

Лабораторна робота № 5

Моделювання та оцінювання гетероскедастичних процесів

Мета роботи: *Навчитися аналізувати та моделювати гетероскедастичні процеси (процеси із змінною дисперсією).*

1 Порядок виконання роботи

1.1 Прочитайте теоретичні відомості (розділ 2), що стосуються аналізу та моделювання гетероскедастичних процесів.

1.2 Візьміть у викладача файл з даними *ARCH_01.DAT*.

1.3 Скористайтесь для аналізу даних пакетом прикладних програм (ППП) *EViews*. Введіть (прочитайте) дані з вказаного вище файла з диска за допомогою ППП, знайдіть середнє значення ряду, стандартне відхилення і побудувати для введених даних модель *AR(1)*.

1.4 Побудуйте автокореляційну функцію (АКФ) для отриманого після побудови моделі *AR(1)* ряду із залишків $\hat{\varepsilon}(k)$. Залишки мають ідентифікатор *resid*, а для побудови АКФ можна скористатися командами: *Quick* —> *Series Statistics* —> *Correlogram* —> *OK* —> *Level* —> *OK*.

1.5 Згенеруйте новий ряд із квадратів залишків: $\{\hat{\varepsilon}^2(k)\} = \{e^2(k)\}$, де $e(k)$ — залишки (похибки) моделі *AR(1)*. Для цього можна скористатися командами: *Proc* —> *Generate Series* —> *resid_2=resid*resid*.

1.6 Обчисліть автокореляційну функцію для ряду із квадратів залишків.

1.7 За допомогою МНК обчисліть оцінки коефіцієнтів рівняння першого порядку для дисперсії залишків:

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + v(k),$$

де $v(k)$ — похибки моделі. Таку модель називають авторегресійною умовно гетероскедастичною або скорочено АРУГ (в даному випадку АРУГ(1)).

1.8 Обчисліть коефіцієнти рівняння АРУГ(4):

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \alpha_2 \hat{\varepsilon}^2(k-2) + \alpha_3 \hat{\varepsilon}^2(k-3) + \alpha_4 \hat{\varepsilon}^2(k-4).$$

1.9 Побудуйте ряд умовних дисперсій для *ARCH_01.DAT* і побудуйте узагальнену авторегресійну умовно гетероскедастичну (УАРУГ) модель процесу у вигляді:

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \hat{\varepsilon}^2(k-i) + \sum_{i=1}^p \beta_i h(k-i) + v(k),$$

де q – визначається за допомогою АКФ процесу $\{\varepsilon^2(k)\}$; p – визначається за допомогою АКФ для $\{h(k)\}$; $h(k)$ – умовна дисперсія процесу, яка розраховується послідовно для досліджуваного ряду за виразом:

$$h(k) = E_k[y^2(k)] = \frac{1}{k-1} \sum_{i=1}^k [y(i) - \bar{y}_k]^2, \quad k = 2, 3, \dots, N,$$

де N – довжина ряду; $\bar{y}_k = \frac{1}{k-1} \sum_{i=1}^k y(i)$, $k = 2, \dots, N$ – умовне математичне сподівання.

1.10 Побудуйте модель гетероскедастичного процесу

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i),$$

де

$$h(k) = E_k[\varepsilon^2(k)] = \frac{1}{k-1} \sum_{i=1}^k [\varepsilon(i) - \bar{\varepsilon}_k]^2, \quad k = 2, 3, \dots, N,$$

$$\bar{\varepsilon}_k = \frac{1}{k-1} \sum_{i=1}^k \varepsilon(i), \quad k = 2, \dots, N.$$

Наведіть графіки процесів: $y(k)$, $h(k)$, $\varepsilon(k)$, $\varepsilon^2(k)$ та автокореляційні функції, використані при побудові моделей. Порівняйте моделі, побудовані на кроках 1.7 – 1.10.

1.11 Дайте відповіді на контрольні запитання, які даються в кінці опису лабораторної роботи.

2 Загальні відомості про гетероскедастичні процеси

Гетероскедастичними називають процеси зі змінною дисперсією, а

гомоскедастичними – процеси із сталою дисперсією.

Формально для гетероскедастичного процесу можна записати

$$\text{var}[\varepsilon(k)] = \sigma_{\varepsilon}^2 \neq \text{const.} \quad (2.1)$$

Припущення *гомоскедастичності* означає, що варіація кожної випадкової величини $\varepsilon(k)$ навколо її математичного сподівання залишається сталою величиною незалежно від значень факторів. Тобто, σ_{ε}^2 не є функцією x_{ij} .

Гетероскедастичність означає, що дисперсія процесу зменшується чи збільшується в часі, або є більш складною функцією часу. Тобто, вона може змінюватися за досить складним законом, який і потрібно знайти при створенні моделі процесу. Іноді використовують припущення, що гетероскедастичність має наступну форму:

$$\sigma_{\varepsilon(k)}^2 = k^2 x^2,$$

де k - константа, яку необхідно оцінити за допомогою експериментальних даних та вибраного методу оцінювання параметрів.

Розглянемо простий випадок, коли

$$y(k+1) = \varepsilon(k+1)x(k),$$

де $y(k+1)$ – основна змінна; $\varepsilon(k+1)$ – збурення у вигляді (наприклад) білого шуму з дисперсією σ^2 ; $x(k)$ – незалежна змінна, виміри якої відомі на відрізку часу, що розглядається. У випадку коли $x(k) = x(k-1) = x(k-2) = \dots = \text{const}$, послідовність $\{y(k)\}$ представляє собою білий шум із постійною дисперсією. Однак, у випадку, коли послідовність $\{x(k)\}$ не є послідовністю однакових чисел, умовна дисперсія $y(k+1)$ при відомих значеннях $x(k)$ визначається як

$$\text{var}[y(k+1) | x(k)] = x^2(k)\sigma^2.$$

Таким чином, умовна дисперсія величини $y(k+1)$ залежить від реалізації $x(k)$. Оскільки $x(k)$ можна виміряти в момент k , то дисперсію величини $y(k)$ можна визначити як умовну при відомому значенні $x(k)$. Якщо величина $(x(k))^2$ велика (мала), то дисперсія $y(k+1)$ також буде великою (малою). З цього витікає, що при наявності позитивної кореляції між значеннями послідовності $\{x(k)\}$ (тобто є тенденція, що за великим значенням $x(k)$ знову іде велике значення $x(k+1)$), то умовна дисперсія

послідовності $\{y(k)\}$ також матиме тенденцію до послідовної позитивної кореляції. На практиці розглядають модифіковану модель типу

$$\ln(y(k)) = a_0 + a_1 \ln(x(k-1)) + \ln e(k), \quad (2.2)$$

де $e(k)$ – похибка моделі. Недоліком такої моделі є те, що потрібно дійсно знайти таку змінну $x(k)$, яка впливає на дисперсію процесу $y(k)$, що не завжди можливо з практичної точки зору. Окрім того, первинний часовий ряд необхідно обробляти таким чином, щоб результуючий ряд (після обробки) мав постійну дисперсію. В наведеному прикладі (2.2) покладається, що послідовність $e(k)$ має постійну дисперсію. Якщо ця умова не виконується, то необхідно зробити додаткову обробку даних.

Замість того, щоб шукати змінну $x(k)$, що впливає на дисперсію $y(k)$, і робити перетворення даних, можна одночасно моделювати (описувати) середнє значення та дисперсію ряду. Перш ніж переходити до описання цієї моделі, зазначимо, що використання умовного прогнозування має значні переваги перед використанням безумовного.

Нехай оцінюється стаціонарна модель авторегресії із ковзним середнім (АРКС) типу $y(k) = a_0 + a_1 y(k-1) + \varepsilon(k)$ і необхідно прогнозувати величину $y(k+1)$. Умовним математичним сподіванням є

$$E_k[y(k+1)] = a_0 + a_1 y(k).$$

Якщо використати це умовне середнє значення для прогнозування значення $y(k+1)$, то похибка прогнозу буде визначатися як

$$E_k\{[y(k+1) - a_0 - a_1 y(k)]^2\} = E_k[\varepsilon^2(k+1)] = \sigma^2.$$

Якщо замість умовного прогнозу використовується безумовний, то він відображає середнє значення для послідовності $\{y(k)\}$ на нескінченному часовому інтервалі, яке дорівнює в даному випадку $a_0/(1-a_1)$. Дисперсія похибки безумовного прогнозу визначається за виразом

$$E\{[y(k+1) - a_0/(1-a_1)]^2\} = E\{[\varepsilon(k+1) + a_1\varepsilon(k) + a_1^2\varepsilon(k-1) + a_1^3\varepsilon(k-2) + \dots]^2\} = \sigma^2/(1-a_1^2).$$

Послідовність $\{\varepsilon(k)\}$ в правій частині останнього виразу з'явилась у зв'язку із використанням розв'язку рівняння першого порядку для знаходження

безумовного середнього. Оскільки $1/(1 - a_1^2) > 1$, то безумовний прогноз має більшу дисперсію ніж умовний. В зв'язку з цим, а також завдяки тому факту, що умовний прогноз базується на відомих поточних та минулих значеннях ряду, віддають перевагу умовному прогнозуванню.

Якщо дисперсія послідовності $\{\varepsilon(k)\}$ непостійна, то тенденцію зміни цього параметру можна описати за допомогою моделі АРКС. Наприклад, позначимо через $\{\hat{\varepsilon}(k)\}$ оцінки залишків (похибок) моделі першого порядку $y(k) = a_0 + a_1 y(k-1) + \varepsilon(k)$. Для такого випадку умовна дисперсія основної змінної визначається як

$$\text{var}[y(k+1) | y(k)] = E_k \{ [y(k+1) - a_0 - a_1 y(k)]^2 \} = E_k [\varepsilon^2(k+1)].$$

До цих пір вважалось, що $E_k[\varepsilon^2(k+1)] = \sigma^2$ є постійною величиною. Покладемо тепер, що умовна дисперсія - величина змінна. Одним із простих підходів до описання такої змінної величини є застосування моделі типу АР(q) до квадратів оцінок залишків. Наприклад,

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \alpha_2 \hat{\varepsilon}^2(k-2) + \dots + \alpha_q \hat{\varepsilon}^2(k-q) + v(k), \quad (2.3)$$

де $v(k)$ – процес білого шуму.

Якщо всі коефіцієнти $\alpha_1, \alpha_2, \dots, \alpha_q$ дорівнюють нулю, то оцінка дисперсії буде просто константою. Інакше, умовна дисперсія для $y(k)$ описується рівнянням (2.3). Це рівняння можна використовувати для прогнозування умовної дисперсії на один крок наступним чином:

$$E_k[\hat{\varepsilon}^2(k+1)] = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \alpha_2 \hat{\varepsilon}^2(k-2) + \dots + \alpha_q \hat{\varepsilon}^2(k+1-q).$$

З цієї причини (2.3) називають **авторегресійним умовно гетероскедастичним (АРУГ) рівнянням**. Оскільки залишки $\varepsilon(k)$, які використовуються у рівнянні (2.3), можуть бути отримані на основі рівнянь регресії, авторегресії або авторегресії із ковзним середнім, то для АРУГ моделі можна знайти багато потенційних можливостей практичного і теоретичного застосування.

Окрім рівняння типу (2.3) можна вибрати і складніші форми описання поведінки дисперсії. Наприклад, збурення можна ввести у мультиплікативній формі:

$$\varepsilon^2(k) = v^2(k)[\alpha_0 + \alpha_1 \varepsilon^2(k-1)], \quad (2.4)$$

де $v(k)$ – мультиплікативне збурення у формі білого шуму, причому $\{v(k)\} \sim (0,1)$, тобто воно має нульове середнє і одиничну дисперсію; змінні $\varepsilon(k-1)$ і $v(k)$ – статистично незалежні (некорельовані) величини. Можна показати, що елементи послідовності $\{\varepsilon(k)\}$ некорельовані між собою та мають нульове середнє. Знайдемо умовне математичне сподівання для $\varepsilon(k)$. Оскільки $E[v(k)] = 0$, то

$$E[\varepsilon(k)] = E[v(k)(\alpha_0 + \alpha_1 \varepsilon^2(k-1))^{1/2}] = E[v(k)] E[\alpha_0 + \alpha_1 \varepsilon^2(k-1)]^{1/2} = 0. \quad (2.5)$$

Оскільки $E[v(k)v(k-1)] = 0$, то

$$E[\varepsilon(k)\varepsilon(k-i)] = 0, \quad i \neq 0. \quad (2.6)$$

Можна знайти також безумовну дисперсію для $\varepsilon(k)$ наступним чином:

$$E[\varepsilon^2(k)] = E[v^2(k)(\alpha_0 + \alpha_1 \varepsilon^2(k-1))] = E[v^2(k)] E[\alpha_0 + \alpha_1 \varepsilon^2(k-1)].$$

Враховуючи те, що $\sigma_v^2 = 1$, і безумовна дисперсія змінної $\varepsilon(k)$ така ж як і для $\varepsilon(k-1)$, тобто $E[\varepsilon^2(k)] = E[\varepsilon^2(k-1)]$, то безумовна дисперсія дорівнює

$$E[\varepsilon^2(k)] = \frac{\alpha_0}{1 - \alpha_1}. \quad (2.7)$$

Таким чином, на безумовне середнє та дисперсію не впливає процес $\varepsilon(k)$, який визначений за допомогою (2.4). По аналогії легко показати, що умовне середнє процесу $\varepsilon(k)$ дорівнює нулю. Якщо $v(k)$ і $\varepsilon(k-1)$ – незалежні величини і $E[v(k)] = 0$, то умовне середнє для $\varepsilon(k)$ можна знайти як

$$E[\varepsilon(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = E[v(k)] E[\alpha_0 + \alpha_1 \varepsilon^2(k-1)]^{1/2} = 0.$$

Очевидно, що процес, який визначається виразом (2.4), впливає на умовну дисперсію. Оскільки дисперсія $\sigma_v^2 = 1$, то умовна дисперсія для $\varepsilon(k)$ визначається як

$$E[\varepsilon^2(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = \alpha_0 + \alpha_1 \varepsilon^2(k-1). \quad (2.8)$$

Звідси видно, що умовна дисперсія змінної $\varepsilon(k)$ залежить від значення $\varepsilon^2(k-1)$. Якщо остання змінна приймає велике значення, то і умовна дисперсія в момент k також буде мати велике значення. Рівняння (2.8) є авторегресією першого порядку, але відрізняється від звичайної авторегресії тим, що коефіцієнти α_0, α_1 повинні мати певні обмеження. Для того щоб забезпечити додатність дисперсії, обидва коефіцієнти α_0, α_1 повинні бути додатними. Крім того, для забезпечення стійкості процесу авторегресії необхідно щоб виконувалась наступна умова: $0 < \alpha_1 < 1$.

Рівняння (2.5)-(2.8) ілюструють загальні властивості будь-якого процесу АРУГ. Структура АРУГ моделі така, що умовне та безумовне середнє значення похибок дорівнює нулю. Окрім того, послідовність $\varepsilon(k)$ є послідовно некорельованою, оскільки $E[\varepsilon(k)\varepsilon(k-s)] = 0$ для $\forall s \neq 0$. Головним моментом в даному випадку є те, що елементи послідовності $\{\varepsilon(k)\}$ зв'язані через їх другий центральний момент, тобто умовну дисперсію. Якщо значення елементу $\varepsilon(k-1)$ суттєво відрізняється від нуля, що приводить до відносно великого значення $\alpha_1 \varepsilon^2(k-1)$, то дисперсія змінної $\varepsilon(k)$ має тенденцію до зростання. Таким чином, умовна гетероскедастичність процесу $\{\varepsilon(k)\}$ свідчить про те, що послідовність $\{y(k)\}$ також відноситься до процесу типу АРУГ.

Яким же чином структура похибки впливає на послідовність $\{y(k)\}$? Суттєвим моментом є те, що структура похибки процесу АРУГ і автокореляційні параметри послідовності $\{y(k)\}$ взаємодіють між собою. Про це свідчать численні обчислювальні експерименти [.....]. Цьому явищу можна дати пояснення на інтуїтивному рівні. Будь-яке різке зростання модуля змінної $v(k)$ викликає збільшення дисперсії послідовності $\{\varepsilon(k)\}$, при цьому чим більше значення має коефіцієнт α_1 , тим більший буде вплив на дисперсію. Більше того, чим більшим буде авторегресійний параметр a_1 , тим більшим буде вплив правої частини на $y(k)$. І чим значнішою буде тенденція до зміни середнього значення послідовності $\{y(k)\}$, тим більшою буде її дисперсія.

Умовні значення середнього та дисперсії для послідовності $\{y(k)\}$ можна записати як

$$E_{k-1}[y(k)] = a_0 + a_1 y(k-1)$$

і

$$\text{Var}[y(k) | y(k-1), y(k-2), \dots] = E_{k-1}[y(k) - a_0 - a_1 y(k-1)]^2 =$$

$$E_{k-1}[\varepsilon^2(k)] = \alpha_0 + \alpha_1 \varepsilon^2(k-1).$$

Оскільки α_1 і $\varepsilon^2(k-1)$ не можуть приймати від'ємних значень, то мінімальним значенням умовної дисперсії буде α_0 . Будь-яке ненульове значення змінної $\varepsilon(k-1)$ позитивно зв'язане з умовною дисперсією змінної $y(k)$ через коефіцієнт α_1 . Безумовні значення середнього та дисперсії для $y(k)$ можна отримати за допомогою розв'язку різницевого рівняння, яке описує процес $\{y(k)\}$ та математичного сподівання. Для рівняння першого порядку (при умові, що процес починається досить далеко в минулому і можна знехтувати довільною константою A) розв'язок має вигляд:

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i). \quad (2.9)$$

Оскільки $E[\varepsilon(k)] = 0, \forall k$, то безумовне математичне сподівання для останнього рівняння матиме вигляд

$$E[y(k)] = \frac{a_0}{1-a_1}.$$

Безумовну дисперсію для $y(k)$ також можна отримати на основі (2.9):

$$\text{var}[y(k)] = \sum_{i=0}^{\infty} a_1^{2i} \text{var}[\varepsilon(k-i)].$$

Якщо врахувати те, що безумовна дисперсія для $\varepsilon(k)$ є величиною постійною, тобто $\text{var}[\varepsilon(k)] = \text{var}[\varepsilon(k-1)] = \text{var}[\varepsilon(k-2)] = \dots = \alpha_0 / (1 - \alpha_1)$, то

$$\text{var}[y(k)] = \frac{\alpha_0}{(1-\alpha_1)} \frac{1}{(1-a_1^2)}.$$

Модель гетероскедастичного процесу (2.4) може бути розширена до довільного порядку і записана у вигляді:

$$\varepsilon(k) = v(k) \left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) \right)^2. \quad (2.10)$$

В цьому рівнянні на змінну $\varepsilon(k)$ впливають всі значення від $\varepsilon(k-1)$ до $\varepsilon(k-q)$, а тому умовну дисперсію можна розглядати як процес авторегресії порядку q .

3 Узагальнена авторегресійна умовно гетероскедастична модель (УАРУГ)

Наступним розширенням АРУГ моделі є описання умовної дисперсії як процесу АРКС. Нехай похибки описуються рівнянням

$$\varepsilon(k) = v(k) [h(k)]^{1/2},$$

де $\sigma_v^2 = 1$, і

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (3.1)$$

Оскільки процес $\{v(k)\}$ визначено як білий шум, який некорельований із значеннями $\varepsilon(k-i)$, то умовне і безумовне середнє для $\varepsilon(k)$ дорівнюють нулю. Очевидно, що безумовне математичне сподівання

$$E[\varepsilon(k)] = E[v(k) (h(k))^{1/2}] = 0.$$

Умовна дисперсія змінної $\varepsilon(k)$ визначається як $E_{k-1}[\varepsilon^2(k)] = h(k)$.

Узагальнена модель АРУГ, яку називають **УАРУГ(p,q)**, складається із двох компонент – авторегресії та ковзного середнього відносно дисперсії гетероскедастичного процесу. Процес першого порядку отримуємо при $p=0, q=1$, його можна формально визначити як процес УАРУГ(0,1). Якщо всі коефіцієнти $\beta_i = 0$, то модель УАРУГ(p,q) еквівалентна моделі АРУГ(p,q). Для того щоб забезпечити скінченність умовної дисперсії, корені характеристичного рівняння, записаного для (3.1), повинні знаходитися всередині кола одиничного радіусу.

Основним характерним моментом УАРУГ моделі є те, що збурення, яке діє на процес $\{y(k)\}$, є процесом АРКС. Тому можна очікувати, що залишки (похибки) моделі АРКС (попередньої моделі процесу) будуть відповідати за своїми характеристиками гетероскедастичному процесу. Це твердження можна пояснити наступним чином. Нехай $\{y(k)\}$ — процес АРКС. Якщо модель АРКС адекватна процесу $\{y(k)\}$, то автокореляційна функція (АКФ) і часткова автокореляційна функція (ЧАКФ) залишків

повинні вказувати на те, що це процес білого шуму. З іншого боку, АКФ *квадратів* залишків можна використати для попереднього визначення порядку процесу УАРУГ. Оскільки $E_{k-1}[\varepsilon(k)] = (h(k))^{1/2}$, то рівняння (3.1) можна записати у вигляді

$$E_{k-1}[\varepsilon^2(k)] = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (3.2)$$

Останнє рівняння формально схоже на процес АРКС(p,q) відносно послідовності $\{\varepsilon^2(k)\}$. Якщо умовна гетероскедастичність в процесі присутня, то корелограма повинна вказувати на цей факт. Алгоритм аналізу корелограми квадратів залишків можна записати наступним чином:

Крок 1:

Побудувати по можливості адекватну модель АРКС (або АР) для послідовності $\{y(k)\}$ і побудувати додатковий ряд із квадратів похибок $\hat{\varepsilon}^2(k)$. Обчислити вибірку (на основі значень елементів ряду) дисперсію залишків

$$\hat{\sigma}^2 = \sum_{k=1}^N \frac{\hat{\varepsilon}^2(k)}{N},$$

де N – кількість елементів ряду похибок.

Крок 2:

Обчислити вибірку автокореляційну функцію квадратів залишків за виразом

$$\rho(s) = \frac{\sum_{k=s+1}^N [\hat{\varepsilon}^2(k) - \hat{\sigma}^2][\hat{\varepsilon}^2(k-s) - \hat{\sigma}^2]}{\sum_{k=1}^N [\hat{\varepsilon}^2(k) - \hat{\sigma}^2]}.$$

Крок 3:

Для вибірок даних досить великої довжини стандартне відхилення для змінної $\rho(s)$ приблизно прирівнюють значенню $T^{1/2}$. Індивідуальні значення змінної $\rho(s)$, які суттєво відрізняються від нуля, вказують на наявність процесу УАРУГ. Для визначення того факту наскільки є відмінними від нуля значення вибіркової АКФ, можна застосувати Q – статистику Лjung-Бокса:

$$Q = N(N + 2) \sum_{i=1}^n \rho(s) / (N - s).$$

Якщо значення $\hat{\varepsilon}^2(k)$ некорельовані між собою, то Q – статистика має в асимптотиці χ^2 розподілення з n степенями свободи. За нульову гіпотезу можна прийняти те, що значення $\hat{\varepsilon}^2(k)$ некорельовані між собою. Відмова від цієї нульової гіпотези еквівалентна відмові від нуль-гіпотези, що процес АРУГ чи УАРУГ в конкретному випадку відсутній.

Формально перевірку на наявність гетероскедастичності можна здійснити за допомогою наступної двокрокової процедури:

Крок 1:

За допомогою МНК побудувати «найкращу» модель $AR(n)$ для основної змінної

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \dots + a_n y(k-n) + \varepsilon(k).$$

Крок 2:

Побудувати додатковий ряд із квадратів похибок $\hat{\varepsilon}^2(k)$. За допомогою МНК оцінити авторегресію для нового ряду:

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \alpha_2 \hat{\varepsilon}^2(k-2) + \dots + \alpha_q \hat{\varepsilon}^2(k-q). \quad (2.13)$$

Якщо ефект гетероскедастичності відсутній, то оцінки коефіцієнтів $\alpha_1, \dots, \alpha_q$ повинні бути нульовими або близькими до нуля. При цьому коефіцієнт детермінації R^2 повинен мати досить мале значення. Якщо вибрати за нуль-гіпотезу, що процес АРУГ відсутній, то при довжині ряду N статистика NR^2 повинна мати розподілення χ_q^2 . Якщо значення TR^2 досить високе, то відмова від нуль-гіпотези, що всі коефіцієнти $\alpha_1, \dots, \alpha_q$ дорівнюють нулю, еквівалентна відмові від нуль-гіпотези, що процес АРУГ (УАРУГ) відсутній. З іншого боку, якщо NR^2 має відносно невелике значення, то можна зробити висновок про відсутність ефекту АРУГ.

4 Приклади застосування моделей АРУГ

Приклад 1. Застосування моделі АРУГ до оцінки інфляції.

АРУГ та УАРУГ зручні тим, що дозволяють оцінити дисперсію процесу в конкретний момент часу.

Фірми зацікавлені у тому, щоб на протязі контракта з найманими робітниками оцінити рівень інфляції. З цим зв'язана величина заробітної плати.

Нехай $E_k[\pi(k+1)]$ - умовне математичне сподівання очікуваної інфляції в момент $k+1$, $\sigma^2_{\pi}(k)$ - умовна дисперсія інфляції. Як правило, міру ризику оцінюють за допомогою умовної дисперсії.

Модель інфляції

Відомі поквартальні дані для Англії за період 1958р. (2 квартал) – 1977р. (2 квартал) (80 точок).

Позначення:

$P(k)$ – \log індексу споживчих цін;

$W(k)$ – \log індексу номінальної (середньої) заробітної плати.

Рівень інфляції : $\pi(k)=p(k)-p(k-1)$.

Реальний прибуток: $r(k)=w(k)-p(k)$.

1-й етап побудови моделі

За часовим рядом отримали наступну модель:

$$\pi(k)=0,0257+0,394 \pi(k-1)+0,408 \pi(k-4)-0,404 \pi(k-5)+0,0559 r(k-1)+\varepsilon_1(k), \quad (4.1)$$

$h(k)=0,000089$ – дисперсія процесу $\{\varepsilon_1(k)\}$; член $0,0559r(k-1)$ збільшує інфляцію в поточний момент часу; $\pi(k-4), \pi(k-5)$ зумовлені сезонними ефектами.

Тестування моделі АРУГ(4) дало величину $NR^2 = 15,2$. Критичне значення величини χ^2 ($\tau=4$) дорівнює 13,28. Тому Engle заключає наявність АРУГ-процесу:

$$h(k) = \varepsilon_1^2(k) = \alpha_0 + \alpha_1 [0,4 \varepsilon_1^2(k-1)+0,3 \varepsilon_1^2(k-2)+0,2 \varepsilon_1^2(k-3)+0,1 \varepsilon_1^2(k-4)] \quad (4.2)$$

для будь-якого $\varepsilon_1(k)>0$.

Два параметри α_0 і α_1 були взяті з умови невід'ємності та стаціонарності. Вибрано: $\alpha_0>0$, $0<\alpha_1<1$.

Необхідно зазначити, що коефіцієнти 0,4; 0,3; 0,2; 0,1 вибрані довільно за припущенням, що найбільше значення має перший коефіцієнт.

2-й етап побудови моделі

Таким чином процедура оцінювання коефіцієнтів має 2-х ступінчатий характер:

1-ий етап:

- спочатку оцінюються коефіцієнти рівняння (2.1) і залишки;
- на основі залишків оцінюються коефіцієнти α_0 і α_1 рівняння (2.2);

2-ий етап:

- уточнення коефіцієнтів моделі (2.1) після знаходження коефіцієнтів моделі (2.2).

Це ітераційна процедура. Знаходять ряд $\{\hat{\varepsilon}(k)\} = \{\sqrt{\varepsilon(k)}\}$, всі $\varepsilon_1^2(k) > 0$ і потім модель (2.1) будується для ряду $\{y(k) - \sqrt{\varepsilon(k)}\}$, $\varepsilon_1^2(k) = h(k)$. Для цього добре підходить процедура максимальної правдоподібності. До неї повернемося дещо пізніше.

Остаточно отримана модель:

$$\pi(k) = 0,0328 + 0,162 \pi(k-1) + 0,264 \pi(k-4) - 0,325 \pi(k-5) + 0,0707 r(k-1) + \varepsilon_2(k), \quad (4.3)$$

$$h(k) = \varepsilon_2^2(k) + 1,4E-5 + 0,955 [0,4\varepsilon_2^2(k-1) + 0,3\varepsilon_2^2(k-2) + 0,2\varepsilon_2^2(k-3) + 0,1\varepsilon_2^2(k-4)];$$

$h(k)$ – передбачена на один крок дисперсія похибки моделі.

$\alpha_1 = 0,955$ – свідчить про те, що дисперсія похибок значна.

$$\sigma_{\varepsilon_2}^2 = \text{const}; E[\varepsilon_2(k)] = 0$$

Повертаємось до довільно вибраних коефіцієнтів 0,4; 0,3; 0,2; 0,1.

Один із авторів стверджує [Bollerslev, 1986]:

Спільним моментом при виборі структури рівнянь є використання лінійно спадаючих коефіцієнтів (структури) рівняння умовної дисперсії з метою обліку довготривалої пам'яті, як правило притаманній емпірично отриманим процесам. Якщо структура цього рівняння оцінюється одним з відомих методів, то це часто призводить до порушення обмеження невід'ємності.

Контрольні запитання

1. Поясніть стандартні статистичні параметри, що характеризують оцінку авторегресії.
2. Як може вплинути позитивна кореляція величин $\{v(k)\}$ моделі

$$y(k) = a_0 + a_1 y(k-1) + v(k)$$

на оцінку коефіцієнта a_1 при короткій вибірці даних?

3. Поясніть відмінність між безумовними та умовними статистичними параметрами: *математичне сподівання та дисперсія*.
4. Використовуючи АКФ та ЧАКФ для залишків та квадратів залишків зробіть висновок про наявність процесу АРУГ в процесі, що розглядається (п. 1.6)?
5. Порівняйте моделі першого та четвертого порядків, знайдені в пп. 1.7 і 1.10.
6. Поясніть на прикладі реального процесу природу гетероскедастичності (тобто звідки вона береться)?
7. Поясніть методику визначення гетероскедастичності.
8. В чому полягає різниця між моделями АРУГ та УАРУГ?
9. Наведіть приклад процесу (ряду), де може бути присутня гетероскедастичність?
10. Наведіть альтернативний вираз для обчислення умовної дисперсії.