Тема 4. Узагальнений метод найменших квадратів. Лекція 5. Узагальнений метод найменших квадратів.

Зміст.

1. Поняття гетероскедастичності.

Передумови використання 1МНК для оцінювання параметрів моделі, що були розглянуті нами раніше, на практиці можуть порушуватися.

Дві лекції попереду було розглянуто проблеми мультиколінеарності, які пов'язані з порушенням умови . Минулого разу ми вивчили автокореляцію.

Однією з чотирьох необхідних передумов для застосування 1МНК при оцінюванні параметрів економетричної моделі є вимоги постійної дисперсії залишків для кожного спостереження вихідної сукупності, тобто $M(uu') = \sigma_u^2 E$, де E – одинична матриця. Ця властивість незмінної дисперсії залишків в спостереженнях називається **гомоскедастичністю**. У практичних дослідженнях вона часто порушується.

Означення: Якщо дисперсія залишків стала для кожного спостереження, тобто $M(uu') = \sigma_u^2 E$, то ця її властивість називається **гомоскедастичністю.**

Означення: Якщо дисперсія залишків змінюється для кожного спостереження або групи спостережень, тобто $M(uu') = \sigma_u^2 S$, то це явище називається **гетероскедастичністю.**

Де σ_u^2 - дисперсія залишків, яка виступає невідомим параметром; S – діагональна додатно визначена матриця розміром $n \times n$. Якщо існує гетероскедастичність залишків, то це спричинюється до того, що оцінки параметрів моделі МНК будуть незміщеними, обґрунтованими, але не ефективними. При цьому формулу для стандартної помилки оцінки, однозначно, застосовувати не можна. Тому завжди постає потреба вивчати це явище і, якщо воно існує, для оцінювання параметрів моделі застосовувати узагальнений метод найменших квадратів (метод Ейткена).

При наявності гетероскедастичності в простій економетричній моделі, тобто $Y = a_0 + a_1 X + u$, щоб оцінити параметри МНК, достатньо змінити специфікацію моделі. Коли будується модель множинної регресії з багатьма змінними, то таке перетворення значно ускладнюється. Тому спочатку треба одним із методів визначити наявність гетероскедастичності, а потім оцінити параметри моделі із застосуванням спеціального підходу, який ми розглянемо.

Припустимо, що дисперсія залишків для моделі $Y = a_0 + a_1 x + u$ пропорційна до величини x. Тоді доцільно виконати перетворення вихідної інформації, поділивши, наприклад, усі змінні на x. Модель набере вигляду

$$\frac{y}{x} = a_0 \frac{1}{x} + a_1 + u \;\rightarrow\; \frac{y}{x} = a_1 + a_0 \frac{1}{x} + u .$$

У результаті для оцінювання параметрів можна застосувати 1МНК. Зауважимо, що параметри a_0 і a_1 помінялися ролями. Вільним членом моделі замість a_0 став параметр a_1.

Приклад 1. Побудуємо економетричну модель, що характеризує залежність між заощадженнями та доходом населення, млрд ф. ст. (табл. 1).

Скориставшись оператором оцінювання 1МНК

$$A = \left(X' X\right)^{-1} X' Y ,$$

Дістанемо $\hat{a}_0 = -1{,}081$; $\hat{a}_1 = 0{,}1178$.

Економетрична модель має вигляд $\hat{y} = -1{,}081 + 0{,}1178x$.

Коефіцієнт детермінації $R^2 = \hat{a}_1 \sum xy \big/ \sum y^2$ для цієї моделі, $R^2 = 0{,}918$ а це означає, що варіація заощаджень Y на 91,8% визначається варіацією доходів населення.

Таблиця 1.

Вхідні дані

Рік	Заощадження	Дохід	Рік	Заощадження	Дохід
1	0,36	8,8	10	0,59	15,5
2	0,2	9,4	11	0,90	16,7
3	0,08	10,0	12	0,95	17,7
4	0,20	10,6	13	0,82	18,6
5	0,10	11,0	14	1,04	19,7
6	0,12	11,9	15	1,53	21,1
7	0,41	12,7	16	1,94	22,8
8	0,50	13,5	17	1,75	23,9
9	0,43	14,3	18	1,99	25,2

На перший погляд, результат наводить на думку, що специфікація моделі не містить помилки.

Але логічно висунути гіпотезу, що відхилення заощаджень від норми можуть бути пропорційними до доходу, тобто для цієї моделі дуже ймовірне існування гетероскедастичності залишків.

Отже, вихідну інформацію доцільно перетворити, поділивши обидві змінні на величину доходу X (табл. 2):

$$. y' = y/x; \quad x' = 1/x$$

Таблиця 2.

Розрахункові дані

Рік	y'	x'	Рік	y'	x'
1	0,041	0,114	10	0,038	0,065
2	0,022	0,106	11	0,054	0,060
3	0,008	0,100	12	0,054	0,056
4	0,019	0,094	13	0,044	0,054
5	0,009	0,091	14	0,053	0,051

6	0,010	0,084	15	0,073	0,047
7	0,032	0,079	16	0,085	0,044
8	0,037	0,074	17	0,073	0,042
9	0,030	0,070	18	0,079	0,040

Нове рівняння зв’язку згідно з даними табл. 2 має вигляд

$$y' = -0{,}854 + 0{,}1026x'.$$

У результаті перетворення вихідних даних практично повністю змінилася специфікація моделі. Оскільки $x' = 1/x$, то цей зв’язок нелінійний. По-друге, $y' = y/x$ характеризує відносний показник — рівень заощаджень, який припадає на одиницю доходу.

Виконавши цю процедуру, дістанемо таке: спостереження з меншими значеннями x' мають відносно більшу питому вагу при оцінюванні параметрів моделі, ніж у першому варіанті.

З наведеного прикладу бачимо, що явище гетероскедастичності не впливатиме на оцінки параметрів 1МНК, якщо певним чином перетворити вихідну інформацію. При цьому якщо економетрична модель має лише дві змінні, то це можна зробити так, як у прикладі 1.

Це перетворення значно ускладнюється, якщо будується економетрична модель з багатьма змінними. У такому разі потрібно з’ясувати зміст гіпотези, згідно з якою $M(uu') = \sigma_u^2 S$, де σ_u^2 лишається невідомим параметром, а S — відома симетрична додатно визначена матриця.

До змісту

2. Методи визначення гетероскедастичності.

Можливість перевірки припущень про наявність гетероскедастичності залежить від природи вихідних даних. В економетрії існують наступні методи перевірки гетероскедастичності для різних вихідних даних:

1. критерій μ – при великій кількості сукупності спостережень n (n – удвічі більше кількості оцінюваних параметрів);

2. параметричний тест Гольдфельда-Квандта – при великій кількості сукупності спостережень n (коли дисперсія залишків зростає пропорційно квадрату однієї з незалежних змінних моделі) ;
3. непараметричний тест Гольдфельда-Квандта – базується на розгляді графічно зображених спостережень за значеннями x;
4. тест Глейсера .

До змісту

3. Узагальнений метод найменших квадратів (метод Ейткена).

Економетрична модель, якій притаманна гетероскедастичність, є узагальненою моделлю, і для оцінювання її параметрів слід скористатися узагальненим методом найменших квадратів. Розглянемо цей метод.

Нехай задано економетричну модель

$$Y = XA + u, \tag{1}$$

коли $M(uu') = \sigma_u^2 S$.

Задача полягає в знаходженні оцінок елементів вектора A в моделі. Для цього використовується матриця S, за допомогою якої коригується вихідна інформація. Ця ідея була покладена в основу методу Ейткена.

Базуючись на особливостях матриць P і S, які були розглянуті в лекції з автокореляції, можна записати співвідношення між цими матрицями та оберненими до них.

Оскільки S — додатно визначена матриця, то вона може бути зображена як добуток PP', де матриця P є невиродженою, тобто:

$$S = PP', \tag{2}$$

коли

$$P^{-1} S P^{-1'} = E; \tag{3}$$

$$P^{-1'} P^{-1} = S^{-1}. \tag{4}$$

Помноживши рівняння (1) ліворуч на матрицю P^{-1}, дістанемо:

$$P^{-1}Y = P^{-1}XA + P^{-1}u. \tag{5}$$

Позначимо $Y^* = P^{-1}Y$;

$$X^* = P^{-1}X;$$

$$u^* = P^{-1}u.$$

Тоді модель матиме вигляд:

$$Y^* = X^*A + u^*. \tag{6}$$

Використовуючи вираз (3), неважко показати, що

$$M(u^*u^{*'}) = \sigma^2 E$$

тобто модель (6) задовольняє умови (2), коли параметри моделі можна оцінити на основі 1МНК.

Звідси

$$\hat{A} = (X^{*'}X^*)^{-1}X^{*'}Y^* = (X'S^{-1}X)^{-1}X'S^{-1}Y. \tag{7}$$

Ця оцінка є незміщеною лінійною оцінкою вектора A, який має найменшу дисперсію і матрицю коваріацій

$$\mathrm{var}(\hat{A}) = \sigma_u^2(X^{*'}X^*)^{-1} = \sigma_u^2(X'S^{-1}X)^{-1} \tag{8}$$

Незміщену оцінку для дисперсії σ_u^2 можна дістати так:

$$\frac{1}{n-m}(Y^* - X^*\hat{A})'(Y^* - X^*\hat{A}) = \frac{1}{n-m}(Y - X\hat{A})'S^{-1}(Y - X\hat{A}) = \\ = \frac{1}{n-m}u'S^{-1}u. \tag{9}$$

Оцінка параметрів $\hat{A}$, яку знайдено за допомогою (7), є оцінкою узагальненого методу найменших квадратів (методу Ейткена).

До змісту

4.Визначення прогнозу за узагальненим методом найменших квадратів.

При заданій матриці S оцінку параметрів моделі можна обчислити згідно із (7), а стандартну помилку — згідно із (8). Тому можна сконструювати звичайні критерії значущості і довірчі інтервали для параметрів $\hat{a}_j$.

Визначивши залишки $u = Y - X\hat{A}$ і помноживши ліворуч на матрицю P^{-1}, дістанемо:

$$P^{-1}u = P^{-1}Y - P^{-1}X\hat{A},$$

або $u^* = Y^* - X^*\hat{A}.$

Звідси $Y^* = X^*A + u^*.$

Тоді $Y^{*\prime}Y^* = (X^*\hat{A} + u^*)'(X^*\hat{A} + u^*).$

Оскільки $\hat{A} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* = (X'S^{-1}X)^{-1}X'S^{-1}Y,$

то $Y'S^{-1}Y = \hat{A}'X'S^{-1}Y + uS^{-1}u$ (10)

Отже, ми розбили загальну суму квадратів для (7.6) на суму квадратів регресії і залишкову. Згідно з цими даними дисперсійний аналіз буде виконано для перетворених вихідних даних. Крім того, коли незалежна змінна Y^* виміряна відносно початку відліку, а не у формі відхилення від середньої, то необхідно визначити її середнє значення і скористатись ним для корекції загальної суми квадратів і суми квадратів регресії.

Модель узагальненого методу найменших квадратів іноді специфікується у вигляді

$$Y = XA + u,$$

$$M(u) = 0, \tag{11}$$

$$M(uu') = V,$$

де $V = \sigma_u^2 S$ – відома симетрична додатно визначена матриця. Тоді вираз для оцінки параметрів згідно з методом Ейткена запишеться так:

$$\hat{A} = (X'V^{-1}X)^{-1}X'V^{-1}Y, \tag{12}$$

а для її коваріаційної матриці

$$\sigma_u^2(\hat{A}) = \sigma_u^2(X'V^{-1}X)^{-1}. \tag{13}$$

До змісту