ДЕЯКІ АСПЕКТИ ЗАСТОСУВАННЯ МЕТОДУ НАЙМЕНШИХ КВАДРАТІВ ПРИ РОЗВ'ЯЗАННІ ЗАДАЧ ЕКОНОМІЧНОГО ЗМІСТУ

Сушко О.С.,
аспірант,
Національний педагогічний університет імені М.П.Драгоманова

У даній статті розглядаються умови порушення передумов використання звичайного методу найменших квадратів та можливості використання узагальненого методу найменших квадратів для отримання більш достовірних результатів. Розкривається суть гетероскедастичності та зваженого методу найменших квадратів.

В данной статье рассматриваются условия нарушения предпосылок использования обычного метода наименьших квадратов и возможности использования обобщённого метода наименьших квадратов для получения более достоверных результатов. Раскрывается суть гетероскедастичности и взвешенного метода наименьших квадратов.

This article focuses on the conditions of use violation preconditions ordinary least squares and the possibility of using a generalized least squares method to obtain more reliable results. Disclosed the nature of heteroscedasticity and the weighted least squares method.

Вступ. Однією з головних задач економетрії в ринковій економіці є ретельне вивчення кількісних зв'язків між показниками для кращого розуміння економічних явищ і процесів, що в свою чергу дозволяє більш обґрунтовано сформулювати управлінські рішення та дати прогнози на майбутнє. Для вирішення цієї задачі потрібна побудова економетричної моделі. Ті з моделей, що базуються на методі найменших квадратів (МНК) при оцінюванні їх параметрів, називаються класичними і вивчаються у класичній економетрії.

Економетрія — це наука, що вивчає кількісні і якісні характеристики та взаємозв'язки економічних об'єктів і процесів за допомогою математичних і статистичних методів і моделей.

З огляду на це, економетрія є однією з найважливіших дисциплін фундаментальної підготовки як бакалаврів з економіки, так і вчителів (викладачів) економічних та економіко-математичних дисциплін. Донедавна економетрія як окремий предмет не вивчалася у вищих навчальних економічних закладах нашої країни. Дисципліни з математичного моделювання економічних процесів та явищ, що входили до програм вузів України, містили лише окремі теми з економетрії, які здебільшого базувалися на класичному регресійному аналізі.

Як відомо, застосувати метод найменших квадратів для оцінювання параметрів економетричної моделі можна лише в разі виконання певних умов, які далеко не завжди виконуються на практиці для вихідної економічної інформації. Якщо ці умови порушуються, доводиться застосовувати інші методи оцінювання параметрів економетричної моделі. Вивчити методи оцінювання параметрів моделі та особливості економічної інформації з метою кількісного вимірювання взаємозв'язку між досліджуваними процесами та явищами — основне завдання курсу «Економетрія».

Огляд сучасних досліджень. Застосуванню методу найменших квадратів присвячено ряд монографій та підручників [3-5, 8], в яких розроблено теоретичні основи та розглянуто прикладні аспекти застосування цього методу до різних економічних процесів і явищ.

Мета статті – розглянути загальні аспекти методу найменших квадратів як одного з основних методів дослідження економічних процесів та явищ, навести приклади порушення передумов його використання та способи їх вирішення.

Основний виклад. Під час реалізації регресійного аналізу за допомогою звичайного МНК особливу увагу необхідно звернути на проблеми, пов'язані з виконанням необхідних умов для випадкових відхилень, оскільки властивості статистичних оцінок параметрів лінійної регресії перебувають у прямій залежності від цих відхилень ε_i.

Для одержання якісних статистичних оцінок потрібно уважно стежити за виконанням передумов, що сформульовані в теоремі Гаусса-Маркова, бо при їх порушенні звичайний МНК дає статистичні оцінки, яким притамані небажані властивості.

Однією із передумов теореми Гаусса-Маркова є:

$$\operatorname{cov}(\varepsilon_i \cdot \varepsilon_j) = \begin{cases} 0, \ i \neq j \\ \sigma_\varepsilon^2 = const, \ i = j \end{cases} \text{ при } i, j = \overline{1, n},$$

де n – число спостережень.

Виконання цієї умови називають *гомоскедастичністю* залишків. У випадку, коли порушується ця передумова, тобто

$$\operatorname{cov}(\varepsilon_i \cdot \varepsilon_j) = \begin{cases} 0, \ i \neq j \\ \sigma_\varepsilon^2 \neq const, \ i = j \end{cases} \text{ при } i, j = \overline{1, n},$$

це є головною ознакою наявності гетероскедастичності моделі.

Таким чином, моделі, для яких не виконуються передумови Гаусса-Маркова, можна умовно поділити на три групи:

До ***першої*** належать такі моделі, для яких виконуються наступні умови стосовно компонент випадкового вектора $\vec{\varepsilon}$:

1) вони мають нульові математичні сподівання:

$$M(\varepsilon_i) = 0, \qquad i = \overline{1, n};$$

2) між собою є попарно некорельовані:

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, \quad при \ i \neq j, \\ \sigma_{\varepsilon_i}^2 \neq const, \ при \ i = j, \ i = \overline{1, n} \end{cases}$$

В цьому випадку коваріаційна матриця випадкового вектора $\vec{\varepsilon}$ буде мати такий вигляд:

$$\operatorname{cov}(\vec{\varepsilon}\cdot\vec{\varepsilon}') = M(\vec{\varepsilon}\cdot\vec{\varepsilon}') = M\left(\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{pmatrix}\cdot\begin{pmatrix}\varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n\end{pmatrix}\right) =$$

$$= M\begin{pmatrix}\varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n\\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2\varepsilon_n\\ \ldots & \ldots & \ldots & \ldots\\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2\end{pmatrix} = \begin{pmatrix}M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \cdots & M(\varepsilon_1\varepsilon_n)\\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \cdots & M(\varepsilon_2\varepsilon_n)\\ \ldots & \ldots & \ldots & \ldots\\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \cdots & M(\varepsilon_n^2)\end{pmatrix} =$$

$$= \begin{pmatrix}\sigma_{\varepsilon_1}^2 & 0 & \cdots & 0\\ 0 & \sigma_{\varepsilon_2}^2 & \cdots & 0\\ \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & \cdots & \sigma_{\varepsilon_n}^2\end{pmatrix}.$$

Отже, $$\operatorname{cov}(\vec{\varepsilon}\cdot\vec{\varepsilon}') = \begin{pmatrix}\sigma_{\varepsilon_1}^2 & 0 & \cdots & 0\\ 0 & \sigma_{\varepsilon_2}^2 & \cdots & 0\\ \ldots & \ldots & \ldots & \ldots\\ 0 & 0 & \cdots & \sigma_{\varepsilon_n}^2\end{pmatrix}$$

Такі моделі називають економетричними моделями з ознакою гетероскедастичності залишків.

До **другої** групи належать моделі, для яких виконуються такі умови:

1) збурення мають нульові математичні сподівання:

$M(\varepsilon_i) = 0, \quad i = \overline{1,n}$;

2) вони є попарно корельованими:

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} k_{ij}, & при\ i \neq j,\\ \sigma_{\varepsilon_i}^2 = \sigma_\varepsilon^2 = const, & при\ i = j,\ i = \overline{1,n}\end{cases}$$

де $k_{ij} = M(\varepsilon_i\cdot\varepsilon_j) - M(\varepsilon_i)\cdot M(\varepsilon_j) = M(\varepsilon_i\cdot\varepsilon_j) \neq 0$.

В цих моделях між випадковими відхиленями $\varepsilon_i, \varepsilon_j$ існує кореляційний зв'язок, хоча дисперсії їх є сталими величинами.

Коваріаційна матриця в цьому випадку матиме вигляд

$$\operatorname{cov}(\vec{\varepsilon}\cdot\vec{\varepsilon}') = \begin{pmatrix}\sigma_\varepsilon^2 & k_{12} & k_{13} & \cdots & k_{1n}\\ k_{21} & \sigma_\varepsilon^2 & k_{23} & \cdots & k_{2n}\\ k_{31} & k_{32} & \sigma_\varepsilon^2 & \cdots & k_{3n}\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ k_{n1} & k_{n2} & k_{n3} & \cdots & \sigma_\varepsilon^2\end{pmatrix}$$

і слід пам'ятати, що $k_{ij} = k_{ji}$, тобто матриця є симетричною. Тому в цих моделях хоча умова гомоскедастичності (сталість дисперсій залишків) і виконується, але використання звичайного МНК не рекомендується внаслідок існування коваріаційних моментів між випадковими залишками.

До **третьої** групи належать моделі, для яких:

1) збурення мають нульові математичні сподівання $M(\varepsilon_i) = 0, \quad i = \overline{1,n}$;

2) елементи $\vec{\varepsilon}$ є попарно корельованими $\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} k_{ij}, & при\ i \neq j, \\ \sigma^2_{\varepsilon_i} \neq const, & при\ i = j, \ i = \overline{1,n}. \end{cases}$

Слід наголосити, що для всіх трьох груп лінійних моделей з порушенням передумов застосування МНК точкові статистичні оцінки β_i^* для теоретичних параметрів β_i будуть незміщеними, але втрачають свою ефективність, тобто вони не матимуть мінімальну дисперсію, що призведе до зниження ймовірності одержання доброякісної оцінки.

Для моделей першої групи статистична оцінка параметрів здійснюється шляхом використання зваженого методу найменших квадратів. Для моделей другої та третьої груп – узагальненого методу найменших квадратів, який ми розглянемо пізніше.

Умова гомоскедастичності є головною для лінійної класичної моделі і записується як

$$\operatorname{cov}(\vec{\varepsilon} \cdot \vec{\varepsilon}\,') = \sigma_\varepsilon^2 I_n$$

Для лінійних моделей з властивістю випадкового вектора $\vec{\varepsilon}$, коли

$$\operatorname{cov}(\vec{\varepsilon} \cdot \vec{\varepsilon}\,') = \begin{pmatrix} \sigma_1^2 & k_{12} & k_{13} & \cdots & k_{1n} \\ k_{21} & \sigma_2^2 & k_{23} & \cdots & k_{2n} \\ k_{31} & k_{32} & \sigma_3^2 & \cdots & k_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & k_{n3} & \cdots & \sigma_n^2 \end{pmatrix} = \Omega$$

(матриця Ω є симетричною, додатньо визначеною матрицею n-го порядку) неможливим є використання звичайного МНК з метою визначення статистичних оцінок, як це було здійснено для лінійної класичної моделі. В такому випадку використовують так званий узагальнений метод найменьших квадратів (УМНК).

Нехай досліджується лінійна модель

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \qquad (1)$$

з порушенням умови гомоскедастичності, а саме

$$\operatorname{cov}(\vec{\varepsilon} \cdot \vec{\varepsilon}\,') = \Omega$$

Тоді додатньо визначена матриця Ω допускає існування такої невиродженої матриці π, що

$$\Omega = \pi \cdot \pi'$$

Звідси буде випливати

$$\pi^{-1} \cdot \Omega \cdot (\pi^{-1})' = I_n$$

Таким чином, одержимо

$$\Omega^{-1} = (\pi^{-1})' \cdot \pi^{-1} \qquad (2)$$

Враховуючи (2) для моделі (1) здійснимо таке перетворення: ліву і праву частини рівняння помножимо зліва на матрицю π^{-1}.

$$\pi^{-1}\vec{y} = \pi^{-1}X\vec{\beta} + \pi^{-1}\vec{\varepsilon}$$

Позначивши

$$\vec{y}^* = \pi^{-1}\vec{y}, \quad X^* = \pi^{-1}X, \quad \vec{\varepsilon}^* = \pi^{-1}\vec{\varepsilon} \tag{3}$$

Одержимо

$$\vec{y}^* = X^*\vec{\beta} + \vec{\varepsilon}^* \tag{4}$$

Здійснивши перевірку моделі на наявність гетероскедастичності, маємо:

1. $M(\vec{\varepsilon}^*) = M(\pi^{-1}\cdot\vec{\varepsilon}) = \pi^{-1}M(\vec{\varepsilon}) = 0$,
2. $\operatorname{cov}(\vec{\varepsilon}^*\cdot(\vec{\varepsilon}^*)') = M(\vec{\varepsilon}^*\cdot(\vec{\varepsilon}^*)') = M\left(\pi^{-1}\cdot\vec{\varepsilon}\cdot(\pi^{-1}\cdot\vec{\varepsilon})'\right) =$

$$= M\left(\pi^{-1}\cdot\vec{\varepsilon}\cdot\vec{\varepsilon}'(\pi^{-1})'\right) = \pi^{-1}M\left(\vec{\varepsilon}\cdot\vec{\varepsilon}'\right)\cdot(\pi^{-1})' = \pi^{-1}\cdot\Omega\cdot(\pi^{-1})' = I_n.$$

Таким чином, виявилось, що перетворена модель (4) є гомоскедастичною, а тому для визначення статистичних оцінок цієї моделі можемо використати звичайний МНК, як для класичної лінійної моделі і одержимо

$$\vec{\beta}^* = \left(\left(X^*\right)'\cdot X^*\right)^{-1}\cdot\left(X^*\right)'\cdot\vec{Y}^*$$

Враховуючи (3) маємо:

$$\vec{\beta}^* = \left(\left(\pi^{-1}X\right)'\cdot\pi^{-1}X\right)^{-1}\cdot\left(\pi^{-1}X\right)'\cdot\pi^{-1}\vec{Y} =$$

$$= \left(X'\left(\pi^{-1}\right)'\cdot\pi^{-1}X\right)^{-1}\cdot X'\left(\pi^{-1}\right)'\cdot\pi^{-1}\vec{Y} = \left(X'\cdot\Omega^{-1}\cdot X\right)^{-1}\cdot X'\cdot\Omega^{-1}\vec{Y}$$

Коваріаційна матриця вектора $\vec{\beta}^*$ буде дорівнювати:

$$\operatorname{cov}\left(\vec{\beta}^*\cdot\left(\vec{\beta}^*\right)'\right) = \left(\left(X^*\right)'\cdot X^*\right)^{-1} = \left(\left(\pi^{-1}X\right)'\cdot\pi^{-1}X\right)^{-1} =$$

$$= \left(X'\left(\pi^{-1}\right)'\cdot\pi^{-1}X\right)^{-1} = \left(X'\cdot\Omega^{-1}\cdot X\right)^{-1}$$

Таким чином, одержали:

$$\vec{\beta}^* = \left(X'\cdot\Omega^{-1}\cdot X\right)^{-1}\cdot X'\cdot\Omega^{-1}\vec{Y}$$

$$\operatorname{cov}\left(\vec{\beta}^*\cdot\left(\vec{\beta}^*\right)'\right) = \left(X'\cdot\Omega^{-1}\cdot X\right)^{-1}$$

Розглянутий метод перетворення початкової моделі (1) із подальшим використанням звичайного МНК до моделі (4) для визначення $\vec{\beta}^*$, $\operatorname{cov}(\vec{\beta}^*\cdot\left(\vec{\beta}^*\right)')$ дістав назву узагальненого методу найменших квадратів (УМНК).

Але при цьому слід наголосити, що для реалізації УМНК необхідно знати елементи матриці Ω, що на практиці є справою дуже складною. А тому цей метод, певною мірою, виконує чисто ілюстративну функцію в економетрії. Для практичного використання цього методу необхідно накласти певні умови на структуру матриці Ω.

Розглянемо моделі, що належать до першої групи моделей з порушенням передумов використання звичайного МНК.

При здійсненні вибірки ми маємо справу з конкретними реалізаціями залежної змінної Y і відповідними значеннями пояснюючих змінних (регресорів), при цьому завжди буде присутній фактор випадкових збурень, що породжують відхилення ε_i.

Випадкові величини ε_i апріорно можуть набувати довільних значень, що підпорядковані певним ймовірнісним розподілам. Однією з головних вимог до цих розподілів є рівність їх дисперсій.

Цю вимогу потрібно розуміти так: не зважаючи на те, що при кожному конкретному спостереженні випадкові відхилення ε_i будуть між собою відрізнятися, не повинно існувати причини, яка б спонукала значну розбіжність між цими величинами. Тобто похибки в середньому для всіх спостережень повинні мало відрізнятися. Звичайно, в певному розумінні, тут припускається ідеалізація ситуації. Така ідеальна ситуація в реальних умовах не спостерігається. Часто, при реалізації спостережень в одних і тих самих умовах, відхилення ε_i будуть суттєво відрізнятися між собою, тобто в одних спостереженнях вони виявляються відносно великими, в інших – малими.

Так, наприклад, нехай залежність витрат на споживання (*Y)* середньостатистичного суб'єкта від його доходів (*X*) описується парною лінійною регресією

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Розглянемо для цієї моделі два випадки (рис.1 а, б):

1) умова гомоскедастичності виконується;

2) умова гомоскедастичності не виконується (наявна гетероскедастичність). В цьому випадку можуть виникати проблеми, пов'язані з ефектом масштабу (різних одиниць виміру). У часових рядах явище гетероскедастичності пов'язане з тим, що одні й ті самі показники розглядаються в різні моменти часу (наприклад чистий експорт, темпи інфляції в певному регіоні за певний проміжок часу).

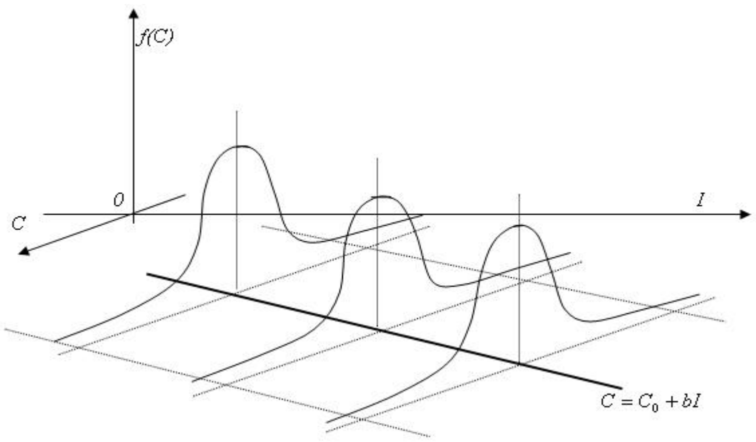

Рис. 1 (а).

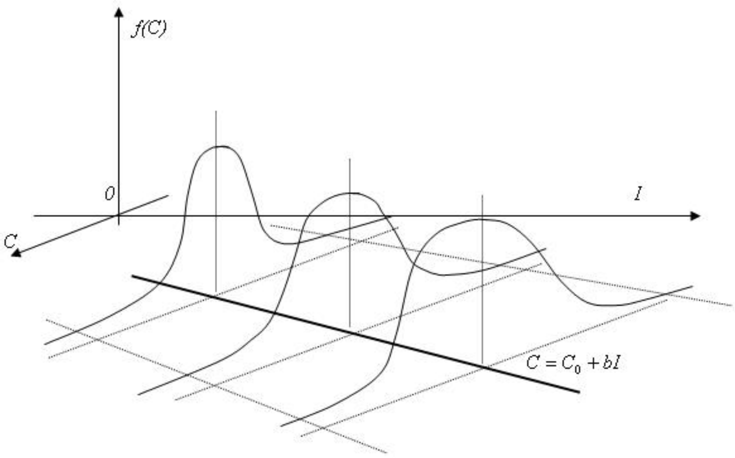

Рис. 1 (б).

За наявності гетероскедастичності (моделі першої групи) статистична оцінка дисперсії σ_ε^2 обчислена за формулою

$$S_\varepsilon^2 = \frac{\vec{e}' \cdot \vec{e}}{n - m - 1},$$

де n – кількість спостережень, m – кількість регресорів в моделі, яка використовується для визначення дисперсій $S_{\beta_i^*}^2$ для всіх емпіричних коефіцієнтів β_i^* не буде незміщеною. Тоді t-статистика, F-статистика, інтервальні оцінки параметрів моделі стануть ненадійними.

Отже, використання звичайного МНК при наявності гетероскедастичності в моделі буде неефективним. Це добре ілюструється на прикладі парної лінійної регресії, графік якої зображено на рис. 2.

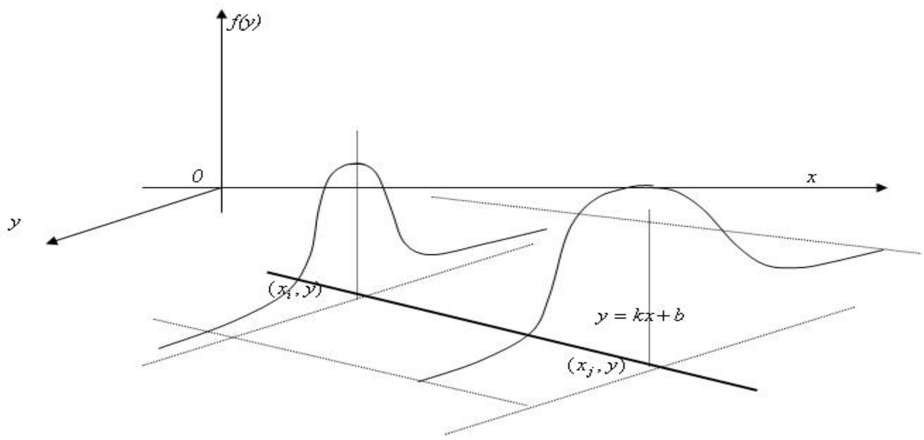

Рис.2.

За МНК маємо суму квадратів похибок

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i^* - \beta_0^* - \beta_1^* x_i \right)^2 . \tag{5}$$

Очевидно, що кожне конкретне значення e_i в наведеній сумі (5) має однакову, так би мовити, “питому вагу”, незалежно *від того, чи одержали його при значенні* $X = x_i$ *(де є мала дисперсія), чи при значенні* $X = x_j$ *(де наявна велика дисперсія),* що звичайно суперечить здоровому глузду, оскільки точка, одержана із розподілу $X = x_i$ точніше визначає напрямок (тенденцію) лінії регресії, ніж точка, одержана при $X = x_j$.

Тому, якщо поталанить врахувати "питому вагу" всіх точок e_i, то це дозволить одержати ефективніші (доброякісні) статистичні оцінки.

Ознаку гетероскедастичності в кожному конкретному випадку виявити складно, оскільки для цього необхідно знати величини σ_{ε_i} для кожного фіксованого значення $X = x_i$. На практиці, як правило, для кожного конкретного значення $X = x_i$ ми маємо в розпорядженні лише одне значення залежної змінної $Y = y_i$, а не цілий ряд розподілу. Це не дозволяє нам оцінити дисперсію випадкової величини Y при фіксованому значенні $X = x_i$.

Існують опробовані тести за допомогою яких можна виявити гетероскедастичність. І, як свідчить практика, їх використання дають позитивні наслідки. Такими є тести Глейзера і Гольдфельда-Квандта.

Тести Глейзера і Гольдфельдта-Квандта можуть виявити присутність ознаки гетероскедастичності в моделі лише у випадку порушення умови $\sigma^2_{\varepsilon_i} = const$, тобто, коли дисперсії залишків ε_i не є сталими величинами. Коли ж умова $\sigma^2_{\varepsilon_i} = const$ виконується, і при цьому $\operatorname{cov}(\varepsilon_i, \varepsilon_j) \neq 0$, то виявити це порушення умови гомоскедаститчності вищенаведеними тестеми неможливо.

У цьому випадку маємо справу із лінійною моделлю з порушенням ознаки гомоскедастичності, яка належить до другої групи. Така ситуація виникає при дослідженні моделей із ознакою автокореляції.

Розглянемо приклад.

Приклад. Розглянемо залежність між прибутком банків України (*Y*, млн.грн.) та величиною їх статутного фонду (*X*, млн.грн.) на прикладі вибіркових даних, приведених в таблиці 1.

Таблиця 1.

№ п/п	1	2	3	4	5	6	7	8	9	10
X	27,54	14,87	11,69	11,49	11,46	11,38	11,07	10,76	10,60	9,73
Y	100,00	102,67	75,00	85,00	70,09	60,00	54,00	57,35	52,30	52,00
№ п/п	11	12	13	14	15	16	17	18	19	20
X	9,48	8,44	8,43	8,19	7,04	6,89	6,04	5,26	5,24	4,39
Y	47,75	46,73	43,42	41,30	41,17	41,04	33,91	33,70	30,01	30,00
№ п/п	21	22	23	24	25	26	27	28	29	30
X	3,47	2,49	2,48	1,41	1,39	1,00	0,53	0,10	0,18	0,15
Y	26,82	24,64	23,37	23,82	22,26	20,50	15,50	14,20	13,54	13,41

Необхідно:

1) обчислити оцінки параметрів моделі за методом найменших квадратів, перевірити суттєвість зв'язку в моделі та статистичну значущість розрахованих оцінок параметрів моделі;

2) проаналізувати доцільність застосування методу найменших квадратів, перевіривши тестами Гольдфельда-Квандта та Глейзера наявність гетероскедастичності залишків;

3) розрахувати оцінки параметрів моделі з урахуванням результатів тестів на наявність гетероскедастичності залишків.

Розв'язання.

1) *Побудова моделі за МНК, перевірка моделі та її параметрів на статистичну значущість.*

Лінійна економетрична модель, оцінки параметрів якої знайдено за методом найменших квадратів для вибіркових даних (табл.1), має вигляд:

$$y_i^* = \text{-}2{,}366 + 0{,}219 x_i \,. \qquad (6)$$

Коефіцієнт детермінації цієї моделі R^2=0,854, між результативною змінною Y і регресором X коефіцієнт парної кореляції r_{xy}=0,924. Перевіримо статистичну значущість коефіцієнта парної кореляції за критерієм Ст'юдента:

$$t_{сп}^* = \sqrt{\frac{0{,}854 \cdot (30-1-1)}{1-0{,}854}} = 12{,}779 \,.$$

При рівні значущості α=0,05 та ступенях свободи $k=n\text{-}m\text{-}1$=28 табличне значення критерію Ст'юдента $t''_{кр}(\frac{\alpha}{2},k)$=2,048, отже, $t_{сп}^* > t''_{кр}(\frac{\alpha}{2},k)$, зв'язок в моделі між Y та X суттєвий. Стандартні помилки оцінок параметрів і відповідні їм значення t-критерію дорівнюють:

$$s_{\beta_0^*} = 0{,}848; \qquad s_{\beta_1^*} = 0{,}017;$$
$$t_{\beta_0^*} = -2{,}789; \qquad t_{\beta_1^*} = 12{,}779.$$

$t_{\beta_0^*}, t_{\beta_1^*} \notin [-2{,}048; 2{,}048]$, це дає змогу зробити висновок про суттєву відмінність від нуля розрахованих оцінок параметрів моделі.

2) *Перевірка на наявність гетероскедастичності залишків.*

Застосуємо тест Гольдфельда-Квандта для перевірки наявності гетероскедастичності залишків. Тестом перевіряється за критерієм Фішера основна гіпотеза H_0: $\sigma_{\varepsilon_1}^2 = \sigma_{\varepsilon 2}^2 = \ldots = \sigma_{\varepsilon_m}^2$ при альтернативній гіпотезі H_α: не H_0. Для цього упорядкуємо вхідні дані в порядку спадання значень пояснюючої змінної X (табл. 2).

Таблиця 2.

№ п/п	Модуль 1: $y_i^{\square} = -0{,}116 + 8{,}187 x_i$					Модуль 2 $y_i^{\square} = -3{,}246 + 0{,}234 x_i$				
	y_i	x_i	$y_i^{\square}$	e_i	e_i^2	y_i	x_i	$y_i^{\square}$	e_i	e_i^2

1	14,87	102,67	19,08	-4,21	17,688	4,39	30,00	3,78	0,61	0,375
2	27,54	100,00	18,58	8,96	80,343	3,47	26,82	3,03	0,44	0,191
3	11,49	85,00	15,77	-4,28	18,342	2,49	24,64	2,52	-0,03	0,001
4	11,69	75,00	13,90	-2,21	4,899	1,41	23,82	2,33	-0,92	0,848
5	11,46	10,09	12,99	-1,53	2,328	2,48	23,37	2,23	0,25	0,065
6	11,38	60,00	11,10	0,28	0,079	1,39	22,26	1,97	-0,58	0,331
7	10,76	57,35	10,60	0,16	0,024	1,00	20,50	1,55	-0,55	0,307
8	11,07	54,00	9,98	1,09	1,192	0,53	15,50	0,38	0,15	0,022
9	10,60	52,30	9,66	0,94	0,883	0,20	14,20	0,08	0,12	0,015
10	9,73	52,00	9,60	0,13	0,016	0,18	13,54	-0,08	0,26	0,065
11	9,48	47,75	8,81	0,67	0,449	0,15	13,41	-0,11	0,26	0,066
$\sum$					126,243					2.285

Відкинемо $c = \frac{4 \cdot n}{15} = 8$ $(i = \overline{12,19})$ середніх спостережень (друга підвибірка), вважаючи, що дисперсія залишків для них постійна. За методом найменших квадратів знайдемо оцінки параметрів спочатку для першої моделі (підвибірка 1) з $n_1 = \frac{n-c}{2} = \frac{30-8}{2} = 11$ найбільшими значеннями регресора *X*, потім для другої – з $n_3 = 11$ найменшими значеннями *X* (третя підвибірка).

Знайдемо значення y_i^* та e_i (табл. 2) для першої та другої моделей та перевіримо наявність гетероскедастичності залишків на основі тесту Гольдфельда-Квандта.

Для цього обчислимо суми квадратів залишків $S_1 = \sum_{i=1}^{n_1} e_i^2$ і $S_3 = \sum_{i=1}^{n_2} e_i^2$ та розрахуємо значення $F_{сп}^*$:

$$F_{сп}^* = \begin{cases} \dfrac{S_3/n_3}{S_1/n_1}, & якщо\ S_3 > S_1; \\ \dfrac{S_1/n_1}{S_3/n_3}, & якщо\ S_1 > S_3. \end{cases}$$

Розраховане значення $F_{сп}^* = \frac{126{,}243}{2{,}285} = 55{,}253$ порівняємо з табличним значенням $F_{кр}(\alpha, k_1, k_2)$ =3,18 із ступенями свободи k_1=11-1-1=9 і k_2=11-1-1=9 при рівні значущості α=0,05. Так як $F_{сп}^* > F_{кр}(\alpha, k_1, k_2)$, гіпотезу H_0 про відсутність гетероскедастичності залишків відхиляємо.

Проаналізуємо наявність гетероскедастичності за тестом Глейзера. Для моделі (5) обчислимо значення залишків та запишемо їх за абсолютною величиною (табл. 3).

Таблиця 3.

№п/п	1	2	3	4	5	6	7	8	9	10
e_i	7,96	-5,29	-2,40	-4,80	-1,55	0,58	1,59	0,54	1,49	0,69
$\lvert e_i \rvert$	7,96	5,29	2,40	4,80	1,55	0,58	1,59	0,54	1,49	0,69
№п/п	11	12	13	14	15	16	17	18	19	20
e_i	1,37	0,55	1,27	1,49	0,37	0,25	0,97	0,23	1,02	0,17
$\lvert e_i \rvert$	1,37	0,55	1,27	1,49	0,37	0,25	0,97	0,23	1,02	0,17
№п/п	21	22	23	24	25	26	27	28	29	30
e_i	-0,05	-0,55	-0,28	-1,45	-1,13	-1,13	-0,51	-0,55	-0,43	-0,43
$\lvert e_i \rvert$	0,05	0,55	0,28	1,45	1,13	1,13	0,51	0,55	0,43	0,43

Коефіцієнт детермінації моделі R^2=0,628, між $\lvert e_i \rvert$ та X коефіцієнт парної кореляції r=0,792. Стандартні помилки оцінок параметрів і відповідні їм значення t-критерію дорівнюють:

$$s_{\alpha_0^*} = 0{,}403; \qquad s_{\alpha_1^*} = 0{,}008;$$
$$t_{\alpha_0^*} = -2{,}604; \qquad t_{\alpha_1^*} = 6{,}874.$$

При рівні значущості α=0,05 і ступенях свободи $k=n-m-1$=28 табличне значення критерію Ст'юдента $t''_{кр}(\frac{\alpha}{2},k)$ $t_{\alpha/2,k}$=2,048. Порівняємо розраховані значення $t_{\alpha_j^*}$ з табличним:

$$t_{\alpha_0^*}, t_{\alpha_1^*} \notin [-2{,}048; 2{,}048].$$

Отже, приймаємо за тестом Глейзера гіпотезу про наявність гетероскедастичності залишків.

3) *Застосуємо зважений метод найменших квадратів для знаходження оцінок параметрів моделі з гетероскедастичними регресійними залишками.*

На основі тесту Гольдфельда-Квандта та Глейзера припускаємо, що $\sigma_{\varepsilon_i}^2 = \sigma^2 x_i^2$. Використовуючи УМНК, одержуємо

$$\vec{\beta}^* = \left(\left(X^*\right)' \cdot X^*\right)^{-1} \cdot \left(X^*\right)' \cdot \vec{y}^*, \quad \text{де } \vec{y}^* = \begin{pmatrix} \frac{y_1}{x_1} \\ \frac{y_2}{x_2} \\ \ldots \\ \frac{y_{30}}{x_{30}} \end{pmatrix}; \; X^* = \begin{pmatrix} 1 & \frac{1}{x_1} 1 \\ 1 & \frac{1}{x_2} \\ \ldots & \ldots \\ 1 & \frac{1}{x_{30}} \end{pmatrix}, \; \vec{\beta}^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \end{pmatrix}.$$

На основі даних таблиці 4, знаходимо, що $\vec{\beta}^* = \begin{pmatrix} -3{,}194 \\ 0{,}241 \end{pmatrix}$.

Таблиця 4.

№ п/п	1	2	3	4	5	6	7	8	9	10
$\frac{y_i}{x_i}$	0,2754	0,1448	0,1558	0,1352	0,1632	0,1897	0,2050	0,1877	0,2026	0,1870
$\frac{1}{x_i}$	0,0100	0,0097	0,0133	0,0118	0,0143	0,0167	0,0185	0,0174	0,0191	0,0192
№ п/п	11	12	13	14	15	16	17	18	19	20
$\frac{y_i}{x_i}$	0,1986	0,1806	0,1941	0,1984	0,1710	0,1678	0,1782	0,1561	0,1747	0,0333
$\frac{1}{x_i}$	0,0209	0,0214	0,0230	0,0242	0,0243	0,0244	0,0295	0,0297	0,0333	0,0333
№ п/п	21	22	23	24	25	26	27	28	29	30
$\frac{y_i}{x_i}$	0,1295	0,1011	0,1060	0,0592	0,0623	0,0487	0,0343	0,0143	0,0132	0,0110
$\frac{1}{x_i}$	0,0373	0,0406	0,0428	0,0420	0,0449	0,0488	0,0645	0,0704	0,0738	0,0746

Для визначення s_{β_0} та s_{β_1} використовується формула:

$$\operatorname{cov}(\vec{\beta}^* \cdot (\vec{\beta}^*)') = s_\varepsilon^2((X^*)' \cdot X^*)^{-1},$$

де $s_\varepsilon^2 = \dfrac{\vec{e}' \cdot \vec{e}}{n-m-1} = \dfrac{\sum_{i=1}^{30}(y_i - (-3{,}194 + 0{,}241 x_i))^2}{28} = \dfrac{0{,}02785}{28} = 0{,}00995$

та $((X^*)' \cdot X^*)^{-1} = \begin{pmatrix} 0{,}131 & -3{,}063 \\ -3{,}063 & 96{,}328 \end{pmatrix}$. Тут *n-m-1*=30-1-1=28.

Таким чином, одержано $s_{\beta_0^*} = 0{,}310$; $s_{\beta_1^*} = 0{,}011$.

Перевіримо статистичну значущість оцінок параметрів на основі *t*-критерію:

$t_{\beta_0^*} = 10{,}311$; $t_{\beta_1^*} = 21{,}146$.

При рівні значущості α=0,05 та ступенях свободи *k=n-m-1*=28 табличне значення критерію Ст'юдента $t''_{кр}(\frac{\alpha}{2}, k)$=2,048. $t_{\beta_0^*}, t_{\beta_1^*} \notin [-2{,}048; 2{,}048]$, розраховані оцінки параметрів моделі суттєво відрізняються від нуля.

В порівнянні з моделлю (6) середньоквадратичні помилки оцінок параметрів зменшились при відповідному збільшенні значення $t_{\beta_j^*}$.

Отже, економетрична модель з урахуванням гетероскедастичності залишків має наступний вигляд:

$$y_i^* = -3{,}194 + 0{,}241 x_i ,$$

тобто майже чверть величини статутного фонду визначає щорічний прибуток банку, що може бути характеристикою ефективності використання початкового (акціонерного) капіталу українських банків.

Висновки. Структура багатьох підручників побудована в такий спосіб, що студенти вчаться застосувати метод найменших квадратів для оцінювання параметрів економетричної моделі без урахування певних умов, які далеко не завжди виконуються на практиці для вихідної економічної інформації. Якщо ці умови порушуються, доводиться застосовувати інші методи оцінювання параметрів економетричної моделі, з якими студентів знайомлять майже на завершення курсу. Пропонуємо підхід, що дає підстави одночасно застосовувати метод найменших квадратів до вивчення класичного регресійного аналізу для побудови економетричної моделі, а далі розглядати економетричні задачі, що відповідають реальним економічним умовам, коли порушуються вихідні гіпотези регресійного аналізу.

Список використаної літератури

1. Гончаренко Я. В. Економетрія як наука і навчальна дисципліна в системі підготовки студентів // Матеріали доповідей звітно-наукової конференції викладачів УАГІ ВМУУ за 2010 рік : научное издание / М-во освіти і науки України, Нац. пед. ун-т ім. М. П. Драгоманова, Українсько-Американський гуманітарний ін-т "Вісконсінський Міжнар. ун-т (США) в Україні". - К. : Вид-во НПУ ім. М. П. Драгоманова, 2011. - С. 22-27 .
2. Гончаренко Я. В., Ляшко О. В. Сучасна економетрія як наука і навчальна дисципліна // Науковий часопис Національного педагогічного університету імені М. П. Драгоманова. Серія 3, Фізика і математика у вищій і середній школі : сборник / М-во освіти і науки України, Нац. пед. ун-т ім. М. П. Драгоманова. - К. : Вид-во НПУ ім. М. П. Драгоманова, 2009. - Вип. 5. - С. 56-62.
3. Гончаренко Я.В. Економетрія: методичні вказівки для виконання розрахункової роботи з економетрії для студентів спеціальності "Економічна теорія" / НПУ ім. М.П.Драгоманова. – К.: НПУ, 2005. – 82 с.
4. Грубер Й. Эконометрия. т.1. - К., 1997. - 422 с.
5. Лугінін О. Є., Білоусова С. В., Білоусов О. М. Економетрія: навчальний посібник для студ. вищ. навч. закладів. - К. : Центр навчальної літератури, 2005. - 252 с.
6. Лук'яненко І.Г., Краснікова Л.І. Економетрика: Практикум з використанням комп'ютера. - К.: "Знання", 1998.
7. Лукьяненко І. Г., Краснікова Л. І. Економетрика. - К.: «Знання», 1998ю - 345с.
8. Наконечний С.І.. Терещенко Т.О., Романюк Т.П. Економетрія. - К.: КНЕУ, 2000.