторної авторегресії першого порядку

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \dots \\ y_{t-(p-1)} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (2.9.7)$$

У цьому випадку $Y_t = (y_t \ y_{t-1} \ \dots \ y_{t-(p-1)})'$ і

$$\Gamma_k = E[Y_t Y_{t-k}'] = \begin{pmatrix} \gamma_k & \gamma_{k+1} & \gamma_{k+2} & \dots & \gamma_{k+p-1} \\ \gamma_{k-1} & \gamma_k & \gamma_{k+1} & \dots & \gamma_{k+p-2} \\ \vdots & \vdots & \dots & \vdots & \vdots \\ \gamma_{k-(p-1)} & \gamma_{k-(p-2)} & \gamma_{k-(p-3)} & \dots & \gamma_k \end{pmatrix}.$$

Згідно з рівнянням (2.9.3) $\Gamma_k = A\Gamma_{k-1}$. Використовуючи перший стовпець матриці Γ_k , матимемо, що автоковаріації γ_k задовольняють рівняння

$$\begin{pmatrix} \gamma_k \\ \gamma_{k-1} \\ \dots \\ \gamma_{k-(p-1)} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \gamma_{k-1} \\ \gamma_{k-2} \\ \dots \\ \gamma_{k-p} \end{pmatrix}, \quad (2.9.8)$$

або в матричному вигляді

$$\tilde{\gamma}_k = A\tilde{\gamma}_{k-1}. \quad (2.9.9)$$

Зокрема, перший елемент вектора $\tilde{\gamma}_k$ є k -ою автоковаріацією y_t і задовольняє одновимірне різницеве рівняння

$$\gamma_k = \phi_1\gamma_{k-1} + \dots + \phi_p\gamma_{k-p}$$

порядку p , яке збігається з рівнянням Юла - Уолкера (2.7.8).

Розв'язок рівняння (2.9.9) має вигляд

$$\tilde{\gamma}_k = A^k\tilde{\gamma}_0 = BD^k B^{-1}\tilde{\gamma}_0 = (b_{ij})D^k(b^{ij})\tilde{\gamma}_0,$$

де b_{ij} – елементи матриці B ; b^{ij} – елементи матриці B^{-1} . Тоді перший елемент вектора $\tilde{\gamma}_k$ матиме вигляд

$$\begin{aligned} \gamma_k &= (b_{11} \dots b_{1p}) \begin{pmatrix} \lambda_1^k & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \lambda_p^k \end{pmatrix} \begin{pmatrix} b^1\tilde{\gamma}_0 \\ \vdots \\ b^p\tilde{\gamma}_0 \end{pmatrix} = \\ &= (b_{11}b^1\tilde{\gamma}_0)\lambda_1^k + (b_{12}b^2\tilde{\gamma}_0)\lambda_2^k + \dots + (b_{1p}b^p\tilde{\gamma}_0)\lambda_p^k, \end{aligned} \quad (2.9.10)$$

де b^i позначає i -ий рядок матриці B^{-1} . Отже, функція автоковаріації γ_k є лінійною комбінацією k -их степенів власних значень матриці A .

Отож, для коваріаційної стаціонарності $VAR(1)$ процесу треба вимагати, щоб модулі всіх власних значень λ_i були меншими за одиницю.

Лема 2.9.1. *Власні значення матриці A , заданої в (2.9.7), є коренями рівняння*

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \dots - \phi_p = 0.$$

Д о в е д е н н я. Для матриці A відповідний визначник дорівнює

$$\begin{aligned} |A - \lambda I| &= \\ &= \left| \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} - \begin{pmatrix} \lambda & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 0 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix} \right| = \\ &= \begin{vmatrix} (\phi_1 - \lambda) & \phi_2 & \phi_3 & \dots & \phi_{p-2} & \phi_{p-1} & \phi_p \\ 1 & -\lambda & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\lambda & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & -\lambda \end{vmatrix} \end{aligned}$$

Помноживши p -ий стовпець на $1/\lambda$ і додавши результат до $(p-1)$ -шого стовпця, матимемо

$$|A - \lambda I| = \begin{vmatrix} \phi_1 - \lambda & \phi_2 & \phi_3 & \dots & \phi_{p-2} & \phi_{p-1} + \frac{\phi_p}{\lambda} & \phi_p \\ 1 & -\lambda & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -\lambda & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & -\lambda \end{vmatrix}.$$

Далі помножимо $(p-1)$ -ший стовпець на $1/\lambda$ і додамо до $(p-2)$ -гого стовпця. Матимемо

$$|A - \lambda I| = \begin{vmatrix} \phi_1 - \lambda & \phi_2 & \phi_3 & \dots & \phi_{p-2} + \frac{\phi_{p-1}}{\lambda} + \frac{\phi_p}{\lambda^2} & \phi_{p-1} + \frac{\phi_p}{\lambda} & \phi_p \\ 1 & -\lambda & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & -\lambda & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & -\lambda \end{vmatrix}$$

Продовжуючи аналогічно, отримаємо

$$\begin{vmatrix} \phi_1 - \lambda + \frac{\phi_2}{\lambda} + \dots + \frac{\phi_p}{\lambda^{p-1}} & \phi_2 + \frac{\phi_3}{\lambda} + \dots + \frac{\phi_p}{\lambda^{p-2}} & \dots & \phi_{p-1} + \frac{\phi_p}{\lambda} & \phi_p \\ 0 & -\lambda & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & -\lambda & 0 \\ 0 & 0 & \dots & 0 & -\lambda \end{vmatrix}$$

Детермінант одержаної матриці дорівнює

$$\begin{aligned} |A - \lambda I| &= (-\lambda)^{p-1} \left[\phi_1 - \lambda + \frac{\phi_2}{\lambda} + \frac{\phi_3}{\lambda^2} + \dots + \frac{\phi_p}{\lambda^{p-1}} \right] = \\ &= (-1)^p [\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \dots - \phi_p]. \end{aligned}$$

Отже, власні значення матриці A є коренями рівняння

$$\lambda^p - \phi_1 \lambda^{p-1} - \phi_2 \lambda^{p-2} - \dots - \phi_p = 0.$$

□

З леми випливає, що власні значення λ_i ($i = 1, 2, \dots, n$) матриці A є оберненими величинами до коренів характеристичного рівняння

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0.$$

Тоді умова $|\lambda_i| < 1$ для всіх $i = 1, 2, \dots, n$ еквівалентна умові, що всі корені характеристичного рівняння лежать поза одиничним кругом.

У цьому випадку функція автоковаріації (2.9.10) має тенденцію прямування до нуля при зростанні k . Отож, при зростанні "відстані" k кореляція між змінними y_t і y_{t-k} прямує до нуля. Зокрема, для часового ряду $y_t = \phi y_{t-1} + \varepsilon_t$ маємо $\gamma_k = \phi^k \gamma_0$. У цьому випадку автоковаріація спадає як геометрична прогресія.

У випадку авторегресії $AR(p)$ при $p > 1$ серед коренів рівняння може бути пара комплексно спряжених коренів, наприклад, $\lambda_1 = a + bi$, $\lambda_2 = a - bi$, $i = \sqrt{-1}$.

Рис. 2.9.22. $AR(2)$: $y_t = -1,8 y_{t-1} - 0,9 y_{t-2} + \varepsilon_t$

Рис. 2.9.23. $AR(3)$: $y_t = 0,3 y_{t-1} + 1,6 y_{t-2} - 0,9 y_{t-3} + \varepsilon_t$

У цьому випадку коефіцієнти $b_{11}b^1\tilde{\gamma}_0$, $b_{12}b^2\tilde{\gamma}_0$ у формулі (2.9.10) теж будуть комплексно спряжені, оскільки сума двох перших доданків цієї формули повинна бути дійсною (функція автоковаріації дійсна). Запишемо ці коефіцієнти у вигляді $se^{i\varphi}$ і $se^{-i\varphi}$, а власні значення λ_1 , λ_2 , відповідно, у вигляді $re^{i\vartheta}$ і $re^{-i\vartheta}$, де $r = \sqrt{a^2 + b^2}$, $\text{tg } \vartheta = b/a$. Підставляючи ці вирази у формулу (2.9.10), одержимо

$$\begin{aligned}\gamma_k &= sr^k(e^{i(k\vartheta+\varphi)} + e^{-i(k\vartheta+\varphi)}) + \dots = \\ &= 2sr^k \cos(k\vartheta + \varphi) + \dots\end{aligned}\quad (2.9.11)$$

Оскільки $r < 1$, то перший доданок у (2.9.11) – згасаюча косинусоїда аргументу k з частотою ϑ , що підтверджує вигляд автокореляцій для AR процесів з комплексними λ_i на рис.2.9.22 – 2.9.25.

Рис. 2.9.24. $AR(3) : y_t = 1,8y_{t-1} - 0,9y_{t-2} + \varepsilon_t$

Рис. 2.9.25. $AR(3) : y_t = -0,6y_{t-1} + 1,3y_{t-2} - 0,7y_{t-3} + \varepsilon_t$

Зведення $ARMA(p, q)$ моделі до векторної авторегресії

Мішаний авторегресійний процес рухомого середнього $ARMA(p, q)$

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.9.12)$$

також можна подати у вигляді векторної авторегресії першого порядку

$$\begin{pmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \\ \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q+1} \end{pmatrix} = \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_p & \theta_1 & \theta_2 & \dots & \theta_{q-1} & \theta_q \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p} \\ \varepsilon_{t-1} \\ \varepsilon_{t-2} \\ \vdots \\ \varepsilon_{t-q} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (2.9.13)$$

яку можна записати у векторному вигляді (2.9.1), зі стовпцями c, Y_t, e_t і матрицею A , визначеними у (2.9.13).

Розділ 3

Властивості $ARMA$ процесів

3.1 Умови стаціонарності $ARMA$ процесів

Коваріаційна стаціонарність стохастичного процесу вимагає, щоб дисперсія і автоковаріації були скінченими і незалежними від часу. Наведемо умови стаціонарності $ARMA$ процесів.

Процеси рухомого середнього (MA процеси) скінченного порядку є стаціонарними за побудовою, оскільки вони утворені як лінійна комбінація фіксованої кількості стаціонарних процесів білого шуму. MA процес буде не стаціонарним лише у випадку, якщо його коефіцієнти змінюються з часом

$$y_t = \mu(t) + \varepsilon_t + g(t) \varepsilon_{t-1},$$

де $\mu(t), g(t)$ – деякі функції від t . У цьому випадку маємо

$$E[y_t] = \mu(t),$$

$$\text{Var}[y_t] = E[y_t^2] = \sigma^2 + g^2(t) \sigma^2,$$

що не є не залежними від t . Очевидно, що цей процес не є стаціонарним.

Стаціонарність AR процесів не така тривіальна. Зокрема, AR(p) процес

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

є стаціонарним тоді і лише тоді, коли корені характеристичного рівняння

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

лежать поза одиничним кругом.

Розглянемо найпростіший випадок, а саме AR(1) процес

$$y_t = \phi y_{t-1} + \varepsilon_t. \quad (3.1.1)$$

У цьому випадку характеристичне рівняння має вигляд

$$1 - \phi z = 0,$$

а його корінь дорівнює $z = 1/\phi$. Відповідно $|z| > 1$, якщо $|\phi| < 1$. Отже, процес (3.1.1) стаціонарний, якщо $|\phi| < 1$. Тоді з (3.1.1) отримаємо дисперсію

$$\text{Var}[y_t] = \frac{\sigma^2}{1 - \phi^2}.$$

При $|\phi| < 1$ (3.1.1) описує нестаціонарний процес. Справді якщо візьмемо будь-яке значення ϕ таке, що $|\phi| \geq 1$, то, припустивши стаціонарність, отримаємо від'ємну дисперсію, що неможливо.

На рис. 3.1.1, 3.1.2 зображено динаміку часових рядів, які описуються AR(1) процесами відповідно з $\phi > 1$ і $\phi < -1$.

Рис. 3.1.1. $AR(1)$: $y_t = 1,1y_{t-1} + \varepsilon_t$

Рис. 3.1.2. $AR(1)$: $y_t = -1,2y_{t-1} + \varepsilon_t$

Детальніше розглянемо приклад $AR(1)$ процесу, для якого $\phi = 1$

$$y_t = y_{t-1} + \varepsilon_t. \quad (3.1.2)$$

У цьому випадку корінь характеристичного рівняння дорівнює одиниці. Тому процес (3.1.2) називається процесом з одиничним коренем.

Обчисливши дисперсію обох частин (3.1.2), матимемо

$$\text{Var}[y_t] = \text{Var}[y_{t-1}] + \sigma^2. \quad (3.1.3)$$

З (3.1.2) видно, що процес (3.1.1) не є стаціонарним, оскільки рівність $\text{Var}[y_t] = \text{Var}[y_{t-1}]$ не виконується, якщо $\sigma^2 \neq 0$. Якщо ж $\sigma^2 = 0$, то в цьому випадку існує нескінченна кількість розв'язків.

$AR(1)$ процес з одиничним коренем (3.1.2) також має назву процесу випадкового блукання. Для нього безумовна дисперсія y_t не існує, тобто є нескінченною, тому процес не є стаціонарним.

На рис. 3.1.3 зображено динаміку часового ряду, що описується процесом з одиничним коренем, а саме динаміку процесу випадкового блукання.

Рис. 3.1.3. $AR(1)$: $y_t = y_{t-1} + \varepsilon_t$

Можемо узагальнити так: $AR(1)$ процес є стаціонарним тоді і тільки тоді, коли $|\phi| < 1$. У цьому випадку корінь характеристичного рівняння $1 - \phi z = 0$ більший за одиницю і оператор $1 - \phi L$ є зворотним.

Цей результат поширюється на довільну $ARMA$ модель. А саме $ARMA(p, q)$ модель

$$\phi(L)y_t = \theta(L)\varepsilon_t$$

відповідає стаціонарному процесу тоді і тільки тоді, коли розв'язки z_1, z_2, \dots, z_p характеристичного рівняння $\phi(z) = 0$ за мо-

дулем більші за одиницю, тобто AR поліном зворотний. Якщо хоча б один з $|z_j|$ дорівнює або менший за одиницю, то процес не є стаціонарним.

Приклад 3.1.1. *ARMA(2, 1) процес*

$$y_t = 1,2 y_{t-1} - 0,2 y_{t-2} + \varepsilon_t - 0,5 \varepsilon_{t-1}$$

нестационарний, бо $z = 1$ розв'язок рівняння

$$1 - 1,2z + 0,2z^2 = 0.$$

3.2 Одиничні корені. *ARIMA* процеси

Розглянемо *ARMA* процес, для якого один з коренів характеристичного рівняння точно дорівнює одиниці, тоді як решта коренів більші за одиницю. У цьому випадку процес y_t можна записати у вигляді

$$\phi^*(L)(1 - L)y_t = \theta(L)\varepsilon_t,$$

де $\phi^*(L)$ – зворотний поліном порядку $p - 1$, або

$$\phi^*(L) \Delta y_t = \theta(L)\varepsilon_t, \quad (3.2.1)$$

де

$$\Delta y_t = (1 - L)y_t = y_t - y_{t-1}.$$

Оскільки корені AR полінома є розв'язками рівняння

$$\phi^*(z)(1 - z) = 0,$$

то маємо, що один розв'язок $z_1 = 1$, а решта коренів z_j лежать ззовні одиничного круга $|z_j| > 1$. У цьому випадку говорять, що заданий процес має простий одиничний корінь.

З рівності (3.2.1) випливає таке: якщо процес y_t має простий одиничний корінь, то часовий ряд Δy_t може бути описаний стаціонарною ARMA моделлю.

Приклад 3.2.1. *Запишемо процес з прикладу 3.1.1 у вигляді*

$$(1 - 0,2L)(1 - L)y_t = (1 - 0,5L)\varepsilon_t.$$

Звідси випливає, що Δy_t описується стаціонарним ARMA(1,1) процесом, який задається рівнянням

$$\Delta y_t = 0,2\Delta y_{t-1} + \varepsilon_t - 0,5\varepsilon_{t-1}.$$

На рис. 3.2.4, 3.2.5 зображено динаміку часового ряду, що описується процесом, наведеним у прикладі 3.1.1 і відповідно динаміку ряду його перших різниць Δy_t .

Рис. 3.2.4. ARMA(2,1) : $y_t = 1,2y_{t-1} - 0,2y_{t-2} + \varepsilon_t - 0,5\varepsilon_{t-1}$

Рис. 3.2.5. *ARMA*(1, 1) : $\Delta y_t = 0,2\Delta y_{t-1} + \varepsilon_t - 0,5\varepsilon_{t-1}$

Часовий ряд, який перетворюється у стаціонарний після обчислення перших різниць, називається інтегровним процесом першого порядку і позначається $I(1)$. Якщо Δy_t описується стаціонарною *ARMA*(p, q) моделлю, то кажуть, що y_t описується авторегресійним інтегровним процесом рухомого середнього порядку $p, 1, q$, або *ARIMA*($p, 1, q$) моделлю.

Стаціонарний часовий ряд є інтегровним порядку нуль, позначаємо $I(0)$, а часові ряди $I(1)$ – це ряди, які перетворюються в стаціонарні після обчислення перших різниць. Білий шум і стаціонарний *AR*(1) процес є прикладами $I(0)$ рядів, тоді як процес випадкового блукання, описаний в (3.1.2), є прикладом $I(1)$ процесу.

Зауважимо, що *AR* поліном має мати точно одиничний корінь. Розглянемо процес, який описується авторегресійною *AR*(1) моделлю з $\phi = 1,01$. Вона є нестаціонарною, оскільки $\phi > 1$, тобто не є $I(0)$. Крім того, у цьому випадку маємо $\Delta y_t = 0,01y_{t-1} + \varepsilon_t$. Тому *AR*(1) процес з $\phi = 1,01$ не є інтегровним процесом порядку один, тобто не є процесом $I(1)$.

Можемо побачити неочікуваний результат від порівняння часових рядів, які мають одиничний корінь і тих, які мають корінь трохи більший за одиницю. Це загальна різниця між

процесами $I(1)$ і $I(0)$. У загальному випадку, головні відмінності між процесами $I(0)$ і $I(1)$ полягають у тому, що $I(0)$ процеси коливаються навколо математичних сподівань з скінченною дисперсією, яка не залежить від часу, тоді як $I(1)$ процеси широко розповсюджені. $I(0)$ процеси є процесами повернення до математичних сподівань, оскільки вони мають тенденцію за довгий період часу повертатись до свого математичного сподівання. $I(0)$ процеси мають обмежену пам'ять минулої поведінки, тоді як $I(1)$ процес має нескінченно велику пам'ять. Цей останній аспект стає зрозумілим з автокореляційних функцій: для $I(0)$ процесу автокореляції зменшуються швидко, зі збільшенням проміжку, тоді як для $I(1)$ процесу оцінка коефіцієнтів автокореляції спадає до нуля дуже повільно. Остання властивість робить одиничний корінь цікавим з економічного погляду. У моделях з одиничними коренями збурення мають стійкий вплив, тоді як у випадку стаціонарних моделей збурення мають лише тимчасовий вплив.

У 80-тих роках з'явилась велика кількість літератури по застосуванню теорії одиничних коренів до багатьох макроекономічних часових рядів. Залежно від того, яку техніку застосовували, часом отримували суперечливі висновки.

Обчислення перших різниць ряду доволі часто може перетворити нестационарний часовий ряд у стаціонарний. Наприклад можемо розглянути випадок, коли y_t вимірює логарифм національного доходу. Тоді Δy_t відповідає зростанню надходжень. У деяких випадках перша різниця недостатня для отримання стаціонарності і треба застосовувати ще одну різницю. У цьому випадку будемо ряд, складений з величин $\Delta(\Delta y_t) = \Delta y_t - \Delta y_{t-1}$, що, зокрема, може відповідати зміні зростання надходжень. Якщо ряд, складений з других різниць $\Delta^2 y_t$, є стаціонарним, то часовий ряд y_t називається інтегровним порядку два і позначається $I(2)$. Такі ряди мають одиничний корінь кратності два. Для часового ряду $I(2)$ ряд Δy_t є

нестационарним, але $\Delta^2 y_t$ – стаціонарний.

На рис. 3.2.6 зображено динаміку часового ряду, що описується процесом, який має одиничний корінь кратності два. На рис. 3.2.7, 3.2.8 зображено динаміку часових рядів, які описують його перші і другі різниці відповідно.

Рис. 3.2.6. $I(2)$: $y_t = 2y_{t-1} - y_{t-2} + \varepsilon_t$

Рис. 3.2.7. $I(1)$: $\Delta y_t = \Delta y_{t-1} + \varepsilon_t$

Рис. 3.2.8. $I(0) : \Delta^2 y_t = \varepsilon_t$

Аналогічно можемо усунути нестационарність перетворенням часового ряду з d -кратним одиничним коренем. Загалом, якщо для перетворення у стаціонарний ряд треба обчислювати різниці до порядку d

$$\Delta^d y_t = (1 - L)^d y_t,$$

причому $\Delta^d y_t$ описується стаціонарним $ARMA(p, q)$ процесом, то процес y_t називається авторегресійним інтегровним процесом рухомого середнього і позначається $ARIMA(p, d, q)$. Перший параметр p належить до порядку авторегресійного оператора, другий d – до порядку інтегрування, третій q позначає порядок оператора рухомого середнього. Обчислюючи різниці порядку d від $ARIMA(p, d, q)$ процесу, породжуємо $ARMA(p, q)$ процес.

Детальніше дослідження процесів з одиничним коренем продовжимо у розділах 7, 11.

3.3 Тотожні корені

Розвинення MA і AR поліномів на добуток лінійних функцій показує проблему тотожних коренів. Це означає, що AR і MA

частини моделей мають однакові корені, яким відповідає лінійна функція від L і яку, очевидно, можна скоротити.

Щоб проілюструвати це, розглянемо процес білого шуму

$$y_t = \varepsilon_t. \quad (3.3.1)$$

Помноживши обидві частини (3.3.1) на $(1 - \rho L)$, матимемо

$$(1 - \rho L)y_t = (1 - \rho L)\varepsilon_t, \quad (3.3.2)$$

або

$$y_t = \rho y_{t-1} + \varepsilon_t - \rho \varepsilon_{t-1}.$$

Зрозуміло, якщо справджується зображення (3.3.1), тоді (3.3.2) виконується для кожного ρ . Отже, (3.3.2) може бути описане як $ARMA(1, 1)$ процес, з $\phi_1 = \rho$ і $\theta_1 = -\rho$. Важливо уникати такої параметризації. Оскільки кожне значення ρ в (3.3.2) описує дані однаково добре, матимемо проблеми з оцінкою параметра ρ . Крім того, теоретичні маніпуляції, які ґрунтуються на зображеннях (3.3.2), можуть пропустити ключові відмінності. Якщо ми використовуємо $ARMA(1, 1)$ модель, в якій θ_1 близьке до $-\phi_1$, то дані могли б краще бути змодельовані як простий білий шум.

Аналогічно можна розглянути $ARMA(2, 1)$ модель, яку запишемо у вигляді

$$(1 - \phi_1 L - \phi_2 L^2) y_t = (1 + \theta L) \varepsilon_t, \quad (3.3.3)$$

або

$$(1 - \varphi_1 L)(1 - \varphi_2 L) y_t = (1 + \theta L) \varepsilon_t.$$

Якщо $\theta = -\varphi_1$, то можемо поділити обидві частини на $(1 - \varphi_1 L)$ і отримаємо

$$(1 - \varphi_2 L) y_t = \varepsilon_t,$$

що є $AR(1)$ процесом, який еквівалентний до (3.3.3).

Приклад 3.3.1. Розглянемо модель

$$y_t = y_{t-1} - 0,25y_{t-2} + \varepsilon_t - 0,5\varepsilon_{t-1}, \quad (3.3.4)$$

яка може бути записана як

$$(1 - 0,5L)(1 - 0,25L)y_t = (1 - 0,5L)\varepsilon_t.$$

Очевидно, що цю модель можна звести до AR(1) моделі

$$y_t = 0,5y_{t-1} + \varepsilon_t,$$

яка описує процес еквівалентний до (3.3.4).

Тотожні корені можуть виникати в ARMA(p, q) моделях. Розглянемо розвинення на множники поліноміальних операторів цього процесу

$$\begin{aligned} (1 - \lambda_1L)(1 - \lambda_2L) \cdots (1 - \lambda_pL)(y_t - \mu) = \\ = (1 - \eta_1L)(1 - \eta_2L) \cdots (1 - \eta_qL)\varepsilon_t. \end{aligned} \quad (3.3.5)$$

Припускаємо, що $|\lambda_i| < 1$ для кожного i , тобто процес є коваріаційно - стаціонарним. Якщо авторегресійний оператор $(1 - \phi_1L - \phi_2L^2 - \dots - \phi_pL^p)$ і оператор рухомого середнього $(1 + \theta_1L + \theta_2L^2 + \dots + \theta_qL^q)$ мають однакові корені, наприклад, $\lambda_i = \eta_j$ для деякого i та j , то обидві частини (3.3.5) можна поділити на $(1 - \lambda_iL)$

$$\prod_{\substack{k=1 \\ k \neq i}}^p (1 - \lambda_kL)(y_t - \mu) = \prod_{\substack{k=1 \\ k \neq j}}^q (1 - \eta_kL)\varepsilon_t,$$

або

$$\begin{aligned} (1 - \phi_1^*L - \phi_2^*L^2 - \dots - \phi_{p-1}^*L^{p-1})(y_t - \mu) = \\ = (1 + \theta_1^*L + \theta_2^*L^2 + \dots + \theta_{q-1}^*L^{q-1})\varepsilon_t, \end{aligned} \quad (3.3.6)$$

де

$$\begin{aligned} & (1 - \phi_1^* L - \phi_2^* L^2 - \dots - \phi_{p-1}^* L^{p-1}) \equiv \\ & \equiv (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_{i-1} L)(1 - \lambda_{i+1} L) \cdots (1 - \lambda_p L), \end{aligned}$$

$$\begin{aligned} & (1 + \theta_1^* L + \theta_2^* L^2 + \dots + \theta_{q-1}^* L^{q-1}) \equiv \\ & \equiv (1 - \eta_1 L)(1 - \eta_2 L) \cdots (1 - \eta_{j-1} L)(1 - \eta_{j+1} L) \cdots (1 - \eta_q L). \end{aligned}$$

Тому у випадку одного тотожного кореня стаціонарна $ARMA(p, q)$ модель може бути записана у вигляді еквівалентної стаціонарної $ARMA(p-1, q-1)$ моделі.

Проблема тотожних коренів ілюструє труднощі оцінювання $ARMA$ моделі з AR і MA частинами високих порядків. Причина складності в тому, що корені MA і AR поліномів є майже однаковими. У цьому випадку спрощена $ARMA(p-1, q-1)$ модель буде майже еквівалентним зображенням. Отже, тотожні корені призводять до надпараметризації моделі.

3.4 Автоковаріаційна генеруюча функція

Нехай $\{\gamma_j\}_{j=-\infty}^{\infty}$ послідовність автоковаріацій коваріаційно-стаціонарного процесу y_t . Припустимо, що ряд, складений з коефіцієнтів автоковаріацій, абсолютно збіжний, тобто

$$\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty.$$

Означення 3.4.1. Функція

$$g_y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j, \quad (3.4.1)$$

аргумент z якої є комплексним скаляром, називається автоковаріаційною генеруючою функцією.

Особливий інтерес викликають ті значення z , які лежать на комплексному одиничному колі

$$z = \cos(\omega) - i \sin(\omega) = e^{-i\omega}.$$

Якщо в автоковаріаційну генеруючу функцію підставити $z = e^{-i\omega}$ і поділити на 2π , то отримаємо функцію від ω

$$s_y(\omega) = \frac{1}{2\pi} g_y(e^{-i\omega}) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j},$$

яка називається спектром ряду y_t . Для процесів з абсолютно сумовними автоковаріаціями функція $s_y(\omega)$ існує і може бути використана для обчислення усіх автоковаріацій. Звідси випливає, що коли два різних процеси мають однакові автоковаріаційні генеруючі функції, то ці два процеси мають однакові послідовності автоковаріацій.

Обчислимо автоковаріаційну генеруючу функцію $MA(1)$ процесу. З формул (2.2.3), (2.2.4), (2.2.5) у розділі 2 випливає, що автоковаріаційна генеруюча функція має вигляд

$$\begin{aligned} g_y(z) &= [\theta\sigma^2] z^{-1} + [(1 + \theta^2)\sigma^2] z^0 + [\theta\sigma^2] z^1 = \\ &= \sigma^2 [\theta z^{-1} + (1 + \theta^2) + \theta z]. \end{aligned} \quad (3.4.2)$$

Зауважимо, що цей вираз можна записати у вигляді

$$g_y(z) = \sigma^2 (1 + \theta z)(1 + \theta z^{-1}). \quad (3.4.3)$$

Подібний вигляд має автоковаріаційна генеруюча функція для $MA(q)$ процесу

$$y_t = \mu + (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t.$$

У цьому випадку маємо

$$g_y(z) = \sigma^2(1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q) \times \\ \times (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \dots + \theta_q z^{-q}). \quad (3.4.4)$$

Перемножимо дужки в (3.4.4) і згрупуємо члени з однаковими степенями z

$$(1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q) (1 + \theta_1 z^{-1} + \theta_2 z^{-2} + \dots + \theta_q z^{-q}) = \\ = (\theta_q) z^q + (\theta_{q-1} + \theta_q \theta_1) z^{(q-1)} + \\ + (\theta_{q-2} + \theta_{q-1} \theta_1 + \theta_q \theta_2) z^{(q-2)} + \dots + \\ + (\theta_1 + \theta_2 \theta_1 + \theta_3 \theta_2 + \dots + \theta_q \theta_{q-1}) z^1 + \\ + (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) z^0 + \\ + (\theta_1 + \theta_2 \theta_1 + \theta_2 \theta_3 + \dots + \theta_q \theta_{q-1}) z^{-1} + \dots + (\theta_q) z^{-q}. \quad (3.4.5)$$

Порівнюючи (3.4.5) з (2.3.4), бачимо, що коефіцієнти біля z^j в (3.4.5) є справді j -ми автоковаріаціями.

Знайдемо автоковаріаційну генеруючу функцію $g_y(z)$ у випадку $MA(\infty)$ процесу. Нехай

$$y_t = \mu + \psi(L) \varepsilon_t, \quad (3.4.6)$$

де

$$\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \dots \quad (3.4.7)$$

і

$$\sum_{j=0}^{\infty} |\psi_j| < \infty, \quad (3.4.8)$$

тоді

$$g_y(z) = \sigma^2 \psi(z) \psi(z^{-1}). \quad (3.4.9)$$

Розглянемо побудову автоковаріаційної генеруючої функції для авторегресійних процесів. Зокрема, стаціонарний $AR(1)$ процес можна записати у вигляді

$$y_t - \mu = (1 - \phi L)^{-1} \varepsilon_t,$$

який збігається з (3.4.6) для $\psi(L) = 1/(1 - \phi L)$. Тоді автоковаріаційну генеруючу функцію для $AR(1)$ процесу можна обчислити за формулою

$$g_y(z) = \frac{\sigma^2}{(1 - \phi z)(1 - \phi z^{-1})}. \quad (3.4.10)$$

Справді, перетворивши (3.4.10)

$$\begin{aligned} & \frac{\sigma^2}{(1 - \phi z)(1 - \phi z^{-1})} = \\ & = \sigma^2(1 + \phi z + \phi^2 z^2 + \phi^3 z^3 + \dots)(1 + \phi z^{-1} + \phi^2 z^{-2} + \phi^3 z^{-3} + \dots), \end{aligned}$$

матимемо, що коефіцієнти біля z^j мають вигляд

$$\sigma^2(\phi^j + \phi^{j+1}\phi + \phi^{j+2}\phi^2 + \dots) = \frac{\sigma^2\phi^j}{1 - \phi^2},$$

тобто збігаються з формулами (2.2.5) для j -их автоковаріацій, які виведено у розділі 2.

У загальному випадку автоковаріаційна генеруюча функція для стаціонарного $ARMA(p, q)$ процесу має вигляд

$$\begin{aligned} g_y(z) &= \quad (3.4.11) \\ &= \frac{\sigma^2(1 + \theta_1 z + \dots + \theta_q z^q)(1 + \theta_1 z^{-1} + \dots + \theta_q z^{-q})}{(1 - \phi_1 z - \dots - \phi_p z^p)(1 - \phi_1 z^{-1} - \dots - \phi_p z^{-p})}. \end{aligned}$$

Фільтрування

Іноді дані відфільтровані, тобто перетворені чи оброблені перед тим, як ми їх аналізуємо. Визначимо наслідки впливу такої обробки на автоковаріації. Ці висновки легко зробити на підставі аналізу автоковаріаційної генеруючої функції.

Припустимо, що початкові дані y_t описуються $MA(1)$ процесом

$$y_t = (1 + \theta L) \varepsilon_t \quad (3.4.12)$$

з автоковаріаційною генеруючою функцією заданою в (3.4.2).

Нехай спостережувані дані x_t зображають перші різниці ряду y_t

$$x_t = y_t - y_{t-1} = (1 - L)y_t. \quad (3.4.13)$$

Підставляючи (3.4.12) в (3.4.13), отримаємо, що спостережувані дані можуть охарактеризувати такий $MA(2)$ процес:

$$\begin{aligned} x_t &= (1 - L)(1 + \theta L) \varepsilon_t = (1 + (\theta - 1)L - \theta L^2) \varepsilon_t = \\ &= (1 + \theta_1 L + \theta_2 L^2) \varepsilon_t, \end{aligned} \quad (3.4.14)$$

де $\theta_1 \equiv (\theta - 1)$ і $\theta_2 = -\theta$. Автоковаріаційну генеруючу функцію спостережуваних даних x_t можна обчислити за допомогою (3.4.4)

$$g_x(z) = \sigma^2(1 + \theta_1 z + \theta_2 z^2)(1 + \theta_1 z^{-1} + \theta_2 z^{-2}). \quad (3.4.15)$$

Згідно з (3.4.14), поліном $(1 + \theta_1 z + \theta_2 z^2)$ можна переписати у вигляді

$$(1 + \theta_1 z + \theta_2 z^2) = (1 - z)(1 + \theta z).$$

Тоді (3.4.15) запишеться як

$$\begin{aligned} g_x(z) &= \sigma^2(1 - z)(1 + \theta z)(1 - z^{-1})(1 + \theta z^{-1}) = \\ &= (1 - z)(1 - z^{-1})g_y(z). \end{aligned} \quad (3.4.16)$$

Вирази (3.4.15) і (3.4.16) зображають ту саму функцію від z , який вираз ми оберемо – це питання зручності. Отже, застосування фільтра $(1 - L)$ до y_t дає той самий результат, що і множення автоковаріаційної генеруючої функції на $(1 - z)(1 - z^{-1})$.

Узагальнимо цей висновок. Припустимо, що початкові дані y_t задовольняють (3.4.6) – (3.4.8), спостережувані дані – це, відповідно, фільтровані значення y_t . Нехай

$$x_t = h(L)y_t, \quad (3.4.17)$$

де

$$h(L) = \sum_{j=-\infty}^{\infty} h_j L^j, \quad \sum_{j=-\infty}^{\infty} |h_j| < \infty.$$

Підставляючи (3.4.6) в (3.4.17), матимемо, що спостережувані дані x_t описуються так:

$$x_t = h(1)\mu + h(L)\psi(L)\varepsilon_t \equiv \mu^* + \phi^*(L)\varepsilon_t,$$

де $\mu^* \equiv h(1)\mu$ і $\phi^*(L) \equiv h(L)\psi(L)$.

Доведемо, що послідовність коефіцієнтів $\{\psi_j^*\}_{j=-\infty}^{\infty}$, які відповідають оператору ψ^* , абсолютно сумовна. Для цього перетворимо

$$\begin{aligned} \psi^*(z) &= \left(\sum_{j=-\infty}^{\infty} h_j z^j \right) \left(\sum_{k=0}^{\infty} \psi_k z^k \right) = \\ &= (\dots + h_{-j} z^{-j} + h_{-j+1} z^{-j+1} + \dots + h_{-1} z^{-1} + \\ &+ h_0 z^0 + h_1 z^1 + \dots + h_j z^j + h_{j+1} z^{j+1} + \dots) \times \\ &\quad \times (\psi_0 z^0 + \psi_1 z^1 + \psi_2 z^2 + \dots), \end{aligned}$$

звідки бачимо, що коефіцієнт біля z^j матиме вигляд

$$\psi_j^* = h_j \psi_0 + h_{j-1} \psi_1 + h_{j-2} \psi_2 + \dots = \sum_{k=0}^{\infty} h_{j-k} \psi_k.$$

Тоді

$$\begin{aligned} \sum_{j=-\infty}^{\infty} |\psi_j^*| &= \sum_{j=-\infty}^{\infty} \left| \sum_{k=0}^{\infty} h_{j-k} \psi_k \right| \leq \sum_{j=-\infty}^{\infty} \sum_{k=0}^{\infty} |h_{j-k} \psi_k| = \\ &= \sum_{k=0}^{\infty} |\psi_k| \sum_{j=-\infty}^{\infty} |h_{j-k}| = \sum_{k=0}^{\infty} |\psi_k| \sum_{j=-\infty}^{\infty} |h_j| < \infty \end{aligned}$$

і автоковаріаційну генеруючу функцію перетворених даних x_t можна обчислити як

$$\begin{aligned} g_x(z) &= \sigma^2 \psi^*(z) \psi^*(z^{-1}) = \sigma^2 h(z) \psi(z) \psi(z^{-1}) h(z^{-1}) = \\ &= h(z) h(z^{-1}) g_y(z). \end{aligned} \quad (3.4.18)$$

Отже, застосування фільтру $h(L)$ до ряду y_t спричинює множення його автоковаріаційної генеруючої функції на $h(z)h(z^{-1})$.

3.5 Зворотність процесів часових рядів

Зворотність авторегресійних процесів еквівалентна стаціонарності цих процесів. А саме, якщо для $AR(p)$ процесу

$$y_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

корені характеристичного рівняння

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

лежать ззовні одиничного круга, то такий процес еквівалентний до процесу $MA(\infty)$, тобто є зворотним. Процеси рухомого середнього стаціонарні за побудовою. Проте для їхньої зворотності треба накладати додаткові умови.

Дослідимо зворотність процесів рухомого середнього.

3.5.1 Зворотність $MA(1)$ процесу

Розглянемо $MA(1)$ процес

$$y_t - \mu = (1 + \theta L)\varepsilon_t, \quad (3.5.1)$$

де

$$E[\varepsilon_t \varepsilon_s] = \begin{cases} \sigma^2 & \text{для } t = s, \\ 0 & \text{для } t \neq s. \end{cases}$$

Зауважимо таке: якщо $|\theta| < 1$, то

$$\begin{aligned} (1 + \theta L)^{-1} &= [1 - (-\theta L)]^{-1} = \\ &= 1 + (-\theta)L + (-\theta)^2 L^2 + (-\theta)^3 L^3 + \dots \end{aligned}$$

Помноживши обидві частини (3.5.1) на $(1 + \theta L)^{-1}$, отримаємо

$$(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \dots)(y_t - \mu) = \varepsilon_t, \quad (3.5.2)$$

що можна розглядати як $AR(\infty)$ процес.

Якщо зображення процесу рухомого середнього (3.5.1) за допомогою оператора, оберненого до оператора $(1 + \theta L)$, може бути записано як $AR(\infty)$ зображення (3.5.2), то процес рухомого середнього називається зворотним.

Отже, для $MA(1)$ процесу зворотність вимагає, щоб $|\theta| < 1$. Якщо $|\theta| \geq 1$, то нескінченну послідовність (3.5.2) не можна визначити.

Дослідимо зворотність у термінах першого і другого моментів. Нехай $MA(1)$ процес (3.5.1) має математичне сподівання μ і автоковаріаційну генеруючу функцію

$$g_y(z) = \sigma^2(1 + \theta z)(1 + \theta z^{-1}). \quad (3.5.3)$$

Розглянемо інший $MA(1)$ процес

$$\tilde{y}_t - \mu = (1 + \tilde{\theta}L)\tilde{\varepsilon}_t, \quad (3.5.4)$$

де

$$E[\tilde{\varepsilon}_t \tilde{\varepsilon}_s] = \begin{cases} \tilde{\sigma}^2 & \text{для } t = s, \\ 0 & \text{для } t \neq s. \end{cases}$$

Зауважимо, що \tilde{y}_t і y_t мають однакові математичні сподівання μ . Автоковаріаційна генеруюча функція для процесу (3.5.4) має вигляд

$$\begin{aligned} g_{\tilde{y}}(z) &= \tilde{\sigma}^2(1 + \tilde{\theta}z)(1 + \tilde{\theta}z^{-1}) = \\ &= \tilde{\sigma}^2\{(\tilde{\theta}^{-1}z^{-1} + 1)(\tilde{\theta}z)\}\{(\tilde{\theta}^{-1}z + 1)(\tilde{\theta}z^{-1})\} = \quad (3.5.5) \\ &= (\tilde{\sigma}^2\tilde{\theta}^2)(1 + \tilde{\theta}^{-1}z)(1 + \tilde{\theta}^{-1}z^{-1}). \end{aligned}$$

Припустимо, що параметри $(\tilde{\theta}, \tilde{\sigma}^2)$ процесу (3.5.4) пов'язані з параметрами (θ, σ) процесу (3.5.1) так:

$$\theta = \tilde{\theta}^{-1}, \quad (3.5.6)$$

$$\sigma^2 = \tilde{\theta}^2\tilde{\sigma}^2. \quad (3.5.7)$$

Тоді автоковаріаційні генеруючі функції (3.5.3) і (3.5.5) будуть однакові, тобто y_t і \tilde{y}_t будуть мати однакові перші та другі моменти.

З (3.5.6) випливає таке: якщо $|\theta| < 1$, то $|\tilde{\theta}| > 1$. Іншими словами, для кожного зворотного $MA(1)$ зображення (3.5.1) можна знайти незворотне $MA(1)$ зображення (3.5.4) з однаковими першими та другими моментами. Навпаки, для заданого незворотного зображення з $|\tilde{\theta}| > 1$ існує зворотне зображення з $\theta = (1/\tilde{\theta})$ з однаковими першими та другими моментами. В граничних випадках, коли $\theta = \pm 1$, існує лише одне зображення процесу – незворотне.

Крім того, що зворотні і незворотні зображення мають ті самі моменти, правильна лема.

Лема 3.5.1. *Будь-яке зображення (3.5.1) чи (3.5.4) можна використати як однаково адекватний опис будь-якого $MA(1)$ процесу.*

Д о в е д е н н я. Нехай маємо нескінченну послідовність \tilde{y}_t побудовану за допомогою (3.5.4) з $\tilde{\theta} > 1$. Отже, ми знаємо,

що ці дані згенеровані $MA(1)$ процесом, який має незворотне зображення.

На підставі даних \tilde{y}_t визначимо ряд $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ так:

$$\begin{aligned}\varepsilon_t &\equiv (1 + \theta L)^{-1}(\tilde{y}_t - \mu) = \\ &= (\tilde{y}_t - \mu) - \theta(\tilde{y}_{t-1} - \mu) + \theta^2(\tilde{y}_{t-2} - \mu) - \theta^3(\tilde{y}_{t-3} - \mu) + \dots,\end{aligned}\quad (3.5.8)$$

де $\theta = 1/\tilde{\theta}$. Зауважимо таке: оскільки $|\theta| < 1$, то послідовність ε_t збігається в середньому квадратичному.

Більше того, послідовність ε_t є білим шумом. Найпростішим шляхом перевірки цього є обчислення автоковаріаційної генеруючої функції для ε_t . Якщо коефіцієнти в $g_\varepsilon(z)$ біля z^j (j -та автоковаріація) дорівнюють 0 для кожного $j \neq 0$, то матимемо, що ε_t є процесом білого шуму. З (3.5.8) і (3.4.18) випливає, що автоковаріаційна генеруюча функція для ε_t має вигляд

$$g_\varepsilon = (1 + \theta z)^{-1}(1 + \theta z^{-1})^{-1}g_{\tilde{y}}(z). \quad (3.5.9)$$

Підставивши (3.5.5) в (3.5.9), маємо

$$\begin{aligned}g_\varepsilon &= (1 + \theta z)^{-1}(1 + \theta z^{-1})^{-1}(\tilde{\sigma}^2\tilde{\theta}^2)(1 + \tilde{\theta}^{-1}z)(1 + \tilde{\theta}^{-1}z^{-1}) = \\ &= \tilde{\sigma}^2\tilde{\theta}^2,\end{aligned}\quad (3.5.10)$$

оскільки за побудовою $\tilde{\theta}^{-1} = \theta$. З (3.5.10) бачимо, що ε_t є процесом білого шуму з дисперсією $\tilde{\sigma}^2\tilde{\theta}^2$.

Помноживши обидві сторони (3.5.8) на $(1 + \theta L)$, отримаємо

$$\tilde{y}_t - \mu = (1 + \theta L)\varepsilon_t,$$

тобто правильне зворотне $MA(1)$ зображення даних, які ми побудували на основі незворотного зображення (3.5.4).

Обернене твердження також правильне. Припустимо, що дані генеруються на основі зворотного зображення (3.5.1) з $|\theta| < 1$. Тоді існує незворотне зображення з $\tilde{\theta} = 1/\theta$, яке описує

ці дані з однаковою точністю. Щоб охарактеризувати це незворотне зображення, розглянемо оператор обернений до $(1 + \tilde{\theta}L)$

$$\begin{aligned} (1 + \tilde{\theta}L)^{-1} &= \tilde{\theta}^{-1}L^{-1}(1 + \tilde{\theta}^{-1}L^{-1})^{-1} = \\ &= (\tilde{\theta}^{-1}L^{-1})[1 - \tilde{\theta}^{-1}L^{-1} + \tilde{\theta}^{-2}L^{-2} - \tilde{\theta}^{-3}L^{-3} + \dots] = \\ &= \theta L^{-1}[1 - \theta L^{-1} + \theta^2 L^{-2} - \theta^3 L^{-3} + \dots]. \end{aligned}$$

Визначимо $\tilde{\varepsilon}_t$ так:

$$\tilde{\varepsilon}_t \equiv \theta(y_{t+1} - \mu) - \theta^2(y_{t+2} - \mu) + \theta^3(y_{t+3} - \mu) - \dots \quad (3.5.11)$$

Зауважимо, що цей ряд збігається для $|\theta| < 1$. Крім того, ряд (3.5.11) є білим шумом, оскільки

$$\begin{aligned} g_{\tilde{\varepsilon}}(z) &= \theta z^{-1} (1 - \theta z^{-1} + \theta^2 z^{-2} - \theta^3 z^{-3} + \dots) \theta z \times \\ &\quad \times (1 - \theta z^1 + \theta^2 z^2 - \theta^3 z^3 + \dots) \sigma^2 (1 + \theta z)(1 + \theta z^{-1}) = \\ &= \theta^2 \sigma^2. \end{aligned}$$

Коефіцієнти біля z^j дорівнюють 0 для $j \neq 0$, отже, всі автоковаріації дорівнюють нулю, а тому $\tilde{\varepsilon}_t$ є білим шумом, що й треба було довести. За побудовою

$$y_t - \mu = (1 + \tilde{\theta}L)\tilde{\varepsilon}_t.$$

Отже, ми знайшли незворотне $MA(1)$ зображення даних, що були фактично згенеровані зворотним $MA(1)$ зображенням (3.5.1). \square

На рис.3.5.9, 3.5.10 зображено динаміку часових рядів, які описуються $MA(1)$ процесами, що мають відповідно зворотне ($\theta < 1$) і незворотне ($\theta > 1$) зображення.

Зворотне або незворотне зображення можуть характеризувати будь-які дані однаково добре, але є практична причина використовувати зворотне зображення. Зокрема, зауважимо

Рис. 3.5.9. $MA(1)$: $y_t = \varepsilon_t + 0,1\varepsilon_{t-1}$

Рис. 3.5.10. $MA(1)$: $y_t = \varepsilon_t + 10\varepsilon_{t-1}$

таке: щоб знайти значення ε в момент часу t пов'язане зі зворотним зображення (3.5.8), нам треба знати попередні значення y_t . Для знаходження значення $\tilde{\varepsilon}$ в момент часу t пов'язане зі зворотним зображенням (3.5.11), нам потрібно використати усі наступні значення y_t : y_{t+1}, y_{t+2}, \dots

Як буде зазначено в наступних розділах, деякі зручні алгоритми оцінювання параметрів і прогнозу можна застосувати тільки тоді, коли використовується зворотне зображення. Значення ε_t , пов'язане зі зворотним зображенням, іноді називають фундаментальним для y_t .

3.5.2 Зворотність $MA(q)$ процесу

Розглянемо $MA(q)$ процес

$$(y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t, \quad (3.5.12)$$

$$E[\varepsilon_t \varepsilon_s] = \begin{cases} \sigma^2 & \text{для } t = s, \\ 0 & \text{для } t \neq s. \end{cases}$$

Якщо корені рівняння

$$(1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q) = 0 \quad (3.5.13)$$

лежать ззовні одиничного круга, то (3.5.12) можна записати як $AR(\infty)$ процес

$$(1 + \eta_1 L + \eta_2 L^2 + \eta_3 L^3 + \dots)(y_t - \mu) = \varepsilon_t,$$

де

$$(1 + \eta_1 L + \eta_2 L^2 + \eta_3 L^3 + \dots) = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)^{-1}.$$

У цьому випадку $MA(q)$ зображення (3.5.12) є зворотним.

Запишемо оператор рухомого середнього у вигляді

$$(1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) = (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_q L). \quad (3.5.14)$$

Якщо $|\lambda_i| < 1$ для кожного i , то всі корені (3.5.13) лежать ззовні одиничного круга і зображення (3.5.12) – зворотне.

У випадку, коли деякі λ_i лежать ззовні, але не на одиничному колі, тобто $|\lambda_i| > 1$ для деяких i , Хенсен і Сарджент пропонували таку процедуру знаходження зворотного зображення. Автоковаріаційну генеруючу функція для y_t можна записати у вигляді

$$g_y(z) = \sigma^2 \{(1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_q z)\} \times \{(1 - \lambda_1 z^{-1})(1 - \lambda_2 z^{-1}) \cdots (1 - \lambda_q z^{-1})\}. \quad (3.5.15)$$

Впорядкуємо λ так, щоб $(\lambda_1, \lambda_2, \dots, \lambda_n)$ були всередині одиничного круга, а $(\lambda_{n+1}, \lambda_{n+2}, \dots, \lambda_q)$ ззовні одиничного круга. Замінімо σ^2 в (3.5.15) на $\sigma^2 \lambda_{n+1}^2 \lambda_{n+2}^2 \cdots \lambda_q^2$. Оскільки комплексні λ_i з'являються як спряжені пари, то такий добуток є додатним дійсним числом. Припустимо також, що $(\lambda_{n+1}, \lambda_{n+2}, \dots, \lambda_q)$ замінено на $(\lambda_{n+1}^{-1}, \lambda_{n+2}^{-1}, \dots, \lambda_q^{-1})$. Результуюча функція матиме вигляд

$$\begin{aligned}
& \sigma^2 \lambda_{n+1}^2 \lambda_{n+2}^2 \cdots \lambda_q^2 \prod_{i=1}^n (1 - \lambda_i z) \prod_{i=n+1}^q (1 - \lambda_i^{-1} z) \times \\
& \quad \times \prod_{i=1}^n (1 - \lambda_i z^{-1}) \prod_{i=n+1}^q (1 - \lambda_i^{-1} z^{-1}) = \\
& = \sigma^2 \prod_{i=1}^n (1 - \lambda_i z) \prod_{i=n+1}^q (\lambda_i z^{-1} (1 - \lambda_i^{-1} z)) \times \\
& \quad \times \prod_{i=1}^n (1 - \lambda_i z^{-1}) \prod_{i=n+1}^q (\lambda_i z (1 - \lambda_i^{-1} z^{-1})) = \\
& = \sigma^2 \prod_{i=1}^n (1 - \lambda_i z) \prod_{i=n+1}^q (\lambda_i z^{-1} - 1) \prod_{i=1}^n (1 - \lambda_i z^{-1}) \prod_{i=n+1}^q (\lambda_i z - 1) = \\
& = \sigma^2 \prod_{i=1}^q (1 - \lambda_i z) \prod_{i=1}^q (1 - \lambda_i z^{-1}),
\end{aligned} \tag{3.5.16}$$

яка збігається з (3.5.15).

Отже, якщо незворотне зображення $MA(q)$ процесу має вигляд

$$y_t = \mu + \prod_{i=1}^q (1 - \lambda_i L) \tilde{\varepsilon}_t, \tag{3.5.17}$$

де

$$|\lambda_i| < 1 \quad \text{для} \quad i = 1, 2, \dots, n,$$

$$|\lambda_i| > 1 \quad \text{для} \quad i = n+1, n+2, \dots, q$$

і

$$E[\tilde{\varepsilon}_t \tilde{\varepsilon}_s] = \begin{cases} \tilde{\sigma}^2 & \text{для } t = s, \\ 0 & \text{для } t \neq s, \end{cases}$$

то еквівалентне до нього зворотне зображення має вигляд

$$y_t = \mu + \prod_{i=1}^n (1 - \lambda_i L) \prod_{i=n+1}^q (1 - \lambda_i^{-1} L) \varepsilon_t, \quad (3.5.18)$$

де

$$E[\varepsilon_t \varepsilon_s] = \begin{cases} \tilde{\sigma}^2 \lambda_{n+1}^2 \lambda_{n+2}^2 \cdots \lambda_q^2 & \text{для } t = s, \\ 0 & \text{для } t \neq s. \end{cases}$$

Тоді (3.5.17), (3.5.18) мають однакові автоковаріаційні генеруючі функції, хоча лише зображення (3.5.18) задовольняє умову зворотності.

Отже, можна зробити висновок, що існує множина альтернативних $MA(q)$ зображень для y_t , пов'язаних з усіма можливими λ_i і λ_i^{-1} , але тільки одне з них має усі λ_i на або всередині одиничного круга.

3.6 Сумування ARMA процесів

3.6.1 Додавання $MA(1)$ процесу і білого шуму

Нехай ряд x_t описується $MA(1)$ процесом з нульовим математичним сподіванням

$$x_t = u_t + \delta u_{t-1}, \quad (3.6.1)$$

де u_t – процес білого шуму, для якого

$$E[u_t u_{t-j}] = \begin{cases} \sigma_u^2 & \text{для } j = 0, \\ 0 & \text{для } j = 1, 2, \dots \end{cases}$$

Тоді автоковаріації ряду x_t мають вигляд

$$E[x_t x_{t-j}] = \begin{cases} (1 + \delta^2)\sigma_u^2 & \text{для } j = 0, \\ \delta\sigma_u^2 & \text{для } j = 1, \\ 0 & \text{для } j = 2, 3, \dots \end{cases} \quad (3.6.2)$$

Нехай ν_t – білий шум, для якого

$$E[\nu_t \nu_{t-j}] = \begin{cases} \sigma_\nu^2 & \text{для } j = 0, \\ 0 & \text{для } j = 1, 2, \dots \end{cases} \quad (3.6.3)$$

Крім того, припустимо, що ν_t і u_t не корелюють між собою в будь-які моменти часу

$$E[u_t \nu_{t-j}] = 0 \quad \text{для всіх } j,$$

звідки випливає, що

$$E[x_t \nu_{t-j}] = 0 \quad \text{для всіх } j. \quad (3.6.4)$$

Нехай спостережуваний ряд y_t зображається як сума $MA(1)$ процесу і процесу білого шуму

$$y_t = x_t + \nu_t = u_t + \delta u_{t-1} + \nu_t. \quad (3.6.5)$$

Дослідимо, якими властивостями володіє часовий ряд y_t .

Очевидно, що y_t має математичне сподівання нуль, а його автоковаріації можна вивести з (3.6.2) – (3.6.4)

$$\begin{aligned} E[y_t y_{t-j}] &= E[(x_t + \nu_t)(x_{t-j} + \nu_{t-j})] = \\ &= E[x_t x_{t-j}] + E[\nu_t \nu_{t-j}] = \\ &= \begin{cases} (1 + \delta^2)\sigma_u^2 + \sigma_\nu^2 & \text{для } j = 0, \\ \delta\sigma_u^2 & \text{для } j = 1, \\ 0 & \text{для } j = 2, 3, \dots \end{cases} \end{aligned} \quad (3.6.6)$$

Отже, сума $x_t + \nu_t$ є коваріаційно - стаціонарним процесом і його автоковаріації порядку два і вище дорівнюють нулю. Тому y_t також є $MA(1)$ процесом.

Визначимо вигляд $MA(1)$ зображення з нульовим математичним сподіванням, за допомогою якого можна описати дані y_t . Нехай

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}, \quad (3.6.7)$$

де

$$E[\varepsilon_t\varepsilon_{t-j}] = \begin{cases} \sigma^2 & \text{для } j = 0, \\ 0 & \text{для } j = 1, 2, \dots \end{cases}$$

Знайдемо значення θ і σ^2 , які відповідають зображенню (3.6.7). Зауважимо, що автоковаріації процесу (3.6.7) мають вигляд

$$E[y_t y_{t-j}] = \begin{cases} (1 + \theta^2)\sigma^2 & \text{для } j = 0, \\ \theta\sigma^2 & \text{для } j = 1, \\ 0 & \text{для } j = 2, 3, \dots \end{cases}$$

Порівнюючи їх з (3.6.6), отримаємо рівняння

$$(1 + \theta^2)\sigma^2 = (1 + \delta^2)\sigma_u^2 + \sigma_v^2 \quad (3.6.8)$$

і

$$\theta\sigma^2 = \delta\sigma_u^2. \quad (3.6.9)$$

Рівняння (3.6.9) можна розв'язати стосовно σ^2 . Матимемо

$$\sigma^2 = \frac{\delta\sigma_u^2}{\theta}. \quad (3.6.10)$$

Підставивши (3.6.10) в (3.6.8), одержимо

$$(1 + \theta^2)(\delta\sigma_u^2/\theta) = (1 + \delta^2)\sigma_u^2 + \sigma_v^2$$

або

$$(1 + \theta^2)\delta = [(1 + \delta^2) + (\sigma_v^2/\sigma_u^2)]\theta,$$

звідки

$$\delta\theta^2 - [(1 + \delta^2) + (\sigma_v^2/\sigma_u^2)]\theta + \delta = 0. \quad (3.6.11)$$

За відомих значень δ , σ_u^2 , σ_v^2 рівняння (3.6.11) є квадратним рівнянням стосовно θ . Тому існує два значення θ , які є коренями рівняння (3.6.11) і відповідно можна обчислити за формулою

$$\theta = \frac{[(1 + \delta^2) + (\sigma_v^2/\sigma_u^2)] \pm \sqrt{[(1 + \delta^2) + (\sigma_v^2/\sigma_u^2)]^2 - 4\delta^2}}{2\delta}. \quad (3.6.12)$$

Зокрема, якщо σ_v^2 дорівнює нулю, то квадратне рівняння (3.6.11) має вигляд

$$\delta\theta^2 - (1 + \delta^2)\theta + \delta = 0 \quad (3.6.13)$$

або

$$\delta(\theta - \delta)(\theta - \delta^{-1}) = 0.$$

Його розв'язками є величини $\theta = \delta$ і $\tilde{\theta} = \delta^{-1}$, тобто параметри процесу рухомого середнього відповідно зворотного і незворотного зображень.

Розглянемо ліві частини рівнянь (3.6.11) і (3.6.13) як функції від θ за припущення додатної автокореляції ($\delta > 0$). Нехай

$$f_1(\theta) = \delta\theta^2 - [(1 + \delta^2) + (\sigma_v^2/\sigma_u^2)]\theta + \delta,$$

$$f_2(\theta) = \delta\theta^2 - (1 + \delta^2)\theta + \delta.$$

На рис. 3.6.1 зображено графіки цих парабол. Значення функції $f_1(\theta)$ в кожній точці є меншими за значення функції $f_2(\theta)$ на величину $(\sigma_v^2/\sigma_u^2)\theta$, відповідно її графік лежить нижче.

Рівняння (3.6.11) має два дійсних розв'язки для θ , які позначимо θ^* і $\tilde{\theta}^*$ і для яких виконуються нерівності

$$0 < |\theta^*| < |\delta| < 1,$$

$$1 < |\delta^{-1}| < |\tilde{\theta}^*|. \quad (3.6.14)$$

Тому θ^* відповідає зворотному зображенню, а $\tilde{\theta}^*$ – незворотному зображенню $MA(1)$ процесу. Тоді з (3.6.10) матимемо $(\sigma^*)^2 = \delta\sigma_u^2/\theta^*$.

Рис. 3.6.11.

Визначивши (θ^*, σ^{*2}) як параметри зворотного зображення, дослідимо чи (3.6.7) і (3.6.5) справді однаково характеризують дані y_t . Для цього треба, щоб виконувалась умова

$$(1 + \theta^* L)\varepsilon_t = (1 + \delta L)u_t + \nu_t \quad (3.6.15)$$

або

$$\begin{aligned} \varepsilon_t &= (1 + \theta^* L)^{-1}[(1 + \delta L)u_t + \nu_t] = \\ &= (u_t - \theta^* u_{t-1} + \theta^{*2} u_{t-2} - \theta^{*3} u_{t-3} + \dots) + \\ &\quad + \delta(u_{t-1} - \theta^* u_{t-2} + \theta^{*2} u_{t-3} - \dots) + \\ &\quad + (\nu_t - \theta^* \nu_{t-1} + \theta^{*2} \nu_{t-2} - \dots). \end{aligned} \quad (3.6.16)$$

Ряд ε_t , визначений в (3.6.16), ґрунтується на попередніх значеннях u і ν , і може здаватися процесом з багатою автокореляційною структурою. Насправді, цей ряд є процесом білого шуму. Щоб довести це, побудуємо автокореляційну генеруючу функцію ряду y_t . Використавши (3.6.6), отримаємо, що автоковаріаційну генеруючу функцію можна записати так:

$$g_y(z) = (1 + \delta z)\sigma_u^2(1 + \delta z^{-1}) + \sigma_\nu^2. \quad (3.6.17)$$

Автоковаріаційна генеруюча функція ряду $\varepsilon_t = (1 + \theta^*L)y_t$ матиме вигляд

$$g_\varepsilon(z) = \frac{(1 + \delta z)\sigma_u^2(1 + \delta z^{-1}) + \sigma_v^2}{(1 + \theta^*z)(1 + \theta^*z^{-1})}. \quad (3.6.18)$$

Але θ^* і σ^{*2} вибрані так, що автоковаріаційна генеруюча функція $(1 + \theta^*L)\varepsilon_t$, а саме

$$(1 + \theta^*z)\sigma^{*2}(1 + \theta^*z^{-1}),$$

збігається з правою частиною рівності (3.6.17). Отож, (3.6.18) дорівнює

$$g_\varepsilon(z) = \sigma^{*2},$$

що і доводить, що ε_t є білим шумом.

Отже, додавання $MA(1)$ процесу до білого шуму, з яким він не корелює, породжує новий $MA(1)$ процес, заданий (3.6.7).

Процеси, що додавалися, обидва мали нульові математичне сподівання. Додавання сталих членів до процесів не змінить результатів у будь-якому напрямі – якщо x_t – $MA(1)$ процес з математичним сподіванням μ_x , а ν_t – білий шум плюс стала μ_ν , тоді $x_t + \nu_t$ буде $MA(1)$ процесом з математичним сподіванням $\mu_x + \mu_\nu$. Таким чином, без втрати загальності можна розглядати сумування процесів з нульовими математичними сподіваннями.

3.6.2 Сумування двох процесів рухомих середніх

Нехай x_t – $MA(q_1)$ процес з нульовим математичним сподіванням

$$x_t = (1 + \delta_1L + \delta_2L^2 + \dots + \delta_{q_1}L^{q_1})u_t \equiv \delta(L)u_t,$$

де

$$E[u_t u_{t-j}] = \begin{cases} \sigma_u^2 & \text{для } j = 0, \\ 0 & \text{для } j = 1, 2, \dots \end{cases}$$

і нехай w_t – $MA(q_2)$ процес з нульовим математичним сподіванням

$$w_t = (1 + k_1 L + k_2 L^2 + \dots + k_{q_2} L^{q_2}) v_t \equiv k(L) v_t.$$

Тут

$$E[v_t v_{t-j}] = \begin{cases} \sigma_v^2 & \text{для } j = 0, \\ 0 & \text{для } j = 1, 2, \dots \end{cases}$$

Тоді автоковаріації $\gamma_0^x, \gamma_1^x, \dots, \gamma_{q_1}^x$ процесу x_t і автоковаріації $\gamma_0^w, \gamma_1^w, \dots, \gamma_{q_2}^w$ процесу w_t мають вигляд, описаний формулами (2.3.4). Припустимо, що x_t і w_t не корелюють за всіма компонентами

$$E[x_t w_{t-j}] = 0 \quad \text{для всіх } j,$$

і припустимо, що ми спостерігаємо

$$y_t = x_t + w_t.$$

Означимо q як максимум q_1 і q_2

$$q \equiv \max\{q_1, q_2\}.$$

Тоді j -та автоковаріація y_t визначається так:

$$\begin{aligned} E[y_t y_{t-j}] &= E[(x_t + w_t)(x_{t-j} + w_{t-j})] = \\ &= E[x_t x_{t-j}] + E[w_t w_{t-j}] = \\ &= \begin{cases} \gamma_j^x + \gamma_j^w & \text{для } j = 0, 1, 2, \dots, q, \\ 0 & \text{для } j = q + 1, q + 2, \dots \end{cases} \end{aligned}$$

Отже, автоковаріації процесу y_t порядку вище за q дорівнюють нулю. Це означає, що y_t можна описати $MA(q)$ процесом.

Переконаємось, що y_t справді описується $MA(q)$ процесом за допомогою автоковаріаційної генеруючої функції. Оскільки

$$\gamma_j^y = \gamma_j^x + \gamma_j^w,$$

то

$$\sum_{j=-\infty}^{+\infty} \gamma_j^y z^j = \sum_{j=-\infty}^{+\infty} \gamma_j^x z^j + \sum_{j=-\infty}^{+\infty} \gamma_j^w z^j.$$

Використовуючи означення відповідних автоковаріаційних генеруючих функцій, матимемо

$$g_y(z) = g_x(z) + g_w(z). \quad (3.6.19)$$

З рівняння (3.6.19) випливає загальний результат – якщо додати два коваріаційно - стаціонарних процеси, які не корелюють між собою за всіма компонентами, то автоковаріаційна генеруюча функція суми є сумою автоковаріаційних генеруючих функцій окремих процесів.

Нехай y_t описується $MA(q)$ процесом

$$y_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t \equiv \theta(L) \varepsilon_t,$$

де

$$E[\varepsilon_t \varepsilon_{t-j}] = \begin{cases} \sigma^2 & \text{для } j = 0, \\ 0 & \text{для } j = 1, 2, \dots \end{cases}$$

Тоді автоковаріаційна генеруюча функція матиме вигляд

$$g_y(z) = \theta(z)\theta(z^{-1})\sigma^2.$$

Завжди існують значення $(\theta_1, \theta_2, \dots, \theta_q, \sigma^2)$ такі, що задовольняють рівняння

$$\theta(z)\theta(z^{-1})\sigma^2 = \delta(z)\delta(z^{-1})\sigma_u^2 + k(z)k(z^{-1})\sigma_v^2. \quad (3.6.20)$$

Отже, якщо додати два процеси рухомих середніх, що не корелюють між собою за всіма компонентами, то результуючим процесом є новий процес рухомого середнього, порядок якого є більшим серед порядків двох рядів

$$MA(q_1) + MA(q_2) = MA(\max\{q_1, q_2\}). \quad (3.6.21)$$

3.6.3 Сумування двох авторегресійних процесів

Нехай тепер x_t і w_t – два $AR(1)$ процеси

$$(1 - \pi L)x_t = u_t, \quad (3.6.22)$$

$$(1 - \rho L)w_t = \nu_t, \quad (3.6.23)$$

де u_t і ν_t – процеси білого шуму, і u_t не корелює з ν_τ для всіх t і τ . Знову припустимо, що ми досліджуємо

$$y_t = x_t + w_t.$$

Якщо x_t і w_t мають однакові авторегресійні параметри, тобто

$$\pi = \rho,$$

то (3.6.22) просто додається до (3.6.23)

$$(1 - \pi L)x_t + (1 - \pi L)w_t = u_t + \nu_t$$

або

$$(1 - \pi L)(x_t + w_t) = u_t + \nu_t.$$

Сума $u_t + \nu_t$ є білим шумом (частковий випадок результату (3.6.21)), тобто y_t має $AR(1)$ зображення

$$(1 - \pi L)y_t = \varepsilon_t.$$

Загалом, коли $\pi \neq \rho$, то помножимо (3.6.22) на $(1 - \rho L)$

$$(1 - \rho L)(1 - \pi L)x_t = (1 - \rho L)u_t, \quad (3.6.24)$$

а (3.6.23) на $(1 - \pi L)$

$$(1 - \pi L)(1 - \rho L)w_t = (1 - \pi L)\nu_t. \quad (3.6.25)$$

Додавши (3.6.24) і (3.6.25), отримаємо

$$(1 - \rho L)(1 - \pi L)(x_t + w_t) = (1 - \rho L)u_t + (1 - \pi L)\nu_t. \quad (3.6.26)$$

Згідно з (3.6.21) права частина (3.6.26) має $MA(1)$ зображення. Тобто (3.6.26) можна записати у вигляді

$$(1 - \phi_1 L - \phi_2 L^2)y_t = (1 + \theta L)\varepsilon_t,$$

де

$$(1 - \phi_1 L - \phi_2 L^2) = (1 - \rho L)(1 - \pi L)$$

і

$$(1 + \theta L)\varepsilon_t = (1 - \rho L)u_t + (1 - \pi L)\nu_t.$$

Іншими словами,

$$AR(1) + AR(1) = ARMA(2, 1). \quad (3.6.27)$$

Цей результат можна узагальнити на випадок додавання $AR(p_1)$ процесу

$$\pi(L)x_t = u_t,$$

і $AR(p_2)$ процесу

$$\rho(L)w_t = \nu_t,$$

які не корелюють між собою. У цьому випадку сумою буде $ARMA(p_1 + p_2, \max\{p_1, p_2\})$ процес

$$\phi(L)y_t = \theta(L)\varepsilon_t,$$

де

$$\phi(L) = \pi(L)\rho(L)$$

і

$$\theta(L)\varepsilon_t = \rho(L)u_t + \pi(L)\nu_t.$$

3.7 Трикутна факторизація матриць

3.7.1 Трикутна факторизація додатно визначених симетричних матриць

Теорема 3.7.1. *Будь-яка додатно визначена симетрична матриця Ω розміру $n \times n$ має єдине зображення вигляду*

$$\Omega = ADA', \quad (3.7.1)$$

де A – нижня трикутна матриця з одиницями на головній діагоналі

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ a_{31} & a_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & 1 \end{pmatrix},$$

а D – діагональна матриця

$$D = \begin{pmatrix} d_{11} & 0 & 0 & \dots & 0 \\ 0 & d_{22} & 0 & \dots & 0 \\ 0 & 0 & d_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & d_{nn} \end{pmatrix},$$

для якої $d_{ii} > 0$ для всіх i .

Зображення (3.7.1) відоме як трикутна факторизація матриці Ω .

Д о в е д е н н я. Розглянемо матрицю

$$\Omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} & \dots & \omega_{1n} \\ \omega_{21} & \omega_{22} & \omega_{23} & \dots & \omega_{2n} \\ \omega_{31} & \omega_{32} & \omega_{33} & \dots & \omega_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \omega_{n1} & \omega_{n2} & \omega_{n3} & \dots & \omega_{nn} \end{pmatrix} \quad (3.7.2)$$

і обчислимо її трикутну факторизацію.

Припускаємо, що Ω додатно визначена матриця, тобто для довільного ненульового вектора x розмірності n виконується нерівність $x'\Omega x > 0$. Також припускаємо, що Ω є симетричною, тобто $\omega_{ij} = \omega_{ji}$.

Матриця Ω може бути перетворена на матрицю з нульовим елементом (2,1) за допомогою множення першого рядка Ω на ω_{21}/ω_{11} і віднімання рядка, що утворився від другого рядка матриці Ω . Елемент (3,1) стане нульовим, якщо від третього рядка відняти перший, помножений на ω_{31}/ω_{11} . Продовжуємо цей процес, послідовно множачи перший рядок матриці Ω на ω_{j1}/ω_{11} і віднімаючи отриманий рядок від j -го рядка. Така сукупність операцій може бути підсумована як множення зліва матриці Ω на матрицю

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\frac{\omega_{21}}{\omega_{11}} & 1 & 0 & \dots & 0 \\ -\frac{\omega_{31}}{\omega_{11}} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -\frac{\omega_{n1}}{\omega_{11}} & 0 & 0 & \dots & 1 \end{pmatrix}. \quad (3.7.3)$$

Матриця (3.7.3) завжди існує, оскільки $\omega_{11} \neq 0$. Це випливає з того, що ω_{11} дорівнює $e_1'\Omega e_1$, де $e_1 = (1 \ 0 \ 0 \ \dots \ 0)$. Оскільки Ω додатно визначена, то $e_1'\Omega e_1$ повинно бути більшим за нуль.

Помноживши Ω зліва на E_1 і справа на E'_1 , отримаємо

$$E_1 \Omega E'_1 = H, \quad (3.7.4)$$

де

$$H = \begin{pmatrix} h_{11} & 0 & 0 & \dots & 0 \\ 0 & h_{22} & h_{23} & \dots & h_{2n} \\ 0 & h_{32} & h_{33} & \dots & h_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & h_{n2} & h_{n3} & \dots & h_{nn} \end{pmatrix} = \quad (3.7.5)$$

$$= \begin{pmatrix} \omega_{11} & 0 & 0 & \dots & 0 \\ 0 & \omega_{22} - \frac{\omega_{21}\omega_{12}}{\omega_{11}} & \omega_{23} - \frac{\omega_{21}\omega_{13}}{\omega_{11}} & \dots & \omega_{2n} - \frac{\omega_{21}\omega_{1n}}{\omega_{11}} \\ 0 & \omega_{32} - \frac{\omega_{31}\omega_{12}}{\omega_{11}} & \omega_{33} - \frac{\omega_{31}\omega_{13}}{\omega_{11}} & \dots & \omega_{3n} - \frac{\omega_{31}\omega_{1n}}{\omega_{11}} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \omega_{n2} - \frac{\omega_{n1}\omega_{12}}{\omega_{11}} & \omega_{n3} - \frac{\omega_{n1}\omega_{13}}{\omega_{11}} & \dots & \omega_{nn} - \frac{\omega_{n1}\omega_{1n}}{\omega_{11}} \end{pmatrix}.$$

Далі продовжуємо так само, використовуючи другий рядок матриці H . Помножимо тепер другий рядок H на h_{32}/h_{22} і віднімемо результуючий рядок від третього. Так само помножимо другий рядок H на h_{42}/h_{22} і віднімемо результат від четвертого рядка, і так далі. Ці операції можна узагальнити за допомогою матриці

$$E_2 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & -\frac{h_{32}}{h_{22}} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & -\frac{h_{n2}}{h_{22}} & 0 & \dots & 1 \end{pmatrix}. \quad (3.7.6)$$

Існування цієї матриці забезпечується тим, що $h_{22} \neq 0$. Доведемо це, враховуючи, що h_{22} може бути обчислене як $h_{22} = e_2' H e_2$, де $e_2' = (0 \ 1 \ 0 \ \dots \ 0)$. Використаємо зображення $H = E_1 \Omega E_1'$, де Ω – додатно визначена і E_1 матриця, задана формулою (3.7.3). Оскільки E_1 нижня трикутна матриця, її визначник дорівнює добутку елементів головної діагоналі, які всі дорівнюють 1. Отже, з того, що E_1 є невиродженою випливає, що $H = E_1 \Omega E_1'$ додатно визначена і, отже, елемент $h_{22} = e_2' H e_2$ повинен бути строго додатним, тобто матриця (3.7.6) може бути обчислена.

Якщо H зліва помножити на E_2 і справа на E_2' , то отримаємо

$$E_2 H E_2' = K,$$

де

$$K = \begin{pmatrix} h_{11} & 0 & 0 & \dots & 0 \\ 0 & h_{22} & 0 & \dots & 0 \\ 0 & 0 & h_{33} - \frac{h_{32}h_{23}}{h_{22}} & \dots & h_{3n} - \frac{h_{32}h_{2n}}{h_{22}} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & h_{n3} - \frac{h_{n2}h_{23}}{h_{22}} & \dots & h_{nn} - \frac{h_{n2}h_{2n}}{h_{22}} \end{pmatrix}.$$

Оскільки H додатно визначена і E_2 невироджена, то K є додатно визначена і k_{33} – додатне. Продовжуючи цю процедуру, послідовно використовуючи третій, четвертий і всі наступні рядки, отримаємо, що для кожної додатно визначеної симетричної матриці Ω існують матриці E_1, E_2, \dots, E_{n-1} такі, що

$$E_{n-1} \cdots E_2 E_1 \Omega E_1' E_2' \cdots E_{n-1}' = D, \quad (3.7.7)$$

де

$$D = \begin{pmatrix} \omega_{11} & 0 & 0 & \dots & 0 \\ 0 & \omega_{22} - \frac{\omega_{21}\omega_{12}}{\omega_{11}} & 0 & \dots & 0 \\ 0 & 0 & h_{33} - \frac{h_{32}h_{23}}{h_{22}} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & c_{nn} - \frac{c_{n,n-1}c_{n-1,n}}{c_{n-1,n-1}} \end{pmatrix},$$

для якої всі діагональні елементи D строго додатні. У формулі (3.7.7) матриці E_1 і E_2 мають вигляд (3.7.3) і (3.7.6) відповідно. У загальному випадку, E_j є матрицею з ненульовими значеннями в j -му стовпці під головною діагоналлю, одиничками на головній діагоналі і решта нулями.

Отож, кожна E_j є нижньою трикутною матрицею з одиничним визначником. Оскільки для кожного $j = 1, \dots, n-1$ існує E_j^{-1} , то існує матриця

$$A = (E_{n-1} \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_{n-1}^{-1}. \quad (3.7.8)$$

Якщо (3.7.7) помножити зліва на A і справа на A' , то отримаємо

$$\Omega = A D A'. \quad (3.7.9)$$

Згадаємо, що матриця E_1 бере участь у перетворенні матриці Ω , при якому використовується множення першого рядка на відомі числа і віднімання результату від кожного з наступних рядків. Якщо використати її обернену матрицю E_1^{-1} , то це призведе до множення першого рядка на ті самі числа і

додавання результатів до наступних рядків. Отже,

$$E_1^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{\omega_{21}}{\omega_{11}} & 1 & 0 & \dots & 0 \\ \frac{\omega_{31}}{\omega_{11}} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{\omega_{n1}}{\omega_{11}} & 0 & 0 & \dots & 1 \end{pmatrix}, \quad (3.7.10)$$

що може бути перевірено множенням (3.7.3) і (3.7.10), яке дасть одиничну матрицю. Подібно

$$E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & \frac{h_{32}}{h_{22}} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & \frac{h_{n2}}{h_{22}} & 0 & \dots & 1 \end{pmatrix}$$

і так далі. Підставивши в (3.7.8), матимемо

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{\omega_{21}}{\omega_{11}} & 1 & 0 & \dots & 0 \\ \frac{\omega_{31}}{\omega_{11}} & \frac{h_{32}}{h_{22}} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{\omega_{n1}}{\omega_{11}} & \frac{h_{n2}}{h_{22}} & \frac{k_{n3}}{k_{33}} & \dots & 1 \end{pmatrix}. \quad (3.7.11)$$

Тобто, j -ий стовпець матриці A є j -им стовпцем E_j^{-1} .

Зазначимо, що простота множення матриць $E_1^{-1}, E_2^{-1}, \dots, E_{n-1}^{-1}$ залежить не тільки від структури матриць E_j^{-1} , а також

від порядку, в якому вони множаться. Зокрема, $A^{-1} = E_{n-1} \cdots E_2 E_1$ не може бути обчислена просто за допомогою використання j -го стовця E_j за j -ий стовець A^{-1} .

Оскільки матриця A в (3.7.11) є нижньою трикутною матрицею з одиницями на головній діагоналі, то зображення (3.7.9) є трикутною факторизацією матриці Ω . \square

Приклад 3.7.1 Розглянемо трикутну факторизацію для матриці Ω розміру 2×2 .

$$\begin{aligned} & \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} = & (3.7.12) \\ & = \begin{pmatrix} 1 & 0 \\ \frac{\omega_{21}}{\omega_{11}} & 1 \end{pmatrix} \begin{pmatrix} \omega_{11} & 0 \\ 0 & \omega_{22} - \frac{\omega_{21}\omega_{12}}{\omega_{11}} \end{pmatrix} \begin{pmatrix} 1 & \frac{\omega_{12}}{\omega_{11}} \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Факторизація матриці розміром 3×3 має вигляд

$$\begin{aligned} & \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{\omega_{21}}{\omega_{11}} & 1 & 0 \\ \frac{\omega_{31}}{\omega_{11}} & \frac{h_{32}}{h_{22}} & 1 \end{pmatrix} \times & (3.7.13) \\ & \times \begin{pmatrix} \omega_{11} & 0 & 0 \\ 0 & h_{22} & 0 \\ 0 & 0 & h_{33} - \frac{h_{32}h_{23}}{h_{22}} \end{pmatrix} \begin{pmatrix} 1 & \frac{\omega_{12}}{\omega_{11}} & \frac{\omega_{13}}{\omega_{11}} \\ 0 & 1 & \frac{h_{23}}{h_{22}} \\ 0 & 0 & 1 \end{pmatrix}, \end{aligned}$$

де

$$\begin{aligned} h_{22} &= \omega_{22} - \frac{\omega_{21}\omega_{12}}{\omega_{11}}, & h_{33} &= \omega_{33} - \frac{\omega_{31}\omega_{13}}{\omega_{11}}, \\ h_{23} &= h_{32} = \omega_{23} - \frac{\omega_{21}\omega_{13}}{\omega_{11}}. \end{aligned}$$

3.7.2 Єдиність трикутної факторизації

Покажемо, що трикутна факторизація єдина. Припустимо, що

$$\Omega = A_1 D_1 A_1' = A_2 D_2 A_2', \quad (3.7.14)$$

де A_1 і A_2 – нижні трикутні матриці з одиницями на головній діагоналі, а D_1 і D_2 – діагональні матриці з додатними діагональними елементами. Тоді до всіх описаних матриць існують обернені. Помноживши (3.7.14) на $D_1^{-1} A_1^{-1}$ зліва і справа на $(A_2')^{-1}$, отримаємо

$$A_1' (A_2')^{-1} = D_1^{-1} A_1^{-1} A_2 D_2. \quad (3.7.15)$$

Оскільки A_2' верхня трикутна матриця з одиницями на головній діагоналі, то і $(A_2')^{-1}$ є верхньою трикутною матрицею з одиницями на головній діагоналі. Позаяк A_1' має таку саму форму, то ліва частина (3.7.15) є верхньою трикутною матрицею з одиницями на головній діагоналі. З подібних міркувань, права частина (3.7.15) є нижньою трикутною матрицею. Єдиний випадок, коли нижня трикутна матриця дорівнює верхній трикутній, виникає тоді, коли всі недіагональні елементи дорівнюють нулю. Оскільки всі діагональні елементи в лівій частині (3.7.15) дорівнюють 1, то ця матриця повинна бути одиничною

$$A_1' (A_2')^{-1} = I_n.$$

Домноживши справа на A_2' , матимемо $A_1' = A_2'$. Подальше домноження (3.7.14) зліва на A_1^{-1} і справа на $(A_1')^{-1}$ доводить, що $D_1 = D_2$.

3.7.3 Факторизація Колеського

Позначимо $D^{1/2}$ діагональну матрицю розміром $n \times n$, елементами якої є квадратні корені елементів матриці D з трикутної

факторизації

$$D^{1/2} = \begin{pmatrix} \sqrt{d_{11}} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{d_{22}} & 0 & \dots & 0 \\ 0 & 0 & \sqrt{d_{33}} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{d_{nn}} \end{pmatrix}.$$

Оскільки матриця D єдина і має строго додатні елементи, то матриця $D^{1/2}$ існує і єдина. Тоді трикутну факторизацію можна записати у вигляді

$$\Omega = AD^{1/2}D^{1/2}A' = (AD^{1/2})(AD^{1/2})',$$

або

$$\Omega = PP', \quad (3.7.16)$$

де

$$\begin{aligned} P &\equiv AD^{1/2} = \\ &= \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{21} & 1 & 0 & \dots & 0 \\ a_{31} & a_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & 1 \end{pmatrix} \begin{pmatrix} \sqrt{d_{11}} & 0 & \dots & 0 \\ 0 & \sqrt{d_{22}} & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sqrt{d_{nn}} \end{pmatrix} = \\ &= \begin{pmatrix} \sqrt{d_{11}} & 0 & 0 & \dots & 0 \\ a_{21}\sqrt{d_{11}} & \sqrt{d_{22}} & 0 & \dots & 0 \\ a_{31}\sqrt{d_{11}} & a_{32}\sqrt{d_{22}} & \sqrt{d_{33}} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ a_{n1}\sqrt{d_{11}} & a_{n2}\sqrt{d_{22}} & a_{n3}\sqrt{d_{33}} & \dots & \sqrt{d_{nn}} \end{pmatrix}. \end{aligned}$$

Зображення (3.7.16) відоме як факторизація Колеського матриці Ω . Зазначимо, що P , як і A , є нижньою трикутною матрицею. Матриця A містить одиниці на головній діагоналі, а множник Колеського P містить квадратні корені елементів матриці D на головній діагоналі.

3.7.4 Трикутна факторизація матриці автокореляцій для $MA(1)$ процесу

Розглянемо $MA(1)$ процес

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}.$$

Позначимо

$$y = (y_1 - \mu \quad \dots \quad y_T - \mu).$$

Матриця автоковаріацій вектора y має вигляд

$$\Omega = E[yy'] = \sigma^2 \begin{pmatrix} 1 + \theta^2 & \theta & 0 & \dots & 0 \\ \theta & 1 + \theta^2 & \theta & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 + \theta^2 \end{pmatrix} = \sigma^2 V.$$

Знайдемо трикутну факторизацію матриці V

$$V = ADA'.$$

Перший діагональний елемент матриці D дорівнює елементу $(1,1)$ матриці V , тобто

$$d_{11} = 1 + \theta^2.$$

Щоб отримати нульовий $(2,1)$ елемент, треба помножити перший рядок матриці V на $\theta/(1 + \theta^2)$ і відняти результат від другого рядка. Одержимо

$$a_{21} = \frac{\theta}{1 + \theta^2}.$$

Тоді $(2,2)$ елемент матриці D матиме вигляд

$$\begin{aligned} d_{22} &= (1 + \theta^2) - \theta \frac{\theta}{1 + \theta^2} = \frac{(1 + \theta^2)^2 - \theta^2}{1 + \theta^2} = \\ &= \frac{1 + \theta^2 + \theta^4}{1 + \theta^2}. \end{aligned}$$

Щоб отримати нуль на місці (3,2), другий рядок матриці треба помножити на θ/d_{22} і відняти результат від третього рядка, тобто

$$a_{32} = \frac{\theta}{d_{22}} = \frac{\theta(1 + \theta^2)}{1 + \theta^2 + \theta^4}.$$

Тоді

$$\begin{aligned} d_{33} &= (1 + \theta^2) - \frac{\theta^2(1 + \theta^2)}{1 + \theta^2 + \theta^4} = \\ &= \frac{(1 + \theta^2)(1 + \theta^2 + \theta^4) - \theta^2(1 + \theta^2)}{1 + \theta^2 + \theta^4} = \\ &= \frac{(1 + \theta^2 + \theta^4) + \theta^2(1 + \theta^2 + \theta^4) - \theta^2(1 + \theta^2)}{1 + \theta^2 + \theta^4} = \\ &= \frac{1 + \theta^2 + \theta^4 + \theta^6}{1 + \theta^2 + \theta^4}. \end{aligned}$$

У загальному випадку

$$d_{ii} = \frac{1 + \theta^2 + \theta^4 + \dots + \theta^{2i}}{1 + \theta^2 + \theta^4 + \dots + \theta^{2(i-1)}}. \quad (3.7.17)$$

Щоб отримати нуль на $(i + 1, i)$ -му місці, множимо i -ий рядок на

$$a_{i+1,i} = \frac{\theta}{d_{ii}} = \frac{\theta(1 + \theta^2 + \theta^4 + \dots + \theta^{2(i-1)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2i}} \quad (3.7.18)$$

і віднімаємо від $(i + 1)$ -го рядка. Отримаємо

$$\begin{aligned} d_{i+1,i+1} &= (1 + \theta^2) - \frac{\theta^2(1 + \theta^2 + \theta^4 + \dots + \theta^{2(i-1)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2i}} = \\ &= \frac{(1 + \theta^2 + \theta^4 + \dots + \theta^{2i}) + \theta^2(1 + \theta^2 + \theta^4 + \dots + \theta^{2i})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2i}} - \\ &\quad - \frac{\theta^2(1 + \theta^2 + \theta^4 + \dots + \theta^{2(i-1)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2i}} = \\ &= \frac{1 + \theta^2 + \theta^4 + \dots + \theta^{2i} + \theta^{2(i+1)}}{1 + \theta^2 + \theta^4 + \dots + \theta^{2i}}. \end{aligned} \quad (3.7.19)$$

Отже, трикутна факторизація матриці Ω має вигляд

$$\Omega = A D A', \quad (3.7.20)$$

де

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \theta & 1 & \dots & 0 & 0 \\ \frac{\theta}{1+\theta^2} & \frac{\theta(1+\theta^2)}{1+\theta^2+\theta^4} & \dots & 0 & 0 \\ 0 & \frac{\theta(1+\theta^2)}{1+\theta^2+\theta^4} & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & \frac{1+\theta^2+\theta^4+\dots+\theta^{2(T-2)}}{1+\theta^2+\theta^4+\dots+\theta^{2(T-1)}} & 1 \end{pmatrix},$$

$$D = \sigma^2 \begin{pmatrix} 1+\theta^2 & 0 & \dots & 0 \\ 0 & \frac{1+\theta^2+\theta^4}{1+\theta^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1+\theta^2+\theta^4+\dots+\theta^{2T}}{1+\theta^2+\theta^4+\dots+\theta^{2(T-1)}} \end{pmatrix}.$$

Розділ 4

Ідентифікація моделі

Процес вибору моделі, яка найкраще відповідає реальному процесу, що розглядається, називається ідентифікацією моделі. Ідентифікація стаціонарного часового ряду полягає у визначенні конкретного вигляду моделі з класу $ARMA(p, q)$ моделей.

Загальна ідея ідентифікації моделі $ARMA(p, q)$ полягає в тому, що властивості реального процесу і властивості побудованої моделі дещо повинні бути близькими одна до одної. Крім того, вибирають модель, яка характеризується найменшою кількістю параметрів (мінімальним порядком) порівняно з іншими моделями, без втрат в точності опису вихідного процесу. Всі ці питання розглянуто в цьому розділі.

Оскільки ми розглядаємо лише стаціонарні процеси, то треба перевірити на стаціонарність дані вихідного ряду. Такі тести подано в пункті 4.1. Важливими при визначенні процесу, що породжує дані, є властивості автокореляційної і часткової автокореляційної функцій. Детальніше властивості цих функцій досліджено у пунктах 4.2, 4.3. У пунктах 4.4, 4.5. описано загальну процедуру моделювання Бокса - Дженкінса.

4.1 Тестування стаціонарності

Лінійні моделі часових рядів здебільшого застосовують до стаціонарних процесів. При цьому зазвичай мають на увазі коваріаційно - стаціонарні процеси, тобто процеси, які характеризуються сталими значеннями всіх моментів порядку два і нижче на всіх часових проміжках, які входять в інтервал $t \in [1, T]$. Тобто, для довільних двох інтервалів часу (T_1, T_2) і (T_3, T_4) для коваріаційно - стаціонарного процесу повинні виконуватись умови, які характеризують рівність на цих інтервалах математичних сподівань, дисперсій і коефіцієнтів кореляцій цього процесу. На практиці це означає, що для відповідних оцінок цих показників повинні виконуватись співвідношення

$$\begin{aligned}\bar{y}_1 &= \sum_{t=T_1+1}^{T_2} \frac{y_t}{T_2 - T_1} = \sum_{t=T_3+1}^{T_4} \frac{y_t}{T_4 - T_3} = \bar{y}_2, \\ \text{Var}_1[y] &= \sum_{t=T_1+1}^{T_2} \frac{(y_t - \bar{y})^2}{T_2 - T_1} = \sum_{t=T_3+1}^{T_4} \frac{(y_t - \bar{y})^2}{T_4 - T_3} = \text{Var}_2[y], \\ r_k^1 &= \sum_{t=T_1+1}^{T_2-k} \frac{(y_t - \bar{y})(y_{t+k} - \bar{y})}{(T_2 - T_1 - k) \text{Var}[y]} = \\ &= \sum_{t=T_3+1}^{T_4-k} \frac{(y_t - \bar{y})(y_{t+k} - \bar{y})}{(T_4 - T_3 - k) \text{Var}[y]} = r_k^2, \quad k = 1, 2, \dots,\end{aligned}$$

де \bar{y}_1 і \bar{y}_2 – оцінки математичних сподівань; $\text{Var}_1[y]$ і $\text{Var}_2[y]$ – оцінки дисперсій; r_k^1 і r_k^2 – оцінки коефіцієнтів автокореляцій k -того порядку процесу y_t на першому і другому інтервалі, відповідно; \bar{y} – середнє значення процесу (оцінка математичного сподівання) на інтервалі $[1, T]$; $\text{Var}[y]$ – оцінка дисперсії процесу на інтервалі $[1, T]$.

Зауважимо, що на практиці рівності розглядають у статистичному сенсі. Для цього зазвичай використовуються відпо-

відні тести. В деяких випадках, перевіряючи одну умову виникає необхідність застосувати декілька тестів, якщо за результатами одного тесту не можна зробити висновок про істинність чи хибність висунутої гіпотези.

Всю множину таких тестів поділяють на три групи: непараметричні, напівпараметричні та параметричні. Розглянемо детальніше деякі тести на стаціонарність, які найчастіше використовують на практиці.

4.1.1 Параметричні тести

З означення коваріаційно - стаціонарного процесу випливає, що тестування стаціонарності еквівалентне тестуванню гіпотез про сталість на інтервалі $t \in [1, T]$, що розглядається, математичного сподівання, коефіцієнтів автокореляції і дисперсії. Будемо використовувати критерії Стьюдента і Фішера, які застосовують за припущення про нормальність розподілу як значень часового ряду y_t , так і його вибіркових параметрів, що справджується для багатьох реальних процесів.

Тестування математичного сподівання

Інтервал часу $[1, T]$ і відповідно часовий ряд y_t ($t = 1, 2, \dots, T$) розбиваємо на дві частини, які не обов'язково містять однакову кількість значень. Нехай перша частина містить T_1 спостережень y_t ($t = 1, 2, \dots, T_1$), а друга T_2 спостережень y_t ($t = T - T_2 + 1, \dots, T$).

Для кожної частини визначаємо оцінки \bar{y}_1 і s_1^2 , \bar{y}_2 і s_2^2 - вибіркових середніх і вибіркових дисперсій змінної y_t відповідно. Далі обчислюємо значення критерію Стьюдента за формулою

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{s_1^2}{T_1} + \frac{s_2^2}{T_2}}}, \quad (4.1.1)$$

якщо припускається, що значення дисперсій на цих інтервалах

не рівні між собою, тобто $\sigma_1^2 \neq \sigma_2^2$, і за формулою

$$\tau = \frac{|\bar{y}_1 - \bar{y}_2|}{s} \sqrt{\frac{T_1 T_2}{T_1 + T_2}}, \quad (4.1.2)$$

якщо $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Величина τ має розподіл

$$\tau \sim t [T_1 + T_2 - 2].$$

Якщо виконується нерівність

$$\tau < t_{\text{кр}}, \quad (4.1.3)$$

де $t_{\text{кр}}$ – критичне значення розподілу Стюдента для вибраного рівня значущості α і ступеня вільності $T_1 + T_2 - 2$, то гіпотезу про сталість математичного сподівання процесу y_t приймаємо. Ймовірність того, що ми при цьому допустили помилку, дорівнює α . Якщо нерівність (4.1.3) не виконується, то гіпотезу відхиляємо, а отже, робимо висновок, що часовий ряд не є стаціонарним.

Для більшої достовірності висновку про сталість математичного сподівання часового ряду y_t , $t = 1, 2, \dots, T$, якщо кількість спостережень достатньо велика, інтервал спостереження може бути розділений на декілька частин. У цьому випадку перевіряється гіпотеза про рівність оцінок середніх значень ряду, які обчислені для цих частин. Для тестування гіпотези використовують критерій Фішера. Значення критерію обчислюємо за формулою

$$F = \frac{\frac{1}{n-1} \sum_{j=1}^n T_j (\bar{y}_j - \bar{y})^2}{\bar{s}^2(n)},$$

де n – кількість частин, на які розбито інтервал $[1, T]$; T_j – кількість спостережень y_t у j -ій частині, $j = 1, 2, \dots, n$; \bar{y}

– середнє значення часового ряду в цілому; $\bar{s}^2(n)$ – середня дисперсія, яка обчислюється за формулою

$$\bar{s}^2(n) = \frac{1}{T-n} \sum_{j=1}^n (T_j - 1) \bar{s}_j^2;$$

а \bar{s}_j^2 – дисперсія, обчислена для j -ої частини вибірки. Статистика F має розподіл Фішера

$$F \sim F[n-1, T_1 + T_2 + \dots + T_n - n]$$

і порівнюється з критичним значенням розподілу Фішера з відповідними ступенями вільності і заданого рівня значущості.

Тестування дисперсії

Перевірка гіпотези про сталість дисперсії часового ряду y_t , $t = 1, 2, \dots, T$ у випадку розбиття часового інтервалу на дві частини відбувається на підставі критерію Фішера. Обов'язковою умовою при цьому також є нормальність розподілу y_t . Спостережуване значення критерію обчислюємо за формулою

$$F = \frac{s_1^2}{s_2^2},$$

де s_1^2 і s_2^2 – оцінки дисперсії першої і другої частини ряду відповідно з кількістю спостережень T_1 і T_2 . Якщо виконується нерівність

$$F \leq F_{\text{кр}} \left[\frac{\alpha}{2}, T_1 - 1, T_2 - 1 \right],$$

то гіпотеза про сталість дисперсії часового ряду приймається, тобто припущення про те, що $s_1^2 = s_2^2 = \sigma^2$ є правильним з ймовірністю $1 - \alpha$.

Для середніх ($40 \leq T \leq 100$) і великих ($T > 100$) обсягів часового ряду замість критерію Фішера рекомендують використовувати стандартний нормальний розподіл. У першому

випадку, тобто для середніх вибірок, використовується критерій

$$\Phi = \frac{\frac{1}{2} \ln \frac{s_1^2}{s_2^2} + \frac{1}{2} \left(\frac{1}{T_1 - 1} - \frac{1}{T_2 - 1} \right)}{\sqrt{\frac{1}{2} \left(\frac{1}{T_1 - 1} - \frac{1}{T_2 - 1} \right)}} \sim N[0, 1]. \quad (4.1.4)$$

У другому випадку (при великих вибірках) використовують критерій

$$\Phi = (s_1 - s_2) \sqrt{\frac{s_1^2}{2T_1} + \frac{s_2^2}{2T_2}}. \quad (4.1.5)$$

В обох випадках, якщо правильна нерівність $|\Phi| < 1,96$, то гіпотеза про сталість дисперсії приймається з коефіцієнтом надійності 0,95.

При розбитті часового ряду y_t , $t = 1, 2, \dots, T$ на декілька частин ($n > 2$) для перевірки гіпотези про сталість дисперсії можна використовувати критерій Кохрейна, який ґрунтується на розподілі Фішера. Цей критерій використовують за припущення, що обсяги частин є рівними між собою, тобто $T_1 = T_2 = \dots = T_n = N$. Значення критерію обчислюють за формулою

$$K = \frac{s_{\max}^2}{s_1^2 + \dots + s_n^2},$$

де $s_{\max} = \max_{1 \leq j \leq n} (s_j^2)$. Критичне значення критерію Кохрейна для рівня надійності γ визначається так:

$$K_{\text{кр}} = \frac{F_{\text{кр}}[N_1, N_2]}{(n-1) + F_{\text{кр}}[N_1, N_2]}.$$

Тут $F_{\text{кр}}[N_1, N_2]$ – критичне значення розподілу Фішера, яке вибрано для рівня значущості $\alpha = 1 - \frac{1-\gamma}{n}$ і ступенів вільності $N_1 = N - 1$ і $N_2 = (n-1)N_1$.

Якщо виконується нерівність $K < K_{кр}$, то гіпотеза про сталість дисперсії часового ряду приймається з рівнем довіри $\gamma = 1 - \alpha$.

Тестування коефіцієнтів автокореляції

Теоретично для перевірки гіпотези про сталість коефіцієнтів автокореляції (автоковаріації) можна використовувати ті самі критерії, що й для перевірки аналогічних гіпотез для математичних сподівань (автокореляції) і дисперсії (автоковаріації). До результатів такої перевірки треба ставитись з обережністю. Це зумовлено тим, що дисперсії вибіркового коефіцієнта автокореляції визначаються з достатньо великою похибкою, яка збільшується зі зростанням значень самого коефіцієнта автокореляції. Зростання похибки зумовлене несиметричністю закону розподілу вибіркового коефіцієнта автокореляції і його розбіжністю з нормальним розподілом. Крім того, між парами коефіцієнтів автокореляції існує достатньо сильний статистичний зв'язок. Наявність такого зв'язку може вносити суттєві зміщення в оцінки значень самих коефіцієнтів автокореляції і їхніх дисперсій.

Загалом величину дисперсії коефіцієнта автокореляції можна оцінити за формулою Бартлетта

$$\text{Var} [\rho_k] = \frac{1}{T} \sum_{i=-j}^{i=j} (\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k\rho_i\rho_{i-k} + 2\rho_i^2\rho_k^2),$$

де індекс j залежить від довжини ряду T і вибирається так, щоб величини в формулі Бартлетта були статистично визначені (передусім це стосується ρ_{j+k}).

Для реальних часових рядів автокореляційна функція часто має визначений вигляд. Зокрема, для MA процесів коефіцієнти автокореляції дорівнюють нулю після деякого лага, тобто $\rho_i = 0$, $i > q$, для $AR(1)$ процесу вони затухають степеневу $\rho_k = \rho^k$.

У першому випадку дисперсію можна наближено оцінити за формулою

$$\text{Var} [\rho_k] = \frac{1}{T} \left(1 + 2 \sum_{i=1}^q \rho_i^2 \right), \quad k > q,$$

де індекси k належать коефіцієнтам автокореляції, які наближаються до нуля після деякого лага q .

У другому випадку дисперсія першого коефіцієнта автокореляції можна наближено визначити за формулою

$$\text{Var} [\rho_1] = \frac{1}{T} (1 - \rho_1^2).$$

Зауважимо, що при невеликих значеннях коефіцієнта автокореляції його розподіл асимптотично нормальний.

На практиці рекомендують застосовувати спрощену формулу для оцінки дисперсії

$$\text{Var} [\rho_k] = \frac{1}{T},$$

яку використовують при тестуванні значущості коефіцієнта кореляції. Статистичним критерієм є статистика Стьюдента, яка визначається так:

$$\tau = \frac{|\rho_k|}{\sigma_k}.$$

4.1.2 Непараметричні тести

Тест Манна - Уїтні

Непараметричний критерій Манна - Уїтні (критерій u^*) використовують для тестування математичного сподівання. Він трохи слабший від критерію Стьюдента у випадку часових рядів з нормальним розподілом, але має переваги над

параметричними критеріями у випадку, якщо розподіл часового ряду не є нормальним.

Критерій u^* застосовують для перевірки ідентичності розподілів двох сукупностей (у нашому випадку часових послідовностей одного часового ряду y_t , які визначені на різних частинах часового інтервалу $t = 1, \dots, T$).

Нехай перша сукупність утворена з T_1 послідовних спостережень y_t , друга – з T_2 його спостережень, причому утворені послідовності не перетинаються.

Всі значення цих сукупностей об'єднуємо в один ряд, в якому вони розташовуються в порядку зростання з першого по $(T_1 + T_2)$ -ий, незалежно від того до якої послідовності вони належать. У цій єдиній послідовності елементи першої послідовності позначимо символом y^1 , другої – символом y^2 . В результаті отримуємо ряд, який складається з $T_1 + T_2$ елементів, в якому символи y^1 (T_1 елементів) і символи y^2 (T_2 елементів) перемішані між собою.

Якщо ряд стаціонарний, то послідовності y^1 і y^2 практично не відрізняються і їхні елементи перемішані між собою. Якщо ж ряд не є стаціонарним, то загальна послідовність буде розділена на однорідні масиви, які складаються здебільшого з елементів певної послідовності. Наприклад, елементи сукупностей будуть нагромаджуватися на різних кінцях загальної послідовності.

Тест Манна - Уїтні перевіряє гіпотезу про стаціонарність часового ряду на підставі обчислення статистики u^* , яка дорівнює кількості випадків, коли елементи з сукупності y^1 передують елементам сукупності y^2 . Тобто значення u^* дорівнює кількості елементів з y^1 , які передують найменшому за величиною елементу з y^2 , плюс кількість елементів з y^1 , які передують наступному за ним елементу з y^2 , враховуючи також раніше вже враховані елементи першої сукупності і так далі, доки не буде включено в суму кількість елементів з y^1 ,

які передують останньому елементу з y^2 .

На практиці значення u^* розраховують або через суму рангів елементів першої сукупності, або через суму рангів елементів другої сукупності відповідно за формулами

$$u^* = R_1 - \frac{T_1(T_1 + 1)}{2},$$

$$u^* = T_1 T_2 - \frac{T_2(T_2 + 1)}{2} - R_2,$$

де R_1 і R_2 – суми рангів елементів першої і другої сукупностей відповідно, які визначають у загальній послідовності.

Для середніх і великих вибірок ($T > 50; 100$) випадкова величина u^* нормально розподілена з математичним сподіванням

$$E[u^*] = \frac{T_1 T_2}{2}$$

і дисперсією

$$\text{Var}[u^*] = \frac{T_1 T_2 (T_1 + T_2 + 1)}{12}.$$

Тоді випадкова величина

$$z = \frac{u^* - \frac{T_1 T_2}{2} \pm \frac{1}{2}}{\sigma(u^*)}$$

має нормальний стандартний розподіл. У цій формулі поправка $1/2$ вводить для забезпечення неперервності величини z . Вона додається, якщо $z < 0$, і віднімається, якщо $z > 0$.

Гіпотеза про стаціонарність процесу може бути прийнята, якщо виконується нерівність

$$x_1 \leq z \leq x_2, \quad (4.1.6)$$

де x_1 і x_2 визначаються з рівності

$$P\{x_1 \leq z \leq x_2\} = p = \int_{x_1}^{x_2} \varphi(x) dx,$$

де

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Зокрема, якщо $p = 0,95$, то обчислене значення z має належати інтервалу

$$-1,96 \leq z \leq 1,96.$$

Тест Сіджела -Тьюкі

Замість параметричного критерію Фішера для перевірки гіпотези про сталість дисперсії часового ряду y_t на інтервалі $t = 1, 2, \dots, T$ можна використовувати непараметричний критерій Сіджела - Тьюкі, який також ґрунтується на зіставленні рангів елементів двох сукупностей.

Для застосування тесту центруємо вихідний ряд, тобто обчислюємо значення $y'_t = y_t - \bar{y}$, де \bar{y} – середнє значення ряду y_t . Далі інтервал $[1, T]$ розділяємо на дві частини (бажано рівні). Елементи першої частини центрованої сукупності позначаємо y^1 , а другої частини – y^2 .

Далі елементи двох центрованих сукупностей об'єднуємо в один ряд, присвоюючи кожному елементу ранг і відповідно їх сортуємо за таким правилом. Ранг 1 присвоюємо найменшому від'ємному елементу і розташовуємо його на першому місці вверху ряду. Ранг 2 присвоюємо найбільшому додатному елементу і розташовуємо його на останньому місці внизу ряду. Ранг 3 присвоюємо значенню, яке слідує за найменшим і розташовуємо його на другому місці вверху ряду. Ранг 4 – значенню, яке передує найбільшому і розташовуємо його на другому місці знизу і так далі.

У загальному ряді номери рангів збільшуються від країв до центра згідно з такою закономірністю: непарні номери (від'ємних елементів) – зверху до центра, а парні (додатних елементів) – знизу до центра.

На підставі цих рангів визначають випадкову величину w^* , яка приблизно розподілена за нормальним законом розподілу з математичним сподіванням

$$E[w^*] = R_1 - \frac{T_1 (T_1 + T_2 + 1)}{2}$$

і дисперсією

$$\text{Var}[w^*] = \frac{T_1 T_2 (T_1 + T_2 + 1)}{12},$$

де R_1 – сума рангів елементів першої сукупності; T_1, T_2 – кількість елементів у першій і другій сукупності відповідно.

Тоді випадкова величина

$$z = \frac{R_1 - \frac{T_1 (T_1 + T_2 + 1)}{2} \pm \frac{1}{2}}{\sqrt{\frac{T_1 T_2 (T_1 + T_2 + 1)}{12}}}$$

має нормальний стандартний розподіл. У цій формулі поправка $1/2$ вводиться для забезпечення неперервності величини z . Вона додається, якщо $z < 0$, і віднімається, якщо $z > 0$.

Гіпотеза про рівність дисперсій побудованих сукупностей приймається, якщо z задовольняє нерівність (4.1.6).

Серійні критерії стаціонарності

Для перевірки гіпотези про стаціонарний характер процесу можна використати універсальні щодо закону розподілу значень $y_t, t = 1, 2, \dots, T$ непараметричні тести, які ґрунтуються на аналізі закономірностей серій цих значень. Це так звані серійні критерії. Необхідною умовою їх застосування є досить значний обсяг часового ряду.

Прикладом серійного критерію є критерій Вальда - Вольфовитца. Згідно з цим критерієм досліджують розташування елементів стосовно медіани. Елементи, які є за рівнем вищі, ніж медіана, утворюють серію з додатним знаком, а до серії з від'ємним знаком входять елементи, рівень яких не перевищує медіани. Середнє значення числа серій визначають так:

$$E[N_s] = \frac{2N_1 N_2}{N_1 + N_2} + 1,$$

а його дисперсію визначають за формулою

$$\text{Var}[N_s] = \frac{2N_1 N_2 (2N_1 N_2 - (N_1 + N_2))}{(N_1 + N_2)^2 (N_1 + N_2 - 1)},$$

де N_1 – кількість елементів з додатним знаком; N_2 – кількість елементів з від'ємним знаком; $N_1 + N_2 = T$ – кількість елементів у часовому ряді; N_s – число серій.

Для великого обсягу часового ряду T нормована змінна

$$z = \frac{N_s - E[N_s] \pm 1/2}{\sigma(N_s)}$$

розподілена за стандартним нормальним розподілом $N[0, 1]$. У цьому випадку для перевірки гіпотези про стаціонарність використовують двосторонній критерій (4.1.6).

4.2 Автокореляційна функція ACF

Близькість реальних даних часового ряду до теоретичної моделі можна визначити на підставі автокореляційної функції. Для цього порівнюють теоретичну автокореляційну функцію ACF моделі й емпіричну функцію, яку визначають на підставі спостережуваних даних за допомогою вибіркового коефіцієнта автокореляції.

У цьому пункті без втрати загальності розглядатимемо ряди, які коливаються навколо нуля, тобто ряди y_t , для яких $E[y_t] = 0$. Якщо в емпіричних застосуваннях це не так, то початковий ряд треба перетворити і досліджувати ряд $y_t - \bar{y}$.

Без втрати загальності будемо розглядати стаціонарні ряди, які коливаються навколо нуля, тобто ряди, для яких $\bar{y} = 0$. Якщо це не так, то початковий ряд треба перетворити і розглядати ряд $y_t - \bar{y}$.

Функція автокореляції описує кореляцію між y_t і його лагом y_{t-k} і залежить від k . Нагадаємо, що коефіцієнти автокореляції k -го порядку визначають за формулами

$$\rho_k = \frac{\text{Cov}[y_t, y_{t-k}]}{\text{Var}[y_t]} = \frac{\gamma_k}{\gamma_0}. \quad (4.2.1)$$

При зміні значення k величини ρ_k є значеннями автокореляційної функції, яку позначають *ACF*.

Загалом у межах регресійної моделі автокореляція задається коефіцієнтом рівняння

$$y_t = \rho_k y_{t-k} + \varepsilon_t. \quad (4.2.2)$$

Справді, у цьому випадку $E[y_t] = 0$. Якщо помножити обидві частини (4.2.2) на y_{t-k} і обчислити математичне сподівання від обох частин, то врахувавши, що $E[\varepsilon_t y_{t-k}] = 0$, матимемо

$$\gamma_k = \rho_k \gamma_0,$$

звідки і випливає (4.2.1).

На підставі спостережуваних даних часового ряду y_t можна обчислити вибірккову функцію автокореляції, яка є функцією від k і дає оцінені автокореляційні коефіцієнти. Вибіркові коефіцієнти автокореляції позначають $\hat{\rho}_k$ і обчислюють за

формулами

$$\hat{\rho}_k = \frac{\frac{1}{T-k} \sum_{t=k+1}^T y_t y_{t-k}}{\frac{1}{T} \sum_{t=1}^T y_t^2}. \quad (4.2.3)$$

Тобто, коваріації у (4.2.1) заміняємо вибірковими оцінками.

У розділі 2 ми визначили вигляд автокореляцій різних порядків для $ARMA(p, q)$ процесів. Зокрема, для $MA(1)$ моделі АСF набуває значень

$$\rho_1 = \frac{\theta}{1 + \theta^2}, \quad \rho_2 = 0, \rho_3 = 0, \dots, \rho_k = 0, \dots,$$

для $MA(2)$ моделі з (2.3.5) впливає, що

$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2}, \quad \rho_2 = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2},$$

$$\rho_3 = 0, \rho_4 = 0, \dots, \rho_k = 0, \dots$$

Для загальної $MA(q)$ моделі АСF дорівнює нулю після q зсувів

$$\rho_{q+1} = 0, \rho_{q+2} = 0, \dots, \rho_k = 0, \dots$$

Теоретична автокореляційна функція моделі $MA(q)$ обривається після зсуву q . Якщо автокореляційна функція реального процесу має аналогічні властивості, то це свідчить про те, що для його опису доцільно використовувати модель рухомого середнього відповідного порядку.

Оскільки вибіркові коефіцієнти автокореляції можуть характеризуватися достатньо великими похибками, і, крім того, сильними кореляційними взаємозв'язками між собою, то на практиці точного збігу між теоретичною і емпіричною автокореляційною функцією не варто чекати, особливо при великих зсувах.

Зокрема, для MA моделі порядку q не завжди отримаємо, що $\hat{\rho}_k$ дорівнюють нулю для $k > q$. Але ми можемо використати $\hat{\rho}_k$ для тестування гіпотези $H_0 : \rho_k = 0$.

Для цього використовуємо асимптотичний розподіл $\hat{\rho}_k$ для $k > q$

$$\sqrt{T}(\hat{\rho}_k - \rho_k) \rightarrow N[0, V_k],$$

де $V_k = 1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_q^2$.

Отже, для тестування гіпотези, що правильна модель $MA(0)$ проти альтернативної $MA(1)$ можна тестувати гіпотезу $H_0 : \rho_1 = 0$ за допомогою тестової статистики

$$\Phi = \sqrt{T}\hat{\rho}_1,$$

яку порівнюємо з критичними значеннями стандартного нормального розподілу.

Тестування гіпотези, що правильна модель $MA(q-1)$ проти альтернативної $MA(q)$ еквівалентне тестуванню гіпотези $H_0 : \rho_q = 0$. Тестовою статистикою у цьому випадку буде величина

$$\Phi = \sqrt{T} \frac{\hat{\rho}_q}{\sqrt{1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_{q-1}^2}}, \quad (4.2.4)$$

яку порівнюємо з критичними значеннями нормального стандартного розподілу. Якщо $|\Phi| < 1,96$, то $MA(q-1)$ модель можна вважати правильною з коефіцієнтом надійності 0,95.

Зокрема, якщо для процесу виявився значущим лише перший коефіцієнт автокореляції $\hat{\rho}_1$ і $\hat{\rho}_1 \leq 0,5$, то доцільно вибрати для опису часового ряду модель $MA(1)$.

ACF авторегресійних моделей менш корисна. Для $AR(1)$ моделі в розділі 2 було показано, що ACF коефіцієнти не закінчуються на зсуві скінченої довжини. Автокореляційна функція моделі $AR(1)$ $\rho_k = \phi^k$ ($|\phi| < 1$) спадає до нуля строго по

експоненті (точніше, цей висновок правильний для абсолютних значень коефіцієнтів автокореляції)

$$|\rho_k| = e^{k \ln |\phi|} = e^{-k \ln |1/\phi|}.$$

Для *AR* моделей вищих порядків *ACF* складніша. Однак для них також характерний плавний характер зменшення. Згідно з пунктом 2.9. залежно від коефіцієнтів моделі в одному випадку спад відбувається або трохи швидше, ніж строго по експоненті, або трохи повільніше, в іншому випадку – за закономірністю, яка відповідає синусоїді, що затухає.

Порівнюючи теоретичні і емпіричні автокореляційні функції ми враховували лише їхні головні властивості. Тому їхній збіг дає змогу лише значно звужити коло прийнятних для опису варіантів моделі. Остаточний вибір на користь певної моделі робиться за результатами наступних етапів оцінювання та діагностики.

4.3 Часткова автокореляційна функція *PACF*

Дуже важлива інформація про порядок моделі міститься в частковій функції автокореляції, яку позначають *PACF*.

Коефіцієнти часткової автокореляції позначають ρ_{kk} і знаходяться з *AR(k)* моделі

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_k y_{t-k} + \varepsilon_t \quad (4.3.1)$$

як МНК-оцінку коефіцієнта при y_{t-k} . Тобто, $\rho_{kk} = \hat{\varphi}_k$. Часткова автокореляція ρ_{kk} вимірює додаткову кореляцію між y_t і y_{t-k} після врахування впливу $y_{t-1}, \dots, y_{t-k-1}$ на y_t .

Якщо y_t – білий шум, то, оскільки кожне значення некорельоване з іншими, всі коефіцієнти в (4.3.1), а отже, і всі

ρ_{kk} будуть дорівнювати нулю. Для інших рядів *PACF* зручно оцінити за допомогою коефіцієнтів автокореляції.

Коефіцієнт часткової автокореляції першого порядку ρ_{11} визначається зі співвідношення

$$y_t = \rho_{11} y_{t-1} + \varepsilon_t. \quad (4.3.2)$$

Порівнявши останню рівність з (4.2.2) для $k = 1$, бачимо, що

$$\rho_{11} = \rho_1. \quad (4.3.3)$$

На практиці оцінка $\hat{\rho}_{11}$ визначається як оцінка коефіцієнта автокореляції першого порядку

$$\hat{\rho}_{11} = \hat{\rho}_1.$$

Значення ρ_{22} знаходять відповідно як коефіцієнт моделі *AR*(2)

$$y_t = \rho_{12} y_{t-1} + \rho_{22} y_{t-2} + \varepsilon_t. \quad (4.3.4)$$

Тут коефіцієнтом часткової автокореляції другого порядку є коефіцієнт при y_{t-2} . Він відображає частковий або додатковий ефект від додавання y_{t-2} до рівняння, в якому є y_{t-1} . Помноживши (4.3.4) на y_{t-1} і обчисливши математичне сподівання, одержимо

$$\gamma_1 = \rho_{12} \gamma_0 + \rho_{22} \gamma_1 + E[\varepsilon_t y_{t-1}].$$

Поділивши на γ_0 і використавши, що $E[\varepsilon_t y_{t-1}] = 0$, матимемо

$$\rho_1 = \rho_{12} + \rho_{22} \rho_1$$

або

$$\rho_{12} = \rho_1(1 - \rho_{22}). \quad (4.3.5)$$

Помноживши (4.3.4) на y_{t-2} , аналогічно одержимо вираз

$$\rho_{22} = \rho_2 - \rho_{12} \rho_1,$$

який, використавши (4.3.5), можна записати у вигляді

$$\rho_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.$$

На практиці оцінку $\hat{\rho}_{22}$ обчислюємо так:

$$\hat{\rho}_{22} = \frac{\hat{\rho}_2 - \hat{\rho}_1^2}{1 - \hat{\rho}_1^2} = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & \hat{\rho}_2 \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 \end{vmatrix}}. \quad (4.3.6)$$

В загальному випадку для моделі $AR(p)$ коефіцієнт $\hat{\rho}_{kk}$ обчислюємо з системи Юла - Уолкера за формулою

$$\hat{\rho}_{kk} = \frac{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \dots & \hat{\rho}_{k-2} & \hat{\rho}_1 \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{k-3} & \hat{\rho}_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \dots & \hat{\rho}_1 & \hat{\rho}_k \end{vmatrix}}{\begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \dots & \hat{\rho}_{k-2} & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{k-3} & \hat{\rho}_{k-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \dots & \hat{\rho}_1 & 1 \end{vmatrix}} = \frac{\Delta_k}{\Delta} \quad (4.3.7)$$

для $k = 1, 2, \dots, p, p+1, \dots$.

На підставі виведених співвідношень обчислимо часткову автокореляційну функцію для авторегресійних моделей. Зокрема, для $AR(1)$ моделі матимемо

$$\rho_{11} = \phi, \quad \rho_{22} = 0, \quad \rho_{33} = 0, \quad \dots, \quad \rho_{kk} = 0, \dots,$$

оскільки перший і останній стовпці визначника

$$\begin{vmatrix} 1 & \phi & \phi^2 & \dots & \phi^{k-2} & \phi \\ \phi & 1 & \phi & \dots & \phi^{k-3} & \phi^2 \\ \phi^2 & \phi & 1 & \dots & \phi^{k-4} & \phi^3 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \phi^{k-1} & \phi^{k-2} & \phi^{k-3} & \dots & \phi & \phi^k \end{vmatrix}$$

пропорційні, тобто $PACF$ дорівнює нулю, крім $PACF$ першого порядку.

Для $AR(2)$ –

$$\rho_{11} = \frac{\phi_1}{1 - \phi_2}, \quad \rho_{22} = \phi_2, \quad \rho_{33} = 0, \quad \rho_{44} = 0, \quad \dots, \quad \rho_{kk} = 0, \quad \dots,$$

тобто $PACF$ дорівнює нулю для зсувів більше двох.

У загальному випадку правильне твердження.

Для $AR(p)$ процесу МНК - оцінки $AR(k)$ моделі для $k \geq p$ є змістовними, тому

$$\text{plim } \rho_{kk} = 0, \quad \text{якщо } k > p.$$

Тобто для моделі $AR(p)$ значення часткової функції автокореляції є значущими (відмінними від нуля) до зсуву порядку p ($\hat{\rho}_{kk} > 0$, при $k \leq p$) і дорівнюють нулю при зсувах, які перевищують порядок моделі ($\hat{\rho}_{kk} = 0$ при $k > p$).

Отже, поведінка часткової функції автокореляції моделей авторегресії аналогічна до поведінки автокореляційної функції моделей рухомого середнього. Для моделі $AR(p)$ її часткова автокореляційна функція $PACF$ обривається після зсуву p , так як це було б для ACF моделі $MA(p)$. Цю властивість часткової функції автокореляції зручно використати для ідентифікації моделі авторегресії.

Використаємо коефіцієнти часткової автокореляції для знаходження порядку AR процесу. Для цього використаємо твердження.

Для $AR(p)$ процесу асимптотичний розподіл $\hat{\rho}_{kk}$ при $k \geq p$ є нормальним, а саме

$$\sqrt{T} \hat{\rho}_{kk} \rightarrow N[0, 1],$$

Тестування гіпотези, що справджується модель $AR(k-1)$ проти альтернативної $AR(k)$ моделі зводиться до тестування

гіпотези $H_0 : \rho_{kk} = 0$. Гіпотеза відхиляється при рівні значущості $\alpha = 0,05$, якщо $|\sqrt{T}\hat{\rho}_{kk}| > 1,96$. Отже, на підставі *PACF* можна протестувати, для яких зсувів коефіцієнти ρ_{kk} відмінні від нуля. Для справжньої *AR(p)* моделі часткові автокореляції будуть близькі до нуля після p -го лага.

Так само як і для авторегресійних моделей часткові функції автокореляції можуть бути побудовані і для моделей рухомого середнього довільного порядку. Для їхньої оцінки використовують формули (4.3.3), (4.3.6), (4.3.7).

Зокрема, для *MA(1)* процесу з (4.3.3), (4.3.6) матимемо

$$\rho_{11} = \frac{\theta}{1 + \theta^2}, \quad \rho_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4}.$$

Оскільки для цієї моделі $\rho_1 = \theta/(1 + \theta^2)$ і $\rho_2 = \rho_3 = \dots = 0$, то можна вивести формулу для значень оцінок часткових коефіцієнтів автокореляції довільного порядку

$$\hat{\rho}_{kk} = (-1)^{k+1} \frac{\theta^k (1 - \theta^2)}{1 - \theta^{2(k+1)}}. \quad (4.3.8)$$

Звідси випливає, що

$$|\hat{\rho}_{kk}| < |\theta|^k,$$

тобто часткова функція автокореляції *PACF* моделі *MA(1)* спадає до нуля за законом близьким до експоненціального. Іншими словами, її поведінка подібна до поведінки *ACF* моделі *AR(1)*.

Можна показати, що аналогічна відповідність простежується у властивостях *PACF* зворотної ($|\theta| < 1$) моделі *MA(2)* і *ACF* моделі *AR(2)*. Вони плавно спадають при зростанні зсуву за експоненціальним законом, або поведуть себе як затухаючі синусоїди. Така відповідність між автокореляційними і частковими автокореляційними функціями характерна також і для моделей авторегресії і рухомого середнього вищих порядків.

Для процесу $ARMA(1, 1)$ коефіцієнти часткової автокореляції мають вигляд

$$\rho_{11} = \phi - \frac{\theta(1 - \phi)}{1 + \theta^2 + 2\theta\phi}, \quad \rho_{22} = \frac{\rho_{11}(\phi - \rho_{11})}{1 - \rho_{11}^2},$$

тобто $PACF$ є спадною. Ці результати також поширюються на $ARMA(p, q)$ модель, для якої $PACF$ є спадною.

4.4 Моделювання Бокса - Дженкінса

Перейдемо до проблеми вибору на підставі сукупності T спостережень y_1, y_2, \dots, y_T відповідної $ARMA$ моделі. Вперше MA , AR і $ARMA$ - моделі розглянули Бокс і Дженкінс у 1976 р. Вони об'єднали кілька напрямів у літературі та запропонували алгоритм для вибору моделі.

Першим кроком є розгляд даних та з'ясування, чи є вибраний ряд стаціонарним (за допомогою методів, описаних у пункті 4.1). Зауважимо таке: для того, щоб робити обґрунтовані висновки, потрібна достатня кількість спостережень.

Властивості нестационарних рядів і питання зведення їх до стаціонарних детально розглянемо в розділі 7. Тут лише зазначимо таке: якщо часовий ряд нестационарний, то Бокс і Дженкінс запропонували знаходити його різниці до одержання стаціонарного процесу, оскільки перехід до перших різниць усуває лінійний тренд, друга різниця усуває квадратичний і т.д. У загальному випадку для одержання стаціонарного ряду ми можемо знаходити різниці d разів. Перетворюючи змінні для одержання стаціонарного ряду важливо уникати знаходження різниць порядку вище потрібного. Це може трапитись при знаходженні різниць стаціонарного ряду. Простою перевіркою є перевірка дисперсії ряду на зростання. Наприклад, припустимо, що ряд є білим шумом з дисперсією σ^2 , тоді обчислення різниці першого порядку дає зростання дисперсії до

$2\sigma^2$. Знаходження різниць нестационарних рядів дає – здебільшого ряди з меншою дисперсією.

Після того, як стаціонарний ряд одержано, на підставі даних вибірки знаходять автокореляційну і часткову автокореляційну функцію, порівнюють з ACF і $PACF$ різних теоретичних моделей у надії знайти подібний випадок. Характеристики ACF і $PACF$ різних моделей підсумовано у таблиці 4.4.1.

Таблиця 4.4.1

Модель	ACF	$PACF$
Білий шум	Всі нулі	Всі нулі
$MA(1)$	Нулі після ρ_1	Спадна після ρ_{11}
$MA(2)$	Нулі після ρ_2	Спадна після ρ_{22}
$MA(q)$	Нулі після ρ_q	Спадна після ρ_{qq}
$AR(1)$	Спадна після ρ_1	Нулі після ρ_{11}
$AR(2)$	Спадна після ρ_2	Нулі після ρ_{22}
$AR(p)$	Спадна після ρ_p	Нулі після ρ_{pp}
$ARMA(1, 1)$	Спадна після ρ_1	Спадна після ρ_{11}
$ARMA(p, q)$	Спадна після ρ_p	Спадна після ρ_{qq}

Зауважимо, що на практиці найчастіше використовують модель $ARMA(1, 1)$, тобто тільки $ARMA(p, q)$ модель першого порядку. Це пов'язано з тим, що частина моделі, яка належить до авторегресії першого порядку поглинає всі процеси рухомого середнього вищих порядків, і навпаки, частина, яка належить до процесу рухомого середнього першого порядку поглинає всі процеси авторегресії вищих порядків. Внаслідок цього поведінка автокореляційної і часткової автокореляційної функції моделі $ARMA(1, 1)$ характеризується комбінацією властивостей, якими володіють ці функції для процесів $AR(1)$ і $MA(1)$. А саме, складова $AR(1)$ сприяє тому, що автокореляційна функція моделі $ARMA(1, 1)$ (абсолютні значення коефіцієнтів автокореляції) затухає експоненціально, але після пер-

шого зсуву. У свою чергу, складова $MA(1)$ визначає закономірність поведінки часткової автокореляційної функції моделі $ARMA(1, 1)$, яка також затухає приблизно експоненціально.

Розглянуті підходи до ідентифікації моделі ґрунтуються на зіставленні властивостей вибіркових автокореляційних і часткових автокореляційних функцій реального стаціонарного процесу і моделі, яку використовують для опису часового ряду. На практиці ідеальне співпадіння трапляється не часто, оскільки реальні процеси не зовсім точно відповідають своїм теоретичним аналогам-моделям, і оцінки їхніх коефіцієнтів автокореляції характеризуються наявністю похибок. Тому точної відповідності між даними і теоретичною моделлю не буде. Це може призвести до вибору на цьому кроці двох або трьох пробних моделей (тобто пар величин p та q).

Наступним кроком є оцінювання параметрів кожної з пробних моделей та перевірка їх на адекватність. Критерієм перевірки значимої відмінності оцінки параметра від нуля є t критерій. Сукупну значимість декількох параметрів перевіряємо за допомогою статистики Вальда. Для прийнятої моделі всі параметри повинні бути значущими. Якщо це не так, то відповідні величини треба прийняти такими, що дорівнюють нулеві та ще раз оцінити параметри моделі.

Якщо на базі ACF і $PACF$ вибрано $ARMA(p, q)$, то треба оцінити і $ARMA(p + 1, q)$ та $ARMA(p, q + 1)$ моделі і протестувати значущість додаткових параметрів.

Крім того, треба перевірити стійкість параметрів за допомогою оцінювання їх на підставі підмножин даних для з'ясування чи змінились параметри. Для цього можна використати критерій правдоподібності для перевірки гіпотези сталості всіх параметрів поза деякими підперіодами всього проміжку часу. Позначимо суму квадратів залишків моделі без обмежень – S_u , а моделі з обмеженнями – S_r . Тоді гіпотеза правдо-

подібності обмежень перевіряється за допомогою статистики

$$T \ln \left(\frac{S_r}{S_u} \right) \sim \chi^2[m],$$

де m – кількість обмежень. Гіпотеза приймається, якщо статистика менша за критичне значення відповідного χ^2 - квадрат розподілу, і робиться висновок, що параметри стійкі.

Далі діагностують кожну модель, тобто аналізують їхні залишки.

4.5 Діагностування моделі

Корисним інструментом у перевірці на адекватність моделі можуть бути залишки $\hat{\varepsilon}_t = y_t - \hat{y}_t$, де \hat{y}_t – значення часового ряду, обчислені за побудованою моделлю.

Залишковий аналіз ґрунтується на тому факті, що залишки адекватної моделі мають бути близькими до білого шуму, – тобто до процесу, який характеризується повною відсутністю будь-яких закономірностей у своїх значеннях, за винятком відомого закону розподілу, який зазвичай припускається нормальним. Це означає, що математичне сподівання фактичної похибки повинно дорівнювати нулю ($E[\hat{\varepsilon}_t] = 0$), її дисперсія стала на будь-якій ділянці виміру ($\sigma_{\hat{\varepsilon}}^2 = const$) і між рядами $\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, \dots$ немає автокореляційної залежності, тобто всі вибіркові коефіцієнти автокореляції ряду $\hat{\varepsilon}_t$ ($t = 1, 2, \dots, T$) повинні бути близькими до нуля.

Значущість залишкових автокореляцій перевіряється порівнянням з двома границями стандартних похибок $\pm 2/\sqrt{T}$,

тобто для всіх k повинні виконуватись нерівності

$$|\text{Corr}[\hat{\varepsilon}_t, \hat{\varepsilon}_{t-k}]| = \left| \frac{\frac{1}{T-k} \sum_{t=k+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2} \right| < \frac{2}{\sqrt{T}}.$$

Якщо обчислено 10 значень АСФ і 10 значень РАСФ, то при 5% рівні значущості в середньому один з 20 коефіцієнтів буде значущим. Цей факт разом з порівняно незначним обсягом вибірки на практиці означає, що критерій на основі окремих коефіцієнтів АСФ чи РАСФ може бути не надійним.

Альтернативою є використання критеріїв, які охоплюють всі оцінені коефіцієнти.

Бокс та Пірс запропонували використати статистику, за допомогою якої оцінюється значимість деякої послідовності вибірових коефіцієнтів автокореляції для $\hat{\varepsilon}_t$

$$BP_K = T \sum_{i=1}^K r_i^2,$$

де r_i – оцінені коефіцієнти автокореляції залишків $\hat{\varepsilon}_t$, $r_i = \text{Corr}[\hat{\varepsilon}_t, \hat{\varepsilon}_{t-i}]$, K – число вибране дослідником. Якщо значення статистики є великим, то гіпотеза про те, що залишки моделі є білим шумом відхиляється.

Однією з модифікацій тесту Бокса - Пірса є статистика Лjung-Бокс (Ljung, Box)

$$LB_K = T(T+2) \sum_{i=1}^K \frac{1}{T-i} r_i^2.$$

При залишковому аналізі обчислюють статистику LB_K для різних K , причому K має бути більшим за кількість параметрів моделі, тобто $K > p + q$.

Для $ARMA(p, q)$ процесу

$$LB_K \sim \chi^2[K - p - q].$$

Отже, якщо $LB_K < \chi_{кр}^2$, то гіпотеза про відсутність у ряді похибок автокореляційних залежностей може бути прийнята. Якщо нерівність не виконується, то модель $ARMA(p, q)$ відхиляється на цій стадії, і цикл будівництва моделі треба повторити.

4.6 Придатність моделі до прогнозування

Для того щоб перевірити придатність оцінених моделей до прогнозування, для вибору і підгонки моделі замість використання всіх зібраних даних деяку частину даних (n останніх спостережень) заощаджують. Ці дані потім використовують для перевірки точності прогнозів.

Отже, оцінимо модель на основі $T - n$ спостережень і побудуємо прогноз на наступні n періодів. Позначимо \hat{y}_t – прогнозовані значення ряду, які обчислені на основі цієї моделі. Нехай $f_t = y_t - \hat{y}_t$ – похибки прогнозу.

Якщо залишки моделі нормально розподілені, то гіпотезу стосовно величини похибок прогнозу перевіряють за допомогою статистики

$$\sum_{t=T-n+1}^T \left(\frac{f_t}{s} \right)^2 \sim \chi^2[n],$$

де s є оцінкою стандартної похибки залишків. Якщо модель задовільна, то значення статистики буде малим.

Для визначення наявності у похибок прогнозу деякої структури можна використати інші критерії, оскільки систематич-

не завищення або заниження прогнозів порівняно з фактичними даними свідчить про незадовільність моделі. Коли модель прийнято, то її параметри ще раз оцінюють на основі всіх доступних даних.

Якщо після діагностування і тестування придатності моделей до прогнозування ще залишилося декілька адекватних моделей, то для вибору однієї з них застосовують статистичні критерії.

4.7 Статистичні критерії для вибору моделі

Раніше було показано, що стаціонарний авторегресійний процес $AR(p)$ може бути зведено до процесу рухомого середнього $MA(\infty)$, і навпаки зворотний процес рухомого середнього $MA(q)$ до авторегресійного $AR(\infty)$ процесу. Звідси випливає важливий висновок. На практиці можна підібрати модель з мінімальною кількістю параметрів, яка описує часовий ряд y_t , який є стаціонарним процесом невисокого порядку, "не гірше", ніж інші варіанти моделей з великою кількістю параметрів. Вираз "не гірше" пов'язують з мінімальною дисперсією моделі і відсутністю автокореляцій в похибках.

Практична цінність цього висновку така. Будуючи моделі часових рядів, треба прагнути до мінімізації їх параметрів, тобто порядку самої моделі. Оскільки параметри моделей часто оцінюють на підставі коефіцієнтів автокореляції вихідного процесу y_t , то зі збільшенням порядку моделі для оцінювання їхніх значень необхідно використовувати як вихідні дані більшу кількість вибірових коефіцієнтів автокореляції. Точність їхніх оцінок при зростанні зсуву падає, а їх абсолютні значення або прямують до нуля, або попадають в ділянку підвищеної невизначеності. Через це знижується і надійність оцінок

коефіцієнтів моделі часових рядів високих порядків, і якість самих моделей. З огляду на це, економетристи шукають для опису реальних процесів моделі часових рядів з мінімальною кількістю параметрів.

Оскільки при моделюванні за допомогою часових рядів економічна теорія не дає умов для вибору моделі з декількох альтернативних, то використовують статистичні методи. Зі статистичного погляду можна використати декілька додаткових критеріїв вибору з альтернативних моделей. Ці критерії враховують кількість параметрів, які входять до моделі. Перевага надається моделі, яка містить меншу кількість параметрів. Наприклад, якщо $MA(2)$ і $AR(10)$ моделі дають подібні результати, то вибирають $MA(2)$ модель, як більш зручну.

Одним з таких критеріїв є інформаційний критерій Акайка (1973) (*Akaike's Information Criterion*)

$$AIC = \ln \hat{s}^2 + 2 \frac{p+q}{T},$$

де \hat{s}^2 – оцінка дисперсії залишків $\hat{\varepsilon}_t$ моделі. Маємо нелінійне компромісне співвідношення між дисперсією залишків та значенням $p+q$, оскільки моделі з більшим $p+q$ може бути надано перевагу лише при пропорційно великому зменшенні \hat{s}^2 .

Гевек та Міз (1981) запропонували Баєсівський інформаційний критерій (*Bayesian Information Criterion*)

$$BIC = \ln \hat{s}^2 + \frac{p+q}{T} \ln T.$$

Він надає більше ваги $p+q$ порівняно з AIC при $T > 7$, тобто зростання $p+q$ потребує більшого зменшення \hat{s}^2 для BIC , ніж для AIC .

Іншу статистику запропонували Хеннан і Квінн

$$HQ = \ln \hat{s}^2 + 2 \frac{p+q}{T} \ln (\ln T).$$

Тут вага при $p + q$ є більшою за 2 при $T > 15$.

Всі ці критерії ґрунтуються на економності, яка вимірюється кількістю вільних параметрів $p + q$. Якщо константа входить в модель, то треба використовувати $p + q + 1$. Кращою є модель, для якої значення AIC чи BIC є меншим.

Вибір між цими критеріями є до деякої міри довільним, оскільки всі статистики змінюються в тому самому напрямі при збільшенні $p + q$. На практиці можна використати один з них. Зазвичай BIC використовують для моделей класу $ARMA(p, q)$ для малих значень p, q . Він має властивість майже точного визначення правильної моделі, якщо $T \rightarrow \infty$. AIC критерій використовується в асимптотичних, перепараметризованих моделях.

Розділ 5

Оцінювання $ARMA$ моделей

Нехай знаємо, що ряд деяких даних y_1, y_2, \dots, y_T описується $ARMA$ процесом порядку (p, q) . Залежно від специфікації моделі і припущень щодо розподілу, можемо оцінювати невідомі параметри за допомогою МНК або за допомогою методу максимальної правдоподібності.

5.1 Метод найменших квадратів

Метод найменших квадратів полягає у мінімізації суми квадратів відхилень. Цей метод легко застосувати для авторегресійних моделей.

Розглянемо $AR(p)$ модель

$$y_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

де ε_t - процес білого шуму. Оскільки ε_t не корелює з довільними ε_j , що передують ε_t , то

$$\text{Cov}[y_{t-j}, \varepsilon_t] = 0 \quad \text{для } j = 1, \dots, p,$$

тобто похибки і пояснювальні змінні моделі не корелюють між собою. Тому для $AR(p)$ моделі МНК дає змістовні оцінки. Отже, оцінювання AR моделі не відрізняється від оцінювання моделі зі лаговими залежними змінними.

Для MA моделей оцінки МНК дещо ускладнені. Розглянемо $MA(1)$ модель

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}.$$

Оскільки ε_{t-1} невідоме, то не можемо застосувати МНК. Теоретично МНК мінімізує функцію

$$F(\theta) = \sum_{t=2}^T (y_t - \theta\varepsilon_{t-1})^2. \quad (5.1.1)$$

У випадку, коли MA поліном зворотний, можемо виразити ε_t через запізнення y_t

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta)^j y_{t-j}.$$

Тоді

$$\varepsilon_{t-1} = \sum_{j=0}^{\infty} (-\theta)^j y_{t-j-1}. \quad (5.1.2)$$

Підставивши (5.1.2) в (5.1.1), отримаємо

$$F(\theta) = \sum_{t=2}^T \left(y_t - \theta \sum_{j=0}^{\infty} (-\theta)^j y_{t-j-1} \right)^2.$$

На практиці, y_t не визначене для $t = 0, -1, -2, \dots$. Тому зменшуємо нескінчену суму в цьому виразі, залишаючи лише доданки з відомими y_t , і отримуємо наближення суми квадратів

$$\tilde{F}(\theta) = \sum_{t=2}^T \left(y_t - \theta \sum_{j=0}^{t-2} (-\theta)^j y_{t-j-1} \right)^2. \quad (5.1.3)$$

Якщо $T \rightarrow \infty$, то різниця між $F(\theta)$ і $\tilde{F}(\theta)$ зникає. Отже, можемо мінімізувати (5.1.3). Але функція $\tilde{F}(\theta)$ є поліномом високого порядку і має багато локальних мінімумів. Тому числова мінімізація (5.1.3) ускладнена. Але якщо вимагати, щоб $\theta \in (-1, 1)$, то можна виконати сітчастий пошук за

$$\theta = \{-0,99; -0,98; \dots; -0,01; 0; 0,01; \dots; 0,98; 0,99\}.$$

Вибираємо те θ , яке дає найменше значення $\tilde{F}(\theta)$. В результаті отримуємо НЛМНК оцінку $\hat{\theta}$, яка є змістовною і асимптотично нормальною.

5.2 Оцінювання параметрів за допомогою коефіцієнтів автокореляції

5.2.1 Оцінювання параметрів $AR(p)$ процесу

Розглянемо авторегресійний процес p -го порядку

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t. \quad (5.2.1)$$

Припустимо, що математичне сподівання ряду y_t дорівнює нулю, тобто $E[y_t] = 0$. Якщо це не так, то замість змінної y_t можна розглядати центровану величину $y'_t = y_t - \bar{y}$, де \bar{y} – оцінка $E[y_t]$. Тоді $E[y'_t] = 0$.

Параметри моделі (5.2.1) можна оцінити за допомогою коефіцієнтів автокореляції. Для цього помножимо (5.2.1) на y_{t-k} і обчислимо математичне сподівання обох частин отриманих рівностей. Матимемо систему

$$E[y_t y_{t-k}] = \phi_1 E[y_{t-1} y_{t-k}] + \phi_2 E[y_{t-2} y_{t-k}] + \dots + \phi_p E[y_{t-p} y_{t-k}] + E[\varepsilon_t y_{t-k}]. \quad (5.2.2)$$

Система (5.2.6) є рівняннями Юла - Уолкера, тому оцінки так отримані називають оцінками Юла - Уолкера. Зокрема, розв'язок системи (5.2.6) можна шукати методом Крамера. Отже, матимемо

$$\hat{\phi}_i = \frac{\Delta_i}{\Delta},$$

де Δ – визначник системи (5.2.6)

$$\Delta = \begin{vmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \dots & \hat{\rho}_{p-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{p-2} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\rho}_{p-1} & \hat{\rho}_{p-2} & \hat{\rho}_{p-3} & \dots & 1 \end{vmatrix}, \quad (5.2.7)$$

а Δ_i – визначник, який одержується з визначника Δ шляхом заміни i -го стовпця на стовпець, який складається з коефіцієнтів автокореляції лівої частини системи (5.2.6) – $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_p$.

На практиці в моделях авторегресії високого порядку властивості незміщеності й ефективності для оцінок Юла - Уолкера можуть не виконуватись. Але при невеликих порядках моделі ($p = 1, 2, 3$) оцінки Юла - Уолкера найчастіше достатньо хороші. У будь-якому випадку їх можна розглядати як перше наближення оптимальних оцінок, які можуть бути отримані за допомогою нелінійних методів, зокрема методом максимальної правдоподібності.

5.2.2 Оцінювання параметрів $MA(q)$ процесу

Розглянемо $MA(q)$ модель

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}. \quad (5.2.8)$$

У розділі 2 було виведено формули для дисперсії і автоковаріацій. А саме

$$\gamma_0 = \sigma^2(1 + \theta_1^2 + \dots + \theta_q^2),$$

$$\gamma_k = \begin{cases} \sigma^2(\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q), & k = 1, 2, \dots, q; \\ 0, & k > q. \end{cases} \quad (5.2.9)$$

Звідси матимемо

$$\rho_k = \begin{cases} \frac{\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2}, & k = 1, 2, \dots, q; \\ 0, & k > q. \end{cases} \quad (5.2.10)$$

Систему з q рівнянь, сформованих для $k = 1, 2, \dots, q$, можна використати для отримання оцінок $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ невідомих параметрів моделі $\theta_1, \theta_2, \dots, \theta_q$. Для цього треба підставити в кожне з її рівнянь замість значень коефіцієнтів автокореляції $\rho_1, \rho_2, \dots, \rho_q$ їхні вибіркові оцінки $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_q$.

На відміну від рівнянь Юла - Уолкера ця система є нелінійною і для її розв'язання треба застосовувати спеціальні ітеративні процедури.

Проте для випадку MA(1) система (5.2.10) розв'язується досить просто. Для цієї моделі дисперсія процесу дорівнює

$$\gamma_0 = (1 + \theta^2)\sigma^2,$$

а її єдиний відмінний від нуля перший коефіцієнт автокореляції виражається через коефіцієнт моделі так:

$$\rho_1 = \frac{\theta}{1 + \theta^2}. \quad (5.2.11)$$

З (5.2.11) можна одержати квадратне рівняння для знаходження оцінки $\hat{\theta}$ невідомого параметра θ

$$\hat{\theta}^2 + \frac{1}{\hat{\rho}_1}\hat{\theta} + 1 = 0, \quad (5.2.12)$$

де $\hat{\rho}_1$ – оцінка коефіцієнта автокореляції першого порядку процесу y_t .

З (5.2.12) випливає існування двох розв'язків цього рівняння

$$\hat{\theta}_{1,2} = \frac{-\frac{1}{\hat{\rho}_1} \pm \sqrt{\left(\frac{1}{\hat{\rho}_1}\right)^2 - 4}}{2}$$

за умови, що

$$\left(\frac{1}{\hat{\rho}_1}\right)^2 - 4 \geq 0,$$

тобто

$$|\hat{\rho}_1| \leq 0,5. \quad (5.2.13)$$

Розв'язки $\hat{\theta}_1, \hat{\theta}_2$ пов'язані між собою співвідношеннями

$$\hat{\theta}_1 \hat{\theta}_2 = 1, \quad \hat{\theta}_1 = \frac{1}{\hat{\theta}_2}.$$

Умову зворотності процесу задовольняє тільки один з розв'язків, модуль якого менший за одиницю.

5.2.3 Оцінювання параметрів $ARMA(p, q)$ процесу

Розглянемо $ARMA(p, q)$ модель

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (5.2.14)$$

Аналогічно як в (5.2.4) можна показати, що при $k > q$ маємо рівності

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}, \quad k \geq q + 1.$$

Тобто при $k > q$ значення коефіцієнтів автоковаріації $ARMA(p, q)$ моделі володіють властивостями коефіцієнтів автоковаріації $AR(p)$ процесу.

Отже, невідомі параметри $\phi_1, \phi_2, \dots, \phi_p$ можна оцінити, розв'язавши систему Юла - Уолкера, яка в цьому випадку має вигляд

$$\begin{cases} \hat{\rho}_{q+1} = \phi_1 \hat{\rho}_q + \phi_2 \hat{\rho}_{q-1} + \dots + \phi_p \hat{\rho}_{q-p+1} \\ \hat{\rho}_{q+2} = \phi_1 \hat{\rho}_{q+1} + \phi_2 \hat{\rho}_q + \dots + \phi_p \hat{\rho}_{q-p+2} \\ \dots \\ \hat{\rho}_{q+p} = \phi_1 \hat{\rho}_{q+p-1} + \phi_2 \hat{\rho}_{q+p-2} + \dots + \phi_p \hat{\rho}_q, \end{cases} \quad (5.2.15)$$

де $\hat{\rho}_i = \hat{\rho}_{-i}$, $\hat{\rho}_0 = 1$.

Використовуючи знайдені з системи (5.2.15) оцінки параметрів $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$, отримуємо процес рухомого середнього q -го порядку

$$y_t - \hat{\phi}_1 y_{t-1} - \dots - \hat{\phi}_p y_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (5.2.16)$$

Подальше оцінювання параметрів $\theta_1, \theta_2, \dots, \theta_q$ проводять нелінійними методами, які дають змогу розв'язати систему вигляду (5.2.10).

5.3 Оцінки максимальної правдоподібності

5.3.1 Функція правдоподібності Гаусівського $AR(1)$ процесу

Точні оцінки максимальної правдоподібності

Гаусівський процес $AR(1)$ має вигляд

$$y_t = c + \phi y_{t-1} + \varepsilon_t, \quad (5.3.1)$$

де $\varepsilon_t \sim N[0, \sigma^2]$. Потрібно оцінити вектор параметрів $\vartheta = [c, \phi, \sigma^2]'$.

Розглянемо ймовірнісний розподіл y_1 першого спостереження вибірки. З рівнянь (2.5.3) і (2.5.4) маємо

$$E[y_1] = \mu = c/(1 - \phi), \quad \text{Var}[y_1] = E[(y_1 - \mu)^2] = \sigma^2/(1 - \phi^2).$$

Оскільки змінні $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ Гауссівські, то змінна y_1 теж є Гауссівська. Отже, щільність першого спостереження має вигляд

$$f(y_1; \vartheta) = \frac{1}{\sqrt{2\pi\sigma^2/(1 - \phi^2)}} \exp \left[\frac{-(y_1 - (c/(1 - \phi)))^2}{2\sigma^2/(1 - \phi^2)} \right]. \quad (5.3.2)$$

Розглянемо друге спостереження y_2 умовно за y_1 . З (5.3.1)

$$y_2 = c + \phi y_1 + \varepsilon_2. \quad (5.3.3)$$

Трактуючи y_1 як детерміновану сталу, матимемо

$$y_2|y_1 \sim N[c + \phi y_1, \sigma^2].$$

Отож,

$$f(y_2|y_1; \vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-(y_2 - c - \phi y_1)^2}{2\sigma^2} \right]. \quad (5.3.4)$$

Функція правдоподібності двох спостережень y_1 і y_2 є добутком (5.3.4) і (5.3.2)

$$L(y_1, y_2; \vartheta) = f(y_1; \vartheta)f(y_2|y_1; \vartheta).$$

Аналогічно для третього спостереження

$$f(y_3|y_2, y_1; \vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-(y_3 - c - \phi y_2)^2}{2\sigma^2} \right],$$

звідки

$$L(y_1, y_2, y_3; \vartheta) = L(y_1, y_2; \vartheta)f(y_3|y_2, y_1; \vartheta).$$

Продовжуючи цей процес, одержуємо

$$\begin{aligned} f(y_t|y_{t-1}, y_{t-2}, \dots, y_1; \vartheta) &= \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[\frac{-(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right], \quad 2 \leq t \leq T \end{aligned} \quad (5.3.5)$$

і

$$L(y_1, \dots, y_T; \vartheta) = f(y_1; \vartheta) \prod_{t=2}^T f(y_t|y_{t-1}, y_{t-2}, \dots, y_1; \vartheta). \quad (5.3.6)$$

Підставляючи (5.3.2) і (5.3.5) у (5.3.6), одержимо логарифм функції правдоподібності, який позначимо $\mathcal{L}(\vartheta) = \ln L(y; \vartheta)$

$$\begin{aligned} \mathcal{L}(\vartheta) &= -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln[\sigma^2/(1-\phi^2)] - \frac{(y_1 - (c/(1-\phi)))^2}{2\sigma^2/(1-\phi^2)} - \\ &- \frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln \sigma^2 - \sum_{t=2}^T \left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right]. \end{aligned}$$

Спростивши подібні доданки, матимемо

$$\begin{aligned} \mathcal{L}(\vartheta) &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 + \frac{1}{2} \ln(1-\phi^2) - \\ &- \frac{(y_1 - (c/(1-\phi)))^2}{2\sigma^2/(1-\phi^2)} - \sum_{t=2}^T \left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right]. \end{aligned} \quad (5.3.7)$$

Розглянемо ще один спосіб обчислення функції правдоподібності. Для цього розглянемо вектор

$$y = [y_1, y_2, \dots, y_T]'. \quad (5.3.8)$$

Для всіх t матимемо

$$E[y_t] = \mu = \frac{c}{1-\phi}, \quad (5.3.9)$$

тобто математичне сподівання кожної компоненти вектора y є однаковим. Коваріаційну матрицю вектора (5.3.8) запишемо у вигляді

$$E[(y - \mu)(y - \mu)'] = \Omega. \quad (5.3.10)$$

Нагадаємо, що елементи ω_{ij} матриці Ω відповідно мають вигляд

$$E[(y_i - \mu)(y_j - \mu)] = \sigma^2 \frac{\phi^{|j-i|}}{1 - \phi^2}. \quad (5.3.11)$$

Отже,

$$\Omega = \sigma^2 V, \quad (5.3.12)$$

де

$$V = \frac{1}{1 - \phi^2} \begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & 1 & \phi & \dots & \phi^{T-2} \\ \phi^2 & \phi & 1 & \dots & \phi^{T-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \phi^{T-1} & \phi^{T-2} & \phi^{T-3} & \dots & 1 \end{pmatrix}. \quad (5.3.13)$$

Розглядаючи змінну y як таку, що має розподіл $N[\mu, \Omega]$, функцію правдоподібності вибірки можна записати, використовуючи формулу багатовимірної Гауссівської щільності

$$L(y; \vartheta) = (2\pi)^{-T/2} |\Omega^{-1}|^{1/2} \exp \left[-\frac{1}{2} (y - \mu)' \Omega^{-1} (y - \mu) \right] \quad (5.3.14)$$

і

$$\begin{aligned} \mathcal{L}(\vartheta) = & -\frac{T}{2} \ln(2\pi) + \frac{1}{2} \ln |\Omega^{-1}| - \\ & - \frac{1}{2} (y - \mu)' \Omega^{-1} (y - \mu). \end{aligned} \quad (5.3.15)$$

Очевидно, формули (5.3.7) і (5.3.15) повинні зображати ту саму функцію правдоподібності. Для того щоб це перевірити,

введемо квадратну матрицю порядку T

$$W = \begin{pmatrix} \sqrt{1-\phi^2} & 0 & 0 & \dots & 0 & 0 \\ -\phi & 1 & 0 & \dots & 0 & 0 \\ 0 & -\phi & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\phi & 1 \end{pmatrix}. \quad (5.3.16)$$

Безпосередньо перевіряємо, що

$$W'W = V^{-1}. \quad (5.3.17)$$

Отже, на підставі (5.3.12)

$$\Omega^{-1} = \sigma^{-2}W'W. \quad (5.3.18)$$

У результаті підстановки (5.3.18) у (5.3.15) одержуємо

$$\begin{aligned} \mathcal{L}(\vartheta) = & -\frac{T}{2} \ln(2\pi) + \frac{1}{2} \ln |\sigma^{-2}W'W| - \\ & - \frac{1}{2} (y - \mu)' \sigma^{-2}W'W (y - \mu). \end{aligned} \quad (5.3.19)$$

Визначимо вектор \tilde{y} так:

$$\tilde{y} = W(y - \mu) = \begin{pmatrix} \sqrt{1-\phi^2}(y_1 - \mu) \\ (y_2 - \mu) - \phi(y_1 - \mu) \\ (y_3 - \mu) - \phi(y_2 - \mu) \\ \vdots \\ (y_T - \mu) - \phi(y_{T-1} - \mu) \end{pmatrix}. \quad (5.3.20)$$

Підставивши у (5.3.20) $\mu = c/(1 - \phi)$, матимемо

$$\tilde{y} = \begin{pmatrix} \sqrt{1-\phi^2}(y_1 - c/(1-\phi)) \\ y_2 - c - \phi y_1 \\ y_3 - c - \phi y_2 \\ \vdots \\ y_T - c - \phi y_{T-1} \end{pmatrix}.$$

Тоді останній доданок формули (5.3.19) можемо записати у вигляді

$$\begin{aligned} \frac{1}{2}(y - \mu)' \sigma^{-2} W' W (y - \mu) &= \frac{1}{2\sigma^2} \tilde{y}' \tilde{y} = \\ &= \frac{1}{2\sigma^2} (1 - \phi^2) [y_1 - c/(1 - \phi)]^2 + \frac{1}{2\sigma^2} \sum_{t=2}^T (y_t - c - \phi y_{t-1})^2. \end{aligned} \quad (5.3.21)$$

Середній доданок у (5.3.19) запишемо у вигляді

$$\frac{1}{2} \ln |\sigma^{-2} W' W| = \frac{1}{2} \ln [\sigma^{-2T} |W' W|] = -\frac{T}{2} \ln \sigma^2 + \ln |W|. \quad (5.3.22)$$

Легко бачити, що $|W| = \sqrt{1 - \phi^2}$. Тому (5.3.22) набере вигляду

$$\frac{1}{2} \ln |\sigma^{-2} W' W| = -\frac{T}{2} \ln \sigma^2 + \frac{1}{2} \ln (1 - \phi^2). \quad (5.3.23)$$

Підставляючи (5.3.21) і (5.3.23) у (5.3.19), одержуємо формулу (5.3.7). Отже, (5.3.15) і (5.3.7) виражають ту саму функцію правдоподібності.

Вираз (5.3.15) вимагає обертання матриці порядку T , тоді як (5.3.7) ні. Тому формула (5.3.7) є зручнішою для обчислення.

Для знаходження $\hat{\vartheta}_{\text{ОМП}}$ треба максимізувати функцію (5.3.7). Для цього можемо використати необхідні умови екстремуму, в результаті чого одержимо систему нелінійних рівнянь, яку можемо розв'язати лише чисельними методами.

Умовні оцінки максимальної правдоподібності

Альтернативним методом до чисельної максимізації логарифма функції правдоподібності (5.3.7) є максимізація умов-

ної правдоподібності вигляду

$$\ln f(y_T, y_{T-1}, \dots, y_2 | y_1; \vartheta) = -\frac{T-1}{2} \ln(2\pi) - \frac{T-1}{2} \ln \sigma^2 - \sum_{t=2}^T \left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^2} \right]. \quad (5.3.24)$$

Максимізація (5.3.24) стосовно c і ϕ еквівалентна мінімізації функції

$$\sum_{t=2}^T (y_t - c - \phi y_{t-1})^2, \quad (5.3.25)$$

що аналогічне до методу найменших квадратів у регресії y_t за сталою і її власним лагом y_{t-1} . Отже, для c і ϕ умовна оцінка максимальної правдоподібності УОМП (CMLE) збігається з МНК оцінкою і має вигляд

$$\begin{pmatrix} \hat{c} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} T-1 & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum y_{t-1} y_t \end{pmatrix},$$

де \sum позначає підсумовування за t від 2 до T .

Для того щоб знайти умовну оцінку максимальної правдоподібності для σ^2 , диференціюємо (5.3.24) стосовно σ^2 і результат прирівнюємо до нуля

$$-\frac{T-1}{2\sigma^2} + \sum_{t=2}^T \left[\frac{(y_t - c - \phi y_{t-1})^2}{2\sigma^4} \right] = 0,$$

звідки

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=2}^T (y_t - \hat{c} - \hat{\phi} y_{t-1})^2.$$

На відміну від оцінок максимальної правдоподібності умовні оцінки максимальної правдоподібності можна легко обчислити. Якщо розмір вибірки T достатньо великий, то перший

доданок робить незначний вклад у загальну правдоподібність. Якщо $|\phi| < 1$, то для великих вибірок оцінки ОМП (MLE) і УОМП (SMLE) мають такий самий розподіл. Для $|\phi| > 1$ оцінки УОМП продовжують забезпечувати змістовність, тоді як максимізація (5.3.7) ні. З цієї причини у більшості застосовань використовують УОМП.

5.3.2 Функція правдоподібності Гауссівського $AR(p)$ процесу

Розглянемо Гауссівський процес

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad (5.3.26)$$

де $\varepsilon_t \sim N[0, \sigma^2]$. У цьому випадку треба оцінити вектор параметрів $\vartheta = [c, \phi_1, \phi_2, \dots, \phi_p, \sigma^2]'$.

Для обчислення функції правдоподібності вибірки розміру T для процесу $AR(p)$ використовують комбінацію двох методів описаних для процесу $AR(1)$. На підставі перших p спостережень (y_1, y_2, \dots, y_p) будуємо $p \times 1$ вектор y^p , який розглядаємо як реалізацію p -вимірної Гауссівської змінної. Математичним сподіванням цього вектора є $p \times 1$ вектор, який позначаємо μ^p і елементи якого мають вигляд

$$\mu = \frac{c}{1 - \phi_1 - \phi_2 - \dots - \phi_p}. \quad (5.3.27)$$

Нехай $\sigma^2 V_p$ позначає $p \times p$ коваріаційну матрицю вектора y^p

$$\begin{aligned} \sigma^2 V_p &= \\ &= \begin{pmatrix} E[(y_1 - \mu)^2] & \dots & E[(y_1 - \mu)(y_p - \mu)] \\ E[(y_2 - \mu)(y_1 - \mu)] & \dots & E[(y_2 - \mu)(y_p - \mu)] \\ \vdots & \dots & \vdots \\ E[(y_p - \mu)(y_1 - \mu)] & \dots & E[(y_p - \mu)^2] \end{pmatrix}. \end{aligned} \quad (5.3.28)$$

Наприклад, для випадку $p = 1$ матриця $V_1 = 1/(1 - \phi^2)$. Для загального випадку

$$\sigma^2 V_p = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & \gamma_0 \end{pmatrix},$$

де γ_j – j -та автоковаріація процесу $AR(p)$. Щільність вектора y^p , який містить перших p спостережень і має розподіл $N[\mu^p, \sigma^2 V_p]$, має вигляд

$$f(y_p, y_{p-1}, \dots, y_1; \vartheta) = (2\pi)^{-p/2} (\sigma^{-2})^{p/2} |V_p^{-1}|^{1/2} \times \\ \times \exp \left[-\frac{1}{2\sigma^2} (y^p - \mu^p)' V_p^{-1} (y^p - \mu^p) \right]. \quad (5.3.29)$$

Умовне за $t - 1$ спостереженнями t -те спостереження є Гауссівським з математичним сподіванням

$$c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p}$$

і дисперсією σ^2 . Тільки p попередніх спостережень мають значення для його розподілу. Отже, для $t > p$

$$f(y_t | y_{t-1}, y_{t-2}, \dots, y_1; \vartheta) = f(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}; \vartheta) = \\ = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p})^2}{2\sigma^2} \right].$$

Тоді функція правдоподібності для цілої вибірки має вигляд

$$L(y_T, y_{T-1}, \dots, y_1; \vartheta) = f(y_p, y_{p-1}, \dots, y_1; \vartheta) \times \\ \times \prod_{t=p+1}^T f(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-p}; \vartheta), \quad (5.3.30)$$

а логарифм функції правдоподібності

$$\begin{aligned}
\mathcal{L}(\vartheta) &= -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln(\sigma^2) + \frac{1}{2} \ln |V_p^{-1}| - \\
&\quad - \frac{1}{2\sigma^2} (y^p - \mu^p)' V_p^{-1} (y^p - \mu^p) - \frac{T-p}{2} \ln(2\pi) - \frac{T-p}{2} \ln(\sigma^2) - \\
&\quad - \sum_{t=p+1}^T \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2} = \\
&= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) + \frac{1}{2} \ln |V_p^{-1}| - \\
&\quad - \frac{1}{2\sigma^2} (y^p - \mu^p)' V_p^{-1} (y^p - \mu^p) - \\
&\quad - \sum_{t=p+1}^T \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2}. \tag{5.3.31}
\end{aligned}$$

Для обчислення (5.3.31) треба знайти обернену матрицю для V_p . Позначимо елемент з номером i, j матриці V_p^{-1} через $\nu^{ij}(p)$. Правильна рівність

$$\nu^{ij}(p) = \sum_{k=0}^{i-1} \phi_k \phi_{k+j-i} - \sum_{k=p+1-j}^{p+i-j} \phi_k \phi_{k+j-i} \tag{5.3.32}$$

для $1 \leq i \leq j \leq p$, де $\phi_0 = -1$. Значення $\nu^{ij}(p)$ для $i > j$ можемо знайти з того, що матриця V_p^{-1} симетрична.

Наприклад, для процесу $AR(1)$ $V_1^{-1} = v^{11} = 1 - \phi^2$. Отже, $\sigma^2 V_1 = \sigma^2 / (1 - \phi^2)$, тобто приходимо до відомої вже формули.

Для $p = 2$ з рівняння (5.3.32) одержимо

$$V_2^{-1} = \begin{bmatrix} (1 - \phi_2^2) & -(\phi_1 + \phi_1 \phi_2) \\ -(\phi_1 + \phi_1 \phi_2) & (1 - \phi_2^2) \end{bmatrix},$$

звідки матимемо

$$\begin{aligned}
|V_2^{-1}| &= \left| (1 + \phi_2) \begin{pmatrix} 1 - \phi_2 & -\phi_1 \\ -\phi_1 & 1 - \phi_2 \end{pmatrix} \right| = \\
&= (1 + \phi_2)^2 [(1 - \phi_2)^2 - \phi_1^2]
\end{aligned}$$

i

$$\begin{aligned}
(y^2 - \mu^2)' V_2^{-1} (y^2 - \mu^2) &= ((y_1 - \mu) \ (y_2 - \mu)) (1 + \phi_2) \times \\
&\times \begin{pmatrix} 1 - \phi_2 & -\phi_1 \\ -\phi_1 & 1 - \phi_2 \end{pmatrix} \begin{pmatrix} y_1 - \mu \\ y_2 - \mu \end{pmatrix} = \\
&= (1 + \phi_2) [(1 - \phi_2)(y_1 - \mu)^2 - 2\phi_1(y_1 - \mu)(y_2 - \mu) + \\
&\quad + (1 - \phi_2)(y_2 - \mu)^2].
\end{aligned}$$

Тоді логарифм функції правдоподібності для Гауссівського $AR(2)$ процесу можемо записати у вигляді

$$\begin{aligned}
\mathcal{L}(\vartheta) &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) + \\
&\quad + \frac{1}{2} \ln \{(1 + \phi_2)^2 [(1 - \phi_2)^2 - \phi_1^2]\} - \\
&\quad - \frac{1 + \phi_2}{2\sigma^2} [(1 - \phi_2)(y_1 - \mu)^2 - \\
&\quad - 2\phi_1(y_1 - \mu)(y_2 - \mu) + (1 - \phi_2)(y_2 - \mu)^2] - \\
&\quad - \sum_{t=3}^T \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2})^2}{2\sigma^2}, \tag{5.3.33}
\end{aligned}$$

де $\mu = c / (1 - \phi_1 - \phi_2)$.

Оцінки максимальної правдоподібності коефіцієнтів $AR(p)$ процесу є точками максимуму логарифма функції правдоподібності (5.3.31). Їхнє чисельне обчислення досить важке. Тому розглядають умовні ОМП (СMLE), використовуючи функцію

$$\begin{aligned}
\ln f(y_T, y_{T-1}, \dots, y_{p+1} | y_p, \dots, y_1; \vartheta) &= \\
&= -\frac{T-p}{2} \ln(2\pi) - \frac{T-p}{2} \ln(\sigma^2) - \\
&\quad - \sum_{t=p+1}^T \frac{(y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2}{2\sigma^2}. \tag{5.3.34}
\end{aligned}$$

Значення параметрів $c, \phi_1, \phi_2, \dots, \phi_p$, які максимізують функцію (5.3.34), такі самі, що мінімізують функцію

$$\sum_{t=p+1}^T (y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2. \quad (5.3.35)$$

Отже, умовні оцінки правдоподібності УОМП цих параметрів збігаються з оцінками, які отримані методом найменших квадратів для регресії y_t стосовно сталої і p її власних лагових значень. Тоді УОМП (СМЛЕ) для σ^2 можемо записати у вигляді

$$\hat{\sigma}^2 = \frac{1}{T-p} \sum_{t=p+1}^T (y_t - \hat{c} - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2} - \dots - \hat{\phi}_p y_{t-p})^2.$$

Зауваження. Якщо процес не є Гауссівським і використано для знаходження оцінок параметрів моделі (5.3.26) логарифм Гауссівської функції правдоподібності, то знайдені оцінки будуть змістовними оцінками параметрів моделі. Такі оцінки називають оцінками квазімаксимальної правдоподібності (QMLE). Проте стандартні похибки оцінених параметрів, які обчислено на підставі припущення про те, що процес є Гауссівський, для негауссівських даних некоректні.

Іноді у випадку негауссівських даних просте перетворення, наприклад логарифмування, призводить до Гауссівських часових рядів. Для додатних випадкових змінних Бокс і Кокс (Box, Cox) запропонували загальний клас перетворень

$$y_t^{(\lambda)} = \begin{cases} \frac{y_t^\lambda - 1}{\lambda}, & \lambda \neq 0, \\ \ln y_t, & \lambda = 0. \end{cases}$$

Одним з підходів є вибір відповідного значення λ і максимізація функції правдоподібності для $y_t^{(\lambda)}$ у припущенні, що $y_t^{(\lambda)}$ є Гауссівським *ARMA* процесом.

5.3.3 Функція правдоподібності Гауссівського $MA(1)$ процесу

Розглянемо Гауссівський $MA(1)$ процес

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1}, \quad (5.3.36)$$

де $\varepsilon_t \sim N[0, \sigma^2]$. Нехай $\vartheta = [\mu, \theta, \sigma^2]'$ позначає вектор параметрів, які треба оцінити.

Умовні оцінки максимальної правдоподібності

Якщо значення ε_{t-1} відоме, то

$$y_t | \varepsilon_{t-1} \sim N[\mu + \theta\varepsilon_{t-1}, \sigma^2]$$

і

$$f(y_t | \varepsilon_{t-1}; \vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_t - \mu - \theta\varepsilon_{t-1})^2}{2\sigma^2}\right]. \quad (5.3.37)$$

Припустимо, що $\varepsilon_0 = 0$. Тоді $(y_1 | \varepsilon_0 = 0) \sim N[\mu, \sigma^2]$. Маючи спостереження y_1 , знайдемо $\varepsilon_1 = y_1 - \mu$ і, застосовуючи (5.3.37), одержимо

$$f(y_2 | y_1, \varepsilon_0 = 0; \vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_2 - \mu - \theta\varepsilon_1)^2}{2\sigma^2}\right].$$

Оскільки ε_1 відоме, то можемо записати ε_2 у вигляді

$$\varepsilon_2 = y_2 - \mu - \theta\varepsilon_1.$$

Отож, якщо відомо, що $\varepsilon_0 = 0$, то можемо обчислити всю множину $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T\}$ за допомогою рекурсії

$$\varepsilon_t = y_t - \mu - \theta\varepsilon_{t-1} \quad (5.3.38)$$

для $t = 1, \dots, T$. На підставі (5.3.37) обчислюємо умовну щільність t -го спостереження

$$\begin{aligned} f(y_t | y_{t-1}, \dots, y_1, \varepsilon_0 = 0; \vartheta) &= f(y_t | \varepsilon_{t-1}; \vartheta) = \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\varepsilon_t^2}{2\sigma^2}\right]. \end{aligned}$$

Функція правдоподібності всієї вибірки матиме вигляд

$$\begin{aligned} L(y_T, y_{T-1}, \dots, y_1 | \varepsilon_0 = 0; \vartheta) &= f(y_1 | \varepsilon_0 = 0; \vartheta) \times \\ &\times \prod_{t=2}^T f(y_t | y_{t-1}, \dots, y_1, \varepsilon_0 = 0; \vartheta), \quad (5.3.39) \end{aligned}$$

а її логарифм –

$$\begin{aligned} \mathcal{L}(\vartheta) &= \ln L(y_T, y_{T-1}, \dots, y_1 | \varepsilon_0 = 0; \vartheta) = \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln \sigma^2 - \sum_{t=1}^T \frac{\varepsilon_t^2}{2\sigma^2}. \quad (5.3.40) \end{aligned}$$

Для заданого значення ϑ можемо обчислити послідовність ε_t на підставі (5.3.38). Тоді логарифм умовної функції правдоподібності (5.3.40) буде сумою їхніх квадратів. Проте як функція змінних μ і θ вона є нелінійною і знаходити точки максимуму цієї функції нелегко. Тому УОМП (СМЛЕ) параметрів процесу $MA(1)$ знаходять за допомогою чисельної оптимізації.

Ітеруючи (5.3.38) для довільного заданого значення ε_0 , одержимо

$$\begin{aligned} \varepsilon_t &= (y_t - \mu) - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu) + \\ &+ \dots + (-1)^{t-1} \theta^{t-1} (y_1 - \mu) + (-1)^t \theta^t \varepsilon_0. \end{aligned}$$

Якщо $|\theta|$ суттєво менше від одиниці, то для достатньо великої вибірки формула (5.3.39) дає добру апроксимацію. У випадку $|\theta| > 1$ обчислення оцінок параметрів на підставі (5.3.39)

не має сенсу. Якщо в результаті чисельної оптимізації (5.3.40) $|\hat{\theta}| > 1$, то обчислення треба повторити, використовуючи обернене значення до $\hat{\theta}$ як початкове наближення.

Точні оцінки максимальної правдоподібності

Для обчислення точної функції правдоподібності Гауссівського процесу $MA(1)$ використовують такий підхід. Нехай

$$\Omega = E[(y - \mu)(y - \mu)'],$$

де $y = (y_1, \dots, y_T)'$, $\mu = (\mu, \dots, \mu)'$. Тоді

$$\Omega = \sigma^2 \begin{pmatrix} (1 + \theta^2) & \theta & 0 & \dots & 0 \\ \theta & (1 + \theta^2) & \theta & \dots & 0 \\ 0 & \theta & (1 + \theta^2) & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & (1 + \theta^2) \end{pmatrix}$$

і функція правдоподібності має вигляд

$$L(y; \vartheta) = (2\pi)^{-T/2} |\Omega|^{-1/2} \exp \left[-\frac{1}{2} (y - \mu)' \Omega^{-1} (y - \mu) \right]. \quad (5.3.41)$$

Згідно з (3.7.20) матрицю Ω можна подати у вигляді

$$\Omega = ADA', \quad (5.3.42)$$

де A – нижня трикутна матриця з одиницями на головній діагоналі

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \frac{\theta}{1 + \theta^2} & 1 & \dots & 0 & 0 \\ 0 & \frac{\theta(1 + \theta^2)}{1 + \theta^2 + \theta^4} & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & \frac{\theta(1 + \theta^2 + \theta^4 + \dots + \theta^{2(T-2)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2(T-1)}} & 1 \end{pmatrix},$$

а D – діагональна матриця

$$D = \sigma^2 \begin{pmatrix} 1 + \theta^2 & 0 & \dots & 0 \\ 0 & \frac{1 + \theta^2 + \theta^4}{1 + \theta^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1 + \theta^2 + \theta^4 + \dots + \theta^{2T}}{1 + \theta^2 + \theta^4 + \dots + \theta^{2(T-1)}} \end{pmatrix}.$$

Підставляючи (5.3.42) у (5.3.41), одержимо

$$L(y; \vartheta) = (2\pi)^{-T/2} |ADA'|^{-1/2} \times \\ \times \exp \left[-\frac{1}{2} (y - \mu)' (A')^{-1} D^{-1} A^{-1} (y - \mu) \right]. \quad (5.3.43)$$

Оскільки A – нижня трикутна матриця з одиницями на головній діагоналі, то $|A| = 1$ і

$$|ADA'| = |A||D||A'| = |D|.$$

Введемо вектор

$$\tilde{y} = A^{-1}(y - \mu). \quad (5.3.44)$$

Тоді функцію (5.3.43) можна записати у вигляді

$$L(y; \vartheta) = (2\pi)^{-T/2} |D|^{-1/2} \exp \left[-\frac{1}{2} \tilde{y}' D^{-1} \tilde{y} \right]. \quad (5.3.45)$$

Зазначимо, що з (5.3.44) випливає рівність

$$A\tilde{y} = y - \mu.$$

Перший рядок цієї системи має вигляд $\tilde{y}_1 = y_1 - \mu$, тоді як t -ий рядок є

$$\tilde{y}_t = y_t - \mu - \frac{\theta(1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-2)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)}} \tilde{y}_{t-1}. \quad (5.3.46)$$

Вектор \tilde{y} можна обчислити, ітеруючи (5.3.46) за $t = 2, 3, \dots, T$. Елемент d_{tt} матриці D має вигляд

$$d_{tt} = E[\tilde{y}_t^2] = \sigma^2 \frac{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}}{1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)}}. \quad (5.3.47)$$

Оскільки D діагональна, то

$$|D| = \prod_{t=1}^T d_{tt} = (\sigma^2)^T (1 + \theta^2 + \theta^4 + \dots + \theta^{2T}), \quad (5.3.48)$$

і обернена матриця матиме на головній діагоналі елементи, обернені до відповідних елементів матриці D . Отже,

$$\tilde{y}' D^{-1} \tilde{y} = \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}. \quad (5.3.49)$$

Підставляючи (5.3.48) і (5.3.49) у (5.3.45), одержимо

$$L(y; \vartheta) = (2\pi)^{-T/2} \left[\prod_{t=1}^T d_{tt} \right]^{-1/2} \exp \left[-\frac{1}{2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}} \right]. \quad (5.3.50)$$

Отож,

$$\begin{aligned} \mathcal{L}(\vartheta) &= \ln L(y; \vartheta) = & (5.3.51) \\ &= -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln d_{tt} - \frac{1}{2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}} = \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \frac{1}{2} \ln(1 + \theta^2 + \theta^4 + \dots + \theta^{2T}) - \\ &\quad - \frac{1}{2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}. \end{aligned}$$

Маючи чисельні значення для μ , θ і σ^2 послідовність \tilde{y}_t можна обчислити ітеруюванням (5.3.46).

Зауважимо, що значення функції (5.3.51) при $\theta = \tilde{\theta}$, $\sigma^2 = \tilde{\sigma}^2$ буде тотожним до його значення при $\theta = \tilde{\theta}^{-1}$, $\sigma^2 = \tilde{\theta}^2 \tilde{\sigma}^2$.

5.3.4 Функція правдоподібності Гауссівського $MA(q)$ процесу

Умовні ОМП

Для процесу $MA(q)$

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (5.3.52)$$

простим підходом є прийняття умови

$$\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-(q-1)} = 0. \quad (5.3.53)$$

Маючи ці початкові значення, за допомогою ітерації

$$\varepsilon_t = y_t - \mu - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (5.3.54)$$

можемо знайти ε_t для $t = 1, 2, \dots, T$. Позначимо через e_0 вектор $(\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(q-1)})'$. Тоді логарифм умовної функції правдоподібності має вигляд

$$\begin{aligned} \mathcal{L}(\vartheta) &= \ln L(y_T, y_{T-1}, \dots, y_1 | e_0; \vartheta) = \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \sum_{t=1}^T \frac{\varepsilon_t^2}{2\sigma^2}, \end{aligned} \quad (5.3.55)$$

де $\vartheta = (\mu, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)'$. Вираз (5.3.55) можна використувати, якщо корені рівняння

$$1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q = 0$$

лежать поза одиничним кругом, тобто $MA(q)$ процес є зворотним.

Точні ОМП

Точна функція правдоподібності має вигляд

$$L(y; \vartheta) = (2\pi)^{-T/2} |\Omega|^{-1/2} \exp \left[-\frac{1}{2} (y - \mu)' \Omega^{-1} (y - \mu) \right], \quad (5.3.56)$$

де як і раніше $y = (y_1, \dots, y_T)'$, $\mu = (\mu, \dots, \mu)'$,

$$\Omega = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \dots & \gamma_q & 0 & \dots & 0 & 0 & 0 \\ \gamma_1 & \gamma_0 & \gamma_1 & \dots & \gamma_{q-1} & \gamma_q & \dots & 0 & 0 & 0 \\ \gamma_2 & \gamma_1 & \gamma_0 & \dots & \gamma_{q-2} & \gamma_{q-1} & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & \gamma_0 & \gamma_1 & \gamma_2 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & \gamma_1 & \gamma_0 & \gamma_1 \\ 0 & 0 & 0 & \dots & 0 & 0 & \dots & \gamma_2 & \gamma_1 & \gamma_0 \end{pmatrix}. \quad (5.3.57)$$

Тут γ_k є k -тою автоковаріацією процесу $MA(q)$

$$\gamma_k = \begin{cases} \sigma^2(\theta_k + \theta_{k+1}\theta_1 + \theta_{k+2}\theta_2 + \dots + \theta_q\theta_{q-k}), & k = 0, 1, \dots, q, \\ 0, & k > q, \end{cases} \quad (5.3.58)$$

де $\theta_0 = 1$.

Функцію правдоподібності (5.3.56) можна обчислити, використовуючи трикутну факторизацію матриці Ω

$$\Omega = ADA', \quad (5.3.59)$$

де A – нижня трикутна матриця з одиницями на головній діагоналі вигляду (3.7.11); D – діагональна матриця вигляду

(3.7.7). Зокрема, для $MA(q)$ процесу

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ a_{21} & 1 & 0 & \dots & 0 & 0 \\ a_{31} & a_{32} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ a_{q+1,1} & a_{q+1,2} & a_{q+1,3} & \dots & 0 & 0 \\ 0 & a_{q+2,2} & a_{q+2,3} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{T,T-1} & 1 \end{pmatrix}.$$

Підставляючи (5.3.59) у (5.3.56), одержимо функцію правдоподібності у вигляді

$$L(y; \vartheta) = (2\pi)^{-T/2} |D|^{-1/2} \exp \left[-\frac{1}{2} \tilde{y}' D^{-1} \tilde{y} \right], \quad (5.3.60)$$

де

$$A\tilde{y} = y - \mu. \quad (5.3.60)$$

Елементи вектора \tilde{y} можна обчислити рекурсивно

$$\begin{aligned} \tilde{y}_1 &= y_1 - \mu, \\ \tilde{y}_2 &= (y_2 - \mu) - a_{21}\tilde{y}_1, \\ \tilde{y}_3 &= (y_3 - \mu) - a_{32}\tilde{y}_2 - a_{31}\tilde{y}_1, \\ &\vdots \\ \tilde{y}_t &= (y_t - \mu) - a_{t,t-1}\tilde{y}_{t-1} - a_{t,t-2}\tilde{y}_{t-2} - \dots - a_{t,t-q}\tilde{y}_{t-q}. \end{aligned}$$

Отже,

$$\begin{aligned} \mathcal{L}(\vartheta) &= \ln L(y; \vartheta) = & (5.3.61) \\ &= -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln d_{tt} - \frac{1}{2} \sum_{t=1}^T \frac{\tilde{y}_t^2}{d_{tt}}. \end{aligned}$$

5.3.5 Функція правдоподібності для Гауссівського ARMA(p, q) процесу

Гауссівський процес ARMA(p, q) має вигляд

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (5.3.62)$$

де $\varepsilon_t \sim N[0, \sigma^2]$. Треба оцінити вектор параметрів $\vartheta = (c, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)'$.

Апроксимацію функції правдоподібності для авторегресії базували на початкових значеннях y . Для апроксимації функції правдоподібності процесу рухомого середнього використано початкові значення ε . Тому апроксимацію функції правдоподібності ґрунтують на початкових значеннях y і ε .

Вибравши початкові значення $y^0 = (y_0, y_{-1}, \dots, y_{-(p-1)})'$, $e_0 = (\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{-(q-1)})'$, послідовність $\{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T\}$ можна обчислити за допомогою ітерацій

$$\varepsilon_t = y_t - c - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}. \quad (5.3.63)$$

Тоді логарифм умовної функції правдоподібності матиме вигляд

$$\begin{aligned} \mathcal{L}(\vartheta) &= \ln L(y_T, y_{T-1}, \dots, y_1 | y^0, e_0; \vartheta) = \\ &= -\frac{T}{2} \ln(2\pi) - \frac{T}{2} \ln(\sigma^2) - \sum_{t=1}^T \frac{\varepsilon_t^2}{2\sigma^2}. \end{aligned} \quad (5.3.64)$$

Один зі способів вибору початкових значень для y і ε є вибір їхніх очікуваних значень, тобто $y_s = c/(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ для $s = 0, -1, \dots, -(p-1)$ і $\varepsilon_s = 0$ для $s = 0, -1, \dots, -(q-1)$.

Альтернативним є такий підхід. Ітерації (5.3.63) починають з $t = p + 1$, використовуючи відомі дані y_1, y_2, \dots, y_p , і

$\varepsilon_p = \varepsilon_{p-1} = \dots = \varepsilon_{p-q+1} = 0$. Тоді

$$\begin{aligned} \mathcal{L}(\vartheta) &= \ln L(y_T, \dots, y_{p+1} | y_p, \dots, y_1, \varepsilon_p = 0, \dots, \varepsilon_{p-q+1} = 0; \vartheta) = \\ &= -\frac{T-p}{2} \ln(2\pi) - \frac{T-p}{2} \ln(\sigma^2) - \sum_{t=p+1}^T \frac{\varepsilon_t^2}{2\sigma^2}. \end{aligned}$$

Зазначимо, що ці апроксимації можливі, якщо корені рівняння

$$1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q = 0$$

лежать поза одиничним кругом.

5.4 Тестування гіпотез на підставі ОМП

5.4.1 Асимптотичні стандартні похибки оцінок максимальної правдоподібності

Якщо розмір T вибірки достатньо великий, то розподіл оцінки максимальної правдоподібності $\hat{\vartheta}$ можна апроксимувати таким розподілом

$$\hat{\vartheta} \sim N[\vartheta_0, T^{-1} \mathcal{I}^{-1}], \quad (5.4.1)$$

де ϑ_0 позначає правильне значення вектора параметрів. Матриця \mathcal{I} відома як інформаційна матриця і може бути оцінена різними способами.

Перша оцінка інформаційної матриці, яка ґрунтується на другій похідній логарифма функції правдоподібності, має вигляд

$$\hat{\mathcal{I}}_{2D} = -T^{-1} \frac{\partial^2 \mathcal{L}(\vartheta)}{\partial \vartheta \partial \vartheta'} \Big|_{\vartheta = \hat{\vartheta}}. \quad (5.4.2)$$

Тут \mathcal{L} позначає логарифм функції правдоподібності

$$\mathcal{L}(\vartheta) = \sum_{t=1}^T \ln L(y_t | \mathcal{Y}_{t-1}; \vartheta)$$

і \mathcal{Y}_t позначає історію спостережень y до періоду t . Матрицю других похідних логарифма функції правдоподібності часто обчислюють чисельно. Підставляючи (5.4.2) у (5.4.1), скорочуємо T . Тоді коваріаційну матрицю $\hat{\vartheta}$ апроксимують виразом

$$E[(\hat{\vartheta} - \vartheta_0)(\hat{\vartheta} - \vartheta_0)'] = \left[- \frac{\partial^2 \mathcal{L}(\vartheta)}{\partial \vartheta \partial \vartheta'} \Big|_{\vartheta = \hat{\vartheta}} \right]^{-1}. \quad (5.4.3)$$

Другою оцінкою інформаційної матриці \mathcal{I} (outer-product estimate) є

$$\hat{\mathcal{I}}_{op} = T^{-1} \sum_{t=1}^T [h(\hat{\vartheta}, \mathcal{Y}_t)] \cdot [h(\hat{\vartheta}, \mathcal{Y}_t)]', \quad (5.4.4)$$

де

$$h(\hat{\vartheta}, \mathcal{Y}_t) = \frac{\partial \ln L(y_t | y_{t-1}, y_{t-2}, \dots; \vartheta)}{\partial \vartheta} \Big|_{\vartheta = \hat{\vartheta}}.$$

У цьому випадку коваріаційну матрицю $\hat{\vartheta}$ апроксимують виразом

$$E[(\hat{\vartheta} - \vartheta_0)(\hat{\vartheta} - \vartheta_0)'] = \left[\sum_{t=1}^T [h(\hat{\vartheta}, \mathcal{Y}_t)] \cdot [h(\hat{\vartheta}, \mathcal{Y}_t)]' \right]^{-1}. \quad (5.4.5)$$

5.4.2 Тест відношення правдоподібності

Іншим популярним підходом тестування гіпотез стосовно параметрів, оцінених за допомогою методу максимальної правдоподібності, є використання тесту відношення правдоподібності (LR - тест). Припустимо, що нульова гіпотеза охоплює

m різних обмежень на значення $a \times 1$ параметрів вектора ϑ . Спочатку максимізуємо функцію правдоподібності, ігноруючи ці обмеження, щоб одержати ОМП-оцінку $\widehat{\vartheta}$. Потім знаходимо таку оцінку $\widetilde{\vartheta}$, щоб функція правдоподібності приймала найбільші значення і виконувалися обмеження. Практично це можна досягти, ввівши новий $(a - m) \times 1$ вектор λ , у термінах якого можна виразити всі елементи ϑ , якщо виконуються обмеження. Наприклад, якщо тестуємо гіпотезу, що останні m елементів вектора ϑ дорівнюють нулю, то вектор λ складається з перших $a - m$ елементів вектора ϑ . Нехай $\mathcal{L}(\widehat{\vartheta})$ позначає значення логарифма функції правдоподібності при відсутності обмежень і $\mathcal{L}(\widetilde{\vartheta})$ позначає значення логарифма функції правдоподібності при наявності обмежень. Очевидно, $\mathcal{L}(\widehat{\vartheta}) > \mathcal{L}(\widetilde{\vartheta})$ і за певних умов можна довести, що

$$LR = 2 [\mathcal{L}(\widehat{\vartheta}) - \mathcal{L}(\widetilde{\vartheta})] \sim \chi^2[m]. \quad (5.4.6)$$

5.4.3 Тест множників Лагранжа

Для того щоб використати стандартні похибки з (5.4.2) або (5.4.4) для тестування гіпотез щодо ϑ , треба знайти ОМП-оцінку $\widehat{\vartheta}$ без обмежень. Щоб використати тест відношення правдоподібності (5.4.6), необхідно обчислити ОМП-оцінку $\widehat{\vartheta}$ без обмежень і ОМП-оцінку з обмеженнями $\widetilde{\vartheta}$. Тест множників Лагранжа забезпечує тестування нульової гіпотези, використовуючи тільки ОМП-оцінку з обмеженнями $\widetilde{\vartheta}$. Він корисний, якщо оцінку $\widetilde{\vartheta}$ легше обчислити, ніж оцінку $\widehat{\vartheta}$.

Нехай ϑ – оцінка ϑ , яка максимізує логарифм функції правдоподібності при наявності m обмежень на ϑ

$$h(\widetilde{\vartheta}, \mathcal{Y}_t) = \left. \frac{\partial \ln L(y_t | y_{t-1}, y_{t-2}, \dots; \vartheta)}{\partial \vartheta} \right|_{\vartheta = \widetilde{\vartheta}}.$$

Тест множників Лагранжа нульової гіпотези, яка полягає у

виконанні обмежень, використовує статистику

$$LM = T^{-1} \left[\sum_{t=1}^T h(\tilde{\vartheta}, \mathcal{Y}_t) \right]' \mathcal{I}^{-1} \left[\sum_{t=1}^T h(\tilde{\vartheta}, \mathcal{Y}_t) \right]. \quad (5.4.7)$$

Якщо нульова гіпотеза правильна, то для великих T статистика LM асимптотично має розподіл $\chi^2[m]$. Інформаційна матриця знову може бути оцінена як у (5.4.2) або (5.4.4), де $\hat{\vartheta}$ замінено на $\tilde{\vartheta}$.

Розділ 6

Прогнозування

6.1 Принципи прогнозування

6.1.1 Прогнози, які ґрунтуються на умовних математичних сподіваннях

Розглянемо прогнозування величини y_{t+1} , що ґрунтується на відомій у момент часу t множині значень X_t . Зокрема, якщо ми хочемо передбачити y_{t+1} , ґрунтуючись на m попередніх значеннях y_t , то у цьому випадку множина X_t буде складатися з $y_t, y_{t-1}, \dots, y_{t-m+1}$. Нехай $y_{t+h|t}$ позначає прогнозоване значення y_t на h періодів вперед, яке ґрунтується на X_t .

Щоб оцінити ефективність прогнозу, треба визначити наскільки наш прогноз близький до реального значення. Мірою такої близькості може бути квадратична функція втрат.

Означення 6.1.1. *Квадратичною функцією втрат називається математичне сподівання квадрата відхилення істинного значення від прогнозованого*

$$E [(y_{t+h} - y_{t+h|t})^2] . \quad (6.1.1)$$

Означення 6.1.2. Точка $y_{t+h|t}^*$, в якій досягається мінімум квадратичної функції втрат, заданої виразом (6.1.1), називається оптимальним MSE - прогнозом, а мінімум квадратичної функції втрат називається середньоквадратичною похибкою прогнозу і позначається

$$MSE(y_{t+h|t}^*) \equiv E [(y_{t+h} - y_{t+h|t}^*)^2] .$$

Лема 6.1.1. Оптимальний прогноз, тобто прогноз з найменшою середньоквадратичною похибкою дорівнює умовному математичному сподіванню y_{t+h} при заданому X_t

$$y_{t+h|t}^* = E[y_{t+h}|X_t] . \quad (6.1.2)$$

Д о в е д е н н я . Нехай $y_{t+h|t}^*$ – це деяка функція $g(X_t)$, яка не є умовним математичним сподіванням

$$y_{t+h|t}^* = g(X_t) . \quad (6.1.3)$$

Тоді середньоквадратична похибка дорівнює

$$\begin{aligned} MSE &= E [(y_{t+h} - g(X_t))^2] = & (6.1.4) \\ &= E [(y_{t+h} - E[y_{t+h}|X_t] + E[y_{t+h}|X_t] - g(X_t))^2] = \\ &= E [(y_{t+h} - E[y_{t+h}|X_t])^2] + \\ &\quad + 2E[(y_{t+h} - E[y_{t+h}|X_t])(E[y_{t+h}|X_t] - g(X_t))] + \\ &\quad + E [(E[y_{t+h}|X_t] - g(X_t))^2] . \end{aligned}$$

Запишемо середній доданок правої частини (6.1.4) так:

$$2E[\eta_{t+h}] , \quad (6.1.5)$$

де

$$\eta_{t+h} \equiv (y_{t+h} - E[y_{t+h}|X_t])(E[y_{t+h}|X_t] - g(X_t)) .$$

Розглянемо спочатку умовне математичне сподівання η_{t+h} , задане на X_t . Зауважимо, що умовне математичне сподівання $E[y_{t+h}|X_t]$ зображає умовний момент випадкової змінної y_{t+h} і само не є функцією випадкової змінної y_{t+h} . Тому множник $E[y_{t+h}|X_t] - g(X_t)$ є сталим і може бути винесений за математичне сподівання

$$\begin{aligned} E[\eta_{t+h}|X_t] &= (E[y_{t+h}|X_t] - g(X_t)) \times E[(y_{t+h} - E[y_{t+h}|X_t])|X_t] = \\ &= (E[y_{t+h}|X_t] - g(X_t)) \times 0 = 0. \end{aligned}$$

Звідси випливає, що

$$E[\eta_{t+h}] = E_{X_t} [E[\eta_{t+h}|X_t]] = 0.$$

Підстановка в (6.1.4) дає

$$\begin{aligned} E[(y_{t+h} - g(X_t))^2] &= E[(y_{t+h} - E[y_{t+h}|X_t])^2] + \\ &+ E[(E[y_{t+h}|X_t] - g(X_t))^2]. \end{aligned} \quad (6.1.6)$$

Другий доданок правої частини (6.1.6) не може бути меншим за нуль

$$E[(E[y_{t+h}|X_t] - g(X_t))^2] \geq 0.$$

Тому

$$E[(y_{t+h} - g(X_t))^2] \geq E[(y_{t+h} - E[y_{t+h}|X_t])^2].$$

Оскільки перший доданок в (6.1.6) не залежить від $g(X_t)$, то функція $g(X_t)$, яка забезпечує мінімум середньо-квадратичної похибки (6.1.6) є функцією, що обертає другий доданок (6.1.6) в нуль, тобто

$$E[y_{t+h}|X_t] = g(X_t). \quad (6.1.7)$$

Отож, прогноз $g(X_t)$, що мінімізує середньо-квадратичну похибку, є умовним математичним сподіванням

$E[y_{t+h}|X_t]$, як і стверджувалося. \square

Середньоквадратична похибка оптимального прогнозу дорівнює

$$\begin{aligned} MSE &= E [(y_{t+h} - g(X_t))^2] = \\ &= E [(y_{t+h} - E[y_{t+h}|X_t])^2] . \end{aligned} \quad (6.1.8)$$

6.1.2 Прогнози, які ґрунтуються на лінійній проекції

Обмежимо клас прогнозів припущенням, що $y_{t+h|t}^*$ є лінійною функцією від X_t

$$y_{t+h|t}^* = \alpha' X_t . \quad (6.1.9)$$

Означення 6.1.3. Якщо існує таке значення α , для якого похибка прогнозування $(y_{t+h} - \alpha' X_t)$ не корелює з X_t , тобто

$$E [(y_{t+h} - \alpha' X_t) X_t'] = 0' , \quad (6.1.10)$$

то прогноз $\alpha' X_t$ називається лінійною проекцією y_{t+h} на X_t .

Лема 6.1.2. Лінійна проекція має найменшу середньоквадратичну похибку в класі лінійних прогнозів.

Іншими словами, лінійна проекція – це оптимальний лінійний прогноз.

Д о в е д е н н я . Нехай $g' X_t$ – деякий довільний лінійний прогноз. Зазначимо, що його MSE дорівнює

$$\begin{aligned} E [(y_{t+h} - g' X_t)^2] &= E [(y_{t+h} - \alpha' X_t + \alpha' X_t - g' X_t)^2] = \\ &= E [(y_{t+h} - \alpha' X_t)^2] + 2E [(y_{t+h} - \alpha' X_t)(\alpha' X_t - g' X_t)] \\ &\quad + E [(\alpha' X_t - g' X_t)^2] . \end{aligned} \quad (6.1.11)$$

Як і в формулі (6.1.4), згідно з (6.1.10), середній член правої частини (6.1.11) дорівнює нулю

$$\begin{aligned} E [(y_{t+h} - \alpha' X_t)(\alpha' X_t - g' X_t)] &= \\ &= E [(y_{t+h} - \alpha' X_t)X_t'] (\alpha - g) = 0'(\alpha - g) = 0. \end{aligned}$$

Отож, (6.1.11) запишеться у вигляді

$$\begin{aligned} E [(y_{t+h} - g' X_t)^2] &= \qquad \qquad \qquad (6.1.12) \\ &= E [(y_{t+h} - \alpha' X_t)^2] + E [(\alpha' X_t - g' X_t)^2]. \end{aligned}$$

Оптимальний лінійний прогноз $g' X_t$ є таким, що перетворює другий доданок у нуль, тобто

$$g' X_t = \alpha' X_t,$$

де $\alpha' X_t$ задовольняє (6.1.10). \square

Для α' , що задовольняє (6.1.10), оптимальний лінійний прогноз, який є лінійною проекцією, позначатимемо

$$\hat{P}(y_{t+h}|X_t) = \alpha' X_t.$$

Зазначимо, що

$$MSE \left(\hat{P}(y_{t+h}|X_t) \right) \geq MSE (E[y_{t+h}|X_t]),$$

оскільки умовне математичне сподівання пропонує найкращий можливий прогноз.

У більшості випадків до проекції вводять сталий член. Будемо використовувати символ \hat{E} для позначення лінійної проекції на вектор випадкових змінних X_t зі сталим членом

$$\hat{E}(y_{t+h}|X_t) \equiv \hat{P}(y_{t+h}|1, X_t), \qquad (6.1.13)$$

або

$$\hat{y}_{t+h|t} = \alpha' X_t.$$

6.1.3 Властивості лінійної проекції

Щоб обчислити коефіцієнт проекції α за допомогою y_{t+h} і X_t , використаємо рівність (6.1.10), яку можна записати у вигляді

$$E[y_{t+h}X_t'] = \alpha' E[X_tX_t'].$$

Звідси, якщо припустити, що $E[X_tX_t']$ невироджена матриця, матимемо

$$\alpha' = E[y_{t+h}X_t'] (E[X_tX_t'])^{-1}. \quad (6.1.14)$$

Якщо $E[X_tX_t']$ є виродженою, то існує ненульовий вектор c такий, що $c'E[X_tX_t']c = E[(c'X_t)^2] = 0$, звідки матимемо, що лінійна комбінація $c'X_t$ дорівнює нулю для всіх t . Наприклад, якщо X_t складається з двох випадкових змінних, то друга змінна повинна бути пропорційна до першої $x_{2t} = cx_{1t}$. Тому можна просто вилучити зайві змінні з системи X_t і обчислити лінійну проекцію y_{t+h} на X_t^* , де X_t^* – вектор, що складається з лінійно незалежних елементів X_t . Така лінійна проекція $(\alpha')^*X_t^*$ може бути однозначно визначена з (6.1.14), де X_t замінено на X_t^* .

Знайдемо середньоквадратичну похибку прогнозу, який є лінійною проекцією

$$\begin{aligned} E[(y_{t+h} - \alpha'X_t)^2] &= \\ &= E[(y_{t+h})^2] - 2E[\alpha'X_t y_{t+h}] + E[\alpha'X_tX_t'\alpha]. \end{aligned} \quad (6.1.15)$$

Підставивши (6.1.14) в (6.1.15), отримаємо

$$\begin{aligned} E[(y_{t+h} - \alpha'X_t)^2] &= \\ &= E[(y_{t+h})^2] - 2E[y_{t+h}X_t'] (E[X_tX_t'])^{-1} E[X_t y_{t+h}] + \\ &+ E[y_{t+h}X_t'] (E[X_tX_t'])^{-1} E[X_tX_t'] (E[X_tX_t'])^{-1} E[X_t y_{t+h}] = \\ &= E[(y_{t+h})^2] - E[y_{t+h}X_t'] (E[X_tX_t'])^{-1} E[X_t y_{t+h}]. \end{aligned}$$

Отже, маємо

$$\begin{aligned} MSE &= \\ &= E[(y_{t+h})^2] - E[y_{t+h}X_t'] (E[X_tX_t'])^{-1} E[X_t y_{t+h}]. \end{aligned} \quad (6.1.16)$$

6.1.4 Прогнозовані вектори

Попередні результати можна узагальнити для прогнозування вектора Y_{t+h} розмірності n за допомогою векторів X_t розмірності m

$$\hat{P}(Y_{t+h}|X_t) = \alpha' X_t \equiv \hat{Y}_{t+h|t}. \quad (6.1.17)$$

Тут α' позначає матрицю розміром $(n \times m)$ коефіцієнтів проєкції, що задовольняє умову

$$E[(Y_{t+h} - \alpha' X_t) X_t'] = 0, \quad (6.1.18)$$

тобто кожен з n елементів $(Y_{t+h} - \hat{Y}_{t+h|t})$ не корелює з кожним з m елементів X_t . Відповідно j -й елемент вектора $\hat{Y}_{t+h|t}$ дає оптимальний лінійний прогноз скаляра $(y_j)_{t+h}$. Щоб передбачити будь-яку лінійну комбінацію елементів Y_{t+h} , наприклад, $z_{t+h} = h' Y_{t+h}$, необхідно вимагати, щоб для оптимального лінійного прогнозу $\hat{z}_{t+h|t}$ величина $(z_{t+h} - \hat{z}_{t+h|t})$ була некорельованою з X_t . Отож, якщо α задовольняє (6.1.18), то $h' \hat{Y}_{t+h|t}$ є оптимальним лінійним прогнозом $h' Y_{t+h}$ для будь-якого вектора h .

Згідно з (6.1.18) матриця коефіцієнтів проєкції визначається за формулою

$$\alpha' = E[Y_{t+h} X_t'] (E[X_t X_t'])^{-1}. \quad (6.1.19)$$

Матричне узагальнення формули (6.1.16) для середньоквадратичної похибки має вигляд

$$\begin{aligned} MSE(\alpha' X_t) &\equiv E[(Y_{t+h} - \alpha' X_t)(Y_{t+h} - \alpha' X_t)'] = \\ &= E[Y_{t+h} Y_{t+h}'] - E[Y_{t+h} X_t'] (E[X_t X_t'])^{-1} E[X_t Y_{t+h}']. \end{aligned} \quad (6.1.20)$$

6.2 Прогнози, які ґрунтуються на нескінченній кількості спостережень

6.2.1 Прогнози, побудовані на запізненних ε

Розглянемо $MA(\infty)$ процес

$$y_t - \mu = \psi(L)\varepsilon_t, \quad (6.2.1)$$

де ε_t – білий шум,

$$\psi(L) \equiv \sum_{j=0}^{\infty} \psi_j L^j, \quad \psi_0 = 1, \quad \sum_{j=0}^{\infty} |\psi_j| < \infty. \quad (6.2.2)$$

Припустимо, що маємо безмежну кількість спостережень ε_t в моменти часу, які передують t : $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$ і, крім того, знаємо значення μ і $\{\psi_1, \psi_2, \dots\}$. Хочемо передбачити значення y_{t+h} , тобто значення, яке набуде y_t через h періодів.

Зазначимо, що з (6.2.1) випливає

$$y_{t+h} = \mu + \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \dots + \psi_{h-1} \varepsilon_{t+1} + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \dots \quad (6.2.3)$$

Оскільки $E[\varepsilon_{t+s}] = 0$ для $s > 0$, то оптимальний прогноз матиме вигляд

$$\begin{aligned} y_{t+h|t}^* &= E[y_{t+h} | \varepsilon_t, \varepsilon_{t-1}, \dots] = \\ &= \mu + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \dots \end{aligned} \quad (6.2.4)$$

Похибка цього прогнозу дорівнює

$$\begin{aligned} y_{t+h} - E[y_{t+h} | \varepsilon_t, \varepsilon_{t-1}, \dots] &= \\ &= \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \dots + \psi_{h-1} \varepsilon_{t+1}. \end{aligned} \quad (6.2.5)$$

З (6.2.4) видно, що $E[y_{t+h}|\varepsilon_t, \varepsilon_{t-1}, \dots]$ є лінійним прогнозом. Покажемо, що (6.2.4) є лінійною проекцією, тобто оптимальним лінійним прогнозом. Для цього згідно з умовою (6.1.10) необхідно, щоб похибка прогнозу мала математичне сподівання нуль і була некорельованою з $\varepsilon_t, \varepsilon_{t-1}, \dots$. Похибка (6.2.5) має ці властивості. Тому (6.2.4) справді є лінійною проекцією.

Оскільки лінійна проекція y_{t+h} здійснюється на $\varepsilon_t, \varepsilon_{t-1}, \dots$ і константу, то згідно з позначенням (6.1.13) матимемо

$$\hat{E}[y_{t+h}|\varepsilon_t, \varepsilon_{t-1}, \dots] = \mu + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \dots$$

Середня квадратична похибка у цьому випадку, оскільки $E[\varepsilon_{t+h} \varepsilon_{t+h-j}] = 0$ для $h > 0, j > 0$, дорівнює

$$\begin{aligned} MSE &= E \left[(y_{t+h} - \hat{E}[y_{t+h}|\varepsilon_t, \varepsilon_{t-1}, \dots])^2 \right] = \\ &= (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \sigma^2. \end{aligned} \quad (6.2.6)$$

Розглянемо тепер процес рухомого середнього скінченного порядку. Для $MA(q)$ процесу

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

одержимо

$$y_{t+h} = \mu + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \theta_2 \varepsilon_{t+h-2} + \dots + \theta_q \varepsilon_{t+h-q}.$$

У цьому випадку треба розглядати два випадки $h \leq q$ і $h > q$.

Для $h = 1, 2, \dots, q$

$$\begin{aligned} y_{t+h} &= \mu + \varepsilon_{t+h} + \theta_1 \varepsilon_{t+h-1} + \dots + \theta_{h-1} \varepsilon_{t+1} + \\ &+ \theta_h \varepsilon_t + \theta_{h+1} \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t+h-q} \end{aligned}$$

і оптимальний лінійний прогноз має вигляд

$$\hat{E}[y_{t+h}|\varepsilon_t, \varepsilon_{t-1}, \dots] = \mu + \theta_h \varepsilon_t + \theta_{h+1} \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-(q-h)}.$$

Для $h = q + 1, q + 2, \dots$ матимемо

$$\hat{E}[y_{t+h} | \varepsilon_t, \varepsilon_{t-1}, \dots] = \mu.$$

Відповідно MSE дорівнює

$$\begin{aligned} \sigma^2 & \text{ для } h = 1, \\ (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_{h-1}^2)\sigma^2 & \text{ для } h = 2, 3, \dots, q, \\ (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma^2 & \text{ для } h = q + 1, q + 2, \dots \end{aligned}$$

Отже, для $h \leq q$ маємо, що MSE збільшується при збільшенні h .

Якщо ми намагаємося прогнозувати $MA(q)$ процес дальше, ніж на q періодів у майбутнє, то прогноз є просто безумовним математичним сподіванням ряду ($E[y_t] = \mu$), а MSE цього прогнозу – безумовною дисперсією ряду ($\text{Var}[y_t] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2)\sigma^2$).

Подібні властивості також характерні і для $MA(\infty)$ процесу, якщо термін прогнозування прямує до безмежності. З (6.2.2) випливає, що при $h \rightarrow \infty$, прогноз (6.2.4) збігається в середньому квадратичному до μ , безумовного математичного сподівання ряду. MSE (6.2.6) також збігається до $\sigma^2 \sum_{j=0}^{\infty} \psi_j^2$, яка є безумовною дисперсією для $MA(\infty)$ процесу (6.2.1).

Для виведення загальної формули використаємо компактний вираз для запізнюючого оператора в (6.2.4). Розглянемо поліном $\psi(L)$, поділений на L^h

$$\begin{aligned} \frac{\psi(L)}{L^h} &= L^{-h} + \psi_1 L^{1-h} + \dots + \psi_{h-1} L^{-1} + \\ &+ \psi_h L^0 + \psi_{h+1} L^1 + \psi_{h+2} L^2 + \dots \end{aligned} \quad (6.2.7)$$

Визначимо знищуючий оператор, який позначимо $[\cdot]_+$, і який знищує ефект від ділення, тобто відкидає доданки, в яких оператор L має від'ємний степінь. Тобто з (6.2.7) випливає, що

$$\left[\frac{\psi(L)}{L^h} \right]_+ = \psi_h + \psi_{h+1} L + \psi_{h+2} L^2 + \dots \quad (6.2.8)$$

Підставивши (6.2.8) в (6.2.4), оптимальний лінійний прогноз може бути записаний за допомогою знищуючого оператора

$$\hat{E}[y_{t+h} | \varepsilon_t, \varepsilon_{t-1}, \dots] = \mu + \left[\frac{\psi(L)}{L^h} \right]_+ \varepsilon_t. \quad (6.2.9)$$

6.2.2 Прогнози, побудовані на запізнених y . Формула прогнозування Вінера - Колмогорова

Формула прогнозування (6.2.9) ґрунтується на припущенні, що ε_t відомі або можуть бути безпосередньо обчислені. У звичайній ситуації для прогнозування відомими є запізнення y_t , а не ε_t .

Для того щоб виразити величину прогнозу через запізнення y_t , розглянемо $AR(\infty)$ процес

$$\eta(L)(y_t - \mu) = \varepsilon_t, \quad (6.2.10)$$

де

$$\eta(L) \equiv \sum_{j=0}^{\infty} \eta_j L^j, \quad \eta_0 = 1, \quad \sum_{j=0}^{\infty} |\eta_j| < \infty.$$

Оскільки тут розглядаємо лише стаціонарні процеси, а для авторегресійних процесів стаціонарність еквівалентна зворотності, то $AR(\infty)$ процес має $MA(\infty)$ зображення

$$y_t - \mu = \psi(L)\varepsilon_t, \quad (6.2.11)$$

причому AR -поліном $\eta(L)$ і MA -поліном $\psi(L)$ зв'язані між собою так:

$$[\eta(L)]^{-1} = \psi(L). \quad (6.2.12)$$

Стаціонарний авторегресійний процес скінченного порядку $AR(p)$ має вигляд

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(y_t - \mu) = \varepsilon_t,$$

або більш компактно

$$\phi(L)(y_t - \mu) = \varepsilon_t.$$

Тому для нього

$$\eta(L) = \phi(L), \quad \psi(L) = [\phi(L)]^{-1}.$$

Для зворотного $MA(q)$ процесу

$$y_t - \mu = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t \quad (6.2.13)$$

або

$$y_t - \mu = \theta(L) \varepsilon_t$$

матимемо

$$\psi(L) = \theta(L), \quad \eta(L) = [\theta(L)]^{-1}.$$

Процес $ARMA(p, q)$ має вигляд (6.2.10) з

$$\eta(L) = \frac{\phi(L)}{\theta(L)},$$

якщо авторегресійний оператор $\phi(L)$ задовольняє умови стаціонарності (корені рівняння $\phi(z) = 0$ лежать ззовні одиничного круга), а оператор рухомого середнього $\theta(L)$ задовольняє умови зворотності (корені рівняння $\theta(z) = 0$ лежать ззовні одиничного круга).

Отже, якщо виконуються умови, за яких процес може бути записаний у вигляді (6.2.10), то спостереження $\{y_t, y_{t-1}, \dots\}$ будуть достатніми для побудови $\{\varepsilon_t, \varepsilon_{t-1}, \dots\}$. З (6.2.10) маємо

$$\varepsilon_t = \eta(L) (y_t - \mu). \quad (6.2.14)$$

Зокрема для $MA(1)$ процесу (6.2.14) запишеться у вигляді

$$\varepsilon_t = (1 + \theta L)^{-1} (y_t - \mu).$$

Отже, при відомих θ , μ і $\{y_t, y_{t-1}, \dots\}$ може бути побудоване ε_t , а саме

$$\varepsilon_t = (y_t - \mu) - \theta(y_{t-1} - \mu) + \theta^2(y_{t-2} - \mu) - \theta^3(y_{t-3} - \mu) + \dots \quad (6.2.15)$$

Підставляючи (6.2.14) у формулу прогнозування (6.2.9) можна отримати прогноз y_{t+h} як функцію від запізнених y

$$\hat{E}[y_{t+h}|y_t, y_{t-1}, \dots] = \mu + \left[\frac{\psi(L)}{L^h} \right]_+ \eta(L)(y_t - \mu),$$

де $\eta(L)$ – авторегресійний поліном $AR(\infty)$ зображення; $\psi(L)$ – поліном $MA(\infty)$ зображення прогнозованого процесу.

Використовуючи (6.2.12), маємо

$$\hat{E}[y_{t+h}|y_t, y_{t-1}, \dots] = \mu + \left[\frac{\psi(L)}{L^h} \right]_+ \frac{1}{\psi(L)}(y_t - \mu). \quad (6.2.16)$$

Рівняння (6.2.16) є відомим як **формула прогнозування Вінера-Колмогорова**.

6.3 Застосування формули Вінера - Колмогорова для прогнозування часових рядів

Використовуючи формулу Вінера - Колмогорова, можна вивести формули прогнозування для конкретних стаціонарних часових рядів.

6.3.1 Прогнозування $AR(1)$ процесу

Для коваріаційно - стаціонарного $AR(1)$ процесу маємо

$$\psi(L) = \frac{1}{(1 - \phi L)} = 1 + \phi L + \phi^2 L^2 + \dots \quad (6.3.1)$$

У цьому випадку знищуючий оператор має вигляд

$$\begin{aligned} \left[\frac{\psi(L)}{L^h} \right]_+ &= \phi^h + \phi^{h+1}L^1 + \phi^{h+2}L^2 + \dots = \\ &= \phi^h(1 + \phi^1L^1 + \phi^2L^2 + \dots) = \frac{\phi^h}{(1 - \phi L)}. \end{aligned} \quad (6.3.2)$$

Підстановка (6.3.2) в (6.2.16) дасть оптимальний прогноз на h періодів вперед для стаціонарного $AR(1)$ процесу

$$\begin{aligned} \hat{E}[y_{t+h}|y_t, y_{t-1}, \dots] &= \mu + \frac{\phi^h}{1 - \phi L}(1 - \phi L)(y_t - \mu) = \\ &= \mu + \phi^h(y_t - \mu). \end{aligned} \quad (6.3.3)$$

Зауважимо, що прогноз ґрунтується лише на значенні y_t і не використовує попередніх значень y_{t-1}, y_{t-2}, \dots . Оскільки для коваріаційно - стаціонарного $AR(1)$ процесу $|\phi| < 1$, то з (6.3.3) також випливає, що при збільшенні h прогнозоване значення ряду геометрично спадає до μ (безумовного математичного сподівання).

Для знаходження середньоквадратичної похибки прогнозу використаємо вираз (6.2.6), в якому згідно з (6.3.1) $\psi_j = \phi^j$. Підставляючи їх у формулу для знаходження MSE (6.2.6) матимемо середньоквадратичну похибку прогнозу на h періодів вперед

$$MSE = [1 + \phi^2 + \phi^4 + \dots + \phi^{2(h-1)}] \sigma^2.$$

Зазначимо, що вона збільшується при збільшенні h і при $h \rightarrow \infty$

$$MSE = [1 + \phi^2 + \phi^4 + \dots] \sigma^2 = \frac{\sigma^2}{1 - \phi^2},$$

тобто MSE асимптотично наближається до безумовної дисперсії y_t .

6.3.2 Прогнозування $AR(p)$ процесу

Розглянемо прогнозування $AR(p)$ процесу

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t.$$

Формула Вінера-Колмогорова (6.2.16) виражає значення $(y_{t+h} - \mu)$ в термінах попередніх значень $\{(y_t - \mu), (y_{t-1} - \mu), \dots\}$ і опускає вирази, які охоплюють майбутні значення похибки $\{\varepsilon_{t+1}, \varepsilon_{t+2}, \dots, \varepsilon_{t+h}\}$.

Для $AR(p)$ процесу значення $y_{t+h} - \mu$ можна отримати як розв'язок різницевого рівняння

$$y_{t+h} - \mu = f_{11}^{(h)}(y_t - \mu) + f_{12}^{(h)}(y_{t-1} - \mu) + \dots + f_{1p}^{(h)}(y_{t-p+1} - \mu) + \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \dots + \psi_{h-1} \varepsilon_{t+1}. \quad (6.3.4)$$

Тут

$$\psi_j = f_{11}^{(j)}. \quad (6.3.5)$$

Зазначимо, що $f_{11}^{(j)}$ позначає $(1, 1)$ -елемент матриці F^j , $f_{12}^{(j)}$ позначає $(1, 2)$ -елемент F^j і т.д., де F -матриця розміру $(p \times p)$

$$F \equiv \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

З (6.3.4) випливає, що оптимальний прогноз на h періодів вперед має вигляд

$$\hat{y}_{t+h|t} = \mu + f_{11}^{(h)}(y_t - \mu) + f_{12}^{(h)}(y_{t-1} - \mu) + \dots + f_{1p}^{(h)}(y_{t-p+1} - \mu). \quad (6.3.6)$$

Зазначимо, що для довільного прогнозу на h періодів оптимальним прогнозом є константа плюс лінійна функція від $\{y_t, y_{t-1}, \dots, y_{t-p+1}\}$. Відповідна похибка дорівнює

$$y_{t+h} - \hat{y}_{t+h|t} = \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \dots + \psi_{h-1} \varepsilon_{t+1}. \quad (6.3.7)$$

Прогноз (6.3.6) можна також отримати за допомогою простої рекурсії. Оптимальним прогнозом на один період є

$$(\hat{y}_{t+1|t} - \mu) = \phi_1(y_t - \mu) + \phi_2(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p+1} - \mu). \quad (6.3.8)$$

Для побудови прогнозу на два періоди вперед використаємо рівність

$$(y_{t+2} - \mu) = \phi_1(y_{t+1} - \mu) + \phi_2(y_t - \mu) + \dots + \phi_p(y_{t-p+2} - \mu). \quad (6.3.9)$$

У момент часу t значення $y_t, y_{t-1}, \dots, y_{t-p+2}$ в (6.3.9) є відомими, а y_{t+1} – невідомі. Приймаючи замість y_{t+1} його прогнозоване значення $\hat{y}_{t+1|t}$, яке обчислюємо за допомогою (6.3.8), отримуємо

$$(\hat{y}_{t+2|t} - \mu) = \phi_1(\hat{y}_{t+1|t} - \mu) + \phi_2(y_t - \mu) + \dots + \phi_p(y_{t-p+2} - \mu). \quad (6.3.10)$$

Підставивши (6.3.8) в (6.3.10), одержимо формулу прогнозування на два періоди вперед для $AR(p)$ процесу

$$\begin{aligned} (\hat{y}_{t+2|t} - \mu) &= & (6.3.11) \\ &= \phi_1 [\phi_1(y_t - \mu) + \phi_2(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p+1} - \mu)] + \\ &\quad + \phi_2(y_t - \mu) + \phi_3(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p+2} - \mu) = \\ &= (\phi_1^2 + \phi_2)(y_t - \mu) + (\phi_1\phi_2 + \phi_3)(y_{t-1} - \mu) + \dots + \\ &\quad + (\phi_1\phi_{p-1} + \phi_p)(y_{t-p+2} - \mu) + \phi_1\phi_p(y_{t-p+1} - \mu). \end{aligned}$$

Значення $\hat{y}_{t+1|t}$, $\hat{y}_{t+2|t}$, обчислені відповідно за (6.3.8), (6.3.11), використовуємо для побудови прогнозу на три періоди вперед

і т.д. Загалом прогнозування на h періодів вперед для $AR(p)$ процесу може бути отримане за допомогою ітерації

$$(\hat{y}_{t+h|t} - \mu) = \phi_1(\hat{y}_{t+h-1|t} - \mu) + \phi_2(\hat{y}_{t+h-2|t} - \mu) + \dots + \phi_p(\hat{y}_{t+h-p|t} - \mu), \quad (6.3.12)$$

причому

$$\hat{y}_{\tau|t} = y_{\tau} \quad \text{для } \tau \leq t.$$

6.3.3 Прогнозування $MA(1)$ процесу

Розглянемо зворотнє $MA(1)$ зображення

$$y_t - \mu = (1 + \theta L)\varepsilon_t, \quad (6.3.13)$$

де $|\theta| < 1$. Для цього процесу $\psi(L) = 1 + \theta L$ і формула Вінера-Колмогорова (6.2.16) набуде вигляду

$$\hat{y}_{t+h|t} = \mu + \left[\frac{1 + \theta L}{L^h} \right]_+ \frac{1}{1 + \theta L} (y_t - \mu). \quad (6.3.14)$$

При прогнозуванні $MA(1)$ процесу на один період ($h = 1$) знищуючий оператор дорівнює

$$\left[\frac{1 + \theta L}{L} \right]_+ = \theta,$$

тобто

$$\hat{y}_{t+1|t} = \mu + \frac{\theta}{1 + \theta L} (y_t - \mu). \quad (6.3.15)$$

Подіавши оператором $\frac{1}{1 + \theta L}$ на $y_t - \mu$, матимемо

$$\hat{y}_{t+1|t} = \mu + \theta(y_t - \mu) - \theta^2(y_{t-1} - \mu) + \theta^3(y_{t-2} - \mu) - \dots$$

Якщо записати (6.3.13) у вигляді

$$\varepsilon_t = \frac{1}{1 + \theta L} (y_t - \mu),$$

то з (6.3.15) отримаємо ще один вираз для знаходження прогнозу на один період

$$\hat{y}_{t+1|t} = \mu + \theta \hat{\varepsilon}_t, \quad (6.3.16)$$

де $\hat{\varepsilon}_t$ знаходимо з безмежної рекурсії

$$\hat{\varepsilon}_t = (y_t - \mu) - \theta \hat{\varepsilon}_{t-1}. \quad (6.3.17)$$

Рівняння (6.3.17), виведене з (6.3.13), є точною оцінкою ε_t .

Обчислимо прогноз $MA(1)$ процесу на $h = 2, 3, \dots$ періодів вперед. Знищуючий оператор у цьому випадку дорівнює нулю

$$\left[\frac{1 + \theta L}{L^h} \right]_+ = 0 \quad \text{для } h = 2, 3, \dots,$$

і отже, з (6.3.14) матимемо

$$\hat{y}_{t+h|t} = \mu \quad \text{для } h = 2, 3, \dots \quad (6.3.18)$$

Тобто, прогноз $MA(1)$ процесу більше, ніж на один період вперед дорівнює безумовному математичному сподіванню ряду.

6.3.4 Прогнозування $MA(q)$ процесу

Для зворотного $MA(q)$ процесу

$$(y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t$$

прогноз (6.2.16) запишемо

$$\begin{aligned} \hat{y}_{t+h|t} = & \mu + \left[\frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{L^h} \right]_+ \times \\ & \times \frac{1}{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q} (y_t - \mu). \end{aligned} \quad (6.3.19)$$

Знищуючий оператор для $MA(q)$ процесу матиме вигляд

$$\left[\frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{L^h} \right]_+ = \begin{cases} \theta_h + \theta_{h+1} L + \dots + \theta_q L^{q-h} & \text{для } h = 1, 2, \dots, q, \\ 0 & \text{для } h = q + 1, q + 2, \dots \end{cases}$$

Отже, для $h = 1, 2, \dots, q$ прогнозоване значення обчислюємо за формулою

$$\hat{y}_{t+h|t} = \mu + (\theta_h + \theta_{h+1} L + \dots + \theta_q L^{q-h}) \hat{\varepsilon}_t, \quad (6.3.20)$$

де $\hat{\varepsilon}_t$ може бути визначене з рекурсії

$$\hat{\varepsilon}_t = (y_t - \mu) - \theta_1 \hat{\varepsilon}_{t-1} - \theta_2 \hat{\varepsilon}_{t-2} - \dots - \theta_q \hat{\varepsilon}_{t-q}. \quad (6.3.21)$$

Прогнозом на більше, ніж q періодів у майбутнє є безумовне математичне сподівання

$$\hat{y}_{t+h|t} = \mu \quad \text{для } h > q.$$

6.3.5 Прогнозування $ARMA(1, 1)$ процесу

Для $ARMA(1, 1)$ процесу

$$(1 - \phi L)(y_t - \mu) = (1 + \theta L)\varepsilon_t,$$

який є стаціонарним ($|\phi| < 1$) і зворотним ($|\theta| < 1$), згідно з формулою Вінера - Колмогорова прогноз на h періодів вперед матиме вигляд

$$\hat{y}_{t+h|t} = \mu + \left[\frac{1 + \theta L}{(1 - \phi L)L^h} \right]_+ \frac{1 - \phi L}{1 + \theta L} (y_t - \mu). \quad (6.3.22)$$

Перетворимо знищуючий оператор

$$\begin{aligned}
& \left[\frac{1 + \theta L}{(1 - \phi L)L^h} \right]_+ = \\
& = \left[\frac{(1 + \phi L + \phi^2 L^2 + \dots)}{L^h} + \frac{\theta L(1 + \phi L + \phi^2 L^2 + \dots)}{L^h} \right]_+ = \\
& = (\phi^h + \phi^{h+1}L + \phi^{h+2}L^2 + \dots) + \theta(\phi^{h-1} + \phi^h L + \phi^{h+1}L^2 + \dots) = \\
& = (\phi^h + \theta\phi^{h-1})(1 + \phi L + \phi^2 L^2 + \dots) = \\
& = \frac{\phi^h + \theta\phi^{h-1}}{1 - \phi L}. \tag{6.3.23}
\end{aligned}$$

Підстановка (6.3.23) в (6.3.22) дає

$$\begin{aligned}
\hat{y}_{t+h|t} &= \mu + \left[\frac{\phi^h + \theta\phi^{h-1}}{1 - \phi L} \right] \frac{1 - \phi L}{1 + \theta L} (y_t - \mu) = \\
&= \mu + \frac{\phi^h + \theta\phi^{h-1}}{1 + \theta L} (y_t - \mu) = \\
&= \mu + \phi \frac{\phi^{h-1} + \theta\phi^{h-2}}{1 + \theta L} (y_t - \mu). \tag{6.3.24}
\end{aligned}$$

З (6.3.24) випливає, що для $h = 2, 3, \dots$ прогноз задовольняє рекурсію

$$(\hat{y}_{t+h|t} - \mu) = \phi(\hat{y}_{t+h-1|t} - \mu),$$

звідки матимемо

$$(\hat{y}_{t+h|t} - \mu) = \phi^{h-1}(\hat{y}_{t+1|t} - \mu), \quad h \geq 2.$$

Отож, при $h \rightarrow \infty$ прогноз геометрично спадає до безумовного математичного сподівання μ .

Одноперіодичний прогноз ($h = 1$) має вигляд

$$\hat{y}_{t+1|t} = \mu + \frac{\phi + \theta}{1 + \theta L} (y_t - \mu). \tag{6.3.25}$$

Перетворимо (6.3.25) так:

$$\begin{aligned} (\hat{y}_{t+1|t} - \mu) &= \frac{\phi(1 + \theta L) + \theta(1 - \phi L)}{1 + \theta L} (y_t - \mu) = \\ &= \phi(y_t - \mu) + \theta \hat{\varepsilon}_t, \end{aligned} \quad (6.3.26)$$

де $\hat{\varepsilon}_t$ знаходимо з загального вигляду $ARMA(1, 1)$ процесу

$$\hat{\varepsilon}_t = \frac{1 - \phi L}{1 + \theta L} (y_t - \mu)$$

або

$$\hat{\varepsilon}_t = (y_t - \mu) - \phi(y_{t-1} - \mu) - \theta \hat{\varepsilon}_{t-1} = y_t - \hat{y}_{t|t-1}. \quad (6.3.27)$$

6.3.6 Прогнозування $ARMA(p, q)$ процесу

Зрештою, розглянемо прогнозування стаціонарного і зворотного $ARMA(p, q)$ процесу

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(y_t - \mu) = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t.$$

Природним узагальненням (6.3.26) і (6.3.27) буде

$$\begin{aligned} (\hat{y}_{t+1|t} - \mu) &= \phi_1(y_t - \mu) + \phi_2(y_{t-1} - \mu) + \dots + \phi_p(y_{t-p+1} - \mu) + \\ &+ \theta_1 \hat{\varepsilon}_t + \theta_2 \hat{\varepsilon}_{t-1} + \dots + \theta_q \hat{\varepsilon}_{t-q+1}, \end{aligned} \quad (6.3.28)$$

де $\{\hat{\varepsilon}_t\}$ обчислюють як

$$\hat{\varepsilon}_t = y_t - \hat{y}_{t|t-1}. \quad (6.3.29)$$

Прогноз на h періодів вперед матиме вигляд

$$\begin{aligned} (\hat{y}_{t+h|t} - \mu) &= \\ &= \begin{cases} \phi_1(\hat{y}_{t+h-1|t} - \mu) + \phi_2(\hat{y}_{t+h-2|t} - \mu) + \dots + \phi_p(\hat{y}_{t+h-p|t} - \mu) + \\ + \theta_h \hat{\varepsilon}_t + \theta_{h+1} \hat{\varepsilon}_{t-1} + \dots + \theta_q \hat{\varepsilon}_{t+h-q} & \text{для } h = 1, 2, \dots, q, \\ \phi_1(\hat{y}_{t+h-1|t} - \mu) + \phi_2(\hat{y}_{t+h-2|t} - \mu) + \dots + \phi_p(\hat{y}_{t+h-p|t} - \mu) & \text{для } h = q + 1, q + 2, \dots, \end{cases} \end{aligned} \quad (6.3.30)$$

де

$$\hat{y}_{\tau|t} = y_{\tau} \quad \text{для} \quad \tau \leq t.$$

Отже, при прогнозуванні на h періодів, де h більше за q , прогнози будують за допомогою рекурсії p -го порядку, яка однозначно визначається авторегресійними параметрами.

6.4 Прогнози, побудовані на скінченій кількості спостережень

Попередні формули ґрунтувались на припущенні, що маємо безмежну кількість спостережень $y: \{y_t, y_{t-1}, \dots\}$ і точно знаємо параметри μ , ϕ і θ . У цьому пункті знову припускатимемо, що параметри є відомими, але будуватимемо прогнози, які ґрунтуються на скінченій кількості спостережень $\{y_t, y_{t-1}, \dots, y_1\}$.

Для прогнозування $AR(p)$ процесу, оптимальний h -періодичний лінійний прогноз (6.3.12), побудований на безмежній кількості спостережень $\{y_t, y_{t-1}, \dots\}$ використовує p найбільш близьких значень $\{y_t, y_{t-1}, \dots, y_{t-p+1}\}$, тобто в дійсності він ґрунтується на скінченій кількості спостережень.

Проте формули прогнозування для MA або $ARMA$ процесів вимагають, щоб були відомими нескінченна кількість значень y_t .

6.4.1 Наближення оптимальних прогнозів

Одним з підходів до прогнозування, яке ґрунтується на скінченій кількості спостережень, є припущення, що всі ε_t до початку спостережень дорівнюють нулю, тобто

$$\varepsilon_{-n} = 0 \quad \text{для} \quad n = 0, 1, 2, \dots$$

Ідея полягає в тому, щоб використати наближення

$$\hat{E}(y_{t+h}|y_t, y_{t-1}, \dots) \approx \hat{E}(y_{t+h}|y_t, y_{t-1}, \dots, y_1, \varepsilon_0 = 0, \varepsilon_{-1} = 0, \dots). \quad (6.4.1)$$

Наприклад, розглянемо прогнозування $MA(q)$ процесу на підставі формул (6.3.20), (6.3.21). Рекурсію (6.3.21) можна почати так:

$$\hat{\varepsilon}_0 = \hat{\varepsilon}_{-1} = \dots = \hat{\varepsilon}_{-(q-1)} = 0 \quad (6.4.2)$$

і за допомогою підстановки в (6.3.21) продовжена так:

$$\begin{aligned} \hat{\varepsilon}_1 &= (y_1 - \mu), \\ \hat{\varepsilon}_2 &= (y_2 - \mu) - \theta_1 \hat{\varepsilon}_1, \\ \hat{\varepsilon}_3 &= (y_3 - \mu) - \theta_1 \hat{\varepsilon}_2 - \theta_2 \hat{\varepsilon}_1, \end{aligned}$$

і т.д. Результуючі значення $(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q+h})$ треба підставити в (6.3.21) для обчислення прогнозу (6.4.1).

Наприклад, у випадку $MA(1)$ процесу для $h = 1$, наближенням прогнозу буде

$$\begin{aligned} \hat{y}_{t+1|t} = & \mu + \theta(y_t - \mu) - \theta^2(y_{t-1} - \mu) + \\ & + \theta^3(y_{t-2} - \mu) + \dots + (-1)^{t-1} \theta^t (y_1 - \mu), \end{aligned} \quad (6.4.3)$$

який є апроксимацією $AR(\infty)$ процесу

$$\hat{y}_{t+1|t} = \mu + \theta(y_t - \mu) - \theta^2(y_{t-1} - \mu) + \theta^3(y_{t-2} - \mu) - \dots \quad (6.4.4)$$

Для великого t і достатньо малого $|\theta|$ (6.4.3) дає чудове наближення. Для $|\theta|$, близьких до одиниці, наближення може бути поганим. Зазначимо таке: якщо оператор рухомого середнього незворотний, то прогноз (6.4.1) невідповідний, і не може бути використаний.

Загалом випадку можна будувати прогнози, які ґрунтуються не на всіх t спостереженнях $\{y_t, y_{t-1}, \dots, y_1\}$, а на деяких

останніх m значеннях $\{y_t, y_{t-1}, \dots, y_{t-m+1}\}$. Тоді використовуємо наближення

$$\begin{aligned} \hat{E}(y_{t+h}|y_t, y_{t-1}, \dots) &\approx \\ &\approx \hat{E}(y_{t+h}|y_t, y_{t-1}, \dots, y_{t-m+1}, \varepsilon_{t-m} = 0, \varepsilon_{t-m-1} = 0, \dots). \end{aligned} \quad (6.4.5)$$

Відповідно для $MA(q)$ процесу припускаємо

$$\hat{\varepsilon}_{t-m} = \hat{\varepsilon}_{t-m-1} = \dots = \hat{\varepsilon}_{t-m-q+1} = 0. \quad (6.4.6)$$

На підставі цих значень знаходимо

$$\begin{aligned} \hat{\varepsilon}_{t-m+1} &= (y_{t-m+1} - \mu), \\ \hat{\varepsilon}_{t-m+2} &= (y_{t-m+2} - \mu) - \theta_1 \hat{\varepsilon}_{t-m+1}, \\ \hat{\varepsilon}_{t-m+3} &= (y_{t-m+3} - \mu) - \theta_1 \hat{\varepsilon}_{t-m+2} - \theta_2 \hat{\varepsilon}_{t-m+1}, \end{aligned}$$

і т.д. Результуючі значення $(\hat{\varepsilon}_t, \hat{\varepsilon}_{t-1}, \dots, \hat{\varepsilon}_{t-q+h})$ підставляємо в (6.3.20).

Зокрема, для $h = q = 1$ матимемо наближення

$$\begin{aligned} \hat{y}_{t+1|t} &= \mu + \theta(y_t - \mu) - \theta^2(y_{t-1} - \mu) + \\ &+ \theta^3(y_{t-2} - \mu) - \dots + (-1)^{m-1} \theta^m (y_{t-m+1} - \mu). \end{aligned} \quad (6.4.7)$$

6.4.2 Точні скінченороджені прогнози

Альтернативним підходом до побудови прогнозів, які ґрунтуються на скінченій кількості спостережень є знаходження точної проєкції y_{t+1} на m найбільш близьких значень (m може дорівнювати t).

Нехай

$$X_t \equiv \begin{pmatrix} 1 \\ y_t \\ y_{t-1} \\ \vdots \\ y_{t-m+1} \end{pmatrix}.$$

Побудуємо лінійний прогноз у вигляді

$$\begin{aligned} \hat{y}_{t+1|t} &= (\alpha^{(m)})' X_t = \\ &= \alpha_0^{(m)} + \alpha_1^{(m)} y_t + \alpha_2^{(m)} y_{t-1} + \dots + \alpha_m^{(m)} y_{t-m+1}. \end{aligned} \quad (6.4.8)$$

Коефіцієнт залежності $\hat{y}_{t+1|t}$ від y_t в проєкції y_{t+1} на m найбільш близьких значень y позначаємо $\alpha_1^{(m)}$. Загалом це буде відрізняти його від коефіцієнта залежності $\hat{y}_{t+1|t}$ від y_t в проєкції y_{t+1} на $m+1$ найбільш близьких значень y , який позначатимемо $\alpha_1^{(m+1)}$.

Якщо y_t коваріаційно-стаціонарний, то з означення коваріації $\gamma_j = E[y_t y_{t-j}] - E[y_t]E[y_{t-j}]$ випливає, що $E[y_t y_{t-j}] = \gamma_j + \mu^2$. Підстановка $X_t = (1, y_t, \dots, y_{t-m+1})'$ в (6.1.14) дає

$$\begin{aligned} (\alpha^{(m)})' &\equiv \left(\alpha_0^{(m)} \quad \alpha_1^{(m)} \quad \dots \quad \alpha_m^{(m)} \right) = \\ &= \left(\mu \quad (\gamma_1 + \mu^2) \quad \dots \quad (\gamma_m + \mu^2) \right) \times \\ &\quad \times \begin{pmatrix} 1 & \mu & \mu & \dots & \mu \\ \mu & \gamma_0 + \mu^2 & \gamma_1 + \mu^2 & \dots & \gamma_{m-1} + \mu^2 \\ \mu & \gamma_1 + \mu^2 & \gamma_0 + \mu^2 & \dots & \gamma_{m-2} + \mu^2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \mu & \gamma_{m-1} + \mu^2 & \gamma_{m-2} + \mu^2 & \dots & \gamma_0 + \mu^2 \end{pmatrix}^{-1}. \end{aligned} \quad (6.4.9)$$

Якщо в X_t включений сталий член, то зручніше виразити змінні у відхиленнях від математичного сподівання. Тоді можна обчислити проєкцію $(y_{t+1} - \mu)$ на $X_t = [(y_t - \mu) \quad (y_{t-1} - \mu) \quad \dots \quad (y_{t-m+1} - \mu)]'$:

$$\hat{y}_{t+1|t} - \mu = \alpha_1^{(m)}(y_t - \mu) + \alpha_2^{(m)}(y_{t-1} - \mu) + \dots + \alpha_m^{(m)}(y_{t-m+1} - \mu). \quad (6.4.10)$$

За такого визначення X_t коефіцієнти можна обчислити з (6.1.14)

таким способом

$$\begin{pmatrix} \alpha_1^{(m)} \\ \alpha_2^{(m)} \\ \vdots \\ \alpha_m^{(m)} \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{m-2} \\ \vdots & \vdots & \dots & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \dots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_m \end{pmatrix}. \quad (6.4.11)$$

Можна показати, що коефіцієнти $(\alpha_1^{(m)}, \alpha_2^{(m)}, \dots, \alpha_m^{(m)})$ у рівняннях (6.4.8) і (6.4.11) є однаковими. Цей результат подібний до результату для звичайної регресії найменших квадратів – нахил коефіцієнтів буде незмінним, якщо всі змінні виражені у відхиленнях від середніх значень, а сталий член не входить до регресії.

Для узагальнення цього результату для h -періодичного прогнозу $\hat{y}_{t+h|t}$ можна використати рівність

$$\hat{y}_{t+h|t} = \mu + \alpha_1^{(m,h)}(y_t - \mu) + \alpha_2^{(m,h)}(y_{t-1} - \mu) + \dots + \alpha_m^{(m,h)}(y_{t-m+1} - \mu),$$

де

$$\begin{pmatrix} \alpha_1^{(m,h)} \\ \alpha_2^{(m,h)} \\ \vdots \\ \alpha_m^{(m,h)} \end{pmatrix} = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{m-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{m-2} \\ \vdots & \vdots & \dots & \vdots \\ \gamma_{m-1} & \gamma_{m-2} & \dots & \gamma_0 \end{pmatrix}^{-1} \begin{pmatrix} \gamma_h \\ \gamma_{h+1} \\ \vdots \\ \gamma_{h+m-1} \end{pmatrix}. \quad (6.4.12)$$

Використання виразів (6.3.11) і (6.3.12) вимагає існування оберненої матриці розміром $m \times m$.

6.5 Уточнення лінійної проєкції

6.5.1 Зв'язок між лінійною проєкцією і трикутною факторизацією матриці

Нехай $y = (y_1, y_2, \dots, y_T)'$ – вектор випадкових змінних розмірності T з нульовим математичним сподіванням і коваріаційною матрицею

$$E[yy'] = \Omega = \begin{pmatrix} \omega_{11} & \dots & \omega_{1T} \\ \vdots & \dots & \vdots \\ \omega_{T1} & \dots & \omega_{TT} \end{pmatrix}. \quad (6.5.1)$$

Нехай $\Omega = ADA'$ – трикутна факторизація матриці Ω , визначена у розділі 3. А саме,

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{\omega_{21}}{\omega_{11}} & 1 & 0 & \dots & 0 \\ \frac{\omega_{31}}{\omega_{11}} & \frac{h_{32}}{h_{22}} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{\omega_{T1}}{\omega_{11}} & \frac{h_{T2}}{h_{22}} & \frac{k_{T3}}{k_{33}} & \dots & 1 \end{pmatrix},$$

$$D = \begin{pmatrix} \omega_{11} & 0 & 0 & \dots & 0 \\ 0 & \omega_{22} - \frac{\omega_{21}\omega_{12}}{\omega_{11}} & 0 & \dots & 0 \\ 0 & 0 & h_{33} - \frac{h_{32}h_{23}}{h_{22}} & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & c_{TT} - \frac{c_{T,T-1}c_{T-1,T}}{c_{T-1,T-1}} \end{pmatrix}.$$

Позначимо

$$\tilde{y} \equiv A^{-1}y. \quad (6.5.2)$$

Тоді коваріаційна матриця перетворених змінних \tilde{y} має вигляд

$$E[\tilde{y}\tilde{y}'] = E[A^{-1}yy'(A')^{-1}] = A^{-1}E[yy'](A')^{-1}. \quad (6.5.3)$$

Підстановка (6.5.1) в (6.5.3) показує, що коваріаційна матриця \tilde{y} є діагональною

$$E[\tilde{y}\tilde{y}'] = A^{-1}\Omega(A')^{-1} = A^{-1}ADA'(A')^{-1} = D. \quad (6.5.4)$$

Тобто,

$$E[\tilde{y}_i\tilde{y}_j] = \begin{cases} d_{ii} & \text{для } i = j, \\ 0 & \text{для } i \neq j. \end{cases} \quad (6.5.5)$$

Отже, \tilde{y} – вектор складений з випадкових змінних, які не корелюють між собою. Помножимо (6.5.2) зліва на A

$$A\tilde{y} = y. \quad (6.5.6)$$

Використовуючи зображення матриці A , запишемо (6.5.6) у вигляді

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \frac{\omega_{21}}{\omega_{11}} & 1 & 0 & \dots & 0 \\ \frac{\omega_{31}}{\omega_{11}} & \frac{h_{31}}{h_{22}} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \frac{\omega_{T1}}{\omega_{11}} & \frac{h_{T2}}{h_{22}} & \frac{k_{T3}}{k_{33}} & \dots & 1 \end{pmatrix} \begin{pmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \tilde{y}_3 \\ \vdots \\ \tilde{y}_T \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_T \end{pmatrix}. \quad (6.5.7)$$

Перше рівняння в (6.5.7) означає, що

$$y_1 = \tilde{y}_1, \quad (6.5.8)$$

тобто перші елементи векторів y і \tilde{y} збігаються.

Друге рівняння (6.5.7) запишеться у вигляді

$$y_2 = \frac{\omega_{21}}{\omega_{11}} \tilde{y}_1 + \tilde{y}_2$$

або, використовуючи (6.5.8),

$$\tilde{y}_2 = y_2 - \frac{\omega_{21}}{\omega_{11}} y_1 = y_2 - \alpha y_1, \quad (6.5.9)$$

де $\alpha \equiv \omega_{21}/\omega_{11}$. Оскільки \tilde{y}_2 не корелює з \tilde{y}_1 , то

$$E[\tilde{y}_2 \tilde{y}_1] = E[(y_2 - \alpha y_1) y_1] = 0. \quad (6.5.10)$$

Згідно з (6.1.10) значення α , яке задовольняє (6.5.10), є коефіцієнтом лінійної проєкції y_2 на y_1 . Отже, з трикутної факторизації матриці Ω можна зробити висновок, що коефіцієнтом лінійної проєкції y_2 на y_1 є $\alpha = \omega_{21}/\omega_{11}$. Позначивши лінійну проєкцію y_2 на y_1 через $\hat{P}(y_2|y_1)$, матимемо

$$\hat{P}(y_2|y_1) = \frac{\omega_{21}}{\omega_{11}} y_1.$$

Загалом, елемент i -го рядка і першого стовпця матриці A , який дорівнює ω_{i1}/ω_{11} , є коефіцієнтом лінійної проєкції y_i на y_1

$$\hat{P}(y_i|y_1) = \frac{\omega_{i1}}{\omega_{11}} y_1.$$

Величину \tilde{y}_2 можна інтерпретувати як різницю між y_2 і проєкцією y_2 на y_1 . Тоді d_{22} з (6.5.5) є середньоквадратичною похибкою цієї проєкції

$$MSE(\hat{P}(y_2|y_1)) = E[\tilde{y}_2^2] = d_{22} = \omega_{22} - \frac{\omega_{21}\omega_{12}}{\omega_{11}}.$$

Третє рівняння в (6.5.7) означає, що

$$y_3 = \frac{\omega_{31}}{\omega_{11}} \tilde{y}_1 + \frac{h_{32}}{h_{22}} \tilde{y}_2 + \tilde{y}_3.$$

Підставивши сюди значення \tilde{y}_1 і \tilde{y}_2 з (6.5.8) і (6.5.9) і спростивши, отримуємо

$$\tilde{y}_3 = y_3 - \frac{\omega_{31}}{\omega_{11}} y_1 - \frac{h_{32}}{h_{22}} \left(y_2 - \frac{\omega_{21}}{\omega_{11}} y_1 \right). \quad (6.5.11)$$

Отже, \tilde{y}_3 є результатом віднімання лінійної комбінації y_1 і y_2 від y_3 . З (6.5.5) випливає, що \tilde{y}_3 не корелює ні з \tilde{y}_1 , ні з \tilde{y}_2

$$E \left[\left(y_3 - \frac{\omega_{31}}{\omega_{11}} y_1 - \frac{h_{32}}{h_{22}} \left(y_2 - \frac{\omega_{21}}{\omega_{11}} y_1 \right) \right) \tilde{y}_j \right] = 0 \quad \text{для } j = 1, 2.$$

Звідси випливає, що \tilde{y}_3 також не корелює ні з y_1 , оскільки $y_1 = \tilde{y}_1$, ні з y_2 , оскільки $y_2 = \tilde{y}_2 + \alpha y_1$. Тоді згідно з (6.5.11) величина

$$\begin{aligned} \hat{P}(y_3|y_2, y_1) &= \frac{\omega_{31}}{\omega_{11}} y_1 + \frac{h_{32}}{h_{22}} \left(y_2 - \frac{\omega_{21}}{\omega_{11}} y_1 \right) = \\ &= \left(\frac{\omega_{31}}{\omega_{11}} - \frac{h_{32}}{h_{22}} \frac{\omega_{21}}{\omega_{11}} \right) y_1 + \frac{h_{32}}{h_{22}} y_2 \end{aligned} \quad (6.5.12)$$

є лінійною проекцією y_3 на y_2 і y_1 , а \tilde{y}_3 – відхилення y_3 від своєї проекції на y_2 і y_1 . MSE лінійної проекції є дисперсія \tilde{y}_3 , яка згідно з (6.5.5) дорівнює d_{33} , тобто

$$\begin{aligned} MSE(\hat{P}(y_3|y_2, y_1)) &= E [(\tilde{y}_3)^2] = \\ &= E \left[(y_3 - \hat{P}(y_3|y_2, y_1))^2 \right] = h_{33} - \frac{h_{32}h_{23}}{h_{22}}. \end{aligned} \quad (6.5.13)$$

Вирази (6.5.9), (6.5.11) дають корисну інформацію для уточнення лінійної проекції. Припустимо, що ми зацікавлені в прогнозі значення y_3 і нехай y_1 – деяка початкова інформація, на якій може ґрунтуватися цей прогноз. Тоді прогноз y_3 , побудований тільки на підставі y_1 , матиме вигляд

$$\hat{P}(y_3|y_1) = \frac{\omega_{31}}{\omega_{11}} y_1.$$

Нехай тепер y_2 містить деяку нову інформацію, за допомогою якої хочемо уточнити цей прогноз. Виразимо прогноз y_2 , побудований тільки на підставі y_1

$$\hat{P}(y_2|y_1) = \frac{\omega_{21}}{\omega_{11}} y_1.$$

Тоді рівняння (6.5.12) можна переписати у вигляді

$$\hat{P}(y_3|y_2, y_1) = \hat{P}(y_3|y_1) + \frac{h_{32}}{h_{22}} [y_2 - \hat{P}(y_2|y_1)]. \quad (6.5.14)$$

Отже, можемо оптимально уточнити початковий прогноз $\hat{P}(y_3|y_1)$ додаванням до нього неврахованої компоненти, яка містить нову інформацію $[y_2 - \hat{P}(y_2|y_1)]$, помножену на (h_{32}/h_{22}) . З (6.5.12) випливає, що множник h_{32}/h_{22} можна інтерпретувати як коефіцієнт при y_2 в лінійній проєкції y_3 на y_2 і y_1 .

Щоб зрозуміти природу множника (h_{32}/h_{22}) , визначимо вектор $\tilde{y}(1)$ так:

$$\tilde{y}(1) \equiv E_1 y, \quad (6.5.15)$$

де E_1 – матриця з (3.7.3), яка має вигляд

$$E_1 = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -\frac{\omega_{21}}{\omega_{11}} & 1 & 0 & \dots & 0 \\ -\frac{\omega_{31}}{\omega_{11}} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -\frac{\omega_{T1}}{\omega_{11}} & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Зазначимо, що коваріаційна матриця вектора $\tilde{y}(1)$ дорівнює

$$E[\tilde{y}(1)(\tilde{y}(1))'] = E[E_1 y y' E_1'] = E_1 \Omega E_1'.$$

З (3.7.4) випливає, що $E_1 \Omega E_1' = H$. Тобто, H інтерпретується як коваріаційна матриця $\tilde{y}(1)$. Підставивши (3.7.3) в (6.5.15), отримаємо

$$\tilde{y}(1) = \begin{pmatrix} y_1 \\ y_2 - \frac{\omega_{21}}{\omega_{11}} y_1 \\ y_3 - \frac{\omega_{31}}{\omega_{11}} y_1 \\ \vdots \\ y_T - \frac{\omega_{T1}}{\omega_{11}} y_1 \end{pmatrix}.$$

Перший елемент $\tilde{y}(1)$ дорівнює y_1 , а j -ий елемент вектора $\tilde{y}(1)$ ($j = 2, 3, \dots, T$) є відхиленням y_j від його проєкції на y_1 . Отже, матриця H – коваріаційна матриця таких відхилень. Тоді h_{22} є MSE проєкції y_2 на y_1

$$h_{22} = E[(y_2 - \hat{P}(y_2|y_1))^2],$$

а значення h_{32} є математичним сподіванням добутку відхилення цієї проєкції і проєкції y_3 на y_1

$$h_{32} = E[(y_3 - \hat{P}(y_3|y_1))(y_2 - \hat{P}(y_2|y_1))].$$

Отож, рівняння (6.5.14) означає, що лінійна проєкція y_3 може бути уточнена за допомогою формули

$$\begin{aligned} \hat{P}(y_3|y_2, y_1) &= \hat{P}(y_3|y_1) + \\ &+ \frac{E[(y_3 - \hat{P}(y_3|y_1))(y_2 - \hat{P}(y_2|y_1))]}{E[(y_2 - \hat{P}(y_2|y_1))^2]} (y_2 - \hat{P}(y_2|y_1)). \end{aligned} \quad (6.5.16)$$

Зокрема, якщо y_1 – стала, то $\hat{P}(y_j|y_1)$ – математичне сподівання y_j : $\hat{P}(y_2|y_1) = \mu_2$ і $\hat{P}(y_3|y_1) = \mu_3$. Тоді рівняння (6.5.16) означатиме таке:

$$\hat{P}(y_3|y_2, 1) = \mu_3 + \frac{\text{Cov}[y_3, y_2]}{\text{Var}[y_2]} (y_2 - \mu_2).$$

Середньоквадратична похибка цього уточнення також може бути обчислена з трикутної факторизації. Згідно з (6.5.5) MSE лінійної проєкції y_3 на y_2 і y_1 можна обчислити таким способом:

$$\begin{aligned} MSE &= E[(y_3 - \hat{P}(y_3|y_2, y_1))^2] = E[\tilde{y}_3^2] = d_{33} = \\ &= h_{33} - \frac{h_{32}h_{23}}{h_{22}}. \end{aligned}$$

У загальному випадку коефіцієнт при y_2 в лінійній проєкції y_i на y_2 і y_1 обчислюється з другого стовпця матриці A . Для будь-якого $i > j$ коефіцієнт при y_j в лінійній проєкції y_i на y_j, y_{j-1}, \dots, y_1 є елементом i -го рядка та j -го стовпця матриці A . Значення d_{ii} дають MSE лінійної проєкції y_i на $y_{i-1}, y_{i-2}, \dots, y_1$.

6.5.2 Уточнення скінченородженого прогнозу для $MA(1)$ процесу

Як приклад застосування цих результатів припустимо, що y_t описується $MA(1)$ процесом

$$y_t = \mu + \varepsilon_t + \theta\varepsilon_{t-1},$$

де ε_t – білий шум з дисперсією σ^2 . Припустимо, що хочемо спрогнозувати значення y_{t+1} на підставі t попередніх спостережень (y_1, y_2, \dots, y_t) . Нехай

$$y' \equiv ((y_1 - \mu) (y_2 - \mu) \dots (y_{t+1} - \mu)),$$

і нехай Ω – коваріаційна матриця вектора y розміром $(t+1) \times (t+1)$

$$\begin{aligned} \Omega &= E[yy'] = \sigma^2 = & (6.5.17) \\ &= \begin{pmatrix} 1 + \theta^2 & \theta & 0 & \dots & 0 \\ \theta & 1 + \theta^2 & \theta & \dots & 0 \\ 0 & \theta & 1 + \theta^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 + \theta^2 \end{pmatrix}. \end{aligned}$$

У розділі 3 було показано, що трикутна факторизація матриці Ω має вигляд

$$\Omega = ADA',$$

де

$$\begin{aligned} A &= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ \theta & 1 & \dots & 0 & 0 \\ \frac{\theta}{1 + \theta^2} & \frac{\theta(1 + \theta^2)}{1 + \theta^2 + \theta^4} & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & \frac{\theta(1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}} & 1 \end{pmatrix}, \\ D &= \sigma^2 \begin{pmatrix} 1 + \theta^2 & 0 & \dots & 0 \\ 0 & \frac{1 + \theta^2 + \theta^4}{1 + \theta^2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1 + \theta^2 + \theta^4 + \dots + \theta^{2(t+1)}}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}} \end{pmatrix}. \end{aligned}$$

Щоб використати трикутну факторизацію для обчислення точного скінченородженого прогнозу, нагадаємо, що j -тий

елемент \tilde{y}_j вектора $\tilde{y} = A^{-1}y$ інтерпретується як віхилення y_j від його лінійної проєкції на попередні значення

$$\tilde{y}_j = y_j - \hat{E}(y_j | y_{j-1}, y_{j-2}, \dots, y_1).$$

Систему рівнянь $A\tilde{y} = y$ можна записати у вигляді

$$\begin{aligned} \tilde{y}_1 &= y_1 - \mu, \\ \frac{\theta}{1 + \theta^2} \tilde{y}_1 + \tilde{y}_2 &= y_2 - \mu, \\ \frac{\theta(1 + \theta^2)}{1 + \theta^2 + \theta^4} \tilde{y}_2 + \tilde{y}_3 &= y_3 - \mu, \\ &\vdots \\ \frac{\theta(1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}} \tilde{y}_t + \tilde{y}_{t+1} &= y_{t+1} - \mu. \end{aligned}$$

Розв'язавши останнє рівняння стосовно \tilde{y}_{t+1} , матимемо

$$\begin{aligned} \tilde{y}_{t+1} &= y_{t+1} - \hat{E}(y_{t+1} | y_t, y_{t-1}, \dots, y_1) = y_{t+1} - \mu - \\ &- \frac{\theta(1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)})}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}} [y_t - \hat{E}(y_t | y_{t-1}, y_{t-2}, \dots, y_1)]. \end{aligned}$$

Звідси випливає, що

$$\begin{aligned} \hat{E}(y_{t+1} | y_t, y_t, \dots, y_1) &= \mu + \\ &+ \frac{\theta[1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)}]}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}} [y_t - \hat{E}(y_t | y_{t-1}, y_{t-2}, \dots, y_1)]. \end{aligned} \quad (6.5.18)$$

MSE цього прогнозу дорівнює $d_{t+1,t+1}$

$$\begin{aligned} MSE[\hat{E}(y_{t+1} | y_t, y_{t-1}, \dots, y_1)] &= \\ &= \sigma^2 \frac{1 + \theta^2 + \theta^4 + \dots + \theta^{2(t+1)}}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}}. \end{aligned} \quad (6.5.19)$$

Похибка цього оптимального прогнозу збільшується з кількістю спостережень t , на яких ґрунтується прогноз. Нехай процес рухомого середнього є зворотним ($|\theta| < 1$). У цьому випадку при $t \rightarrow \infty$ коефіцієнт в (6.5.18) прямує до θ

$$\begin{aligned} \frac{\theta[1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)}]}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}} &= \frac{\theta[1 - \theta^{2t}]/(1 - \theta^2)}{(1 - \theta^{2(t+1)})/(1 - \theta^2)} = \\ &= \theta \frac{1 - \theta^{2t}}{1 - \theta^{2(t+1)}} \rightarrow \theta, \end{aligned} \quad (6.5.20)$$

а MSE (6.5.19) прямує до дисперсії похибки σ^2

$$\sigma^2 \frac{[1 - \theta^{2(t+2)}]/(1 - \theta^2)}{(1 - \theta^{2(t+1)})/(1 - \theta^2)} = \sigma^2 \frac{1 - \theta^{2(t+2)}}{1 - \theta^{2(t+1)}} \rightarrow \sigma^2.$$

Отже, оптимальний прогноз, який ґрунтується на скінченній кількості спостережень (6.5.18), зрештою прямує до прогнозу, який ґрунтується на безмежній кількості спостережень (6.3.16).

Аналогічні обчислення, зроблені в (6.5.18), можна застосувати до незворотного зображення з $|\theta| > 1$. У цьому випадку коефіцієнт в (6.5.18) прямує до θ^{-1}

$$\begin{aligned} \frac{\theta[1 + \theta^2 + \theta^4 + \dots + \theta^{2(t-1)}]}{1 + \theta^2 + \theta^4 + \dots + \theta^{2t}} &= \frac{\theta[1 - \theta^{2t}]/(1 - \theta^2)}{(1 - \theta^{2(t+1)})/(1 - \theta^2)} = \\ &= \frac{\theta(\theta^{-2(t+1)} - \theta^{-2})}{\theta^{-2(t+1)} - 1} \rightarrow \frac{\theta(-\theta^{-2})}{-1} = \theta^{-1}. \end{aligned}$$

Отож, коефіцієнт в (6.5.18) прямує до θ^{-1} у випадку, якщо θ є коефіцієнтом незворотного зображення процесу рухомого середнього. MSE (6.5.19) прямує до $\sigma^2\theta^2$

$$\sigma^2 \frac{[1 - \theta^{2(t+2)}]/(1 - \theta^2)}{(1 - \theta^{2(t+1)})/(1 - \theta^2)} = \sigma^2 \frac{\theta^{-2(t+2)} - 1}{\theta^{-2(t+2)} - \theta^2} \rightarrow \sigma^2\theta^2.$$

Корисно розглянути граничний випадок $\theta = 1$. Оптимальний скінченнопороджений прогноз для $MA(1)$ процесу з $\theta = 1$ згідно з (6.5.18) має вигляд

$$\begin{aligned} \hat{E}(y_{t+1}|y_t, y_{t-1}, \dots, y_1) = & \quad (6.5.21) \\ & + \frac{t}{t+1} [y_t - \hat{E}(y_t|y_{t-1}, y_{t-2}, \dots, y_1)], \end{aligned}$$

який після послідовної заміни можна перетворити так:

$$\begin{aligned} \hat{E}(y_{t+1}|y_t, y_{t-1}, \dots, y_1) = & \quad (6.5.22) \\ = & \mu + \frac{t}{t+1} \left[y_t - \left(\mu + \frac{t-1}{t} \left[y_{t-1} - \hat{E}(y_{t-1}|y_{t-2}, \dots, y_1) \right] \right) \right] = \\ = & \mu + \frac{t}{t+1} [y_t - \mu] - \frac{t-1}{t+1} \left[y_{t-1} - \hat{E}(y_{t-1}|y_{t-2}, y_{t-3}, \dots, y_1) \right] = \\ = & \mu + \frac{t}{t+1} (y_t - \mu) - \frac{t-1}{t+1} (y_{t-1} - \mu) + \\ & + \frac{t-2}{t+1} (y_{t-2} - \mu) - \dots + (-1)^{t+1} \frac{1}{t+1} (y_1 - \mu). \end{aligned}$$

MSE цього прогнозу обчислюється з (6.5.19)

$$\sigma^2 \frac{t+2}{t+1} \rightarrow \sigma^2.$$

Отож, дисперсія похибки прогнозу знову прямує до дисперсії ε_t . Зазначимо, що є різниця між оптимальним прогнозом (6.5.21) і прогнозом, який ґрунтується на застосуванні (6.4.3) при $\theta = 1$

$$\mu + (y_t - \mu) - (y_{t-1} - \mu) + \dots + (-1)^{t+1} (y_1 - \mu). \quad (6.5.23)$$

Наближення (6.4.3) було отримане за припущення, що процес рухомого середнього має зворотне зображення, а для граничного випадку $\theta = 1$ MA поліном не є зворотним. Для цього

випадку (6.5.22) не збігається до прогнозу (6.5.18) при зростанні t . Коли $\theta = 1$, то

$$y_t = \mu + \varepsilon_t + \varepsilon_{t-1}$$

і (6.5.23) може бути записане у вигляді

$$\begin{aligned} \hat{E}(y_{t+1}|y_t, y_{t-1}, \dots, y_1) &= \\ &= \mu + (\varepsilon_t + \varepsilon_{t-1}) - (\varepsilon_{t-1} + \varepsilon_{t-2}) + \dots + (-1)^{t+1}(\varepsilon_1 + \varepsilon_0) = \\ &= \mu + \varepsilon_t + (-1)^{t+1}\varepsilon_0. \end{aligned}$$

Різниця між прогнозом і безумовним математичним сподіванням дорівнює $\varepsilon_t + (-1)^{t+1}\varepsilon_0$, яка має MSE $2\sigma^2$ для всіх t . Отже, тоді як (6.5.18) збігається до оптимального прогнозу при $t \rightarrow \infty$, то (6.5.22) не збігається.

6.5.3 Блочна трикутна факторизація

Припустимо, що маємо спостереження за двома множинами змінної. Перша множина змінної зібрана у вектор Y_1 розмірності n_1 , а друга множина змінної зібрана у вектор Y_2 розмірності n_2 . Їхню коваріаційну матрицю можна записати в такому вигляді:

$$\Omega \equiv \begin{pmatrix} E[Y_1 Y_1'] & E[Y_1 Y_2'] \\ E[Y_2 Y_1'] & E[Y_2 Y_2'] \end{pmatrix} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix},$$

де Ω_{11} – матриця розміром $n_1 \times n_1$, Ω_{22} – матриця розміром $n_2 \times n_2$, а матриця Ω_{12} розміром $n_1 \times n_2$ є транспонованою матрицею до Ω_{21} .

Можемо зробити, щоб усі елементи лівого знизу ($n_2 \times n_1$) - блоку матриці Ω були нульовими. Для цього використаємо матрицю

$$\bar{E}_1 = \begin{pmatrix} I_{n_1} & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & I_{n_2} \end{pmatrix}.$$

Помноживши Ω зліва на \bar{E}_1 і справа на \bar{E}'_1 , отримаємо

$$\begin{aligned} \begin{pmatrix} I_{n_1} & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & I_{n_2} \end{pmatrix} \begin{pmatrix} \Omega_{11} & \Omega_{21} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} I_{n_1} & -\Omega_{11}^{-1}\Omega_{12} \\ 0 & I_{n_2} \end{pmatrix} = \\ = \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \end{pmatrix}. \end{aligned} \quad (6.5.24)$$

Позначимо

$$\bar{A} \equiv \bar{E}_1^{-1} = \begin{pmatrix} I_{n_1} & 0 \\ \Omega_{21}\Omega_{11}^{-1} & I_{n_2} \end{pmatrix}.$$

Якщо (6.5.24) помножити зліва на \bar{A} і справа на \bar{A}' , то матимемо

$$\begin{aligned} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} = \begin{pmatrix} I_{n_1} & 0 \\ \Omega_{21}\Omega_{11}^{-1} & I_{n_2} \end{pmatrix} \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \end{pmatrix} \times \\ \times \begin{pmatrix} I_{n_1} & \Omega_{11}^{-1}\Omega_{12} \\ 0 & I_{n_2} \end{pmatrix} = \bar{A}\bar{D}\bar{A}'. \end{aligned} \quad (6.5.25)$$

Зображення є подібним до трикутної факторизації матриці $\Omega = ADA'$, тільки тепер \bar{D} – блочно-діагональна матриця

$$\bar{D} = \begin{pmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \end{pmatrix}.$$

Як і в попередньому випадку, \bar{D} можна інтерпретувати як коваріаційну матрицю вектора $\tilde{Y} = \bar{A}^{-1}Y$

$$\begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix} = \begin{pmatrix} I_{n_1} & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & I_{n_2} \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix},$$

тобто $\tilde{Y}_1 = Y_1$, $\tilde{Y}_2 = Y_2 - \Omega_{21}\Omega_{11}^{-1}Y_1$ і i -тий елемент \tilde{Y}_2 дорівнює y_{2i} мінус лінійна комбінація елементів Y_1 . Блочна діагональність \bar{D} означає, що математичне сподівання добутку будь-якого елемента \tilde{Y}_2 з будь-яким елементом \tilde{Y}_1 , а отже, і з Y_1 ,

дорівнює нулю. Тому $\Omega_{21}\Omega_{11}^{-1}$ є матрицею коефіцієнтів лінійної проєкції вектора Y_2 на вектор Y_1

$$\hat{P}(Y_2|Y_1) = \Omega_{21}\Omega_{11}^{-1}Y_1. \quad (6.5.26)$$

Ці результати легко узагальнити на випадок трьох векторів. Нехай Y_1, Y_2, Y_3 – вектори розмірності n_1, n_2 і n_3 . Блочнотрикутну факторизацію їх коваріаційної матриці отримують з узагальнення рівняння (3.7.13)

$$\begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix} = \begin{pmatrix} I_{n_1} & 0 & 0 \\ \Omega_{21}\Omega_{11}^{-1} & I_{n_2} & 0 \\ \Omega_{31}\Omega_{11}^{-1} & H_{32}H_{22}^{-1} & I_{n_3} \end{pmatrix} \times \quad (6.5.27)$$

$$\times \begin{pmatrix} \Omega_{11} & 0 & 0 \\ 0 & H_{22} & 0 \\ 0 & 0 & H_{33} - H_{32}H_{22}^{-1}H_{23} \end{pmatrix} \begin{pmatrix} I_{n_1} & \Omega_{11}^{-1}\Omega_{12} & \Omega_{11}^{-1}\Omega_{13} \\ 0 & I_{n_2} & H_{22}^{-1}H_{23} \\ 0 & 0 & I_{n_3} \end{pmatrix},$$

де

$$H_{22} = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}, \quad H_{33} = \Omega_{33} - \Omega_{31}\Omega_{11}^{-1}\Omega_{13},$$

$$H_{23} = H'_{32} = \Omega_{23} - \Omega_{21}\Omega_{11}^{-1}\Omega_{13}.$$

Це спонукає нас узагальнити попередній результат (6.5.12) на уточнення лінійної проєкції. Оптимальний прогноз Y_3 , який ґрунтується на Y_2 і Y_1 , можна отримати з рядка останнього блока матриці \bar{A}

$$\begin{aligned} \hat{P}(Y_3|Y_2, y_1) &= \Omega_{31}\Omega_{11}^{-1}Y_1 + H_{32}H_{22}^{-1}(Y_2 - \Omega_{21}\Omega_{11}^{-1}Y_1) = \\ &= \hat{P}(Y_3|Y_1) + H_{32}H_{22}^{-1}[Y_2 - \hat{P}(Y_2|Y_1)], \end{aligned} \quad (6.5.28)$$

де

$$H_{22} = E \left[(Y_2 - \hat{P}(Y_2|Y_1)) (Y_2 - \hat{P}(Y_2|Y_1))' \right],$$

$$H_{32} = E \left[(Y_3 - \hat{P}(Y_3|Y_1)) (Y_2 - \hat{P}(Y_2|Y_1))' \right].$$

MSE цього прогнозу є матричним узагальненням (6.5.13)

$$\begin{aligned} E \left[(Y_3 - \hat{P}(Y_3|Y_2, Y_1)) (Y_3 - \hat{P}(Y_3|y_2, Y_1))' \right] &= \quad (6.5.29) \\ &= H_{33} - H_{32}H_{22}^{-1}H_{23}, \end{aligned}$$

де

$$H_{33} = E \left[(Y_3 - \hat{P}(Y_3|Y_1)) (Y_3 - \hat{P}(Y_3|Y_1))' \right].$$

Правило ітераційних проєкцій

Інший корисний результат, правило ітераційних проєкцій можна вивести безпосередньо з (6.5.28). Визначимо, що відбудеться, якщо проєкцію $\hat{P}(y_3|Y_2, Y_1)$ спроектувати на Y_1 . Правило ітераційних проєкцій стверджує, що ця проєкція дорівнює звичайній проєкції Y_3 на Y_1

$$\hat{P}(\hat{P}(Y_3|Y_2, Y_1)|Y_1) = \hat{P}(Y_3|Y_1). \quad (6.5.30)$$

Щоб перевірити це твердження, нам треба показати, що різниця між $\hat{P}(Y_3|Y_2, Y_1)$ і $\hat{P}(Y_3|Y_1)$ не корелює з Y_1 . Але з (6.5.28) ця різниця дорівнює

$$\hat{P}(Y_3|Y_2, Y_1) - \hat{P}(Y_3|Y_1) = H_{32}H_{22}^{-1}[Y_2 - \hat{P}(Y_2|Y_1)],$$

яка справді не корелює з Y_1 згідно з означенням лінійної проєкції $\hat{P}(Y_2|Y_1)$.

6.6 Оптимальні прогнози для Гауссівських процесів

Формули прогнозування, які розглянуто в цьому розділі, описують оптимальні прогнози в класі лінійних функцій від змінних, на яких ґрунтуються ці прогнози. Для Гауссівських процесів можна отримати сильніший результат.

Лема 6.6.1. Якщо серед змінних, на яких ґрунтується прогноз, є константа, то оптимальний MSE - прогноз має лінійну форму, і отже, він є лінійною проекцією.

Д о в е д е н н я. Нехай Y_1 – вектор розмірності n_1 з математичним сподіванням μ_1 , Y_2 – вектор розмірності n_2 з математичним сподіванням μ_2 , а їхня коваріаційна матриця Ω має вигляд

$$\begin{pmatrix} E(Y_1 - \mu_1)(Y_1 - \mu_1)' & E(Y_1 - \mu_1)(Y_2 - \mu_2)' \\ E(Y_2 - \mu_2)(Y_1 - \mu_1)' & E(Y_2 - \mu_2)(Y_2 - \mu_2)' \end{pmatrix} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

Якщо Y_1 і Y_2 є Гауссівськими, то їхня сукупна щільність дорівнює

$$\begin{aligned} f(Y_1, Y_2) &= \frac{1}{(2\pi)^{(n_1+n_2)/2}} \left| \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right|^{-1/2} \times \\ &\times \exp \left\{ -\frac{1}{2} ((Y_1 - \mu_1)'(Y_2 - \mu_2)') \times \right. \\ &\quad \left. \times \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}^{-1} \begin{pmatrix} Y_1 - \mu_1 \\ Y_2 - \mu_2 \end{pmatrix} \right\}. \end{aligned} \quad (6.6.1)$$

Обернену до Ω знаходимо з зображення (6.5.25)

$$\begin{aligned} \Omega^{-1} &= (\bar{A}\bar{D}\bar{A}')^{-1} = (\bar{A}')^{-1}\bar{D}^{-1}\bar{A}^{-1} = \\ &= \begin{pmatrix} I_{n_1} & -\Omega_{11}^{-1}\Omega_{12} \\ 0 & I_{n_2} \end{pmatrix} \begin{pmatrix} \Omega_{11}^{-1} & 0 \\ 0 & (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} \end{pmatrix} \times \\ &\times \begin{pmatrix} I_{n_1} & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & I_{n_2} \end{pmatrix}. \end{aligned} \quad (6.6.2)$$

Визначник Ω дорівнює

$$|\Omega| = |\bar{A}||\bar{D}||\bar{A}'|.$$

Оскільки \bar{A} – нижня трикутна матриця, то її визначник дорівнює добутку елементів головної діагоналі, які всі дорівнюють

1. Тобто, $|\bar{A}| = 1$ і $|\Omega| = |\bar{D}|$. Тоді

$$\begin{aligned} |\Omega| &= \begin{vmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{vmatrix} = \begin{vmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12} \end{vmatrix} = \\ &= |\Omega_{11}| |\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}|. \end{aligned} \tag{6.6.3}$$

Підставивши (6.6.2) і (6.6.3) в (6.6.1), отримаємо щільність розподілу

$$\begin{aligned} f(Y_1, Y_2) &= \frac{1}{(2\pi)^{(n_1+n_2)/2}} |\Omega_{11}|^{-1/2} |\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}|^{-1/2} \times \\ &\times \exp \left\{ -\frac{1}{2} ((Y_1 - \mu_1)' \quad (Y_2 - \mu_2)') \begin{pmatrix} I_{n_1} & -\Omega_{11}^{-1}\Omega_{12} \\ 0 & I_{n_2} \end{pmatrix} \times \right. \\ &\quad \times \begin{pmatrix} \Omega_{11}^{-1} & 0 \\ 0 & (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} \end{pmatrix} \times \\ &\quad \left. \times \begin{pmatrix} I_{n_1} & 0 \\ -\Omega_{21}\Omega_{11}^{-1} & I_{n_2} \end{pmatrix} \begin{pmatrix} Y_1 - \mu_1 \\ Y_2 - \mu_2 \end{pmatrix} \right\} = \\ &= \frac{1}{(2\pi)^{(n_1+n_2)/2}} |\Omega_{11}|^{-1/2} |\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}|^{-1/2} \times \\ &\times \exp \left\{ -\frac{1}{2} ((Y_1 - \mu_1)' \quad (Y_2 - m)') \times \right. \\ &\quad \left. \times \begin{pmatrix} \Omega_{11}^{-1} & 0 \\ 0 & (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} \end{pmatrix} \begin{pmatrix} Y_1 - \mu_1 \\ Y_2 - m \end{pmatrix} \right\} = \\ &= \frac{1}{(2\pi)^{(n_1+n_2)/2}} |\Omega_{11}|^{-1/2} |\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12}|^{-1/2} \times \tag{6.6.4} \\ &\quad \times \exp \left\{ -\frac{1}{2} (Y_1 - \mu_1)' \Omega_{11}^{-1} (Y_1 - \mu_1) - \right. \\ &\quad \left. -\frac{1}{2} (Y_2 - m)' (\Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{12})^{-1} (Y_2 - m) \right\}, \end{aligned}$$

де

$$m \equiv \mu_2 + \Omega_{21}\Omega_{11}^{-1}(Y_1 - \mu_1).$$

Умовну щільність Y_2 за даним Y_1 знаходять діленням щільності (6.6.4) на щільність

$$f(Y_1) = \frac{1}{(2\pi)^{n_1/2}} |\Omega_{11}|^{-1/2} \exp \left\{ -\frac{1}{2} (Y_1 - \mu_1)' \Omega_{11}^{-1} (Y_1 - \mu_1) \right\}.$$

У результаті отримаємо

$$\begin{aligned} f(Y_2|Y_1) &= \frac{f(Y_1, Y_2)}{f(Y_1)} && (6.6.5) \\ &= \frac{1}{(2\pi)^{n_2/2}} |H|^{-1/2} \exp \left\{ -\frac{1}{2} (Y_2 - m)' H^{-1} (Y_2 - m) \right\}, \end{aligned}$$

де

$$H \equiv \Omega_{22} - \Omega_{21} \Omega_{11}^{-1} \Omega_{12}. \quad (6.6.6)$$

Іншими словами,

$$\begin{aligned} Y_2|Y_1 &\sim N[m, H] \sim && (6.6.7) \\ &\sim N[(\mu_2 + \Omega_{21} \Omega_{11}^{-1} (Y_1 - \mu_1)), (\Omega_{22} - \Omega_{21} \Omega_{11}^{-1} \Omega_{12})]. \end{aligned}$$

У пункті 6.6.1 було показано, що оптимальним MSE - прогнозом є умовне математичне сподівання. Для Гауссівського процесу оптимальний MSE - прогноз дорівнює

$$E[Y_2|Y_1] = \mu_2 + \Omega_{21} \Omega_{11}^{-1} (Y_1 - \mu_1).$$

З іншого боку, для будь-якого розподілу лінійною проекцією вектора Y_2 на вектор Y_1 і константу є величина

$$\hat{E}(Y_2|Y_1) = \mu_2 + \Omega_{21} \Omega_{11}^{-1} (Y_1 - \mu_1).$$

Отже, для Гауссівського процесу лінійна проекція є оптимальним MSE - прогнозом. \square

Розділ 7

Нестационарні часові ряди

7.1 Процеси з детермінованим часовим трендом і процеси з одиничним коренем

Модель стаціонарного часового ряду можна записати у вигляді

$$y_t = \mu + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \mu + \psi(L) \varepsilon_t, \quad (7.1.1)$$

причому $\sum_{j=0}^{\infty} |\psi_j| < \infty$, корені характеристичного рівняння $\psi(z) = 0$ лежать ззовні одиничного круга, а ε_t є білим шумом з нульовим математичним сподіванням і дисперсією σ^2 .

Такий стаціонарний процес має дві важливі властивості. По-перше, безумовне математичне сподівання є сталою, яка не залежить від даних

$$E[y_t] = \mu.$$

По-друге, якщо прогнозувати ряд в майбутнє, то прогнозоване значення

$$\hat{y}_{t+h|t} = \hat{E}[y_{t+h} | y_t, y_{t-1}, \dots]$$

при $h \rightarrow \infty$ збігається до безумовного математичного сподівання

$$\lim_{h \rightarrow \infty} \hat{y}_{t+h|t} = \mu.$$

Ці припущення не завжди виконуються на практиці для багатьох економічних і фінансових часових рядів. Зокрема, немає сумніву, що рівень валового внутрішнього продукту країни від часів Другої світової війни зріс, а тому ряд, який описує цей показник, відхилився вгору протягом часу. Таке відхилення повинно відобразитись у теоретичній моделі і впливати на будь-які прогнози цього ряду.

Існують дві відомі специфікації для опису таких відхилень. Перша з них містить детермінований часовий тренд

$$y_t = \alpha + \delta t + \psi(L) \varepsilon_t. \quad (7.1.2)$$

У цьому випадку математичне сподівання μ стаціонарного процесу (7.1.1) замінюється лінійною функцією від t . Оцінка параметра δ дає середній абсолютний приріст y за одиницю часу. Такий процес називатимемо трендово-стаціонарним, оскільки після виділення з (7.1.2) часового тренду δt , отримуємо стаціонарний процес.

Тренд може бути визначений як "загальний рух ряду в певному напрямі". Це поняття відображає довготермінові зміни, які покладено в основу процесу, на відміну від тимчасових коливань. Наприклад, національний дохід зростає в результаті дії таких факторів, як технічний прогрес, зростання населення та зростання інвестицій в інфраструктуру; об'єм продажу нового товару збільшується в міру того, як споживачі дізнаються про його існування через рекламу або іншим шляхом. Оскільки названі фактори важко виміряти, то їх можна подати у вигляді незалежної змінної t , яка зростає на одиницю з кожним наступним спостереженням.

Другою специфікацією є процес з одиничним коренем

$$(1 - L) y_t = \delta + \psi(L) \varepsilon_t, \quad (7.1.3)$$

де $\psi(1) \neq 0$. Нагадаємо, що для процесу з одиничним коренем (7.1.3) перші різниці процесу Δy_t описуються стаціонарним процесом вигляду (7.1.1). В означенні процесу з одиничним коренем ми припускаємо, що $\psi(1)$ є відмінним від нуля, де $\psi(1)$ позначає значення полінома

$$\psi(z) = 1 + \psi_1 z + \psi_2 z^2 + \psi_3 z^3 + \dots$$

у точці $z = 1$. Щоб зрозуміти для чого ця умова введена в означення ряду з одиничним коренем, припустимо, що початковий ряд є стаціонарним і має зображення

$$y_t = \mu + \chi(L) \varepsilon_t.$$

Якщо обчислити різниці такого ряду, то матимемо

$$(1 - L) y_t = (1 - L) \chi(L) \varepsilon_t \equiv \psi(L) \varepsilon_t,$$

де $\psi(L) = (1 - L) \chi(L)$. Тобто, якщо початковий ряд y_t стаціонарний, то оператор рухомого середнього для ряду Δy_t має одиничний корінь, тобто є незворотним. Так формулюючи умову $\psi(1) \neq 0$, виключаємо можливість того, що початковий ряд y_t є стаціонарним.

Найпростішим процесом з одиничним коренем є процес

$$y_t = \delta + y_{t-1} + \varepsilon_t,$$

який має назву процесу випадкового блукання зі зміщенням δ .

На рис.7.1.1., 7.1.2. зображено динаміку трендово - стаціонарного процесу і процесу випадкового блукання зі зміщенням.

Рис. 7.1.1. $y_t = 0,8 + y_{t-1} + \varepsilon_t$

Трендово - стаціонарний процес і процес з одиничним коренем можна об'єднати за допомогою такої специфікації:

$$y_t = \alpha + \delta t + u_t, \quad (7.1.4)$$

де похибка u_t описується стаціонарним $ARMA(p, q)$ процесом з нульовим математичним сподіванням і зворотним оператором рухомого середнього

$$\begin{aligned} (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) u_t &= \\ &= (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t. \end{aligned} \quad (7.1.5)$$

Розкладемо авторегресійний оператор в (7.1.5) на множники

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) = (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L).$$

Якщо всі значення λ_i лежать всередині одиничного круга, то існує обернений оператор і (7.1.5) можна записати у вигляді

$$u_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{(1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L)} \varepsilon_t \equiv \psi(L) \varepsilon_t,$$

Рис. 7.1.2. $y_t = 0,1 + 0,8t + \varepsilon_t$

причому $\sum_{j=0}^{\infty} |\psi_j| < \infty$ і корені характеристичного рівняння $\psi(z) = 0$ лежать ззовні одиничного круга. Отже, якщо $|\lambda_i| < 1$ для всіх $i = 1, \dots, p$, то процес (7.1.4) набуде вигляду

$$y_t = \alpha + \delta t + \psi(L) \varepsilon_t, \quad (7.1.6)$$

тобто буде трендово - стаціонарним процесом.

Нехай тепер $\lambda_1 = 1$ і $|\lambda_i| < 1$ для $i = 2, 3, \dots, p$. Тоді (7.1.5) матиме вигляд

$$\begin{aligned} (1 - L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L) u_t &= \\ &= (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q). \end{aligned} \quad (7.1.7)$$

Звідси випливає, що

$$(1 - L) u_t = \frac{1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q}{(1 - \lambda_2 L) \cdots (1 - \lambda_p L)} \varepsilon_t \equiv \psi^*(L) \varepsilon_t,$$

де $\sum_{j=0}^{\infty} |\psi_j^*| < \infty$ і корені характеристичного рівняння $\psi^*(z) = 0$ лежать ззовні одиничного круга.

Отже, якщо взяти перші різниці ряду (7.1.4), то отримаємо ряд

$$\begin{aligned}(1 - L)y_t &= (1 - L)\alpha + (\delta t - \delta(t - 1)) + (1 - L)u_t = \\ &= \delta + \psi^*(L)\varepsilon_t,\end{aligned}$$

який є процесом з одиничним коренем.

Логарифмічно - лінійні часові тренди

Може виникнути питання, чому відхилення в (7.1.2) беруть як лінійну функцію часу δt , а не квадратичну $\delta t + \gamma t^2$ або експоненціальну $e^{\delta t}$. Насправді, зростання реальних даних багатьох економічних і фінансових часових рядів краще описується за допомогою експоненціального відхилення, оскільки експоненціальне відхилення має сталий темп росту.

Нагадаємо, що темп росту визначають як $r = y'/y$. Зокрема, населення збільшується зі сталим темпом росту, якщо відношення кількості новонароджених дітей до поточного населення буде сталим.

Обчисливши темп росту r для ряду, який описується за допомогою експоненціального відхилення

$$y_t = e^{\delta t},$$

матимемо, що $r = \delta$. Отже, якщо дані характеризуються сталим темпом росту, то для опису тренда треба використати експоненціальну функцію часу.

Експоненціальне зростання часто підтверджується візуальним виглядом ряду. З цієї причини багато економістів зазвичай припускають, що зростання має експоненціальну форму.

Зауважимо таке: якщо обчислити натуральні логарифми від даних y_t , які мають експоненціальне зростання, то ряд $\ln(y_t)$ матиме лінійне зростання, тобто описуватиметься лінійним трендом

$$\ln y_t = \delta t.$$

Тому, загальноприйнятим є використання натуральних логарифмів даних перед тим як моделювати їх за допомогою лінійного часового тренда.

Подібні аргументи застосовують також при моделюванні натуральних логарифмів даних за допомогою процесу з одиничним коренем. Це пояснюється тим, що перші різниці логарифма змінних приблизно дорівнюють відносній зміні змінної

$$(1 - L) \ln(y_t) = \ln\left(\frac{y_t}{y_{t-1}}\right) = \ln\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right) \approx \frac{y_t - y_{t-1}}{y_{t-1}}.$$

Тут ми використали те, що для x близького до нуля $\ln(1 + x) \approx x$. Отже, якщо описуємо логарифм змінної як процес з одиничним коренем, то при цьому припускаємо, що відносна зміна ряду є стаціонарним стохастичним процесом.

7.2 Тести на одиничний корінь

Дуже важливо знати чи нестационарність у даних залежить від детермінованого часового тренда чи від наявності одиничного кореня. Зокрема, макроекономісти завжди цікавляться чи економічний спад є постійним (що відповідає часовому тренду), чи це тимчасовий спад з втраченим випуском, який випадково виник (що відповідає процесу з одиничним коренем). Нельсон і Плосер (1982) показали, що багато економічних рядів краще описуються процесом з одиничним коренем, а ніж детермінованим часовим трендом.

Розглянемо питання: чи часовий ряд має одиничний корінь. З деяких публікацій видно, що на це питання неможливо відповісти на підставі скінченої вибірки. Одним з аргументів цього є те, що для будь-якого процесу з одиничним коренем

існує стаціонарний процес, який неможливо відрізнити від зображення з одиничним коренем для довільної скінченої вибірки обсягом T . Такий процес знаходимо досить легко, якщо прийняти один з характеристичних коренів близьким до одиниці, наприклад, 0,9999.

Нехай вибірка містить T спостережень, які були згенеровані процесом випадкового блукання без зміщення

$$y_t = y_{t-1} + \varepsilon_t. \quad (7.2.1)$$

достовірний процес (з одиничним коренем)

Порівняємо модель (7.2.1) з таким стаціонарним процесом:

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1. \quad (7.2.2)$$

недостовірна модель (стаціонарна)

З формули (6.3.3) випливає, що прогноз на h періодів вперед для (7.2.1) має вигляд

$$\hat{y}_{t+h|t} = y_t, \quad (7.2.3)$$

а відповідна середньоквадратична похибка цього прогнозу дорівнює

$$MSE = E [(y_{t+h} - \hat{y}_{t+h|t})^2] = h \sigma^2. \quad (7.2.4)$$

Відповідним прогнозом для (7.2.2) є

$$\hat{y}_{t+h|t} = \phi^h y_t \quad (7.2.5)$$

з

$$\begin{aligned} MSE &= E [(y_{t+h} - \hat{y}_{t+h|t})^2] = \\ &= (1 + \phi^2 + \phi^4 + \dots + \phi^{2(h-1)}) \sigma^2. \end{aligned} \quad (7.2.6)$$

Існує величина ϕ достатньо близька до одиниці, за якої результати прогнозування (7.2.5), (7.2.6) для стаціонарного

зображення будуть близькими до результатів прогнозування (7.2.3), (7.2.4) процесу з одиничним коренем для скінченного обсягу вибірки T .

Загалом функція правдоподібності для Гауссівського процесу є неперервною за параметром λ_1 . Тому для довільної специфікації з одиничним коренем $\lambda_1 = 1$ для заданого фіксованого обсягу вибірки T і будь-яких малих чисел η і ε існує стаціонарна специфікація з $\lambda_1 < 1$ така, що ймовірність того, що величина функції правдоподібності процесу з одиничним коренем відрізняється від значення функції правдоподібності стаціонарної специфікації на величину η є меншою за ε .

Правильне також обернене твердження: для будь-якого стаціонарного процесу і заданого обсягу вибірки T існує процес з одиничним коренем, який неможливо відрізнити від стаціонарного зображення.

Наприклад, припустимо, що достовірним є процес білого шуму

$$y_t = \varepsilon_t. \quad (7.2.7)$$

достовірна модель (стаціонарна)

Порівняємо її з процесом з одиничним коренем

$$(1 - L)y_t = (1 + \theta L)\varepsilon_t, \quad |\theta| < 1, \quad y_0 = \varepsilon_0 = 0. \quad (7.2.8)$$

недостовірна модель (з одиничним коренем)

Прогноз на h періодів вперед для (7.2.7) має вигляд

$$\hat{y}_{t+h|t} = 0 \quad (7.2.9)$$

з

$$MSE = E[(y_{t+h} - \hat{y}_{t+h|t})^2] = \sigma^2. \quad (7.2.10)$$

Відповідним прогнозом для (7.2.8) буде величина

$$\begin{aligned}\hat{y}_{t+h|t} &= y_t + \theta \varepsilon_t = \\ &= (\Delta y_t + \Delta y_{t-1} + \dots + \Delta y_2 + y_1) + \theta \varepsilon_t = \\ &= [(\varepsilon_t + \theta \varepsilon_{t-1}) + (\varepsilon_{t-1} + \theta \varepsilon_{t-2}) + \dots + (\varepsilon_2 + \theta \varepsilon_1) + \varepsilon_1] + \theta \varepsilon_t = \\ &= (1 + \theta)(\varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1).\end{aligned}\tag{7.2.11}$$

Середня квадратична похибка у цьому випадку матиме вигляд

$$\begin{aligned}MSE &= E[(y_{t+h} - \hat{y}_{t+h|t})^2] = \\ &= (1 + (h-1)(1+\theta)^2) \sigma^2.\end{aligned}\tag{7.2.12}$$

Знову бачимо, що для заданого фіксованого обсягу вибірки T існує величина θ достатньо близька до -1 така, що процес з одиничним коренем (7.2.8) буде давати майже такий самий результат, як і стаціонарний процес (7.2.7).

Порівняємо процес з одиничним коренем і стаціонарний процес на нескінченному часовому проміжку. Для цього розглянемо автоковаріаційну генеруючу функцію.

Для процесу $(1-L)y_t$ вона матиме вигляд

$$g_{\Delta y}(z) = \psi(z) \sigma^2 \psi(z^{-1}),$$

причому в точці $z = 1$ набуде значення

$$g_{\Delta y}(1) = (\psi(1))^2 \sigma^2.\tag{7.2.13}$$

Враховуючи, що спектр Δy з частотою ω визначається формулою

$$s_{\Delta y}(\omega) = \frac{1}{2\pi} g_{\Delta y}(e^{-i\omega}),$$

вираз (7.2.13) можна записати у вигляді

$$g_{\Delta y}(1) = 2\pi s_{\Delta y}(0),$$

де $s_{\Delta y}(0)$ спектр з нульовою частотою

$$s_{\Delta y}(0) = \frac{1}{2\pi} (\phi(1))^2 \sigma^2.$$

Для трендово - стаціонарного зображення (7.1.2) автокореляційна генеруюча функція для Δy може бути обчислена за формулою

$$g_{\Delta y}(z) = (1 - z)\psi(z) \sigma^2 \psi(z^{-1})(1 - z^{-1})$$

і в точці $z = 1$ дорівнює нулю.

Отже, якщо достовірний процес є трендово - стаціонарним, то спектр Δy з нульовою частотою – нульовий. Якщо ж достовірний процес характеризується процесом з одиничним коренем, то такий спектр додатний. Отже, для того щоб перевірити чи процес містить одиничний корінь, треба визначити чи спектр Δy з нульовою частотою є нульовим. Проте у вибірці обсягом T немає інформації про круги з періодом більшим за T , тому практично визначити це неможливо.

Отже, як бачимо, процес з одиничним коренем і близький до нього стаціонарний процес можна розрізнити на нескінченному часовому проміжку, але для будь-якої скінченної кількості спостережень часового ряду існує представник з іншого класу моделей, який дає тотожні результати. Тому необхідно бути уважним, приймаючи гіпотезу про одиничний корінь.

Для подання тестових процедур на одиничний корінь ми зосередимось на авторегресійних моделях. Це можна зробити, оскільки $ARMA$ модель завжди має AR -зображення, якщо припустити, що MA -поліном є зворотним.

Детальні теоретичні виведення асимптотичних розподілів статистик буде досліджено у розділі 9. Тут ми розглянемо практичні застосування таких тестів.

7.2.1 Тестування одиничного кореня

в AR(1) моделі

Розглянемо $AR(1)$ процес з нульовим математичним сподіванням

$$y_t = \phi y_{t-1} + \varepsilon_t. \quad (7.2.14)$$

Тест на одиничний корінь – це тестування гіпотези $H_0 : \phi = 1$. Зазвичай використовують МНК оцінку $\hat{\phi}$ і відповідну стандартну похибку.

Для тестування гіпотези використовуємо статистику

$$\hat{\tau} = \frac{\hat{\phi} - 1}{s_{\hat{\phi}}},$$

де $s_{\hat{\phi}}$ позначає стандартну МНК-похибку оцінки $\hat{\phi}$.

Статистика τ є стандартним t -відношенням, проте вона не має t розподілу навіть асимптотично. Причиною цього є те, що нестационарність процесу спотворює МНК оцінку. Розподіл статистики $\hat{\tau}$ нестандартний, тому ми не можемо взяти ні критичні значення розподілу Стьюдента, ні критичні значення нормального розподілу. Розподіл $\hat{\tau}$ асиметричний справа, а його критичні значення менші за критичні значення для нормального наближення.

Для рівня значущості $\alpha = 0,05$ при тестуванні гіпотези $H_0 : \phi = 1$ (одиничний корінь) проти $H_1 : |\phi| < 1$ (стаціонарність) відповідним критичним значенням буде $-1,95$, тоді як для нормального наближення $-1,96$. Якщо використовувати стандартні t -таблиці, то можна відкидати одиничні корені занадто часто. В таблиці 7.2.1 наведено 1 і 5 процентні критичні значення цього тесту, який відомий як тест Дікі - Фуллера (Dickey-Fuller).

Таблиця 7.2.1.

Розмір вибірки	Без константи без тренда		З константою без тренда		З константою з трендом	
	1%	5%	1%	5%	1%	5%
$T = 25$	-2,66	-1,95	-3,75	-3,00	-4,38	-3,60
$T = 50$	-2,62	-1,95	-3,58	-2,93	-4,15	-3,50
$T = 100$	-2,58	-1,95	-3,51	-2,89	-4,04	-3,45
$T = 250$	-2,58	-1,95	-3,46	-2,88	-3,99	-3,43
$T = 500$	-2,58	-1,95	-3,44	-2,87	-3,98	-3,42
$T = \infty$	-2,58	-1,95	-3,43	-2,86	-3,96	-3,41

Для тестування одиничного кореня використовують також більш загальноприйнятну регресійну процедуру. В цьому випадку модель переписують у вигляді

$$\Delta y_t = (\phi - 1)y_{t-1} + \varepsilon_t,$$

в якій t - статистика для $\phi = 1$ збігається з попередньою статистикою $\hat{\tau}$. Причина цього полягає в тому, що МНК інваріантний стосовно лінійних перетворень моделі.

При виконанні гіпотези H_0 часовий ряд y_t описує процес випадкового блукання $y_t = y_{t-1} + \varepsilon_t$, тоді як в протилежному випадку він є стаціонарною авторегресійною моделлю першого порядку з нульовим математичним сподіванням (7.2.14).

Якщо розглянути ряд y_t , який має багато ненульових значень, то треба ввести константу в регресію Дікі-Фуллера. Оскільки константа в стаціонарному $AR(1)$ процесі задовольняє рівність $c = (1 - \phi)\mu$, де μ - математичне сподівання ряду, то у випадку одиничного кореня варто вимагати, щоб c дорівнювало нулю. Тестова регресія у цьому випадку буде така

$$\Delta y_t = c + (\phi - 1)y_{t-1} + \varepsilon_t, \quad (7.2.15)$$

де нульова гіпотеза - це об'єднання гіпотез $H_0 : c = 0, \phi - 1 = 0$. Можна тестувати ці два обмеження разом, але легше тестувати лише гіпотезу $H_0 : \phi - 1 = 0$. t - статистику для цієї

гіпотези позначимо $\hat{\tau}_\mu$. Вона також має нестандартний розподіл. Критичні значення для $\hat{\tau}_\mu$ наведено у других двох стовпцях таблиці 7.2.1 і є меншими за відповідні критичні значення для $\hat{\tau}$. Для моделей з великою кількістю спостережень гіпотеза про одиничний корінь відхиляється при рівні значущості $\alpha = 0,05$, якщо $\hat{\tau}_\mu < -2,86$.

Якщо (7.2.15) виконується при $\phi = 1$ і ненульовому $c \neq 0$, то в цьому випадку c не може дорівнювати $(1 - \phi)\mu$ і (7.2.15) не може бути отримане з моделі (7.2.14). Вихідним процесом у цьому випадку є

$$\Delta y_t = c + \varepsilon_t, \quad (7.2.16)$$

тобто процес випадкового блукання зі зміщенням, де c - параметр зміщення. Оскільки з (7.2.16) випливає, що $E[\Delta y_t] = c$, то, задаючи початкове значення y_0 , матимемо

$$E[y_t] = y_0 + ct.$$

Звідси видно, що інтерпретація константи c сильно залежить від наявності одиничного кореня. У стаціонарному випадку c відображає ненульове математичне сподівання ряду, а у випадку одиничного кореня – стохастичний тренд для y_t . У другому випадку перші різниці призводять до стаціонарності часового ряду і процес y_t розглядається як інтегровано стаціонарний.

Отже, нестационарні процеси можуть бути охарактеризовані присутністю детермінованого тренда, наприклад δt , або стохастичного тренда, тобто наявністю простого чи кратного одиничного кореня.

У випадку, коли нестационарність процесу зумовлена присутністю детермінованого часового тренда, а не присутністю одиничного кореня, модель має вигляд

$$y_t = c + \phi y_{t-1} + \delta t + \varepsilon_t, \quad (7.2.17)$$

де $|\phi| < 1$ і $\delta \neq 0$. Процес (7.2.17) є нестационарним через наявність лінійного тренда δt . Нестационарність такого вигляду можна усунути за допомогою використання регресії y_t над константою і t і подальшого моделювання залишків цієї регресії, або просто за допомогою включення t , як додаткової змінної в модель.

Можна протестувати чи y_t відповідає випадковому блуканню чи, як альтернатива, трендово стаціонарному процесу (7.2.17). Це можна перевірити за допомогою регресії

$$\Delta y_t = c + (\phi - 1)y_{t-1} + \delta t + \varepsilon_t. \quad (7.2.18)$$

Тестування гіпотези, що процес є процесом випадкового блукання, відповідає тестуванню гіпотези $H_0 : c = \delta = \phi - 1 = 0$. Замість того, щоб тестувати сукупність гіпотез можна використати t -відношення, яке відповідає гіпотезі $H_0 : \hat{\phi} - 1$ (позначимо його $\hat{\tau}_\tau$), прийнявши, що інші обмеження в гіпотезі задовольняються.

Так як і в двох попередніх тестах на одиничний корінь, розподіл тестової статистики відрізняється від розподілу Стьюдента. Критичні значення для $\hat{\tau}_\tau$, подані в останніх двох стовпцях таблиці 7.2.1, і є меншими за відповідні значення $\hat{\tau}_\mu$. Якщо модель містить константу і детермінований тренд, то ймовірність того, що $\hat{\phi} - 1 > 0$ (якщо справжнє значення $\phi - 1$ дорівнює 0), є дуже малою.

Зауважимо, якщо гіпотеза про одиничний корінь $\phi - 1 = 0$ відхилена, то ми не можемо зробити висновок, що процес y_t – стаціонарний. В альтернативній гіпотезі δ може бути ненульовим і процес y_t не буде стаціонарним, а лише трендово-стаціонарним.

Процедуру тесту Дікі-Фуллера використовують для будь-яких описаних тестів і може ґрунтуватися на регресіях зі сталою або без неї, з трендом або без нього. У більшості випадків константа входить у регресію. Важливо зауважити, що гіпотеза на одиничний корінь відповідає нульовій гіпотезі. Якщо ми

не можемо заперечити присутність одиничного кореня, то це не означає що він є, можливо просто не вистачає інформації в даних. Звичайно, це загальна різниця між прийняттям і відхиленням гіпотези. Оскільки довготривалі властивості процесів залежать від присутності одиничного кореня, то треба пам'ятати, що не всі ряди, для яких не можемо відхилити гіпотезу про одиничний корінь, є інтегровними порядку 1.

Щоб обійти проблему слабкості тесту на одиничний корінь, Квіатовський, Філіпс, Шмідт і Шін (Kwiatowski, Phillips, Shmidt, Shin) (1992) запропонували альтернативний тест, де стаціонарність – це нульова гіпотеза, а існування одиничного кореня – альтернативна. Цей тест називається KPSS тестом. Базова ідея полягає в тому, що часовий ряд розкладений на суму детермінованого часового тренда, випадкового блукання і стаціонарної похибки, яка зазвичай є білим шумом. Нульова гіпотеза (стаціонарність) визначає, що дисперсія компоненти випадкового блукання дорівнює 0. Тест фактично є тестом множників Лагранжа і розрахунки в тестовій статистиці досить прості. Спершу розглядають допоміжну регресію для y_t від сталої і часового тренда t . Далі обчислюємо МНК-залишки цієї моделі e_t і часткові суми $S_t = \sum_{i=1}^t e_i$ для всіх t .

Тоді тестову статистику обчислюють за формулою

$$KPSS = \sum_{t=1}^T \frac{S_t^2}{\hat{\sigma}^2},$$

де $\hat{\sigma}^2$ – оцінка дисперсії похибки. Асимптотичний розподіл цієї статистики нестандартний. Зокрема, для рівня значущості $\alpha = 0,05$ критичне значення дорівнює 0,146.

7.2.2 Тестування на одиничний корінь в авторегресійних моделях вищих порядків

Для AR процесів вищих порядків тест на простий одиничний корінь можна отримати за допомогою розширення тестової процедури Дікі-Фуллера. Основною ідеєю є те, що лагові різниці $\Delta y_{t-1}, \Delta y_{t-2}, \dots$ включені в регресію, а похибки відповідають білому шуму. Це призводить до так званого розширеного тесту Дікі-Фуллера (ADF), для якого правильні ті самі асимптотичні критичні значення з таблиці 7.2.1.

Розглянемо $AR(2)$ модель

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \quad (7.2.19)$$

яка може бути записана у вигляді

$$(1 - \lambda_1 L)(1 - \lambda_2 L)y_t = \varepsilon_t. \quad (7.2.20)$$

Умова стаціонарності вимагає, щоб λ_1 і λ_2 були за модулем менші за 1, але якщо $\lambda_1 = 1$ і $|\lambda_2| < 1$, то маємо простий одиничний корінь, причому тоді $\phi_1 + \phi_2 = 1$ і $\phi_2 = -\lambda_2$.

Рівність (7.2.19) можна використати для тестування гіпотези одиничного кореня $H_0 : \phi_1 + \phi_2 = 1$ при заданому $|\phi_2| < 1$. Перетворимо (7.2.19) до вигляду

$$\Delta y_t = (\phi_1 + \phi_2 - 1) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t. \quad (7.2.21)$$

Коефіцієнти у (7.2.21) можуть бути оцінені за МНК. Оцінку коефіцієнта при y_{t-1} використовують для тестування нульової гіпотези $H_0 : \pi = \phi_1 + \phi_2 - 1 = 0$. Результуюче t -відношення $\hat{\pi}/s_{\hat{\pi}}$ має такий самий розподіл, як і $\hat{\tau}$.

У процедурі Дікі-Фуллера до тестової регресії можна додати константу або часовий тренд. Залежно від використаного варіанта, результуюча тестова статистика порівнюється з

критичними значеннями, які відповідно наведено у других чи третіх двох стовпцях таблиці 7.2.1.

Цю процедуру можна легко узагальнити для тестування простого одиничного кореня в $AR(p)$ процесі. Кожен $AR(p)$ процес можна записати у вигляді

$$\Delta y_t = \pi y_{t-1} + C_1 \Delta y_{t-1} + \dots + C_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \quad (7.2.22)$$

де $\pi = \phi_1 + \dots + \phi_p - 1$ і C_1, \dots, C_{p-1} – відповідно підібрані константи. Оскільки з $\pi = 0$ випливає, що $\phi(1) = 0$, то $z = 1$ є розв'язком характеристичного рівняння $\phi(z) = 0$. Як і раніше, гіпотеза $H_0 : \pi = 0$ відповідає одиничному кореню і можемо протестувати її, використавши відповідні t -відношення.

Якщо $AR(p)$ - зображення є коректним, то при нульовій гіпотезі одиничного кореня асимптотичні розподіли $\hat{\tau}$ -, $\hat{\tau}_\mu$ - або $\hat{\tau}_\tau$ -статистик, які обчислені для (7.2.22), мають такий самий розподіл як і раніше.

Отже, якщо y_t описується за допомогою $AR(p)$ процесу, то тест на простий одиничний корінь може бути сконструйований з регресії Δy_t стосовно y_{t-1} і $\Delta y_{t-1}, \dots, \Delta y_{t-p+1}$, тестуючи коефіцієнт при змінній y_{t-1} (використовують односторонні критичні значення). Зауважимо, що за гіпотези простого одиничного кореня всі змінні є стаціонарними, крім y_{t-1} . Тому рівність (7.2.22) має зміст лише тоді, коли y_{t-1} не з'являється і $\pi = 0$, що пояснює чому гіпотеза одиничного кореня відповідає $\pi = 0$.

Регресія вигляду (7.2.22) також може бути використана для тесту на одиничний корінь у загальній зворотній $ARMA$ моделі. Теоретично, коли кількість лагів у регресії зростає в однакових розмірах, то той самий асимптотичний розподіл залишається і ADF тест також підходить і для $ARMA$ моделі з MA - компонентами.

Якщо тест ADF не дає заперечення нульової гіпотези простого одиничного кореня, то присутність другого одиничного

кореня може бути протестовано за допомогою оцінювання регресії $\Delta^2 y_t$ стосовно Δy_{t-1} , $\Delta^2 y_{t-1}$, ..., $\Delta^2 y_{t-p+1}$ і порівнянням t - відношення коефіцієнта при Δy_{t-1} з відповідними критичними значеннями з таблиці 7.2.1.

Крім того, присутність двох одиничних коренів можна також протестувати сукупно, за допомогою оцінки регресії $\Delta^2 y_t$ стосовно y_{t-1} , Δy_{t-1} , $\Delta^2 y_{t-1}$, ..., $\Delta^2 y_{t-p}$ і обчисленням звичайної F - статистики для сукупного тестування коефіцієнтів при y_{t-1} і Δy_{t-1} . Ця тестова статистика за виконання нульової гіпотези наявності одиничного кореня кратності два має нестандартний розподіл.

7.3 Порівняння трендово - стаціонарних процесів і процесів з одиничним коренем

7.3.1 Порівняння прогнозованих значень

Розглянемо прогнозоване значення ряду, який описується трендово - стаціонарною моделлю

$$y_t = \alpha + \delta t + \psi(L) \varepsilon_t. \quad (7.3.1)$$

У цьому випадку детермінована компонента, яка містить тренд $\alpha + \delta t$, додається до прогнозу стаціонарної стохастичної компоненти

$$\hat{y}_{t+h|t} = \alpha + \delta(t+h) + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \psi_{h+2} \varepsilon_{t-2} + \dots \quad (7.3.2)$$

Якщо довжина прогнозованого інтервалу h збільшується, то з абсолютної збіжності ряду $\{\psi_j\}$ випливає, що прогнозоване значення збігається в середньому квадратичному до часового

тренда

$$E[(\hat{y}_{t+h|t} - \alpha - \delta(t+h))^2] \rightarrow 0 \quad \text{при } h \rightarrow \infty.$$

Розглянемо прогнозування процесу з одиничним коренем

$$(1 - L)y_t = \delta + \psi(L)\varepsilon_t. \quad (7.3.3)$$

Для цього процесу перші різниці є стаціонарним процесом, який можна спрогнозувати методами, описаними у розділі 6. Зокрема, матимемо

$$\Delta\hat{y}_{t+h|t} = \delta + \psi_h\varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + \psi_{h+2}\varepsilon_{t-2} + \dots \quad (7.3.4)$$

Для знаходження прогнозованого значення $\hat{y}_{t+h|t}$ на h періодів вперед зробимо перетворення

$$\begin{aligned} y_{t+h} &= (y_{t+h} - y_{t+h-1}) + (y_{t+h-1} - y_{t+h-2}) + \\ &+ \dots + (y_{t+1} - y_t) + y_t = \\ &= \Delta y_{t+h} + \Delta y_{t+h-1} + \dots + \Delta y_{t+1} + y_t. \end{aligned} \quad (7.3.5)$$

Підставляючи (7.3.4) в (7.3.5), отримаємо прогноз у випадку процесу з одиничним коренем

$$\begin{aligned} \hat{y}_{t+h|t} &= \Delta\hat{y}_{t+h|t} + \Delta\hat{y}_{t+h-1|t} + \dots + \Delta\hat{y}_{t+1|t} + y_t = \\ &= (\delta + \psi_h\varepsilon_t + \psi_{h+1}\varepsilon_{t-1} + \psi_{h+2}\varepsilon_{t-2} + \dots) + \\ &+ (\delta + \psi_{h-1}\varepsilon_t + \psi_h\varepsilon_{t-1} + \psi_{h+1}\varepsilon_{t-2} + \dots) + \\ &+ \dots + (\delta + \psi_1\varepsilon_t + \psi_2\varepsilon_{t-1} + \psi_3\varepsilon_{t-2} + \dots) + y_t \end{aligned}$$

або

$$\begin{aligned} \hat{y}_{t+h|t} &= h\delta + y_t + (\psi_h + \psi_{h-1} + \dots + \psi_1)\varepsilon_t + \\ &+ (\psi_{h+1} + \psi_h + \dots + \psi_2)\varepsilon_{t-1} + \dots \end{aligned} \quad (7.3.6)$$

Розглянемо прогнозування деяких рядів з одиничним коренем, які найчастіше використовують на практиці. Проаналізуємо процес випадкового блукання зі зміщенням

$$\Delta y_t = \delta + \varepsilon_t.$$

Цей процес є частковим випадком (7.3.3), в якому $\psi_1 = \psi_2 = \dots = 0$. У цьому випадку прогнозоване значення обчислюють за формулою

$$\hat{y}_{t+h|t} = h\delta + y_t. \quad (7.3.7)$$

З (7.3.7) видно, що прогнозоване значення для процесу випадкового блукання зі збільшенням довжини прогнозованого інтервалу зростає на сталу величину δ від своєї заданої величини y_t .

Як другий приклад прогнозування процесів з одиничним коренем розглянемо $ARIMA(0, 1, 1)$ процес

$$\Delta y_t = \delta + \varepsilon_t + \theta\varepsilon_{t-1}.$$

Для цього процесу $\psi_1 = \theta, \psi_2 = \psi_3 = \dots = 0$, тоді формула (7.3.7) набуде вигляду

$$\hat{y}_{t+h|t} = h\delta + y_t + \theta\varepsilon_t.$$

Тут базу, від якої прогноз зростатиме на сталу величину δ , визначає задане значення ряду y_t разом з заданою похибкою прогнозу

$$\varepsilon_t = y_t - \hat{y}_{t|t-1}.$$

Нехай $\delta = 0$. Тоді прогноз на один період вперед $h = 1$ матиме вигляд

$$\hat{y}_{t+1|t} = y_t + \theta(y_t - \hat{y}_{t|t-1}) \quad (7.3.8)$$

або

$$\hat{y}_{t+1|t} = (1 + \theta)y_t - \theta\hat{y}_{t|t-1}. \quad (7.3.9)$$

Рівняння (7.3.9) є різницевим рівнянням першого порядку. Якщо $|\theta| < 1$, то існує обернений оператор до $1 + \theta L$ і (7.3.9)

можна записати вигляді

$$\begin{aligned}\hat{y}_{t+1|t} &= (1 + \theta)y_t + (-\theta)(1 + \theta)y_{t-1} + (-\theta)^2(1 + \theta)y_{t-2} + \\ &+ (-\theta)^3(1 + \theta)y_{t-3} + \dots = \\ &= (1 + \theta) \sum_{j=0}^{\infty} (-\theta)^j y_{t-j}.\end{aligned}\tag{7.3.10}$$

Вираз (7.3.9) має назву адаптивного очікування, а його запис у формі (7.3.10) називається експоненціальним згладжуванням. Тут додатково припускають, що $-1 < \theta < 0$.

Для процесу $ARIMA(0, 1, q)$ величина y_t і q останніх значень ε_t визначають базове значення, від якого прогнози $\hat{y}_{t+1|t}$, $\hat{y}_{t+2|t}$, \dots , $\hat{y}_{t+q|t}$ зростають на сталу величину δ . Для $ARIMA(p, 1, q)$ величина, на яку зростає прогноз, наближається до δ .

Отож, параметр δ в процесі з одиничним коренем (7.3.3) відіграє подібну роль, що і δ в процесі з детермінованим часовим трендом (7.3.1). В обох специфікаціях прогнозоване значення $\hat{y}_{t+h|t}$ збігається до лінійної функції від ширини прогнозованого інтервалу h і відхилення δ (рис.7.3.3, 7.3.4). Відмінність полягає в перетині прямої. Для трендово - стаціонарного процесу прогноз $\hat{y}_{t+h|t}$ збігається до прямої, перетин якої не залежить від величини y_t . На противагу, перетин граничного значення прогнозу для процесу з одиничним коренем неперервно змінюється при зміні y_t .

7.3.2 Порівняння похибок прогнозу

Трендово - стаціонарні процеси і процеси з одиничним коренем дають різні результати похибок прогнозу. Для трендово - стаціонарного процесу (7.3.1) похибка прогнозу на h періодів

вперед має вигляд

$$\begin{aligned} y_{t+h} - \hat{y}_{t+h|t} &= (\alpha + \delta(t+h) + \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + \dots + \\ &\quad + \psi_{h-1} \varepsilon_{t+1} + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \dots) - \\ &\quad - (\alpha + \delta(t+h) + \psi_h \varepsilon_t + \psi_{h+1} \varepsilon_{t-1} + \psi_{h+2} \varepsilon_{t-2} + \dots) = \\ &= \varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + \dots + \psi_{h-1} \varepsilon_{t+1}. \end{aligned}$$

Середньоквадратична похибка прогнозу дорівнює

$$MSE = E[(y_{t+h} - \hat{y}_{t+h|t})^2] = (1 + \psi_1^2 + \psi_2^2 + \dots + \psi_{h-1}^2) \sigma^2.$$

Звідси видно, що MSE збільшується зі збільшенням ширини прогнозованого інтервалу h . Якщо h велике, то різниця між прогнозами в далекому майбутньому стає незначною і

$$\lim_{h \rightarrow \infty} MSE = \lim_{h \rightarrow \infty} E[(y_{t+h} - \hat{y}_{t+h|t})^2] = (1 + \psi_1^2 + \psi_2^2 + \dots) \sigma^2.$$

Зауважимо, що границя MSE є безумовною дисперсією стаціонарної компоненти $\psi(L) \varepsilon_t$.

На противагу для процесу з одиничним коренем (7.3.3) похибка прогнозу на h періодів вперед має вигляд

$$\begin{aligned} y_{t+h} - \hat{y}_{t+h|t} &= \\ &= (\Delta y_{t+h} + \Delta y_{t+h-1} + \dots + \Delta y_{t+1} + y_t) - \\ &\quad - (\hat{\Delta} y_{t+h|t} + \hat{\Delta} y_{t+h-1|t} + \dots + \hat{\Delta} y_{t+1|t} + y_t) = \\ &= (\varepsilon_{t+h} + \psi_1 \varepsilon_{t+h-1} + \psi_2 \varepsilon_{t+h-2} + \dots + \psi_{h-1} \varepsilon_{t+1}) + \\ &\quad + (\varepsilon_{t+h-1} + \psi_1 \varepsilon_{t+h-2} + \psi_2 \varepsilon_{t+h-3} + \dots + \psi_{h-2} \varepsilon_{t+1}) + \\ &\quad + \dots + (\varepsilon_{t+1}) = \\ &= \varepsilon_{t+h} + (1 + \psi_1) \varepsilon_{t+h-1} + (1 + \psi_1 + \psi_2) \varepsilon_{t+h-2} + \dots + \\ &\quad + (1 + \psi_1 + \psi_2 + \dots + \psi_{h-1}) \varepsilon_{t+1} \end{aligned}$$

з MSE

$$\begin{aligned} MSE &= E[(y_{t+h} - \hat{y}_{t+h|t})^2] = \\ &= (1 + (1 + \psi_1)^2 + (1 + \psi_1 + \psi_2)^2 + \dots + \\ &\quad + (1 + \psi_1 + \psi_2 + \dots + \psi_{h-1})^2) \sigma^2. \end{aligned}$$

MSE зростатиме разом з довжиною прогнозуючого інтервалу h , але на противагу до випадку трендово - стаціонарного процесу, MSE не збігається до фіксованої величини, якщо h прямує до безмежності. У цьому випадку вона асимптотично наближається до лінійної функції від h з відхиленням $(1 + \psi_1 + \psi_2 + \dots)^2 \sigma^2$.

Зокрема, для $ARIMA(0, 1, 1)$ процесу матимемо

$$MSE = E[(y_{t+h} - \hat{y}_{t+h|t})^2] = (1 + (h - 1)(1 + \theta)^2) \sigma^2.$$

Отже, можемо зробити висновок, що для трендово - стаціонарного процесу, якщо прогнозуючий інтервал збільшується, то границя MSE досягає скінченої межі, а для процесу з одиничним коренем MSE зростає лінійно разом із зростанням ширини прогнозуючого інтервалу. Ці висновки показано на рис.7.3.3, 7.3.4.

Рис. 7.3.3.

Зауважимо, що MSE зростає лінійно стосовно ширини прогнозованого інтервалу h , а стандартне відхилення прогнозуючої похибки зростає як квадратний корінь від h . З іншого боку,

Рис. 7.3.4.

якщо $\delta > 0$, то прогнозоване значення зростає лінійно стосовно h . отож, 95% довірчий інтервал для y_{t+h} розширюється повільніше за значення ряду.

7.3.3 Порівняння динамічних множників

Ще однією відмінністю між трендово - стаціонарним процесом і процесом з одиничним коренем є різна поведінка стосовно збурення в похибках, тобто різна стійкість до похибок.

Нехай ε_t зростає на одиницю, а всі інші дані є незмінними. Розглянемо як тоді зміниться послідовність y_t . Величина

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t}$$

називається динамічним множником.

Для трендово - стаціонарного процесу динамічний множник матиме вигляд

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \psi_h$$

і при $h \rightarrow \infty$ вплив збурення в похибці на стохастичний розподіл пом'якшується

$$\lim_{h \rightarrow \infty} \frac{\partial y_{t+h}}{\partial \varepsilon_t} = 0.$$

На противагу, для процесу з одиничним коренем динамічний множник має вигляд

$$\frac{\partial y_{t+h}}{\partial \varepsilon_t} = \frac{\partial y_t}{\partial \varepsilon_t} + \phi_h + \phi_{h-1} + \dots + \psi_1 = 1 + \psi_1 + \psi_2 + \dots + \psi_h.$$

Звідси видно, що збурення в похибці має постійний вплив на стохастичний розподіл

$$\lim_{h \rightarrow \infty} \frac{\partial y_{t+h}}{\partial \varepsilon_t} = 1 + \psi_1 + \psi_2 + \dots = \psi(1).$$

7.3.4 Різниця у двох специфікаціях стосовно перетворення для досягнення стаціонарності

Ще однією відмінністю між трендово - стаціонарним процесом і процесом з одиничним коренем, що заслуговує на дослідження, є перетворення даних, які є потрібні, щоб утворити стаціонарні часові ряди.

Якщо процес насправді трендово - стаціонарний і описується рівністю (7.3.1), то для перетворення його до стаціонарного необхідно виділити δt .

Дослідимо, що відбудеться, якщо виділити тренд у моделі з одиничним коренем. Для цього розглянемо процес випадкового блукання зі зміщенням

$$y_t = \delta + y_{t-1} + \varepsilon_t,$$

який можна подати у вигляді

$$y_t = y_0 + \delta t + \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_1.$$

Виділивши тренд, матимемо

$$y_t - \delta t = y_0 + \varepsilon_t + \dots + \varepsilon_1 = y_0 + u_t.$$

У цьому випадку дисперсія залишку u_t дорівнює $t\sigma^2$ і бачимо, що вона зростає з початку спостережень. Звідси видно таке: якщо дані описуються процесом з одиничним коренем, то, виділяючи δt з y_t , отримаємо зміну часової залежності у математичному сподіванні, а не у дисперсії.

Отож, виділення часового тренда з процесу з одиничним коренем не є достатнім для отримання стаціонарного часового ряду. Правильною поведінкою для процесу з одиничним коренем є обчислення різниць ряду. Тому процес (7.3.3) часто називають процесом стаціонарних різниць.

Якщо ж обчислити різниці трендово - стаціонарного процесу (7.3.1), то в результаті отримаємо ряд

$$\Delta y_t = \delta + (1 - L)\psi(L)\varepsilon_t.$$

Цей ряд є стаціонарним, але містить одиничний корінь у частині рухомого середнього, тобто процес не є зворотним.

Зокрема, для процесу

$$y_t = \alpha + \delta t + \varepsilon_t$$

матимемо

$$\Delta y_t = \delta + \varepsilon_t - \varepsilon_{t-1},$$

який є стаціонарним, але не є зворотним. Крім того, якщо $\text{Var}[y_t] = \sigma^2$, то $\text{Var}[\Delta y_t] = 2\sigma^2$, тобто взяття різниць стаціонарного процесу збільшує дисперсію.

7.4 Інші підходи до моделювання нестационарних часових рядів

Розглянемо ще дві альтернативні специфікації моделювання нестационарності: дробово - інтегровні процеси і процеси з випадковими дискретними змінами в часових трендах.

7.4.1 Дробове інтегрування

Інтегровний процес $I(d)$ порядку d можна зобразити у вигляді

$$(1 - L)^d y_t = \psi(L) \varepsilon_t,$$

де $\sum_{j=0}^{\infty} |\psi^j| < \infty$. Раніше ми припускали, що d - ціле число.

Гренджер, Джоекс і Госкінг (Granger, Joyeux, Hosking) показали, що можна також використовувати дробові d .

Розглянемо процес

$$(1 - L)^d y_t = \varepsilon_t, \quad (7.4.1)$$

де d - дробове число. Застосовуючи обернений оператор, матимемо зображення

$$y_t = (1 - L)^{-d} \varepsilon_t. \quad (7.4.2)$$

Для опису дії оператора $(1 - L)^{-d}$ використаємо розвинення в степеневий ряд в околі точки $z = 0$ функції $f(z) = (1 - z)^{-d}$

$$(1 - z)^{-d} = 1 + dz + \frac{(d+1)d}{2!} z^2 + \frac{(d+2)(d+1)d}{3!} z^3 + \dots$$

Тоді матимемо зображення

$$\begin{aligned} (1 - L)^{-d} &= 1 + dL + \frac{(d+1)d}{2!} L^2 + \frac{(d+2)(d+1)d}{3!} L^3 + \dots = \\ &= \sum_{j=0}^{\infty} h_j L^j, \end{aligned}$$

де $h_0 = 0$ і

$$h_j = \frac{(d+j-1)(d+j-2)\dots(d+1)d}{j!}. \quad (7.4.3)$$

Покажемо таке: якщо $d < 1$, то для великих j

$$h_j \approx (j+1)^{d-1}.$$

Для цього запишемо (7.4.3) у вигляді

$$\begin{aligned} h_j &= \binom{d+j-1}{j} \binom{d+j-2}{j-1} \binom{d+j-3}{j-2} \dots \binom{d+1}{2} \binom{d}{1} = \\ &= \binom{j+d-1}{j} \binom{j-1+d-1}{j-1} \binom{j-2+d-1}{j-2} \times \\ &\quad \times \dots \times \binom{2+d-1}{j-(j-2)} \binom{1+d-1}{j-(j-1)} = \quad (7.4.4) \\ &= \left(1 + \frac{d-1}{j}\right) \left(1 + \frac{d-1}{j-1}\right) \left(1 + \frac{d-1}{j-2}\right) \times \\ &\quad \times \dots \times \left(1 + \frac{d-1}{j-(j-2)}\right) \left(1 + \frac{d-1}{j-(j-1)}\right). \end{aligned}$$

Використовуючи розвинення функції $g(x) = (1+x)^{d-1}$ в ряд Тейлора

$$(1+x)^{d-1} = 1 + (d-1)x + \frac{1}{2}(d-1)(d-2)(1+\delta)^{d-3}x^2$$

для деякого $\delta \in (0, x)$, для $d < 1$ матимемо

$$(1+x)^{d-1} \geq 1 + (d-1)x.$$

Прийнявши $x = \frac{1}{j}$, отримаємо

$$1 + \frac{d-1}{j} \leq \left(1 + \frac{1}{j}\right)^{d-1}$$

для всіх $j > 0$ і $d < 1$. Підставляючи тепер останній вираз у (7.7.4), одержимо

$$\begin{aligned} h_j &\approx \left(1 + \frac{1}{j}\right)^{d-1} \left(1 + \frac{1}{j-1}\right)^{d-1} \left(1 + \frac{1}{j-2}\right)^{d-1} \times \\ &\quad \times \cdots \times \left(1 + \frac{1}{j-(j-2)}\right)^{d-1} \left(1 + \frac{1}{j-(j-1)}\right)^{d-1} = \\ &= \left(\frac{j+1}{j}\right)^{d-1} \left(\frac{j}{j-1}\right)^{d-1} \left(\frac{j-1}{j-2}\right)^{d-1} \cdots \left(\frac{3}{2}\right)^{d-1} \left(\frac{2}{1}\right)^{d-1} = \\ &= (j+1)^{d-1}, \end{aligned}$$

що і треба було довести.

Отже, модель

$$y_t = (1 - L)^{-d} \varepsilon_t = h_0 \varepsilon_t + h_1 \varepsilon_{t-1} + h_2 \varepsilon_{t-2} + \dots \quad (7.4.5)$$

має $MA(\infty)$ зображення, динамічний множник якого для великих j поводить ся як $(j+1)^{d-1}$. Для порівняння нагадаємо, що динамічний множник $AR(1)$ процесу $y_t = (1 - \phi L)\varepsilon_t$ дорівнює ϕ^j . Отже, бачимо, що динамічний множник стаціонарного $ARMA$ процесу спадає геометрично, а для процесу (7.4.1) він має повільний спад. Через цей повільний порядок затухання Гренгер і Джоекс запропонували дробово-інтегровний процес, як підхід до моделювання довгих запізнень у часових рядах.

На кінець зауважимо, що послідовність коефіцієнтів процесу рухомого середнього $\{h_j\}_{j=0}^{\infty}$, яка задана в (7.4.3), збігається в середньому квадратичному, якщо виконується умова $d < 1/2$, тобто

$$\sum_{j=0}^{\infty} h_j^2 < \infty \quad \text{для} \quad d < \frac{1}{2}.$$

Це випливає з того, що

$$\begin{aligned} \sum_{j=0}^{N-1} (j+1)^{2(d-1)} &= \sum_{j=1}^N j^{2(d-1)} < 1 + \int_1^N x^{2(d-1)} dx = \\ &= 1 + \frac{1}{2d-1} x^{2d-1} \Big|_{x=1}^N = 1 + \frac{1}{2d-1} (N^{2d-1} - 1) \end{aligned}$$

при $N \rightarrow \infty$ збігається до $1 - (1/(2d-1))$, якщо $d < 1/2$.

Отже, (7.4.5) визначає коваріаційно - стаціонарний процес, якщо виконується умова $d < 1/2$. Якщо $d > 1/2$, то треба обчислити перші різниці ряду перед тим як використовувати зображення (7.4.2). Наприклад, якщо $d = 0,7$, то процес (7.4.1) можна записати у вигляді

$$(1-L)^{-0,3}(1-L)y_t = \varepsilon_t.$$

Тоді Δy_t є дробово - інтегровним процесом з параметром $d = -0,3 < 1/2$.

7.4.2 Випадкові розриви в трендах

Специфікація з одиничним коренем (7.1.3) передбачає, що події відбуваються протягом часу, що постійно впливає на поведінку y_t . Перрон, Раппопорт і Райхлін (Perron, Rappoport, Reichlin)(1989) дослідили, що економічні події, які мають довгий постійний вплив, є рідкісними.

Цю ідею можна проілюструвати за допомогою моделі, в якій y_t є стаціонарним стосовно тренда з одним розривом

$$y_t = \begin{cases} \alpha_1 + \delta t + \varepsilon_t & \text{для } t < T_0, \\ \alpha_2 + \delta t + \varepsilon_t & \text{для } t \geq T_0. \end{cases} \quad (7.4.6)$$

Процес (7.4.6) можна зобразити також за допомогою зображення

$$\Delta y_t = \xi_t + \delta + \varepsilon_t - \varepsilon_{t-1}, \quad (7.4.7)$$

де величина $\xi_t = (\alpha_1 - \alpha_2)$ при $t = T_0$ і дорівнює нулю в інших випадках.

Припустимо, що випадкова змінна ξ_t має деякий ймовірнісний розподіл. Зокрема, нехай

$$\xi_t = \begin{cases} \alpha_2 - \alpha_1 & \text{з ймовірністю } p, \\ 0 & \text{з ймовірністю } 1 - p. \end{cases}$$

Тоді рівняння (7.4.7) можна записати у вигляді

$$\Delta y_t = \mu + \eta_t, \quad (7.4.8)$$

де

$$\begin{aligned} \mu &= p(\alpha_2 - \alpha_1) + \delta, \\ \eta_t &= \xi_t - p(\alpha_2 - \alpha_1) + \varepsilon_t - \varepsilon_{t-1}. \end{aligned}$$

Тут похибка η_t є сумою процесу білого шуму $\xi_t - p(\alpha_2 - \alpha_1)$ з нульовим математичним сподіванням і незалежного з ним $MA(1)$ процесу $\varepsilon_t - \varepsilon_{t-1}$. Тому для η_t існує $MA(1)$ зображення.

Отже, (7.4.8) можна зобразити як $ARIMA(0, 1, 1)$ процес

$$\Delta y_t = \mu + \nu_t + \theta \nu_{t-1} \quad (7.4.9)$$

з негауссівським розподілом залишків ν_t .

Оптимальним лінійним прогнозом у цьому випадку буде

$$\hat{E}[y_{t+h} | y_t, y_{t-1}, \dots] = \mu h + y_t + \theta \nu_t.$$

Бачимо, що у цьому випадку похибка має вплив на всі прогнози і її вплив не зникає при $h \rightarrow \infty$.

Розділ 8

Процеси з детермінованим часовим трендом

Коефіцієнти регресійних моделей з одиничним коренем та моделей з детермінованим часовим трендом зазвичай оцінюють звичайним методом найменших квадратів. Асимптотичні розподіли оцінок параметрів не можуть бути обчислені так само, як для регресійної моделі, яка містить звичайні змінні. У цьому випадку оцінки різних параметрів будуть мати різну асимптотичну поведінку.

8.1 Асимптотичний розподіл МНК оцінки простої моделі часового тренда

Розглянемо просту модель часового тренда

$$y_t = \alpha + \delta t + \varepsilon_t, \quad (8.1.1)$$

де ε_t – процес білого шуму. Якщо $\varepsilon_t \sim N[0, \sigma^2]$, то модель (8.1.1) задовольняє класичні припущення і стандартні МНК-

статистики t і F для малих вибірок будуть мати точні розподіли Стьюдента і Фішера відповідно. Якщо ж ε_t не є нормально розподіленою, то використовують асимптотичні розподіли МНК-оцінок.

Нагадаємо, що для звичайної регресії

$$y = X\beta + \varepsilon, \quad (8.1.2)$$

МНК-оцінка обчислюється за формулою

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (8.1.3)$$

Підставляючи (8.1.2) в (8.1.3), матимемо

$$\hat{\beta} = \beta + (X'X)^{-1}X'\varepsilon. \quad (8.1.4)$$

Тоді відхилення МНК-оцінки $\hat{\beta}$ від істинного значення β має вигляд

$$\hat{\beta} - \beta = (X'X)^{-1}X'\varepsilon \quad (8.1.5)$$

або

$$\hat{\beta} - \beta = \left(\frac{X'X}{n} \right)^{-1} \left(\frac{1}{n}X'\varepsilon \right),$$

де n – обсяг вибірки, на основі якої обчислено $\hat{\beta}$.

Для отримання граничного розподілу оцінок параметрів моделі (8.1.2) треба вимагати, щоб виконувалась умова

$$\lim_{n \rightarrow \infty} \frac{X'X}{n} = Q. \quad (8.1.6)$$

Тут Q – додатно визначена симетрична матриця. Застосувавши центральну граничну теорему для вектора

$$\bar{w} = \frac{1}{n}X'\varepsilon,$$

матимемо

$$\sqrt{n}\bar{w} \xrightarrow{d} N[0, \sigma^2 Q].$$

Звідси отримуємо асимптотично нормальний розподіл оцінки $\hat{\beta}$, а саме

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N[0, \sigma^2 Q^{-1}]. \quad (8.1.7)$$

Отже, щоб знайти асимптотичний розподіл МНК-оцінок коефіцієнтів моделі зі звичайними змінними, треба (8.1.5) помножити на \sqrt{T} .

Застосуємо цю методику для визначення асимптотичних розподілів МНК-оцінок параметрів простої моделі часового тренда.

Запишемо (8.1.1) у формі стандартної регресійної моделі

$$y_t = x_t \beta + \varepsilon_t, \quad (8.1.8)$$

де

$$x_t = (1 \ t), \quad \beta = \begin{pmatrix} \alpha \\ \delta \end{pmatrix}.$$

Тоді

$$X'X = \sum_{t=1}^T x'_t x_t = \begin{pmatrix} T & \sum t \\ \sum t & \sum t^2 \end{pmatrix}, \quad X'\varepsilon = \sum_{t=1}^T x'_t \varepsilon_t$$

і (8.1.5) матиме вигляд

$$(\hat{\beta}_T - \beta) = \begin{pmatrix} \hat{\alpha}_T - \alpha \\ \hat{\delta}_T - \delta \end{pmatrix} = \begin{pmatrix} T & \sum t \\ \sum t & \sum t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum \varepsilon_t \\ \sum t \varepsilon_t \end{pmatrix}, \quad (8.1.9)$$

де \sum позначає суму за t від 1 до T .

Для моделі з часовим трендом умова (8.1.6) не виконується, тому множення відхилення $(\hat{\beta}_T - \beta)$ на \sqrt{T} не дає результату, подібного до (8.1.7).

Доведемо це. Легко показати, що

$$\begin{aligned}\sum_{t=1}^T t &= \frac{T(T+1)}{2}, \\ \sum_{t=1}^T t^2 &= \frac{T(T+1)(2T+1)}{6}.\end{aligned}\tag{8.1.10}$$

Отож, ведучим доданком у $\sum_{t=1}^T t \in T^2/2$ і

$$\frac{1}{T^2} \sum_{t=1}^T t = \frac{1}{T^2} \left(\frac{T^2}{2} + \frac{T}{2} \right) = \frac{1}{2} + \frac{1}{2T} \rightarrow \frac{1}{2}.\tag{8.1.11}$$

Так само, ведучим членом в $\sum_{t=1}^T t^2 \in T^3/3$ і

$$\begin{aligned}\frac{1}{T^3} \sum_{t=1}^T t^2 &= \frac{1}{T^3} \left(\frac{2T^3}{6} + \frac{3T^2}{6} + \frac{T}{6} \right) = \\ &= \frac{1}{3} + \frac{1}{2T} + \frac{1}{6T^2} \rightarrow \frac{1}{3}.\end{aligned}\tag{8.1.12}$$

Покажемо, що загалом випадку ведучим доданком у $\sum_{t=1}^T t^k \in T^{k+1}/k+1$ і

$$\frac{1}{T^{k+1}} \sum_{t=1}^T t^k \rightarrow \frac{1}{k+1}.\tag{8.1.13}$$

Для цього запишемо (8.1.13) у вигляді

$$\frac{1}{T^{k+1}} \sum_{t=1}^T t^k = \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^k.\tag{8.1.14}$$

Праву частину (8.1.14) можна розглядати як апроксимацію кривої

$$f(r) = r^k$$

в області $0 < r < 1$. Добуток $\frac{1}{T} \left(\frac{t}{T}\right)^k$ зображає площу прямокутника шириною $1/T$ і висотою r^k при $r = t/T$. Отож, (8.1.14) сума площ прямокутників відповідно при $r = 1/T, 2/T, \dots, 1$. При $T \rightarrow \infty$ ця сума збігається до площі під кривою $f(r)$, тобто

$$\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T}\right)^k \rightarrow \int_0^1 r^k dr = \frac{r^{k+1}}{k+1} \Big|_{r=0}^1 = \frac{1}{k+1}.$$

Для заданого x_t з (8.1.10) випливає

$$\begin{aligned} X'X &= \sum_{t=1}^T x'_t x_t = \begin{pmatrix} T & \sum t \\ \sum t & \sum t^2 \end{pmatrix} = \\ &= \begin{pmatrix} T & T(T+1)/2 \\ T(T+1)/2 & T(T+1)(2T+1)/6 \end{pmatrix}. \end{aligned} \tag{8.1.15}$$

На відміну від результату для звичайних регресорів, матриця $(1/T)X'X'$ у (8.1.15) при $T \rightarrow \infty$ розбігається. Для отримання збіжної матриці, (8.1.15) треба поділити на T^3 , а не на T

$$\frac{1}{T^3} X'X \rightarrow \begin{pmatrix} 0 & 0 \\ 0 & 1/3 \end{pmatrix}.$$

Проте в цьому випадку матриця, яку ми отримали, в границі не є невивірженою. Отже, для моделі з часовим трендом потрібний інший підхід для розрахунку асимптотичного розподілу $\hat{\beta}_T$.

Грунтуватимемось на тому, що МНК-оцінки $\hat{\alpha}_T$ і $\hat{\delta}_T$ мають різну асимптотичну поведінку. Для досягнення невідродженості граничної матриці, $\hat{\alpha}_T$ множимо на \sqrt{T} , тоді як $\hat{\delta}_T$ треба помножити на $T^{3/2}$. Можемо виконати це регулювання, як попереднє перемноження (8.1.6) або (8.1.8) на матрицю

$$M_T = \begin{pmatrix} \sqrt{T} & 0 \\ 0 & T^{3/2} \end{pmatrix}. \quad (8.1.16)$$

Лема 8.1.1. *Для моделі (8.1.1)*

$$M_T^{-1} (X'\varepsilon) \xrightarrow{d} N [0, \sigma^2 Q],$$

де

$$Q = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}.$$

Д о в е д е н н я. Помноживши (8.1.8) на (8.1.16), матимемо

$$\begin{aligned} \begin{pmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{pmatrix} &= M_T \left[\sum_{t=1}^T x'_t x_t \right]^{-1} \left[\sum_{t=1}^T x'_t \varepsilon_t \right] = \quad (8.1.17) \\ &= M_T \left[\sum_{t=1}^T x'_t x_t \right]^{-1} M_T M_T^{-1} \left[\sum_{t=1}^T x_t \varepsilon_t \right] = \\ &= \left(M_T^{-1} \left[\sum_{t=1}^T x'_t x_t \right] M_T^{-1} \right)^{-1} M_T^{-1} \left[\sum_{t=1}^T x'_t \varepsilon_t \right]. \end{aligned}$$

Розглянемо перший множник у (8.1.17). Підставивши (8.1.16) і (8.1.15), одержимо

$$\begin{aligned} \left(M_T^{-1} \left[\sum_{t=1}^T x'_t x_t \right] M_T^{-1} \right) &= \\ &= \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix} \begin{pmatrix} T & \sum t \\ \sum t & \sum t^2 \end{pmatrix} \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix} = \\ &= \begin{pmatrix} 1 & (1/T^2) \sum t \\ (1/T^2) \sum t & (1/T^3) \sum t^2 \end{pmatrix}. \end{aligned}$$

Отже, з (8.1.11) і (8.1.12) випливає, що

$$\left(M_T^{-1} \left[\sum_{t=1}^T x_t' x_t \right] M_T^{-1} \right) \rightarrow Q, \quad (8.1.18)$$

де

$$Q = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}. \quad (8.1.19)$$

Розглянемо другий множник у (8.1.17)

$$M_T^{-1} \left[\sum_{t=1}^T x_t' \varepsilon_t \right] = \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix} \begin{pmatrix} \sum_{t=1}^T \varepsilon_t \\ \sum_{t=1}^T t \varepsilon_t \end{pmatrix}. \quad (8.1.20)$$

Перший елемент вектора в (8.1.20) задовольняє центральну граничну теорему

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t \xrightarrow{d} N[0, \sigma^2].$$

Розглянемо другий елемент вектора в (8.1.20), який можна записати у вигляді

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{t}{T} \varepsilon_t.$$

Для знаходження його граничного розподілу застосуємо теорему.

Теорема. Нехай $\{X_t\}$ – послідовність випадкових величин, $\bar{X}_T = 1/T \sum_{t=1}^T X_t$ і виконуються умови:

- 1) $E[X_t^2] = \sigma_t^2 > 0$, $\frac{1}{T} \sum_{t=1}^T \sigma_t^2 \rightarrow \sigma^2 > 0$;
- 2) $E[|X_t|^r] < \infty$ для деякого $r > 2$ і для всіх t ;

$$3) \quad \text{plim} \frac{1}{T} \sum_{t=1}^T X_t^2 = \sigma^2.$$

Тоді

$$\sqrt{TX_T} \xrightarrow{d} N [0, \sigma^2].$$

Позначимо $u_t = \frac{t}{T} \varepsilon_t$. Обчислимо дисперсію u_t

$$\sigma_t^2 = \text{Var}[u_t] = E [u_t^2] = \sigma^2 \frac{t^2}{T^2}.$$

Тоді

$$\frac{1}{T} \sum_{t=1}^T \sigma_t^2 = \sigma^2 \frac{1}{T^3} \sum_{t=1}^T t^2 \rightarrow \frac{\sigma^2}{3}. \quad (8.1.21)$$

Крім того,

$$\text{plim} \left(\frac{1}{T} \sum_{t=1}^T u_t^2 \right) = \frac{\sigma^2}{3}.$$

Доведемо це. З (8.1.21) випливає, що

$$\lim_{T \rightarrow \infty} E \left[\frac{1}{T} \sum_{t=1}^T u_t^2 \right] = \frac{\sigma^2}{3}. \quad (8.1.22)$$

Знайдемо $\lim_{T \rightarrow \infty} \text{Var} \left[\frac{1}{T} \sum_{t=1}^T u_t^2 \right]$. Для цього обчислимо

$$\begin{aligned} E \left[\left(\frac{1}{T} \sum_{t=1}^T u_t^2 - \frac{1}{T} \sum_{t=1}^T \sigma_t^2 \right)^2 \right] &= \\ &= E \left[\left(\frac{1}{T} \sum_{t=1}^T u_t^2 - \frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^2 \sigma^2 \right)^2 \right] = \end{aligned}$$

$$\begin{aligned}
 &= E \left[\left(\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \right)^2 (\varepsilon_t^2 - \sigma^2) \right)^2 \right] = \quad (8.1.23) \\
 &= \left(\frac{1}{T} \right)^2 \sum_{t=1}^T \left(\frac{t}{T} \right)^4 E \left[(\varepsilon_t^2 - \sigma^2)^2 \right].
 \end{aligned}$$

Але з (8.1.13) матимемо, що (8.1.23) збігається до

$$\left(\frac{1}{T} \right)^2 \sum_{t=1}^T \left(\frac{t}{T} \right)^4 E \left[(\varepsilon_t^2 - \sigma^2)^2 \right] \rightarrow \frac{1}{5T} E \left[(\varepsilon_t^2 - \sigma^2)^2 \right], \quad (8.1.24)$$

тобто (8.1.23) безпосередньо збігається до нуля. Тобто,

$$\text{Var} \left[\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \varepsilon_t \right)^2 \right] \rightarrow 0.$$

З (8.1.22) і (8.1.24) матимемо збіжність за ймовірністю

$$\frac{1}{T} \sum_{t=1}^T \left(\frac{t}{T} \varepsilon_t \right)^2 \xrightarrow{p} \frac{\sigma^2}{3}.$$

Тому величина $\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \varepsilon_t \right)$ задовольняє центральну граничну теорему

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\frac{t}{T} \varepsilon_t \right) \xrightarrow{d} N \left[0, \frac{\sigma^2}{3} \right].$$

Нарешті, розглянемо об'єднання цих двох елементів у вектор, описаний у (8.1.20). Довільна лінійна комбінація елементів цього вектора має вигляд

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \left(\lambda_1 + \lambda_2 \frac{t}{T} \right) \varepsilon_t.$$

Дисперсія окремого доданка $(\lambda_1 + \lambda_2(t/T)) \varepsilon_t$ дорівнює $\sigma^2 (\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2)$ і

$$\begin{aligned} E \left[\frac{1}{T} \sum_{t=1}^T (\lambda_1^2 + 2\lambda_1\lambda_2(t/T) + \lambda_2^2(t/T)^2) \varepsilon_t^2 \right] &= \\ &= \frac{1}{T} \sum_{t=1}^T \sigma^2 \left(\lambda_1^2 + 2\lambda_1\lambda_2 \frac{t}{T} + \lambda_2^2 \left(\frac{t}{T} \right)^2 \right) \rightarrow \\ &\rightarrow \sigma^2 \left(\lambda_1^2 + 2\lambda_1\lambda_2 \frac{1}{2} + \lambda_2^2 \frac{1}{3} \right) = \sigma^2 \lambda' Q \lambda, \end{aligned}$$

де $\lambda = (\lambda_1, \lambda_2)'$ і Q – матриця з (8.1.19). Крім того, можна показати, що

$$\frac{1}{T} \sum_{t=1}^T \left(\lambda_1 + \lambda_2 \frac{t}{T} \right)^2 \varepsilon_t^2 \xrightarrow{p} \sigma^2 \lambda' Q \lambda, \quad (8.1.25)$$

Отож, довільна лінійна комбінація елементів вектора (8.1.20) має асимптотично нормальний розподіл. Отже,

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{t}{T} \varepsilon_t \end{pmatrix} \xrightarrow{d} N [0, \sigma^2 Q], \quad (8.1.26)$$

що і треба було довести. \square

Теорема 8.1.1. *Нехай y_t генерується простою моделлю детермінованого часового тренда (8.1.1), де ε_t – послідовність випадкових величин, яка має такі характеристики: $E[\varepsilon_t] = 0$, $E[\varepsilon_t^2] = \sigma^2$, $E[\varepsilon_t^4] < \infty$. Тоді*

$$\begin{pmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{pmatrix} \xrightarrow{d} N [0, \sigma^2 Q^{-1}].$$

Д о в е д е н н я. З (8.1.18) і (8.1.26) випливає, що асимптотичний розподіл (8.1.17) можна розрахувати так:

$$\begin{aligned} \begin{pmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{pmatrix} &\xrightarrow{d} N[0, Q^{-1}\sigma^2 Q Q^{-1}] = \\ &= N[0, \sigma^2 Q^{-1}]. \end{aligned} \quad (8.1.27)$$

Отже,

$$\begin{pmatrix} \sqrt{T}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\delta}_T - \delta) \end{pmatrix} \xrightarrow{d} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}^{-1} \right]. \quad (8.1.28)$$

Теорему доведено. \square

Опишемо різні множники збіжності в термінах ймовірності.

Означення. *Послідовність випадкових змінних $\{X_T\}_{T=1}^{\infty}$ є $O_p(T^{-1/2})$, якщо для кожного $\varepsilon > 0$ існує $M > 0$ таке, що нерівність*

$$P \left\{ |X_T| > \frac{M}{\sqrt{T}} \right\} < \varepsilon \quad (8.1.29)$$

виконується для всіх T .

Іншими словами, випадкова змінна $\sqrt{T} X_T$ майже завжди потрапляє в інтервал $(-M, M)$ для будь-яких T . Більшість оцінюваних функцій, на які натрапляють при дослідженні стаціонарних часових рядів, є $O_p(T^{-1/2})$. Наприклад, нехай \bar{x}_t середнє вибірки обсягом T

$$\bar{x}_T = \frac{1}{T} \sum_{t=1}^T y_t,$$

де y_t має нульове математичне сподівання і дисперсію σ^2 . Тоді дисперсією \bar{x}_T є σ^2/T . Тому згідно з нерівністю Чебишева

матимемо

$$P \left\{ |X_T| > \frac{M}{\sqrt{T}} \right\} < \frac{\sigma^2/T}{M^2/T} = \left(\frac{\sigma}{M} \right)^2$$

для будь-якого M . Вибираючи M , так що $(\sigma/M)^2 < \varepsilon$, ми забезпечимо виконання умови (8.1.29).

У загальному випадку, послідовність випадкових величин $\{X_T\}_{T=1}^{\infty} \in O_p(T^{-k})$, якщо для кожного $\varepsilon > 0$ існує $M > 0$ таке, що

$$P \left\{ |X_T| > \frac{M}{T^k} \right\} < \varepsilon. \quad (8.1.30)$$

Отож, наприклад, оцінка $\hat{\delta}_T$ в (8.1.26) є $O_p(T^{-3/2})$.

8.2 Тестування гіпотез для простої моделі часового тренда

Покажемо, що для простої моделі часового тренда (8.1.1) незважаючи на те, що оцінки $\hat{\alpha}_T$ і $\hat{\delta}_T$ мають різні асимптотичні поведінки, оскільки стандартні похибки параметрів $\hat{\alpha}_T$ і $\hat{\delta}_T$ також мають різні асимптотичні поведінки, то МНК t -статистика $(\hat{\delta}_T - \delta_0)/\hat{\sigma}_{\hat{\delta}_T}$ матиме асимптотичний розподіл $N[0, 1]$, у випадку коли похибка моделі є нормальною.

Розглянемо тестування нульової гіпотези $H_0 : \alpha = \alpha_0$. У випадку моделі часового тренда t статистику можна записати у вигляді

$$t_T = \frac{\hat{\alpha}_T - \alpha_0}{\left[s_T^2 (1 \ 0) (X'X)^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right]^{1/2}}, \quad (8.2.1)$$

де $X'X = \sum_{t=1}^T x_t x_t'$ – матриця, описана рівнянням (8.1.15), а s_T^2

позначає МНК-оцінку невідомої дисперсії σ^2

$$s_T^2 = \frac{1}{T-2} \sum_{t=1}^T (y_t - \hat{\alpha}_T - \hat{\delta}_T t)^2.$$

Лема 8.2.1. Для моделі (8.1.1)

$$t_T \xrightarrow{d} N[0, 1].$$

Д о в е д е н н я.

Помножимо чисельник і знаменник (8.2.1) на \sqrt{T} , матимемо

$$t_T = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\left[s_T^2 (\sqrt{T} \ 0) (X'X)^{-1} \begin{pmatrix} \sqrt{T} \\ 0 \end{pmatrix} \right]^{1/2}}. \quad (8.2.2)$$

Зауважимо, що з (8.1.16) випливає

$$(\sqrt{T} \ 0) = (1 \ 0) M_T. \quad (8.2.3)$$

Тоді, підставивши (8.2.3) у (8.2.2), матимемо

$$t_T = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\left[s_T^2 (1 \ 0) M_T (X'X)^{-1} M_T \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right]^{1/2}}. \quad (8.2.4)$$

Використавши (8.1.18), отримаємо

$$M_T (X'X)^{-1} M_T = [Y_T^{-1} (X'X)^{-1} M_T^{-1}]^{-1} \rightarrow Q^{-1}, \quad (8.2.5)$$

і крім того, можна безпосередньо довести, що $s_T^2 \xrightarrow{p} \sigma^2$.

Тоді з (8.2.4), матимемо

$$t_T \xrightarrow{p} \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\left[\sigma^2 (1 \ 0) Q^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right]^{1/2}} = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha_0)}{\sigma \sqrt{q^{11}}}, \quad (8.2.6)$$

де q^{11} елемент матриці Q^{-1} .

З теореми 8.1.1 випливає, що

$$\sqrt{T}(\hat{\alpha}_T - \alpha_0) \xrightarrow{d} N [0, \sigma^2 q^{11}] .$$

Звідси випливає, що статистика (8.2.7) дорівнює асимптотично нормальній випадковій величині, яка поділена на квадратний корінь з дисперсії. Тому (8.2.6) матиме асимптотичний розподіл $N[0, 1]$, що і треба було довести. \square

Отже, звичайний МНК t тест для $H_0 : \alpha = \alpha_0$ буде має такий самий асимптотичний результат, як і у випадку звичайних регресорів.

Розглянемо тепер тестування гіпотези $H_0 : \delta = \delta_0$. У цьому випадку МНК t -статистика має вигляд

$$t_T = \frac{\hat{\delta}_T - \delta_0}{\left[s_T^2 (0 \ 1) (X'X)^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]^{1/2}} .$$

Помножимо чисельник і знаменник на $T^{3/2}$

$$\begin{aligned} t_T &= \frac{T^{3/2}(\hat{\delta}_T - \delta_0)}{\left[s_T^2 (0 \ T^{3/2}) (X'X)^{-1} \begin{pmatrix} 0 \\ T^{3/2} \end{pmatrix} \right]^{1/2}} = \\ &= \frac{T^{3/2}(\hat{\delta}_T - \delta_0)}{\left[s_T^2 (0 \ 1) M_T (X'X)^{-1} M_T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right]^{1/2}} \xrightarrow{p} \quad (8.2.7) \\ &\xrightarrow{p} \frac{T^{3/2}(\hat{\delta}_T - \delta_0)}{\sigma \sqrt{q^{22}}} , \end{aligned}$$

де q^{22} елемент матриці Q^{-1} . Враховуючи результати теореми 8.1.1, матимемо

$$\sqrt{T}^{3/2}(\delta - \delta_0) \xrightarrow{d} N [0, \sigma^2 q^{22}] .$$

Отже, матимемо асимптотичний розподіл $N[0, 1]$. Отож, хоч $\hat{\alpha}_T$ і $\hat{\delta}_T$ збігаються з різними коефіцієнтами, відповідні середньоквадратичні похибки $\hat{\alpha}_T$ і $\hat{\delta}_T$ також включають різні степені T . Отже, звичайні МНК t -статистики мають асимптотичний стандартний нормальний розподіл.

Розглянемо гіпотезу

$$H_0 : r_1\alpha + r_2\delta = r,$$

де r_1, r_2 і r – задані параметри. t -тест для H_0 можна отримати як квадратний корінь з F -статистики, а саме

$$t_T = \frac{(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\left[s_T^2 (r_1 \ r_2) (X'X)^{-1} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \right]^{1/2}}.$$

У цьому випадку помножимо чисельник і знаменник на \sqrt{T}

$$\begin{aligned} t_T &= \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\left[s_T^2 \sqrt{T} (r_1 \ r_2) (X'X)^{-1} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \sqrt{T} \right]^{1/2}} = \\ &= \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\left[s_T^2 \sqrt{T} (r_1 \ r_2) M_T^{-1} M_T (X'X)^{-1} M_T M_T^{-1} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \sqrt{T} \right]^{1/2}} = \\ &= \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r)}{\left[s_T^2 r_T' (M_T (X'X)^{-1} M_T) r_T \right]^{1/2}}, \end{aligned} \quad (8.2.8)$$

де

$$r_T \equiv M_T^{-1} \begin{pmatrix} r_1 \\ r_2 \end{pmatrix} \sqrt{T} = \begin{pmatrix} r_1 \\ r_2/T \end{pmatrix} \rightarrow \begin{pmatrix} r_1 \\ 0 \end{pmatrix}. \quad (8.2.9)$$

З (8.1.28) випливає, що $\hat{\delta}_T$ є змістовною, тобто

$$\sqrt{T}(r_1\hat{\alpha}_T + r_2\hat{\delta}_T - r) \xrightarrow{p} \sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r). \quad (8.2.10)$$

Повторне використання (8.2.6) дасть

$$t_T \xrightarrow{p} \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r)}{\left[\sigma^2 \begin{pmatrix} r_1 & 0 \end{pmatrix} Q^{-1} \begin{pmatrix} r_1 \\ 0 \end{pmatrix}\right]^{1/2}} = \frac{\sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r)}{[r_1^2 \sigma^2 q^{11}]^{1/2}}. \quad (8.2.11)$$

Зауважимо, що за нульової гіпотези

$$\begin{aligned} \sqrt{T}(r_1\hat{\alpha}_T + r_2\delta - r) &= \sqrt{T}(r_1(\hat{\alpha}_T - \alpha) + r_1\alpha + r_2\delta - r) = \\ &= \sqrt{T}r_1(\hat{\alpha}_T - \alpha). \end{aligned}$$

Відповідно

$$t_T \xrightarrow{p} \frac{\sqrt{T}r_1(\hat{\alpha}_T - \alpha)}{(r_1^2 \sigma^2 q^{11})^{1/2}} = \frac{\sqrt{T}(\hat{\alpha}_T - \alpha)}{(\sigma^2 q^{11})^{1/2}}$$

асимптотично має $N[0, 1]$ - розподіл. Отож, звичайний МНК t -тест для H_0 має той самий асимптотичний розподіл.

Нарешті, розглянемо об'єднане дослідження окремих гіпотез стосовно $\hat{\alpha}_T$ і $\hat{\delta}_T$

$$H_0 : \begin{pmatrix} \alpha \\ \delta \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \delta_0 \end{pmatrix},$$

або у векторній формі

$$H_0 : \beta = \beta_0.$$

Тестова статистика Вальда матиме такий вигляд:

$$\begin{aligned} \chi_T^2 &= (\hat{\beta}_T - \beta_0)' [s_T^2 (X'X)^{-1}]^{-1} (\hat{\beta}_T - \beta_0) = \\ &= (\hat{\beta}_T - \beta_0)' M_T [M_T s_T^2 (X'X)^{-1} M_T]^{-1} M_T (\hat{\beta}_T - \beta_0) \xrightarrow{p} \\ &\xrightarrow{p} [M_T (\hat{\beta}_T - \beta_0)]' [\sigma^2 Q^{-1}]^{-1} [M_T (\hat{\beta}_T - \beta_0)], \end{aligned}$$

причому

$$\chi_T^2 \xrightarrow{d} \chi^2[2].$$

8.3 Асимптотичні висновки для авторегресійного $AR(p)$ процесу з детермінованим часовим трендом

Узагальнимо отримані результати на випадок авторегресійного процесу $AR(p)$ з детермінованим часовим трендом

$$y_t = \alpha + \delta t + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t. \quad (8.3.1)$$

Припустимо, що ε_t має нульове математичне сподівання, дисперсію σ^2 , скінчений четвертий момент і розв'язки рівняння

$$1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p = 0$$

лежать ззовні одиничного круга. Розглянемо вибірку обсягом $T + p$ спостережень $\{y_{-p+1}, y_{-p+2}, \dots, y_1, \dots, y_T\}$ і нехай $\hat{\alpha}_T, \hat{\delta}_T, \hat{\phi}_{1,T}, \dots, \hat{\phi}_{p,T}$ – оцінки коефіцієнтів (8.3.1), які знайдено за допомогою методу найменших квадратів і ґрунтуються на даних для $t = 1, 2, \dots, T$.

8.3.1 Перетворення регресії

Додамо і віднімемо $\phi_j (\alpha + \delta(t - j))$ ($j = 1, 2, \dots, p$) у правій частині регресійної моделі (8.3.1). В результаті отримаємо еквівалентний запис

$$\begin{aligned} y_t = & \alpha (1 + \phi_1 + \phi_2 + \dots + \phi_p) + \delta (1 + \phi_1 + \phi_2 + \dots + \phi_p) t - \\ & - \delta (\phi_1 + 2\phi_2 + \dots + p\phi_p) + \phi_1 [y_{t-1} - \alpha - \delta(t - 1)] + \\ & + \phi_2 [y_{t-2} - \alpha - \delta(t - 2)] + \dots + \\ & + \phi_p [y_{t-p} - \alpha - \delta(t - p)] + \varepsilon_t \end{aligned} \quad (8.3.2)$$

або

$$y_t = \alpha^* + \delta^* t + \phi_1^* y_{t-1}^* + \phi_2^* y_{t-2}^* + \dots + \phi_p^* y_{t-p}^* + \varepsilon_t, \quad (8.3.3)$$

де

$$\begin{aligned} \alpha^* &= \alpha (1 + \phi_1 + \phi_2 + \dots + \phi_p) - \delta (\phi_1 + 2\phi_2 + \dots + p\phi_p), \\ \delta^* &= \delta (1 + \phi_1 + \phi_2 + \dots + \phi_p), \\ \phi_j^* &= \phi_j \quad \text{для } j = 1, 2, \dots, p \end{aligned}$$

і

$$y_{t-j}^* = y_{t-j} - \alpha - \delta(t-j) \quad \text{для } j = 1, 2, \dots, p. \quad (8.3.4)$$

Ідея перетворення регресії до вигляду (8.3.3) належить Сіму, Стоку і Уотсону (Sims, Stock, Watson)(1990). Мета полягає у тому, що регресори (8.3.1) перетворюються у коваріаційно - стаціонарні випадкові змінні y_{t-j}^* ($j = 1, 2, \dots, p$) з нульовим математичним сподіванням. Таке перетворення регресорів відокремлює компоненти МНК-оцінки вектора коефіцієнтів з різними поведінками збіжності і забезпечує загальну методику для визначення асимптотичного розподілу. Загальний результат полягає в такому: якщо перетворене рівняння (8.3.3) оцінити за МНК, то коефіцієнти при коваріаційно - стаціонарних випадкових змінних $\hat{\phi}_j^*$ збігаються за розподілом з множником \sqrt{T} до нормального розподілу. Коефіцієнти $\hat{\alpha}_T^*$ і $\hat{\delta}_T^*$, розраховані за МНК у (8.3.3), поведуть себе асимптотично так само як $\hat{\alpha}_T$ і $\hat{\delta}_T$ для простої моделі часового тренда.

Початкову регресійну модель (8.3.1) можна записати у вигляді

$$y_t = x_t \beta + \varepsilon_t, \quad (8.3.5)$$

де

$$x_t = (y_{t-1} \ y_{t-2} \ \dots \ y_{t-p} \ 1 \ t), \quad \beta = \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \\ \alpha \\ \delta \end{pmatrix}. \quad (8.3.6)$$

За допомогою алгебраїчного перетворення (8.3.3), регресію (8.3.5) можна записати у вигляді

$$y_t = x_t G'(G')^{-1} \beta + \varepsilon_t = x_t^* \beta^* + \varepsilon_t. \quad (8.3.7)$$

Тут

$$G' \equiv \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ -\alpha + \delta & -\alpha + 2\delta & \dots & -\alpha + p\delta & 1 & 0 \\ -\delta & -\delta & \dots & -\delta & 0 & 1 \end{pmatrix}, \quad (8.3.8)$$

$$(G')^{-1} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ \alpha - \delta & \alpha - 2\delta & \dots & \alpha - p\delta & 1 & 0 \\ \delta & \delta & \dots & \delta & 0 & 1 \end{pmatrix},$$

$$(x_t^*)' = G x_t' = \begin{pmatrix} y_{t-1}^* \\ y_{t-2}^* \\ \vdots \\ y_{t-p}^* \\ 1 \\ t \end{pmatrix}, \quad (8.3.9)$$

$$\beta^* = (G')^{-1} \beta = \begin{pmatrix} \phi_1^* \\ \phi_2^* \\ \vdots \\ \phi_p^* \\ \alpha^* \\ \delta^* \end{pmatrix}. \quad (8.3.10)$$

Система (8.3.7) – лише алгебраїчний еквівалент початкової регресійної моделі (8.3.5). МНК-оцінка вектора β_T^* , яка обчислена на підставі регресії y_t стосовно x_t^* , обчислюється за формулою

$$\begin{aligned} \hat{\beta}_T^* &= \left[\sum_{t=1}^T (x_t^*)' x_t^* \right]^{-1} \left(\sum_{t=1}^T (x_t^*)' y_t \right) = \\ &= \left[G \left(\sum_{t=1}^T x_t' x_t \right) G' \right]^{-1} G \left(\sum_{t=1}^T x_t' y_t \right) = \\ &= (G')^{-1} \left(\sum_{t=1}^T x_t' x_t \right)^{-1} G^{-1} G \left(\sum_{t=1}^T x_t' y_t \right) = \\ &= (G')^{-1} \left(\sum_{t=1}^T x_t' x_t \right)^{-1} \left(\sum_{t=1}^T x_t' y_t \right) = (G')^{-1} \hat{\beta}_T, \end{aligned} \quad (8.3.11)$$

де $\hat{\beta}_T$ – МНК-оцінка вектора коефіцієнтів регресії y_t стосовно x_t . Отож, оцінки коефіцієнтів перетвореної регресії є лінійною

комбінацією оцінок коефіцієнтів початкової моделі. Значення y_t для заданого t пов'язане з перетвореною регресією так:

$$x_t^* \hat{\beta}_T^* = (Gx_t')'(G)^{-1} \hat{\beta}_T = x_t \hat{\beta}_T.$$

Отже, величина y_t для перетвореної регресії кількісно дорівнює величині з початкової регресії.

Звичайно, враховуючи лише дані y_t , ми фактично не зможемо оцінити перетворену регресію за МНК, тому що для структури x_t^* треба знати значення параметрів $\hat{\alpha}$ і $\hat{\delta}$. Проте корисно використати властивості гіпотетичної МНК-оцінки вектора параметрів у (8.3.7), оскільки перетворена регресія (8.3.7) легша для аналізу, ніж початкова регресія (8.3.5).

Крім того, якщо ми знайшли асимптотичний розподіл $\hat{\beta}_T^*$, асимптотичний розподіл $\hat{\beta}_T$ може бути виведений за допомогою оберненого перетворення у (8.3.11)

$$\hat{\beta}_T = G' \hat{\beta}_T^*. \tag{8.3.12}$$

8.3.2 Асимптотичний розподіл МНК оцінок параметрів перетвореної регресії

Теорема 8.2.1.

$$M_T (\hat{\beta}_T^* - \beta^*) \xrightarrow{d} N[0, \sigma(Q^*)^{-1}], \tag{8.3.13}$$

де

$$M_T = \begin{pmatrix} \sqrt{T} & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \sqrt{T} & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{T} & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & \sqrt{T} & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & T^{3/2} \end{pmatrix}, \tag{8.3.14}$$

$$Q^* = \begin{pmatrix} \gamma_0^* & \gamma_1^* & \gamma_2^* & \dots & \gamma_{p-1}^* & 0 & 0 \\ \gamma_1^* & \gamma_0^* & \gamma_1^* & \dots & \gamma_{p-2}^* & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \gamma_{p-1}^* & \gamma_{p-2}^* & \gamma_{p-3}^* & \dots & \gamma_0^* & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 1/2 \\ 0 & 0 & 0 & \dots & 0 & 1/2 & 1/3 \end{pmatrix} \quad (8.3.15)$$

і $\gamma_j^* = E[y_t^* y_{t-j}^*]$. Іншими словами, МНК-оцінка $\hat{\beta}_T^*$ є асимптотично нормальною, причому оцінка коефіцієнта часового тренда $\hat{\delta}_T^*$ збігається з множником $T^{3/2}$, а оцінки всіх інших коефіцієнтів збігаються з множниками \sqrt{T} .

Д о в е д е н н я. З (8.1.6) маємо

$$\hat{\beta}_T^* - \beta^* = \left[\sum_{t=1}^T (x_t^*)' x_t^* \right]^{-1} \left(\sum_{t=1}^T (x_t^*)' \varepsilon_t \right).$$

Помноживши ліву частину на M_T , матимемо

$$M_T(\hat{\beta}_T^* - \beta^*) = \left[M_T^{-1} \sum_{t=1}^T (x_t^*)' x_t^* M_T^{-1} \right]^{-1} \left(M_T^{-1} \sum_{t=1}^T (x_t^*)' \varepsilon_t \right). \quad (8.3.16)$$

З (8.3.9) одержимо

$$\begin{aligned} & \sum_{t=1}^T (x_t^*)' x_t^* = \\ & = \begin{pmatrix} \sum (y_{t-1}^*)^2 & \dots & \sum y_{t-1}^* y_{t-p}^* & \sum y_{t-1}^* & \sum t y_{t-1}^* \\ \sum y_{t-2}^* y_{t-1}^* & \dots & \sum y_{t-2}^* y_{t-p}^* & \sum y_{t-2}^* & \sum t y_{t-2}^* \\ \dots & \dots & \dots & \dots & \dots \\ \sum y_{t-p}^* y_{t-1}^* & \dots & \sum (y_{t-p}^*)^2 & \sum y_{t-p}^* & \sum t y_{t-p}^* \\ \sum y_{t-1}^* & \dots & \sum y_{t-p}^* & T & \sum t \\ \sum t y_{t-1}^* & \dots & \sum t y_{t-p}^* & \sum t & \sum t^2 \end{pmatrix} \end{aligned}$$

i

$$M_T^{-1} \sum_{t=1}^T (x_t^*)' x_t^* M_T^{-1} = T^{-1} \times$$

$$\times \begin{pmatrix} \sum (y_{t-1}^*)^2 & \dots & \sum y_{t-1}^* y_{t-p}^* & \sum y_{t-1}^* & T^{-1} \sum t y_{t-1}^* \\ \sum y_{t-2}^* y_{t-1}^* & \dots & \sum y_{t-2}^* y_{t-p}^* & \sum y_{t-2}^* & T^{-1} \sum t y_{t-2}^* \\ \dots & \dots & \dots & \dots & \dots \\ \sum y_{t-p}^* y_{t-1}^* & \dots & \sum (y_{t-p}^*)^2 & \sum y_{t-p}^* & T^{-1} \sum t y_{t-p}^* \\ \sum y_{t-1}^* & \dots & \sum y_{t-p}^* & T & T^{-1} \sum t \\ T^{-1} \sum t y_{t-1}^* & \dots & T^{-1} \sum t y_{t-p}^* & T^{-1} \sum t & T^{-2} \sum t^2 \end{pmatrix}.$$

Для перших p рядків і стовпчиків, (i, j) -ті елементи цієї матриці мають вигляд

$$T^{-1} \sum_{t=1}^T y_{t-i}^* y_{t-j}^*.$$

Оскільки y_t^* є стаціонарним $AR(p)$ процесом з нульовим математичним сподіванням, то ці елементи збігаються за ймовірністю до $\gamma_{|i-j|}^*$. Перші p елементів рядка $p + 1$ (або перші p елементів стовбця $p + 1$) мають вигляд

$$T^{-1} \sum_{t=1}^T y_{t-i}^*$$

і збігаються за ймовірністю до нуля. Перші p елементів рядка $p + 2$ (або перші p елементів стовбця $p + 2$) виглядають так:

$$T^{-1} \sum_{t=1}^T \frac{t}{T} y_{t-i}^*$$

і також збігаються до нуля. Матриця порядку (2×2) у нижньому правому куті збігається до

$$\begin{pmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{pmatrix}.$$

Отож,

$$M_T^{-1} \sum_{t=1}^T (x_t^*)' x_t^* M_T^{-1} \xrightarrow{p} Q^* \quad (8.3.17)$$

з матрицею Q^* з (8.3.15).

Далі розглянемо другий множник у (8.3.16)

$$M_T^{-1} \sum_{t=1}^T (x_t^*)' \varepsilon_t = \begin{pmatrix} T^{-1/2} \sum y_{t-1}^* \varepsilon_t \\ T^{-1/2} \sum y_{t-2}^* \varepsilon_t \\ \dots \\ T^{-1/2} \sum y_{t-p}^* \varepsilon_t \\ T^{-1/2} \sum \varepsilon_t \\ T^{-1/2} \sum (t/T) \varepsilon_t \end{pmatrix} = T^{-1/2} \sum_{t=1}^T \xi_t,$$

де

$$\xi_t = \begin{pmatrix} y_{t-1}^* \varepsilon_t \\ y_{t-2}^* \varepsilon_t \\ \dots \\ y_{t-p}^* \varepsilon_t \\ \varepsilon_t \\ (t/T) \varepsilon_t \end{pmatrix}.$$

Визначимо коваріаційну матрицю вектора ξ_t

$$E[\xi_t \xi_t'] = \sigma^2 Q_t^*,$$

де

$$Q_t^* = \begin{pmatrix} \gamma_0^* & \gamma_1^* & \gamma_1^* & \dots & \gamma_{p-1}^* & 0 & 0 \\ \gamma_1^* & \gamma_0^* & \gamma_1^* & \dots & \gamma_{p-2}^* & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ \gamma_{p-1}^* & \gamma_{p-2}^* & \gamma_{p-3}^* & \dots & \gamma_0^* & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & t/T \\ 0 & 0 & 0 & \dots & 0 & t/T & t^2/T^2 \end{pmatrix}$$

i

$$\frac{1}{T} \sum_{t=1}^T Q_t^* \rightarrow Q^*.$$

Використовуючи (8.1.26) можна показати, що

$$M_T^{-1} \sum_{t=1}^T x_t^* \varepsilon_t \xrightarrow{d} N[0, \sigma^2 Q^*]. \quad (8.3.18)$$

З (8.3.16), (8.3.17), (8.3.18) випливає, що

$$M_T(\hat{\beta}_T^* - \beta^*) \xrightarrow{d} N[0, (Q^*)^{-1} \sigma^2 Q^* (Q^*)^{-1}] = N[0, \sigma^2 (Q^*)^{-1}].$$

8.3.3 Асимптотичний розподіл МНК оцінки для початкової регресії

Використовуючи (8.3.8), з (8.3.12) маємо

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \vdots \\ \hat{\phi}_p \\ \hat{\alpha} \\ \hat{\delta} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 \\ -\alpha + \delta & -\alpha + 2\delta & \dots & -\alpha + p\delta & 1 & 0 \\ -\delta & -\delta & \dots & -\delta & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{\phi}_1^* \\ \hat{\phi}_2^* \\ \vdots \\ \hat{\phi}_p^* \\ \hat{\alpha}^* \\ \hat{\delta}^* \end{pmatrix}. \quad (8.3.19)$$

Оцінка МНК $\hat{\phi}_j$ початкової регресії дорівнює відповідній оцінці коефіцієнта перетвореної регресії $\hat{\phi}_j^*$, тому асимптотичний розподіл $\hat{\phi}_j$ описується (8.3.13). Оцінка $\hat{\alpha}_T = g'_\alpha \hat{\beta}_T^*$, де

$$g'_\alpha \equiv (-\alpha + \delta \quad -\alpha + 2\delta \quad \dots \quad -\alpha + p\delta \quad 1 \quad 0)$$

є лінійною комбінацією змінних, що збігаються до нормального розподілу з множником \sqrt{T} . Тому $\hat{\alpha}_T$ має той самий розподіл. З (8.3.13) випливає, що

$$\sqrt{T}(\hat{\alpha}_T - \alpha) \xrightarrow{d} N[0, \sigma^2 g'_\alpha (Q^*)^{-1} g'_\alpha]. \quad (8.3.20)$$

Нарешті, оцінка $\hat{\delta}_T$ це лінійна комбінація змінних, які збігаються з різними множниками

$$\hat{\delta}_T = g'_\delta \hat{\beta}_T + \hat{\delta}_T^*,$$

де

$$g'_\delta \equiv (-\delta \quad -\delta \quad \dots \quad -\delta \quad 0 \quad 0).$$

Асимптотичний розподіл суми визначається змінними з найменшою швидкістю збіжності

$$\begin{aligned} \sqrt{T}(\hat{\delta}_T - \delta) &= \sqrt{T}(\hat{\delta}_T^* + g'_\delta \hat{\beta}_T - \delta^* - g'_\delta \beta^*) \xrightarrow{p} \\ &\xrightarrow{p} \sqrt{T}(g'_\delta \hat{\beta}_T + \delta^* - \delta^* - g'_\delta \beta^*) = g'_\delta \sqrt{T}(\hat{\beta}_T - \beta^*) \xrightarrow{d} \\ &\xrightarrow{d} N[0, \sigma^2 g'_\delta (Q^*)^{-1} g'_\delta]. \end{aligned}$$

Отож, кожен окремих елемент з β_T є асимптотично нормальним і $O_p(T^{-1/2})$. Асимптотичний розподіл повного вектора $\sqrt{T}(\hat{\beta}_T - \beta)$ є нормальним. Зауважимо, що лінійна комбінація елементів $\hat{\beta}_T$ дає коефіцієнт часового тренда гіпотетичної регресії $\hat{\delta}_T^*$

$$\hat{\delta}_T^* = -g'_\delta \hat{\beta}_T + \hat{\delta}_T = \delta \hat{\phi}_{1,T} + \delta \hat{\phi}_{2,T} + \dots + \delta \hat{\phi}_{p,T} + \hat{\delta}_T$$

і при домноженні на \sqrt{T} збігається до δ^*

$$\sqrt{T}(\hat{\delta}_T^* - \delta^*) \xrightarrow{p} 0.$$

Тоді з (8.3.13) випливає, що

$$T^{3/2}(\hat{\delta}_T - \delta) \xrightarrow{d} N[0, \sigma^2 (q^*)^{p+2, p+2}],$$

де $(q^*)^{p+2, p+2}$ нижній правий елемент матриці $(Q^*)^{-1}$.

8.3.4 Тестування гіпотез

Розглянемо тестування гіпотези стосовно параметрів початкової моделі

$$H_0 : R\beta = r. \quad (8.3.21)$$

Тут R відома матриця $(m \times (p+2))$, r – відомий вектор $(m \times 1)$, і m – кількість обмежень. Вальдівська форма χ^2 тесту для гіпотези H_0 має вигляд

$$\chi_T^2 = (R\hat{\beta}_T - r)' \left[s_T^2 R \left(\sum_{t=1}^T x_t' x_t \right)^{-1} R' \right]^{-1} (R\hat{\beta}_T - r), \quad (8.3.22)$$

де $\hat{\beta}_T$ – МНК-оцінка β , яка ґрунтується на спостереженнях $\{y_{-p+1}, y_{-p+2}, \dots, y_0, y_1, \dots, y_T\}$ і

$$s_T^2 = \frac{1}{T - p - 2} \sum_{t=1}^T (y_t - x_t' \hat{\beta}_T)^2.$$

Статистику для тестування нульової гіпотези (8.3.21), яка задана в (8.3.22), можна переписати у вигляді

$$\begin{aligned} \chi_T^2 &= (R(\hat{\beta}_T - \beta))' \left[s_T^2 R \left(\sum_{t=1}^T x_t' x_t \right)^{-1} R' \right]^{-1} (R(\hat{\beta}_T - \beta)) = \\ &= (R G' (G')^{-1} (\hat{\beta}_T - \beta))' \times \\ &\quad \times \left[s_T^2 R G' (G')^{-1} \left(\sum_{t=1}^T x_t' x_t \right)^{-1} G^{-1} G R' \right]^{-1} \times \\ &\quad \times (R G' (G')^{-1} (\hat{\beta}_T - \beta)). \end{aligned} \quad (8.3.23)$$

Зауважимо, що

$$\begin{aligned} (G')^{-1} \left(\sum_{t=1}^T x'_t x_t \right)^{-1} G^{-1} &= \left[G \sum_{t=1}^T x'_t x_t G' \right]^{-1} = \\ &= \left(\sum_{t=1}^T (x_t^*)' x_t^* \right)^{-1} \end{aligned}$$

для x_t^* визначеного в (8.3.9). Аналогічно, з (8.3.10) і (8.3.11) маємо

$$(\hat{\beta}_T^* - \beta^*) = (G')^{-1} (\hat{\beta}_T - \beta).$$

Якщо прийняти

$$R^* \equiv R G',$$

то величина (8.3.23) може бути записана

$$\begin{aligned} \chi_T^2 &= (R^* (\hat{\beta}_T^* - \beta^*))' \times \\ &\times \left[s_T^2 R^* \left(\sum_{t=1}^T (x_t^*)' x_t^* \right)^{-1} (R^*)' \right]^{-1} (R^* (\hat{\beta}_T^* - \beta^*)). \end{aligned} \quad (8.3.24)$$

Величина (8.3.24) матиме розподіл $\chi^2[m]$. Похибка у перетвореній і початковій регресіях є однаковими, тому s_T^2 буде така сама для початкової і перетвореної моделей. Звідси випливає, що величина χ_T^2 з (8.3.22) також має розподіл $\chi^2[m]$.

Розділ 9

Одновимірні процеси з одиничним коренем

9.1 Асимптотичні властивості МНК оцінки параметра Гауссівського AR(1) процесу

Розглянемо МНК оцінку параметра Гауссівського AR(1) процесу з нульовим математичним сподіванням

$$y_t = \rho y_{t-1} + u_t, \quad (9.1.1)$$

де $u_t \sim N[0, \sigma^2]$, $y_0 = 0$. МНК оцінює ρ так:

$$\hat{\rho}_T = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2}. \quad (9.1.2)$$

З (9.1.2), використовуючи (9.1.1), випливає таке: якщо $|\rho|$ є меншим за одиницю, то

$$\sqrt{T}(\hat{\rho}_T - \rho) \xrightarrow{d} N[0, (1 - \rho^2)]. \quad (9.1.3)$$

Для $\rho = 1$ з (9.1.3) отримаємо

$$\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{p} 0. \quad (9.1.4)$$

тобто, якщо використовувати множник \sqrt{T} , то не матимемо невивродженого розподілу.

Покажемо таке: для того щоб отримати невивроджений асимптотичний розподіл $\hat{\rho}_T$ у випадку процесу з одиничним коренем $\hat{\rho}_T - 1$, треба домножувати на T , а не на \sqrt{T} .

Отож, оцінка коефіцієнта процесу з одиничним коренем збігається з швидшим темпом (T), ніж оцінка коефіцієнта для звичайної регресії (яка збігається з \sqrt{T}).

Якщо істинне значення ρ дорівнює одиниці, то різниця між оціненим $\hat{\rho}_T$ і істинним значенням буде

$$\hat{\rho}_T - 1 = \frac{\sum_{t=1}^T y_{t-1} u_t}{\sum_{t=1}^T y_{t-1}^2}. \quad (9.1.5)$$

Помножимо (9.1.5) на T

$$T(\hat{\rho}_T - 1) = \frac{(1/T) \sum_{t=1}^T y_{t-1} u_t}{(1/T^2) \sum_{t=1}^T y_{t-1}^2}. \quad (9.1.6)$$

Розглянемо чисельник рівності (9.1.6). Якщо істинне значення ρ дорівнює одиниці, то за припущення $y_0 = 0$, рівність (9.1.1) виражає процес

$$y_t = u_t + u_{t-1} + \dots + u_1. \quad (9.1.7)$$

З (9.1.7) випливає

$$y_t \sim N[0, \sigma^2 t]. \quad (9.1.8)$$

Зауважимо, що

$$y_t^2 = (y_{t-1} + u_t)^2 = y_{t-1}^2 + 2y_{t-1}u_t + u_t^2,$$

звідки

$$y_{t-1} u_t = \frac{1}{2} (y_t^2 - y_{t-1}^2 - u_t^2). \quad (9.1.9)$$

Підсумувавши (9.1.9) за від 1 до T , матимемо

$$\sum_{t=1}^T y_{t-1} u_t = \frac{1}{2} (y_T^2 - y_0^2) - \frac{1}{2} \sum_{t=1}^T u_t^2. \quad (9.1.10)$$

Враховуючи, що $y_0 = 0$, перетворимо рівність (9.1.10)

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} u_t = \frac{1}{2T} y_T^2 - \frac{1}{2T} \sum_{t=1}^T u_t^2 \quad (9.1.11)$$

і, поділивши обидві частини (9.1.11) на σ^2 , отримаємо

$$\begin{aligned} \left(\frac{1}{\sigma^2 T} \right) \sum_{t=1}^T y_{t-1} u_t &= \\ &= \left(\frac{1}{2} \right) \left(\frac{y_T}{\sigma \sqrt{T}} \right)^2 - \left(\frac{1}{2\sigma^2} \right) \left(\frac{1}{T} \sum_{t=1}^T u_t^2 \right). \end{aligned} \quad (9.1.12)$$

З (9.1.8) випливає, що $y_T/(\sigma\sqrt{T})$ має нормальний стандартний розподіл, тому його квадрат має розподіл $\chi^2[1]$

$$\left(\frac{y_T}{\sigma \sqrt{T}} \right)^2 \sim \chi^2[1]. \quad (9.1.13)$$

Сума $\sum_{t=1}^T u_t^2$ є сумою T незалежних однаково розподілених випадкових змінних, кожна з яких має математичне сподівання σ^2 , і тому за законом великих чисел

$$\frac{1}{T} \sum_{t=1}^T u_t^2 \xrightarrow{p} \sigma^2. \quad (9.1.14)$$

Використовуючи (9.1.13) та (9.1.14), з (9.1.12) випливає

$$\frac{1}{\sigma^2 T} \sum_{t=1}^T y_{t-1} u_t \xrightarrow{d} \frac{1}{2} (X - 1), \quad (9.1.15)$$

де $X \sim \chi^2[1]$.

Повертаючись до знаменника (9.1.6), розглянемо

$$\sum_{t=1}^T y_{t-1}^2. \quad (9.1.16)$$

Враховуючи з (9.1.8), що $y_{t-1} \sim N[0, \sigma^2(t-1)]$, матимемо $E[y_{t-1}^2] = \sigma^2(t-1)$. Розглянемо математичне сподівання суми (9.1.16)

$$E \left[\sum_{t=1}^T y_{t-1}^2 \right] = \sigma^2 \sum_{t=1}^T (t-1) = \frac{\sigma^2(T-1)T}{2}. \quad (9.1.17)$$

Звідси видно таке: для того щоб величина в (9.1.16) збігалася за розподілом, треба її поділити на T^2 , що і було зроблено в знаменнику (9.1.6).

Отже, якщо істинний процес є процесом випадкового блукання, то відхилення МНК-оцінки від істинного значення ($\hat{\rho}_T - 1$) треба помножити на T , а не на \sqrt{T} , для того, щоб отримати необхідний асимптотичний розподіл. Цей асимптотичний розподіл не є звичайним нормальним розподілом, а відношенням $\chi^2[1]$ розподілу в чисельнику і нестандартного розподілу в знаменнику.

9.2 Броунівський рух

Розглянемо процес

$$y_t = y_{t-1} + \varepsilon_t, \quad (9.2.1)$$

де ε_t – білий шум, причому

$$\varepsilon_t \sim i.i.d. N[0, 1].$$

Якщо процес починається з $y_0 = 0$, то аналогічно до (9.1.7) та (9.1.8) матимемо

$$y_t = \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t, \quad y_t \sim N[0, t].$$

Крім того, приріст величини y за проміжок часу від t до s

$$y_s - y_t = \varepsilon_{t+1} + \varepsilon_{t+2} + \dots + \varepsilon_s$$

має $N[0, (s-t)]$ розподіл і є незалежним від приросту величини y за проміжок часу від r до q для довільних $t < s < r < q$.

Розглянемо різницю між y_t і y_{t-1} : $\varepsilon_t = y_t - y_{t-1}$. Нехай ε_t є сумою двох незалежних гауссівських змінних

$$\varepsilon_t = e_{1t} + e_{2t}$$

з $e_{it} \sim i.i.d. N\left[0, \frac{1}{2}\right]$. Можемо пов'язати e_{1t} з різницею y_t і значенням y в деякий момент часу (скажімо, $y_{t-(1/2)}$), тобто

$$y_t - y_{t-(1/2)} = e_{1t}, \quad (9.2.2)$$

і e_{2t} – із різницею між $y_{t-(1/2)}$ і y_{t-1}

$$y_{t-(1/2)} - y_{t-1} = e_{2t}. \quad (9.2.3)$$

Процес, який є сумою процесів (9.2.2) та (9.2.3), матиме ті самі властивості як і (9.2.1), оскільки

$$y_t - y_{t-1} = e_{1t} + e_{2t} \sim i.i.d. N[0, 1].$$

Зауважимо, що процеси (9.2.2) та (9.2.3) визначені на множині не цілих чисел $\{t + \frac{1}{2}\}_{t=0}^{\infty}$. Крім того,

$$y_s - y_t \sim N[0, s - t]$$

і приріст $y_s - y_t$ не залежить від приросту на довільному іншому інтервалі, який не перетинається з цим.

Аналогічно, можемо уявити поділ різниці між y_{t-1} і y_t на N окремих субперіодів

$$y_t - y_{t-1} = e_{1t} + e_{2t} + \dots + e_{Nt},$$

де $e_{it} \sim i.i.d. N[0, 1/N]$. Результатом буде процес з тими самими властивостями як і в (9.2.1), визначений в густішій сітці даних, оскільки збільшуємо N . Описаний процес при $N \rightarrow \infty$ називається стандартним Броунівським рухом. Значення цього процесу в точці t позначаємо $W(t)$.

Неперервний процес є випадковою величиною, яка визначена при довільних невід'ємних дійсних значеннях t , на відміну від дискретного процесу, який визначений лише для цілих значень t . Щоб підкреслити відмінність, помістимо t в круглі дужки, коли описуватимемо значення неперервної змінної в точці t (як в $W(t)$) і будемо використовувати індекси для дискретної змінної (як в y_t). Дискретний процес зобразили як злічену послідовність випадкових величин $\{y_t\}_{t=1}^{\infty}$. Реалізацію неперервного процесу можна розглядати як стохастичну функція $W(\cdot)$, де $W : t \in [0, \infty) \rightarrow \mathbb{R}^1$.

Часто зручно обмежити аналіз значень t в межах замкненого інтервалу $[0, 1]$. Внаслідок цього розглядаємо стандартний $W(\cdot)$, як функцію на проміжку $t \in [0, 1]$.

Означення 9.2.1. *Стандартним Броунівським рухом називається неперервний стохастичний процес $(W(t), t \in [0, 1])$, який пов'язує кожну величину $t \in [0, 1]$ з скалярною випадковою величиною $W(t)$ і володіє властивостями:*

- 1) $W(0) = 0$;
- 2) для довільних значень $0 \leq t_1 < t_2 < \dots < t_k \leq 1$ прирости $W(t_2) - W(t_1)$, $W(t_3) - W(t_2)$, \dots , $W(t_k) - W(t_{k-1})$ є незалежним багатовимірним гауссіаном, причому

$$W(s) - W(t) \sim N[0, s - t];$$

- 3) $W(t)$ є неперервним за t з ймовірністю 1.

Процес Броунівського руху також називають вінеровим процесом.

Означення 9.2.2. *Нехай X_t – випадковий процес. Якщо для будь-якого набору моментів часу $t_1 < t_2 < \dots < t_n$ і $\forall n \in N$ випадкові величини $X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ взаємно незалежні, то процес X_t називається процесом з незалежними приростами.*

Означення 9.2.3. *Якщо розподіл випадкової величини $X_{t+h} - X_t$ залежить лише від h і не залежить від t , то процес X_t називається однорідним процесом.*

Отже, стандартний вінеровий процес є однорідним процесом з незалежними приростами, для якого різниці мають нормальний розподіл з нульовим математичним сподіванням і дисперсією $t - s$ ($s < t$).

Розглянемо процеси другого порядку, тобто процеси, які мають скінчені моменти другого порядку. Позначимо через H сукупність усіх випадкових величин, які мають скінчені моменти другого порядку: $H = \{\xi : E[|\xi|^2] < \infty\}$. H є лінійним простором. Якщо ввести скалярний добуток $(\xi, \eta) = E[\xi\eta]$ і не вважати різними випадкові величини, які співпадають з ймовірністю одиниця, то H є гільбертовим простором.

Означення 9.2.4. Нехай кожному $t \in T$ поставлено у відповідність випадкову величину $X_t \in H$. Тоді X_t називається випадковим процесом другого порядку, заданим на T .

Для процесів другого порядку визначені числові характеристики $\mu(t) = E[X_t]$ і коваріаційна функція

$$C(t, s) = E[(X_t - \mu(t))(X_s - \mu(s))].$$

Коваріаційну функцію $C(t, s)$ можна записати також у вигляді

$$C(t, s) = E[X_t X_s] - \mu(t)\mu(s).$$

Означення 9.2.5. Процес $X_t = W_t - tW_1$, де W_t - вінеровий процес, називається вінеровим мостом.

Вінеровий міст є процесом другого порядку, числові характеристики якого відповідно дорівнюють

$$\mu(t) = E[X_t] = 0, \quad C(t, s) = E[X_t X_s] = \min\{t, s\} - st.$$

Означення 9.2.6. Процес другого порядку X_t називається неперервним у середньому квадратичному в точці t_0 , якщо *l.i.m.* $t \rightarrow t_0 X_t = X_{t_0}$, тобто

$$\|X_t - X_{t_0}\|^2 = E[(X_t - X_{t_0})^2] \rightarrow 0.$$

Теорема 9.2.1. Процес другого порядку X_t є неперервним в середньому квадратичному в точці t_0 тоді і лише тоді, коли функція $\mu(t)$ неперервна в точці t_0 , а функція $C(t, s)$ неперервна в точці (t_0, t_0) .

Означення 9.2.7. Процес X_t називається диференційовним у середньому квадратичному в точці t_0 , якщо існує границя

$$l.i.m. \ h \rightarrow 0 \frac{X_{t_0+h} - X_{t_0}}{h} = X'_{t_0}.$$

Процес диференційовний на $[a, b]$, якщо він диференційовний у кожній точці відрізка $[a, b]$.

Теорема 9.2.2. Для того щоб процес X_t був диференційовним в середньому квадратичному на $[a, b]$ необхідно, щоб існували похідні

$$\mu'(t), \quad \frac{\partial C(t, s)}{\partial t}, \quad \frac{\partial C(t, s)}{\partial s}, \quad \frac{\partial^2 C(t, s)}{\partial t \partial s}, \quad t \in (a, b), s \in (a, b)$$

і достатньо, щоб існували $\mu'(t)$ і

$$\lim_{h \rightarrow 0, h_1 \rightarrow 0} \frac{C(t+h, t+h_1) - C(t, t+h_1) - C(t+h_1, t) + C(t, t)}{h h_1}.$$

Означення 9.2.8. Випадковий процес X_t називається інтегрованим у середньому квадратичному на відрізку $[a, b]$, якщо існує в H границя інтегральних сум

$$\sum_{k=0}^{n-1} X_{c_k} (t_{k+1} - t_k),$$

де $a = t_0 \leq c_0 \leq t_1 \leq \dots \leq t_{n-1} \leq c_{n-1} \leq t_n$, коли $\max_k (t_{k+1} - t_k) \rightarrow 0$ і границя не залежить від способу вибору

точок c_k . Ця границя позначається $\int_a^b X_t dt$.

Теорема 9.2.3. Якщо існують інтеграли Рімана

$$\int_a^b \mu(t) dt, \quad \int_a^b \int_a^b C(t, s) dt ds,$$

то процес X_t інтегрований на $[a, b]$ і

$$E \left[\int_a^b X_t dt \right] = \int_a^b \mu(t) dt,$$
$$E \left[\int_a^b X_t dt \int_a^b X_s ds \right] = \int_a^b \int_a^b C(t, s) dt ds.$$

Від стандартного Броунівського руху можуть бути утворені інші неперервні процеси. Зокрема, процес

$$Z(t) = \sigma W(t)$$

є однорідним процесом з незалежними приростами, причому $z_t \sim N[0, \sigma^2 t]$. Такий процес називається Броунівським рухом з відхиленням σ . Отож, стандартний Броунівський рух можна називати Броунівським рухом з одиничним відхиленням. Константу σ^2 також називають коефіцієнтом дифузії.

9.3 Асимптотичні властивості авторегресії першого порядку з одиничним коефіцієнтом

Визначимо асимптотичні розподіли деяких простих регресій, які містять одиничний корінь.

Твердження 9.3.1. *Нехай ξ_t описується процесом*

$$\xi_t = \xi_{t-1} + u_t,$$

де $\xi_0 = 0$ і $\{u_t\}$ є *i.i.d.* послідовністю з математичним сподіванням нуль і дисперсією σ^2 . Тоді:

- а) $\frac{1}{\sqrt{T}} \sum_{t=1}^T u_t \xrightarrow{d} \sigma W(1);$
- б) $\frac{1}{T} \sum_{t=1}^T \xi_{t-1} u_t \xrightarrow{d} \frac{1}{2} \sigma^2 ([W(1)]^2 - 1);$
- в) $\frac{1}{T^{3/2}} \sum_{t=1}^T t u_t \xrightarrow{d} \sigma W(1) - \sigma \int_0^1 W(r) dr;$
- г) $\frac{1}{T^{3/2}} \sum_{t=1}^T \xi_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r) dr;$
- д) $\frac{1}{T^2} \sum_{t=1}^T \xi_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2 dr;$
- е) $\frac{1}{T^{5/2}} \sum_{t=1}^T t \xi_{t-1} \xrightarrow{d} \sigma \int_0^1 r W(r) dr;$
- є) $\frac{1}{T^3} \sum_{t=1}^T t \xi_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 r [W(r)]^2 dr;$

h) $\frac{1}{T^{k+1}} \sum_{t=1}^T t^k \rightarrow \frac{1}{k+1}$ для $k = 0, 1, \dots$

Твердження 9.3.2. Нехай $u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, де $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ і $\{\varepsilon_t\}$ є *i.i.d.* послідовністю з математичним сподіванням нуль, дисперсією σ^2 і скінченими четвертими моментами. Позначимо

$$\gamma_j \equiv E[u_t u_{t-j}] = \sigma^2 \sum_{s=0}^{\infty} \psi_s \psi_{s+j}, \quad j = 0, 1, 2, \dots,$$

$$\lambda \equiv \sigma \sum_{j=0}^{\infty} \psi_j = \sigma \psi(1).$$

Тоді

a) $\frac{1}{T^{1/2}} \sum_{t=1}^T u_t \xrightarrow{d} \lambda W(1);$

б) $\frac{1}{T^{1/2}} \sum_{t=1}^T u_{t-j} \varepsilon_t \xrightarrow{d} N[0, \sigma^2 \gamma_0], \quad j = 1, 2, \dots;$

в) $\frac{1}{T} \sum_{t=1}^T u_t u_{t-j} \xrightarrow{p} \gamma_j, \quad j = 0, 1, 2, \dots;$

г) $\frac{1}{T} \sum_{t=1}^T \xi_{t-1} \varepsilon_t \xrightarrow{d} \frac{1}{2} \sigma \lambda ([W(1)]^2 - 1);$

д) $\frac{1}{T} \sum_{t=1}^T \xi_{t-1} u_{t-j} \xrightarrow{d} \begin{cases} (1/2) (\lambda^2 [W(1)]^2 - \gamma_0) & j = 0 \\ (1/2) (\lambda^2 [W(1)]^2 - \gamma_0) + \\ + \gamma_0 + \gamma_1 + \dots + \gamma_{j-1} & j = 1, 2, \dots, \end{cases};$

е) $\frac{1}{T^{3/2}} \sum_{t=1}^T \xi_{t-1} \xrightarrow{d} \lambda \int_0^1 W(r) dr;$

$$\epsilon) \frac{1}{T^{3/2}} \sum_{t=1}^T t u_{t-j} \xrightarrow{d} \lambda \left(W(1) - \int_0^1 W(r) dr \right), \quad j = 0, 1, 2, \dots;$$

$$\text{ж) } \frac{1}{T^2} \sum_{t=1}^T \xi_{t-1}^2 \xrightarrow{d} \lambda^2 \int_0^1 [W(r)]^2 dr;$$

$$\text{з) } \frac{1}{T^{5/2}} \sum_{t=1}^T t \xi_{t-1} \xrightarrow{d} \lambda \int_0^1 r W(r) dr;$$

$$\text{и) } \frac{1}{T^3} \sum_{t=1}^T t \xi_{t-1}^2 \xrightarrow{d} \lambda^2 \int_0^1 r [W(r)]^2 dr;$$

$$\text{і) } \frac{1}{T^{k+1}} \sum_{t=1}^T t^k \rightarrow \frac{1}{\nu + 1} \quad k = 0, 1, \dots$$

Зауважимо таке: оскільки $W(1) \sim N[0, 1]$, а $[W(1)]^2 \sim \chi^2[1]$, то деякі з цих результатів можна спростити. А саме, величина в пункті (а) має $N[0, \lambda^2]$ розподіл, в (г) – $(1/2)\sigma\lambda [\chi^2(1) - 1]$, в (е) і (є) – $N[0, \lambda^2/3]$ розподіл.

Наведені твердження можна використати при обчисленні асимптотичних розподілів статистик для тестування регресій, які містять одиничні корені.

9.4 Тести Дікі - Фуллера

9.4.1 Регресія без сталої і часового тренда у випадку, коли істинний процес є випадковим блуканням

Розглянемо спочатку МНК-оцінку коефіцієнта ρ для $AR(1)$ процесу

$$y_t = \rho y_{t-1} + u_t, \quad (9.4.1)$$

де u_t є білим шумом з дисперсією σ^2 . Дослідимо властивості МНК-оцінки параметра ρ

$$\hat{\rho}_T = \frac{\sum_{t=1}^T y_{t-1} y_t}{\sum_{t=1}^T y_{t-1}^2} \quad (9.4.2)$$

у випадку, якщо істинне значення ρ є одиниця. З (9.1.6) отримаємо стандартне відхилення оцінки від істинного значення

$$T(\hat{\rho}_T - 1) = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} u_t}{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2}. \quad (9.4.3)$$

Якщо істинне значення ρ одиниця, то

$$y_t = y_0 + u_1 + u_2 + \dots + u_t. \quad (9.4.4)$$

За винятком початкового значення y_0 (яке не впливає на асимптотичний розподіл), величина y_t є такою ж самою, як і величина ξ_t з твердження 9.2.1. З результату (б) цього твердження маємо

$$\frac{1}{T} \sum_{t=1}^T y_{t-1} u_t \xrightarrow{d} \frac{1}{2} \sigma^2 ([W(1)]^2 - 1). \quad (9.4.5)$$

З результату (д), отримаємо

$$\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 [W(r)]^2 dr. \quad (9.4.6)$$

Оскільки (9.4.3) є неперервною функцією від (9.4.5) і (9.4.6), то за нульової гіпотези $H_0 : \rho = 1$ матимемо асимптотичний

розподіл МНК-відхилення оцінки $\hat{\rho}_T$ від одиниці

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{(1/2) ([W(1)]^2 - 1)}{\int_0^1 [W(r)]^2 dr}. \quad (9.4.7)$$

Згадаємо, що $[W(1)]^2 \in \chi^2[1]$ розподілом. Ймовірність того, що $\chi^2[1]$ змінна набуває значення менші за нуль дорівнює 0,68, а оскільки знаменник (9.4.7) додатний, то і ймовірність того, що $\hat{\rho}_T - 1 \in$ від'ємним, наближається до 0,68 при $T \rightarrow \infty$. Іншими словами, для двох третіх вибірок, які породжені процесом випадкового блукання, оцінка $\hat{\rho}_T$ буде менша, ніж істинне значення 1. Це є результатом того, що граничний розподіл $T(\hat{\rho}_T - 1)$ зміщений вліво.

У стаціонарному випадку, коли $|\rho| < 1$ граничний розподіл $\sqrt{T}(\hat{\rho}_T - \rho)$ є симетричним стосовно нуля. На відміну від цього, у випадку, коли дійсне значення $\rho = 1$, граничний розподіл $T(\hat{\rho}_T - 1)$ є асиметричний, для якого від'ємні значення є майже вдвічі ймовірніші, ніж додатні.

Доведемо, що $\hat{\rho}_T$ є змістовною оцінкою істинного значення ($\rho = 1$). Для цього поділимо (9.4.3) на \sqrt{T}

$$\sqrt{T}(\hat{\rho}_T - 1) = \frac{\frac{1}{T^{3/2}} \sum_{t=1}^T y_{t-1} u_t}{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2}. \quad (9.4.8)$$

За твердженням 9.2.1(b) чисельник в (9.4.8) збігається за розподілом до $\frac{1}{2T^{1/2}} \sigma^2 (X - 1)$, де X - випадкова величина, яка має $\chi^2[1]$ - розподіл. Оскільки $\chi^2[1]$ має скінченну дисперсію, то дисперсія чисельника в (9.4.8) має порядок $1/T$. Тому чисельник збігається за ймовірністю до нуля. Звідси, оскільки з (9.1.17) випливає, що знаменник збігається за ймовірністю до

σ^2 , то

$$\sqrt{T}(\hat{\rho}_T - 1) \xrightarrow{p} 0.$$

Результат (9.4.7) дає підстави використовувати оцінку $\hat{\rho}_T$ для того, щоб перевірити нульову гіпотезу на одиничний корінь окремо без обчислення його стандартної похибки.

Інший відомий спосіб перевірки нульової гіпотези $H_0 : \rho = 1$ ґрунтується на звичайному МНК t - тесті цієї гіпотези

$$t_T = \frac{(\hat{\rho}_T - 1)}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{(\hat{\rho}_T - 1)}{\left(s_T^2 / \sum_{t=1}^T y_{t-1}^2\right)^{1/2}}, \quad (9.4.9)$$

де $\hat{\sigma}_{\hat{\rho}_T}$ є звичайною стандартною похибкою коефіцієнта, оціненого за МНК

$$\hat{\sigma}_{\hat{\rho}_T} = \left(s_T^2 / \sum_{t=1}^T y_{t-1}^2\right)^{1/2},$$

а s_T^2 – оцінка дисперсії залишків

$$s_T^2 = \frac{1}{T-1} \sum_{t=1}^T (y_t - \hat{\rho}_T y_{t-1})^2.$$

Хоча t - статистику (9.4.9) обчислюють звичайним шляхом, проте у випадку, коли істинний процес охарактеризований $\rho = 1$, вона не має граничного асимптотично нормального розподілу. Щоб знайти відповідний граничний розподіл, зауважимо, що (9.4.9) можна зобразити як

$$t_T = \frac{T(\hat{\rho}_T - 1) \left(\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2\right)^{1/2}}{(s_T^2)^{1/2}}. \quad (9.4.10)$$

Підставивши в останню рівність (9.4.3), матимемо

$$t_T = \frac{\frac{1}{T} \sum_{t=1}^T y_{t-1} u_t}{\left(\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2 \right)^{1/2} (s_T^2)^{1/2}}. \quad (9.4.11)$$

Із змістовності $\hat{\rho}_T$ випливає $s_T^2 \xrightarrow{p} \sigma^2$. Тоді з (9.4.5) та (9.4.6) матимемо, що при $T \rightarrow \infty$

$$\begin{aligned} t_T &\xrightarrow{d} \frac{(1/2) \sigma^2 ([W(1)]^2 - 1)}{\left(\sigma^2 \int_0^1 [W(r)]^2 dr \right)^{1/2} (\sigma^2)^{1/2}} = \\ &= \frac{(1/2) ([W(1)]^2 - 1)}{\left(\int_0^1 [W(r)]^2 dr \right)^{1/2}}. \end{aligned} \quad (9.4.12)$$

Зауважимо, що тестові статистики (9.4.7) та (9.4.12) є прикладами тесту Дікі - Фуллера для одиничного кореня у випадку регресії $y_t = \rho y_{t-1} + u_t$, яка не містить сталої і часового тренда.

9.4.2 Регресія, яка містить сталу, але не містить часового тренда, у випадку, коли істинний процес є випадковим блуканням

Як і в попередньому випадку, досліджуємо дані, які описуються процесом

$$y_t = y_{t-1} + u_t,$$

де u_t – білий шум з дисперсією σ^2 . Припустимо тепер, що сталу включено в AR(1) - специфікацію, яку оцінюємо за МНК

$$y_t = \alpha + \rho y_{t-1} + u_t \quad (9.4.13)$$

Дослідимо властивості МНК оцінок

$$\begin{pmatrix} \hat{\alpha}_T \\ \hat{\rho}_T \end{pmatrix} = \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum y_{t-1}y_t \end{pmatrix} \quad (9.4.14)$$

згідно з нульовою гіпотезою $H_0 : \alpha = 0, \rho = 1$ (тут Σ позначає суму за t від 1 до T). Опишемо відхилення вектора МНК-оцінки $\hat{\beta}_T = (\hat{\alpha}_T, \hat{\rho}_T)'$ від істинного значення $\beta = (0, 1)$. Із загальної формули

$$\hat{\beta}_T - \beta = \begin{pmatrix} T & \sum x'_t x_t \\ \sum x'_t x_t & \sum x'_t u_t \end{pmatrix}^{-1} \begin{pmatrix} \sum u_t \\ \sum x'_t u_t \end{pmatrix} \quad (9.4.15)$$

для цього випадку матимемо

$$\begin{pmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{pmatrix} = \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum u_t \\ \sum y_{t-1}u_t \end{pmatrix}. \quad (9.4.16)$$

Як і у пункті 9.4.1, y_t має такі самі властивості як і змінна ξ_t , описана у твердженні 9.2.1. Результат (г) цього твердження засвідчує, що суму $\sum y_{t-1}$ треба поділити на $T^{3/2}$, для того щоб отримати випадкову змінну, яка збігається за розподілом

$$\frac{1}{T^{3/2}} \sum_{t=1}^T y_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r) dr. \quad (9.4.17)$$

Іншими словами,

$$\sum_{t=1}^T y_{t-1} = O_p(T^{3/2}).$$

Аналогічно з (9.4.5) і (9.4.6) впливає

$$\sum_{t=1}^T y_{t-1}u_t = O_p(T), \quad \sum_{t=1}^T y_{t-1}^2 = O_p(T^2),$$

а з твердження 9.2.1 (а)

$$\sum u_t = O_p(T^{1/2}).$$

Отож, порядок збіжності окремих членів рівності (9.4.16) такий:

$$\begin{pmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{pmatrix} = \begin{pmatrix} O_p(T) & O_p(T^{3/2}) \\ O_p(T^{3/2}) & O_p(T^2) \end{pmatrix}^{-1} \begin{pmatrix} O_p(T^{1/2}) \\ O_p(T) \end{pmatrix}. \quad (9.4.18)$$

З (9.4.18) видно, що оцінки $\hat{\alpha}_T$ і $\hat{\rho}_T$ мають різні степені збіжності тому для опису їхніх граничних розподілів використаємо розрахункову матрицю M_T . Помножимо (9.4.15) на M_T

$$\begin{aligned} M_T(\hat{\beta}_T - \beta) &= M_T \left[\sum_{t=1}^T x_t x_t' \right]^{-1} M_T M_T^{-1} \left[\sum_{t=1}^T x_t u_t \right] = \quad (9.4.19) \\ &= \left(M_T^{-1} \left[\sum_{t=1}^T x_t x_t' \right] M_T^{-1} \right)^{-1} \left(M_T^{-1} \left[\sum_{t=1}^T x_t u_t \right] \right). \end{aligned}$$

Враховуючи (9.4.18), робимо висновок, що матриця M_T повинна мати вигляд

$$M_T \equiv \begin{pmatrix} T^{1/2} & 0 \\ 0 & T \end{pmatrix}. \quad (9.4.20)$$

Підставивши (9.4.20) в (9.4.19), матимемо

$$\begin{aligned} &\begin{pmatrix} T^{1/2} & 0 \\ 0 & T \end{pmatrix} \begin{pmatrix} \hat{\alpha}_T \\ \hat{\rho}_T - 1 \end{pmatrix} = \\ &= \left[\begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{pmatrix} \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix} \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{pmatrix} \right]^{-1} \times \\ &\quad \times \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-1} \end{pmatrix} \begin{pmatrix} \sum u_t \\ \sum y_{t-1} u_t \end{pmatrix} \end{aligned}$$

або

$$\begin{aligned} \begin{pmatrix} T^{1/2} \hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{pmatrix} &= \quad (9.4.21) \\ &= \begin{pmatrix} 1 & T^{-3/2} \sum y_{t-1} \\ T^{-3/2} \sum y_{t-1} & T^{-2} \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2} \sum u_t \\ T^{-1} \sum y_{t-1} u_t \end{pmatrix}. \end{aligned}$$

Розглянемо перший член правої частини рівності (9.4.21). Результати (9.4.6) і (9.4.17) зумовлюють збіжність

$$\begin{aligned} \begin{pmatrix} 1 & T^{-3/2} \sum y_{t-1} \\ T^{-3/2} \sum y_{t-1} & T^{-2} \sum y_{t-1}^2 \end{pmatrix} &\xrightarrow{d} \quad (9.4.22) \\ &\xrightarrow{d} \begin{pmatrix} 1 & \sigma \int W(r) dr \\ \sigma \int W(r) dr & \sigma^2 \int [W(r)]^2 dr \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & \sigma \end{pmatrix} \begin{pmatrix} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \sigma \end{pmatrix}, \end{aligned}$$

де знак інтеграла означає інтегрування за r від 0 до 1.

Аналогічно, результат (а) твердження 9.2.1 та (9.4.5) визначає асимптотичний розподіл другого члена рівності (9.4.21)

$$\begin{aligned} \begin{pmatrix} T^{-1/2} \sum u_t \\ T^{-1} \sum y_{t-1} u_t \end{pmatrix} &\xrightarrow{d} \begin{pmatrix} \sigma W(1) \\ (1/2) \sigma^2 ([W(1)]^2 - 1) \end{pmatrix} = \quad (9.4.23) \\ &= \sigma \begin{pmatrix} 1 & 0 \\ 0 & \sigma \end{pmatrix} \begin{pmatrix} W(1) \\ (1/2) ([W(1)]^2 - 1) \end{pmatrix}. \end{aligned}$$

Підставивши (9.4.22) та (9.2.23) в (9.4.21), отримаємо

$$\begin{aligned}
 & \left(\begin{array}{c} T^{1/2} \hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{array} \right) \xrightarrow{d} \\
 & \xrightarrow{d} \sigma \left(\begin{array}{cc} 1 & 0 \\ 0 & \sigma \end{array} \right)^{-1} \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \times \\
 & \quad \times \left(\begin{array}{cc} 1 & 0 \\ 0 & \sigma \end{array} \right)^{-1} \left(\begin{array}{c} W(1) \\ (1/2) ([W(1)]^2 - 1) \end{array} \right) = \\
 & = \left(\begin{array}{cc} \sigma & 0 \\ 0 & 1 \end{array} \right) \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \times \\
 & \quad \times \left(\begin{array}{c} W(1) \\ (1/2) ([W(1)]^2 - 1) \end{array} \right). \tag{9.4.24}
 \end{aligned}$$

Правильна рівність

$$\begin{aligned}
 & \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} = \\
 & = \Delta^{-1} \left(\begin{array}{cc} \int [W(r)]^2 dr & - \int W(r) dr \\ - \int W(r) dr & 1 \end{array} \right), \tag{9.4.25}
 \end{aligned}$$

де

$$\Delta \equiv \int [W(r)]^2 dr - \left(\int W(r) dr \right)^2. \tag{9.4.26}$$

Отож, другий елемент у векторному рівнянні (9.4.24) матиме розподіл

$$T(\hat{\rho}_T - 1) \xrightarrow{d} \frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\int [W(r)]^2 dr - [\int W(r) dr]^2}. \tag{9.4.27}$$

З (9.4.24) видно, що ні оцінка $\hat{\alpha}_T$, ні $\hat{\rho}_T$ не мають асимптотично нормального розподілу. Крім того, асимптотичний розподіл оцінки $\hat{\rho}$ в (9.4.27) не є такий як в (9.4.7), тобто, якщо

стала включена в регресію, то треба бути використати іншу таблицю критичних значень.

Зауважимо, що цей розподіл є більш асиметричний, ніж розподіл, який використовували у пункті 9.4.1. Тому у випадку, коли стала включено в регресію, оцінений коефіцієнт при y_{t-1} повинен бути далеким від одиниці, щоб відхилити нульову гіпотезу на одиничний корінь. Зокрема, якщо обсяг вибірки для $T > 25$, то 95% оцінених значень $\hat{\rho}_T$ будуть меншими за одиницю. Наприклад, якщо оцінене значення $\hat{\rho}_T = 0,999$ для вибірки обсягу $T = 100$, то нульова гіпотеза $\rho = 1$ буде відхилена на користь альтернативи $\rho > 1$. Якщо істинне значення ρ дорівнює одиниці, то не отримаємо оцінку таку велику, як 0,999.

Дікі та Фуллер (Dickey, Fuller) також запропонували альтернативний тест, що ґрунтується на МНК t -статистиці тестування нульової гіпотези $H_0 : \rho = 1$

$$t_T = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}_T}}, \quad (9.4.28)$$

де

$$\hat{\sigma}_{\hat{\rho}_T}^2 = s_T^2 (0 \ 1) \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (9.4.29)$$

$$s_T^2 = \frac{1}{T-2} \sum_{t=1}^T (y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2.$$

Зазначимо таке: якщо обидві частини рівності (9.4.29) помножити на T^2 , то отримаємо

$$\begin{aligned} T^2 \hat{\sigma}_{\hat{\rho}_T}^2 &= s_T^2 (0 \ T) \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ T \end{pmatrix} = \\ &= s_T^2 (0 \ 1) M_T \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix}^{-1} M_T \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \end{aligned} \quad (9.4.30)$$

де M_T матриця з (9.4.20). Використавши (9.4.22), матимемо

$$\begin{aligned}
 M_T \left(\begin{array}{cc} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{array} \right)^{-1} M_T &= \quad (9.4.31) \\
 = \left(M_T^{-1} \left[\begin{array}{cc} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{array} \right] M_T^{-1} \right)^{-1} &= \\
 = \left(\begin{array}{cc} 1 & T^{-3/2} \sum y_{t-1} \\ T^{-3/2} \sum y_{t-1} & T^{-2} \sum y_{t-1}^2 \end{array} \right) \xrightarrow{d} & \\
 \xrightarrow{d} \left(\begin{array}{cc} 1 & 0 \\ 0 & \sigma \end{array} \right)^{-1} \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \left(\begin{array}{cc} 1 & 0 \\ 0 & \sigma \end{array} \right)^{-1}. &
 \end{aligned}$$

Отож, з (9.4.30) випливає, що

$$\begin{aligned}
 T^2 \hat{\sigma}_{\hat{\rho}_T}^2 &\xrightarrow{p} \quad (9.4.32) \\
 &\xrightarrow{p} s_T^2 (0 \ \sigma^{-1}) \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \left(\begin{array}{c} 0 \\ \sigma^{-1} \end{array} \right).
 \end{aligned}$$

Легко показати, що

$$s_T^2 \xrightarrow{p} \sigma^2. \quad (9.4.33)$$

Тоді з (9.4.32) отримаємо

$$\begin{aligned}
 T^2 \hat{\sigma}_{\hat{\rho}_T}^2 &\xrightarrow{d} (0 \ 1) \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \left(\begin{array}{c} 0 \\ 1 \end{array} \right) = \\
 &= \frac{1}{\int [W(r)]^2 dr - \left[\int W(r) dr \right]^2}. \quad (9.4.34)
 \end{aligned}$$

Отож, асимптотичний розподіл МНК t -статистики в (9.4.28) такий:

$$\begin{aligned}
 t_T &= \frac{T(\hat{\rho}_T - 1)}{(T^2 \hat{\sigma}_{\hat{\rho}_T}^2)^{1/2}} \xrightarrow{p} \\
 &\xrightarrow{p} T(\hat{\rho}_T - 1) \left(\int [W(r)]^2 dr - \left[\int W(r) dr \right]^2 \right)^{1/2}.
 \end{aligned}$$

Звідси, використовуючи (9.4.27), матимемо

$$t_T \xrightarrow{d} \frac{(1/2) ([W(r)]^2 - 1) - W(1) \int W(r) dr}{\left(\int [W(r)]^2 dr - [\int W(r) dr]^2 \right)^{1/2}}. \quad (9.4.35)$$

Розподіл (9.4.35) використовують для тестування нульової гіпотези $H_0 : \rho = 1$. Припущення, на якому ґрунтувалося твердження (9.4.24), полягає в тому, що істинне значення $\alpha = 0$. Отож, у цій специфікації доцільно перевірити на одиничний корінь за допомогою об'єднання гіпотез $\alpha = 0$ та $\rho = 1$. Діккі і Фуллер, використавши метод Монте Карло, розраховали розподіл F -статистики Вальда для цієї гіпотези.

9.4.3 Регресія, яка містить сталу, але не містить часового тренда у випадку, коли істинний процес є випадковим блуканням зі зміщенням

Оцінимо регресію (9.4.13) так само, як і у пункті 9.4.2, але припустивши, що істинний процес є випадковим блуканням із зміщенням

$$y_t = \alpha + y_{t-1} + u_t, \quad (9.4.36)$$

де істинне значення α не дорівнює 0. Хоча ця зміна може здатися не значною, проте вона матиме радикальний вплив на асимптотичний розподіл $\hat{\alpha}$ і $\hat{\rho}$. У цьому випадку рівність (9.4.4) має вигляд

$$y_t = y_0 + \alpha t + (u_1 + u_2 + \dots + u_t) = y_0 + \alpha t + \xi_t, \quad (9.4.37)$$

де

$$\xi_t \equiv u_1 + u_2 + \dots + u_t$$

для $t = 1, 2, \dots, T$ і $\xi_0 \equiv 0$. Розглянемо поведінку суми

$$\sum_{t=1}^T y_{t-1} = \sum_{t=1}^T [y_0 + \alpha(t-1) + \xi_{t-1}]. \quad (9.4.38)$$

Якщо перший член Ty_0 у рівності (9.4.38) поділити на T , то в результаті отримаємо сталу величину. Другий член $\sum \alpha(t-1)$ поділимо на T^2 , тоді згідно з твердженням 9.2.1(h) матимемо

$$\frac{1}{T^2} \sum_{t=1}^T \alpha(t-1) \rightarrow \frac{\alpha}{2}.$$

Третій член збіжний, якщо його поділити на $T^{3/2}$

$$\frac{1}{T^{3/2}} \sum_{t=1}^T \xi_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r) dr$$

згідно з твердженням 9.2.1(d). Порядок збіжності трьох окремих членів рівності (9.4.38) такий

$$\sum_{t=1}^T y_{t-1} = \underbrace{\sum_{t=1}^T y_0}_{O_p(T)} + \underbrace{\sum_{t=1}^T \alpha(t-1)}_{O_p(T^2)} + \underbrace{\sum_{t=1}^T \xi_{t-1}}_{O_p(T^{3/2})}.$$

Часовий тренд $\alpha(t-1)$ асимптотично домінує над іншими двома компонентами, тому поділимо (9.4.38) на T^2

$$\begin{aligned} \frac{1}{T^2} \sum_{t=1}^T y_{t-1} &= \frac{1}{T} y_0 + \frac{1}{T^2} \sum_{t=1}^T \alpha(t-1) + \frac{1}{T^{1/2}} \left(\frac{1}{T^{3/2}} \sum_{t=1}^T \xi_{t-1} \right) \xrightarrow{p} \\ &\xrightarrow{p} 0 + \alpha/2 + 0. \end{aligned} \quad (9.4.39)$$

Аналогічно, маємо, що

$$\begin{aligned} \sum_{t=1}^T y_{t-1}^2 &= \sum_{t=1}^T [y_0 + \alpha(t-1) + \xi_{t-1}]^2 = \\ &= \underbrace{\sum_{t=1}^T y_0^2}_{O_P(T)} + \underbrace{\sum_{t=1}^T \alpha^2(t-1)^2}_{O_P(T^3)} + \underbrace{\sum_{t=1}^T \xi_{t-1}^2}_{O_P(T^2)} + \\ &+ \underbrace{\sum_{t=1}^T 2y_0 \alpha(t-1)}_{O_P(T^2)} + \underbrace{\sum_{t=1}^T 2y_0 \xi_{t-1}}_{O_P(T^{3/2})} + \underbrace{\sum_{t=1}^T 2\alpha(t-1)\xi_{t-1}}_{O_P(T^{5/2})}. \end{aligned}$$

Звідси випливає таке: якщо поділити на T^3 , то всі доданки, крім доданка, який належить часовому тренду $\alpha^2(t-1)^2$, асимптотично наближаються до нуля. Тому

$$\frac{1}{T^3} \sum_{t=1}^T y_{t-1}^2 \xrightarrow{p} \frac{\alpha^2}{3}. \quad (9.4.40)$$

Розглянемо тепер

$$\begin{aligned} \sum_{t=1}^T y_{t-1} u_t &= \sum_{t=1}^T [y_0 + \alpha(t-1) + \xi_{t-1}] u_t \\ &= \underbrace{y_0 \sum_{t=1}^T u_t}_{O_P(T^{1/2})} + \underbrace{\sum_{t=1}^T \alpha(t-1) u_t}_{O_P(T^{3/2})} + \underbrace{\sum_{t=1}^T \xi_{t-1} u_t}_{O_P(T)}, \end{aligned}$$

звідки

$$\frac{1}{T^{3/2}} \sum_{t=1}^T y_{t-1} u_t \xrightarrow{p} \frac{1}{T^{3/2}} \sum_{t=1}^T \alpha(t-1) u_t. \quad (9.4.41)$$

Результати (9.4.39) - (9.4.41) означають таке: якщо істинний процес є випадковим блуканням зі зміщенням, то оцінені МНК коефіцієнти в (9.4.14) мають властивість

$$\begin{pmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{pmatrix} = \begin{pmatrix} O_p(T) & O_p(T^2) \\ O_p(T^2) & O_p(T^3) \end{pmatrix}^{-1} \begin{pmatrix} O_p(T^{1/2}) \\ O_p(T^{3/2}) \end{pmatrix}.$$

Отже, для цього випадку розрахункова матриця матиме вигляд

$$M_T \equiv \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{pmatrix},$$

для якої рівність (9.4.19) перетворимо у такий спосіб:

$$\begin{aligned} & \begin{pmatrix} T^{1/2} & 0 \\ 0 & T^{3/2} \end{pmatrix} \begin{pmatrix} \hat{\alpha}_T - \alpha \\ \hat{\rho}_T - 1 \end{pmatrix} = \\ & = \left[\begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix} \begin{pmatrix} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{pmatrix} \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix} \right]^{-1} \times \\ & \times \begin{pmatrix} T^{-1/2} & 0 \\ 0 & T^{-3/2} \end{pmatrix} \begin{pmatrix} \sum u_t \\ \sum y_{t-1} u_t \end{pmatrix}, \end{aligned}$$

тобто

$$\begin{aligned} & \begin{pmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\rho}_T - 1) \end{pmatrix} = \tag{9.4.42} \\ & = \begin{pmatrix} 1 & T^{-2} \sum y_{t-1} \\ T^{-2} \sum y_{t-1} & T^{-3} \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2} \sum u_t \\ T^{-3/2} \sum y_{t-1} u_t \end{pmatrix}. \end{aligned}$$

Згідно з (9.4.39) і (9.4.40) перший множник у рівності (9.4.42) збігається до

$$\begin{pmatrix} 1 & T^{-2} \sum y_{t-1} \\ T^{-2} \sum y_{t-1} & T^{-3} \sum y_{t-1}^2 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{pmatrix} \equiv Q. \tag{9.4.43}$$

З (9.4.41) випливає, що другий множник (9.4.42) задовольняє

$$\begin{aligned} \begin{pmatrix} T^{-1/2} \sum u_t \\ T^{-3/2} \sum y_{t-1} u_t \end{pmatrix} &\xrightarrow{p} \begin{pmatrix} T^{-1/2} \sum u_t \\ T^{-3/2} \sum \alpha(t-1) u_t \end{pmatrix} \xrightarrow{d} \\ &\xrightarrow{d} N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & \alpha/2 \\ \alpha/2 & \alpha^2/3 \end{pmatrix} \right] = N[0, \sigma^2 Q]. \end{aligned} \quad (9.4.44)$$

Поєднавши (9.4.42) - (9.4.44), отримаємо

$$\begin{aligned} \begin{pmatrix} T^{1/2}(\hat{\alpha}_T - \alpha) \\ T^{3/2}(\hat{\rho}_T - 1) \end{pmatrix} &\xrightarrow{d} N[0, Q^{-1} \sigma^2 Q Q^{-1}] = \\ &= N[0, \sigma^2 Q^{-1}]. \end{aligned} \quad (9.4.45)$$

Отже, для цього випадку обидва оцінені коефіцієнти є асимптотично нормальними, тому для тестування гіпотези використовують таблиці стандартного нормального розподілу.

9.4.4 Регресія, яка містить сталий член і часовий тренд. Істинний процес є випадковим блуканням зі зміщенням або без нього

Припустимо, як і у попередньому випадку, що істинною моделлю є

$$y_t = \alpha + y_{t-1} + u_t,$$

де u_t є білим шумом з дисперсією σ^2 . У цьому випадку істинне значення α не має значення для асимптотичного розподілу. На відміну від попереднього випадку, припустимо, що часовий тренд є включеним в регресію, яка оцінюється за МНК

$$y_t = \alpha + \rho y_{t-1} + \delta t + u_t. \quad (9.4.46)$$

Якщо $\alpha \neq 0$, то y_{t-1} буде асимптотично еквівалентним часовому тренду. Оскільки часовий тренд є вже включеним у регресію як окрема змінна, то це зумовить колінеарність пояснювальних змінних у великих вибірках. Модель регресії (9.4.46) можна записати в еквівалентному вигляді

$$\begin{aligned} y_t &= (1 - \rho)\alpha + \rho [y_{t-1} - \alpha(t - 1)] + (\delta + \rho\alpha)t + u_t \equiv \\ &\equiv \alpha^* + \rho^*\xi_{t-1} + \delta^*t + u_t, \end{aligned} \quad (9.4.47)$$

де $\alpha^* \equiv (1 - \rho)\alpha$, $\rho^* \equiv \rho$, $\delta^* \equiv (\delta + \rho\alpha)$ і $\xi_t \equiv y_t - \alpha t$. Крім того, врахувавши нульову гіпотезу $\rho = 1$ і $\delta = 0$, матимемо

$$\xi_t = y_0 + u_1 + u_2 + \dots + u_t,$$

тобто ξ_t – випадковий процес, описаний у твердженні 9.2.1. Розглянемо перетворену регресію y_t за сталою, ξ_{t-1} і часовим трендом, що дає таку МНК-оцінку:

$$\begin{pmatrix} \hat{\alpha}_T^* \\ \hat{\rho}_T^* \\ \hat{\delta}_T^* \end{pmatrix} = \begin{pmatrix} T & \sum \xi_{t-1} & \sum t \\ \sum \xi_{t-1} & \sum \xi_{t-1}^2 & \sum t \xi_{t-1} \\ \sum t & \sum t \xi_{t-1} & \sum t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum y_t \\ \sum \xi_{t-1} y_t \\ \sum t y_t \end{pmatrix}. \quad (9.4.48)$$

За обраної гіпотези $\alpha = \alpha_0$, $\rho = 1$ і $\delta = 0$ у перетвореній системі будемо мати $\alpha^* = 0$, $\rho^* = 1$ і $\delta^* = \alpha_0$. Відхилення МНК-оцінки від цих істинних значень можна зобразити

$$\begin{aligned} &\begin{pmatrix} \hat{\alpha}_T^* \\ \hat{\rho}_T^* - 1 \\ \hat{\delta}_T^* - \alpha_0 \end{pmatrix} = \\ &= \begin{pmatrix} T & \sum \xi_{t-1} & \sum t \\ \sum \xi_{t-1} & \sum \xi_{t-1}^2 & \sum t \xi_{t-1} \\ \sum t & \sum t \xi_{t-1} & \sum t^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum u_t \\ \sum \xi_{t-1} u_t \\ \sum t u_t \end{pmatrix}. \end{aligned} \quad (9.4.49)$$

Враховуючи порядок збіжностей з твердження 9.2.1, у цьому випадку розрахункову матрицю M_T треба вибрати так:

$$M_T = \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{pmatrix}.$$

Тоді рівність (9.4.49) набуде вигляду

$$\begin{aligned} & \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{pmatrix} \begin{pmatrix} \hat{\alpha}_T^* \\ \hat{\rho}_T^* - 1 \\ \hat{\delta}_T^* - \alpha_0 \end{pmatrix} = \\ & = \left[\begin{pmatrix} T^{-1/2} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & T^{-3/2} \end{pmatrix} \begin{pmatrix} T & \sum \xi_{t-1} & \sum t \\ \sum \xi_{t-1} & \sum \xi_{t-1}^2 & \sum t \xi_{t-1} \\ \sum t & \sum t \xi_{t-1} & \sum t^2 \end{pmatrix} \right] \times \\ & \times \begin{pmatrix} T^{-1/2} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & T^{-3/2} \end{pmatrix}^{-1} \times \\ & \times \begin{pmatrix} T^{-1/2} & 0 & 0 \\ 0 & T^{-1} & 0 \\ 0 & 0 & T^{-3/2} \end{pmatrix} \begin{pmatrix} \sum u_t \\ \sum \xi_{t-1} u_t \\ \sum t u_t \end{pmatrix} \end{aligned}$$

або

$$\begin{aligned} & \begin{pmatrix} T^{1/2} \hat{\alpha}_T^* \\ T (\hat{\rho}_T^* - 1) \\ T^{3/2} (\hat{\delta}_T^* - \alpha_0) \end{pmatrix} = \\ & = \begin{pmatrix} 1 & T^{-3/2} \sum \xi_{t-1} & T^{-2} \sum t \\ T^{-3/2} \sum \xi_{t-1} & T^{-2} \sum \xi_{t-1}^2 & T^{-5/2} \sum t \xi_{t-1} \\ T^{-2} \sum t & T^{-5/2} \sum t \xi_{t-1} & T^{-3} \sum t^2 \end{pmatrix}^{-1} \times \\ & \times \begin{pmatrix} T^{-1/2} \sum u_t \\ T^{-1} \sum \xi_{t-1} u_t \\ T^{-3/2} \sum t u_t \end{pmatrix}. \end{aligned}$$

З твердження 9.2.1 можемо знайти асимптотичний розподіл

$$\begin{aligned}
 & \begin{pmatrix} T^{1/2} \hat{\alpha}_T^* \\ T(\hat{\rho}_T^* - 1) \\ T^{3/2}(\hat{\delta}_T^* - \alpha_0) \end{pmatrix} \xrightarrow{d} \quad (9.4.50) \\
 & \xrightarrow{d} \begin{pmatrix} 1 & \sigma \int W(r) dr & 1/2 \\ \sigma \int W(r) dr & \sigma^2 \int [W(r)]^2 dr & \sigma \int r W(r) dr \\ 1/2 & \sigma \int r W(r) dr & 1/3 \end{pmatrix}^{-1} \times \\
 & \times \begin{pmatrix} \sigma W(1) \\ (1/2) \sigma^2 ([W(1)]^2 - 1) \\ \sigma (W(1) - \int W(r) dr) \end{pmatrix} = \\
 & = \sigma \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \times \\
 & \times \begin{pmatrix} 1 & \int W(r) dr & 1/2 \\ \sigma \int W(r) dr & \int [W(r)]^2 dr & \int r W(r) dr \\ 1/2 & \int r W(r) dr & 1/3 \end{pmatrix}^{-1} \times \\
 & \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} W(1) \\ (1/2) ([W(1)]^2 - 1) \\ W(1) - \int W(r) dr \end{pmatrix} = \\
 & = \begin{pmatrix} \sigma & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sigma \end{pmatrix} \begin{pmatrix} 1 & \int W(r) dr & 1/2 \\ \int W(r) dr & \int [W(r)]^2 dr & \int r W(r) dr \\ 1/2 & \int r W(r) dr & 1/3 \end{pmatrix}^{-1} \\
 & \times \begin{pmatrix} W(1) \\ (1/2) ([W(1)]^2 - 1) \\ W(1) - \int W(r) dr \end{pmatrix}
 \end{aligned}$$

Зауважимо, що МНК-оцінка $\hat{\rho}_T^*$, що ґрунтується на (9.4.47) співпадає з МНК-оцінкою $\hat{\rho}_T$, яка ґрунтується на (9.4.46). Отож, асимптотичний розподіл $T(\hat{\rho}_T - 1)$ задається за допомогою

(9.4.50). Цей розподіл не залежить ні від σ , ні від α , а також не має значення чи α дорівнює нулю, чи ні.

Асимптотичний розподіл $\hat{\sigma}_{\hat{\rho}_T}$ стандартної МНК-похибки для $\hat{\rho}_T$ можна знайти, використовуючи подібні обчислення як і для (9.4.30) та (9.4.31). Отже,

$$\begin{aligned}
 T^2 \hat{\sigma}_{\hat{\rho}_T}^2 &= \tag{9.4.51} \\
 &= T^2 s_T^2 (0 \ 1 \ 0) \begin{pmatrix} T & \sum \xi_{t-1} & \sum t \\ \sum \xi_{t-1} & \sum \xi_{t-1}^2 & \sum t \xi_{t-1} \\ \sum t & \sum t \xi_{t-1} & \sum t^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \\
 &= s_T^2 (0 \ 1 \ 0) \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{pmatrix} \times \\
 &\times \begin{pmatrix} T & \sum \xi_{t-1} & \sum t \\ \sum \xi_{t-1} & \sum \xi_{t-1}^2 & \sum t \xi_{t-1} \\ \sum t & \sum t \xi_{t-1} & \sum t^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{1/2} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{3/2} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \\
 &= s_T^2 (0 \ 1 \ 0) \begin{pmatrix} 1 & T^{-3/2} \sum \xi_{t-1} & T^{-2} \sum t \\ T^{-3/2} \sum \xi_{t-1} & T^{-2} \sum \xi_{t-1}^2 & T^{-5/2} \sum t \xi_{t-1} \\ T^{-2} \sum t & T^{-5/2} \sum t \xi_{t-1} & T^{-3} \sum t^2 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
 &\xrightarrow{d} \sigma^2 (0 \ 1 \ 0) \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \times \\
 &\times \begin{pmatrix} 1 & \int W(r) dr & 1/2 \\ \int W(r) dr & \int [W(r)]^2 dr & \int r W(r) dr \\ 1/2 & \int r W(r) dr & 1/3 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sigma & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \\
 &= (0 \ 1 \ 0) \begin{pmatrix} 1 & \int W(r) dr & 1/2 \\ \int W(r) dr & \int [W(r)]^2 dr & \int r W(r) dr \\ 1/2 & \int r W(r) dr & 1/3 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \equiv \\
 &\equiv Q.
 \end{aligned}$$

З цього випливає, що асимптотичний розподіл МНК t -статистики для тестування гіпотези $\rho = 1$ є таким:

$$t_T = \frac{T(\hat{\rho}_T - 1)}{(T^2 \hat{\sigma}_{\hat{\rho}_T}^2)^{1/2}} \xrightarrow{p} \frac{T(\hat{\rho}_T - 1)}{\sqrt{Q}}. \quad (9.4.52)$$

Знову цей розподіл не залежить ні від α , ні від σ .

Припущення про те, що істинне значення δ дорівнює нулю, є допоміжною гіпотезою, від якої залежать асимптотичні властивості тесту. Отож, як і у пункті 9.4.2, треба розглянути МНК F -тест об'єднання гіпотез $\delta = 0$ і $\rho = 1$. Хоча F -статистику обчислюють звичайним методом, її асимптотичний розподіл є нестандартним.

9.5 Тест Філіпса - Перрона

Щоб відобразити основну ідею тесту Філіпса - Перрона (Phillips, Perron) на одиничний корінь, розглянемо детальніше регресію, яка досліджувалась у пункті 9.4.2.

Розглянемо МНК-оцінки параметрів α і ρ регресійної моделі

$$y_t = \alpha + \rho y_{t-1} + u_t \quad (9.5.1)$$

згідно з припущенням, що істинні $\alpha = 0$, $\rho = 1$ і u_t є білим шумом.

При виведенні асимптотичного розподілу у пункті 9.4.2 припускали, що збурення не корелюють. Філіпс і Перрон узагальнили ці результати для випадку, коли u_t послідовно корелюють і можливо, гетероскедастичні. Нехай істинний процес має вигляд

$$y_t - y_{t-1} = u_t = \psi(L)\varepsilon_t,$$

де $\psi(L)$ та ε_t задовольняють умови твердження 9.2.2.

Якщо рівність (9.5.1) є стаціонарною авторегресією з $|\rho| < 1$, то у випадку, якщо u_t послідовно корелюють МНК-оцінка $\hat{\rho}_T$, яка визначена в (9.4.14), не є змістовною оцінкою параметра ρ . Однак, якщо $\rho = 1$, то матимемо збіжність $\hat{\rho}_T \rightarrow 1$ навіть і при послідовно корельованих u_t . Тому Філіпс і Перрон запропонували використовувати МНК-оцінку (9.5.1) також і у випадку послідовно корельованих u_t .

Нехай МНК-оцінки $\hat{\alpha}_T$ і $\hat{\rho}_T$ параметрів моделі (9.5.1) за відсутності послідовної кореляції, тобто $\hat{\alpha}_T$ і $\hat{\rho}_T$ – величини, визначені в (9.4.14). Якщо істинні значення $\alpha = 0$ і $\rho = 1$, то згідно з (9.4.21) маємо

$$\begin{aligned} \begin{pmatrix} T^{1/2}\hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{pmatrix} &= \\ &= \begin{pmatrix} 1 & T^{-3/2} \sum y_{t-1} \\ T^{-3/2} \sum y_{t-1} & T^{-2} \sum y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} T^{-1/2} \sum u_t \\ T^{-1} \sum y_{t-1} u_t \end{pmatrix}, \end{aligned} \quad (9.5.2)$$

де Σ означає $\sum_{t=1}^T$. Також, згідно з нульовою гіпотезою $H_0 : \alpha = 0, \rho = 1$ аналогічно як і в (9.4.4) одержимо

$$y_t = y_0 + u_1 + u_2 + \dots + u_t.$$

Якщо $u_t = \psi(L)\varepsilon_t$, то y_t є змінною, яка має ті самі властивості, що і ξ_t у твердженні 9.2.2. Використовуючи результати (е) і (ж) цього твердження, маємо

$$\begin{aligned} &\begin{pmatrix} 1 & T^{-3/2} \sum y_{t-1} \\ T^{-3/2} \sum y_{t-1} & T^{-2} \sum y_{t-1}^2 \end{pmatrix}^{-1} \xrightarrow{d} \\ &\xrightarrow{d} \begin{pmatrix} 1 & \lambda \int W(r) dr \\ \lambda \int W(r) dr & \lambda^2 \int [W(r)]^2 dr \end{pmatrix}^{-1} = \\ &= \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}^{-1} \begin{pmatrix} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}^{-1}, \end{aligned} \quad (9.5.3)$$

де інтегральний знак означає інтеграл за r від 0 до 1. Аналогічно результати (а) і (д) твердження 9.2.2 дають

$$\begin{aligned}
 & \left(\begin{array}{c} T^{-1/2} \sum u_t \\ T^{-1} \sum y_{t-1} u_t \end{array} \right) \xrightarrow{d} \quad (9.5.4) \\
 & \xrightarrow{d} \left(\begin{array}{c} \lambda W(1) \\ (1/2) (\lambda^2 [W(1)]^2 - \gamma_0) \end{array} \right) \left(\begin{array}{c} \lambda W(1) \\ (1/2) (\lambda^2 [W(1)]^2 - 1) \end{array} \right) + \\
 & + \left(\begin{array}{c} 0 \\ (1/2) (\lambda^2 - \gamma_0) \end{array} \right) = \\
 & = \lambda \left(\begin{array}{cc} 1 & 0 \\ 0 & \lambda \end{array} \right) \left(\begin{array}{c} W(1) \\ (1/2) ([W(1)]^2 - 1) \end{array} \right) + \left(\begin{array}{c} 0 \\ (1/2) (\lambda^2 - \gamma_0) \end{array} \right).
 \end{aligned}$$

Підставивши (9.5.3) та (9.5.4) у (9.5.2), матимемо

$$\begin{aligned}
 & \left(\begin{array}{c} T^{1/2} \hat{\alpha}_T \\ T(\hat{\rho}_T - 1) \end{array} \right) \xrightarrow{d} \quad (9.5.5) \\
 & \xrightarrow{d} \left(\begin{array}{cc} 1 & 0 \\ 0 & \lambda \end{array} \right)^{-1} \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \left(\begin{array}{cc} 1 & 0 \\ 0 & \lambda \end{array} \right)^{-1} \times \\
 & \times \left[\lambda \left(\begin{array}{cc} 1 & 0 \\ 0 & \lambda \end{array} \right) \left(\begin{array}{c} W(1) \\ \frac{1}{2} ([W(1)]^2 - 1) \end{array} \right) + \left(\begin{array}{c} 0 \\ \frac{1}{2} (\lambda^2 - \gamma_0) \end{array} \right) \right] = \\
 & = \left(\begin{array}{cc} 1 & 0 \\ 0 & \lambda \end{array} \right) \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \left(\begin{array}{c} W(1) \\ \frac{1}{2} ([W(1)]^2 - 1) \end{array} \right) + \\
 & + \left(\begin{array}{cc} 1 & 0 \\ 0 & \lambda^{-1} \end{array} \right) \left(\begin{array}{cc} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{array} \right)^{-1} \left(\begin{array}{c} 0 \\ \frac{1}{2} \frac{\lambda^2 - \gamma_0}{\lambda} \end{array} \right).
 \end{aligned}$$

Тому другий елемент цього вектора має такий асимптотичний

розподіл:

$$\begin{aligned}
& T(\hat{\rho}_T - 1) \xrightarrow{d} \\
& \xrightarrow{d} (0 \ 1) \begin{pmatrix} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{pmatrix}^{-1} \begin{pmatrix} W(1) \\ \frac{1}{2} ([W(1)]^2 - 1) \end{pmatrix} + \\
& + \frac{\lambda^2 - \gamma_0}{2\lambda^2} (0 \ 1) \begin{pmatrix} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \\
& = \frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\int [W(r)]^2 dr - [\int W(r) dr]^2} + \\
& + \frac{(1/2) (\lambda^2 - \gamma_0)}{\lambda^2 \left(\int [W(r)]^2 dr - [\int W(r) dr]^2 \right)}. \tag{9.5.6}
\end{aligned}$$

Перший член останньої рівності такий самий, як і в (9.4.27). Він описує асимптотичний розподіл $T(\hat{\rho}_T - 1)$, якщо $u_t \in \text{i.i.d.}$ Останній член цієї рівності скорегований згідно з послідовною кореляцією. Якщо u_t не є послідовно корельованими, тоді $\psi_0 = 1$ і $\psi_j = 0$ для $j = 1, 2, \dots$. Отож, якщо u_t не є послідовно корельованими, тоді $\lambda^2 = \sigma^2 [\psi(1)]^2 = \sigma^2$ і $\gamma_0 = E[u_t^2] = \sigma^2$.

Зручно використати МНК-похибку $\hat{\sigma}_{\hat{\rho}_T}$ оцінки $\hat{\rho}_T$ у випадку послідовної кореляції. Нехай M_T – матриця, визначена в (9.4.20), а s_T^2 – МНК-оцінка дисперсії u_t . Тоді

$$s_T^2 = \frac{1}{T-2} \sum_{t=1}^T (y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2.$$

Асимптотичний розподіл $T^2 \cdot \hat{\sigma}_{\hat{\rho}_T}$ можна знайти, використовуву-

ючи аналогічний підхід як і в (9.4.30) - (9.4.32)

$$\begin{aligned}
 T^2 \hat{\sigma}_{\hat{\rho}_T}^2 &= s_T^2 (0 \ 1) M_T \left(\begin{array}{cc} T & \sum y_{t-1} \\ \sum y_{t-1} & \sum y_{t-1}^2 \end{array} \right)^{-1} M_T \begin{pmatrix} 0 \\ 1 \end{pmatrix} \xrightarrow{p} \\
 &\xrightarrow{p} s_T^2 (0 \ 1) \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}^{-1} \begin{pmatrix} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{pmatrix}^{-1} \times \\
 &\times \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \\
 &= \frac{s_T^2}{\lambda^2} (0 \ 1) \begin{pmatrix} 1 & \int W(r) dr \\ \int W(r) dr & \int [W(r)]^2 dr \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \\
 &= \frac{s_T^2}{\lambda^2} \frac{1}{\int [W(r)]^2 dr - [\int W(r) dr]^2}. \tag{9.5.7}
 \end{aligned}$$

З рівності (9.5.6) випливає, що

$$\begin{aligned}
 T(\hat{\rho}_T - 1) - \frac{1}{2} \frac{T^2 \hat{\sigma}_{\hat{\rho}_T}^2}{s_T^2} (\lambda^2 - \gamma_0) &\xrightarrow{p} \tag{9.5.8} \\
 \xrightarrow{p} T(\hat{\rho}_T - 1) - \frac{1}{2} \left(\frac{1}{\lambda^2} \right) \frac{\lambda^2 - \gamma_0}{\int [W(r)]^2 dr - [\int W(r) dr]^2} &\xrightarrow{d} \\
 \xrightarrow{d} \frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\int [W(r)]^2 dr - [\int W(r) dr]^2}.
 \end{aligned}$$

Отже, (9.5.8) має такий самий асимптотичний розподіл як (9.4.27). Результат (9.5.8) можна використати для пошуку асимптотичного розподілу МНК t - статистики

для $H_0 : \rho = 1$

$$\begin{aligned}
 t_T &= \frac{(\hat{\rho}_T - 1)}{\hat{\sigma}_{\hat{\rho}_T}} = \frac{T(\hat{\rho}_T - 1)}{(T^2 \hat{\sigma}_{\hat{\rho}_T}^2)^{1/2}} \xrightarrow{d} \\
 &\xrightarrow{d} \frac{1}{(T^2 \hat{\sigma}_{\hat{\rho}_T}^2)^{1/2}} \left(\frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\int [W(r)]^2 dr - [\int W(r) dr]^2} + \right. \\
 &\quad \left. + \frac{1}{2} \frac{T^2 \hat{\sigma}_{\hat{\rho}_T}^2}{s_T^2} (\lambda^2 - \gamma_0) \right) = \\
 &= \frac{1}{(T^2 \hat{\sigma}_{\hat{\rho}_T}^2)^{1/2}} \frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\int [W(r)]^2 dr - [\int W(r) dr]^2} + \\
 &\quad + \frac{(\lambda^2 - \gamma_0)}{2s_T} \left(\frac{T^2 \hat{\sigma}_{\hat{\rho}_T}^2}{s_T^2} \right)^{1/2} \xrightarrow{p} \\
 &\xrightarrow{p} \left(\frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\int [W(r)]^2 dr - [\int W(r) dr]^2} \right) \left(\frac{\lambda^2}{s_T^2} \right)^{1/2} \times \\
 &\quad \times \left(\int [W(r)]^2 dr - [\int W(r) dr]^2 \right)^{1/2} + \frac{(\lambda^2 - \gamma_0)}{2s_T} \left(\frac{T^2 \hat{\sigma}_{\hat{\rho}_T}^2}{s_T^2} \right)^{1/2}. \tag{9.5.9}
 \end{aligned}$$

Крім того,

$$s_T^2 = \frac{1}{T-2} \sum_{t=1}^T (y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1})^2 \xrightarrow{p} E[u_t^2] = \gamma_0. \tag{9.5.10}$$

Отже, (9.5.9) означає, що

$$\begin{aligned}
 \left(\frac{\gamma_0}{\lambda^2} \right)^{1/2} t_T &\xrightarrow{p} \frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\left(\int [W(r)]^2 dr - [\int W(r) dr]^2 \right)^{1/2}} + \\
 &\quad + \frac{1}{2} \frac{\lambda^2 - \gamma_0}{\lambda} \frac{T \hat{\sigma}_{\hat{\rho}_T}}{s_T}. \tag{9.5.11}
 \end{aligned}$$

Отже,

$$\begin{aligned} & \left(\frac{\gamma_0}{\lambda^2}\right)^{1/2} t_T - \frac{1}{2} \frac{\lambda^2 - \gamma_0}{\lambda} \frac{T \hat{\sigma}_{\hat{\rho}_T}}{s_T} \xrightarrow{d} \\ & \xrightarrow{d} \frac{(1/2) ([W(1)]^2 - 1) - W(1) \int W(r) dr}{\left(\int [W(r)]^2 dr - \left[\int W(r) dr\right]^2\right)^{1/2}}, \end{aligned} \quad (9.5.12)$$

що є тим самим граничним розподілом, який одержано в (9.4.35).

Формули (9.5.8) та (9.5.12) потребують знань параметрів γ_0 і λ^2 . Хоча ці моменти є невідомими, їх легко послідовно оцінити. Оскільки $\gamma_0 = E[u_t^2]$, то маємо першу оцінку

$$\hat{\gamma}_0 = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2, \quad (9.5.13)$$

де $\hat{u}_t = y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1}$ є оцінкою МНК-залишка. Замість цього Філіпс і Перрон (1988) використали стандартну МНК-оцінку

$$\hat{\gamma}_0 = \frac{1}{T-2} \sum_{t=1}^T \hat{u}_t^2 = s_T^2.$$

Аналогічно на підставі результату (а) твердження 9.2.2 λ є асимптотичним середньоквадратичним відхиленням середнього значення \bar{u} . Тому маємо

$$\sqrt{T} \bar{u} = \frac{1}{T^{1/2}} \sum_{t=1}^T u_t \xrightarrow{d} N[0, \lambda^2]. \quad (9.5.14)$$

Цю величину можна записати у вигляді

$$\lambda^2 = \sigma^2 [\psi(1)]^2 = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j = 2\pi s_u(0), \quad (9.5.15)$$

де γ_j – j -та автоковаріація u_t , а $s_u(0)$ – спектр u_t з нульовою частотою. Отже, цю величину можна оцінити декількома способами. Наприклад, якщо лише перші q автоковаріації вважають суттєвими, то можна використати оцінку Нев'є-Веста (Newey-West)

$$\hat{\lambda}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^q [1 - j/(q+1)] \hat{\gamma}_j, \quad (9.5.16)$$

де

$$\hat{\gamma}_j = \frac{1}{T} \sum_{t=j+1}^T \hat{u}_t \hat{u}_{t-j} \quad (9.5.17)$$

і $\hat{u}_t = y_t - \hat{\alpha}_T - \hat{\rho}_T y_{t-1}$.

Згідно з нульовою гіпотезою, за якої перше спостереження y_t є нулем, тобто математичним сподіванням коваріаційно-стаціонарного процесу, щоб підсумувати підхід Філіпса і Перрона треба оцінити рівність (9.5.1) за допомогою МНК і використати стандартні формули МНК для обчислення $\hat{\rho}$, його стандартної похибки $\hat{\sigma}_{\hat{\rho}}$, а також стандартної похибки регресії s . j -та автоковаріація $\hat{u}_t = y_t - \hat{\alpha} - \hat{\rho} y_{t-1}$ тоді обчислюються згідно з (9.5.17).

Список літератури

1. *Hamilton, James D.* Times series analysis. — Published by Princeton University Press, Princeton, New Jersey, 1994.
2. *Verbeek, Marno.* A guide to modern econometrics. — John Wiley and Sons, Ltd, 2000 — UD, England.
3. *Mills, Terence C.* The econometric modelling of financial time series. — Cambridge University Press, 1999.
4. *Леоненко М.М., Мішура Ю.С., Пархоменко В.М., Ядренко М.Й.* Теоретико-ймовірнісні та статистичні методи в економетриці та фінансовій математиці. — К.: Інформтехніка, 1995.
5. *Тихомиров Н.П., Дорохина Е.Ю.* Економетрика. — М.: Экзамен, 2003.
6. *Холден К., Піл Д.А., Томсон Дж.Л.* Економічне прогнозування: Вступ. — К.: Інформтехніка, 1996.
7. *Бокс Дж., Дженкінс Г.* Анализ временных рядов. — М.: Мир, 1974.