МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

ЕКОНОМЕТРИКА

Методичні рекомендації і завдання до самостійної роботи за темою "Проблеми в побудові лінійних множинних регресійних моделей: гетероскедастичність"

для студентів усіх спеціальностей першого (бакалаврського) рівня

Харків
ХНЕУ ім. С. Кузнеця
2019

УДК 519.862(07.034)
Е45

Укладачі: І. Л. Лебедєва
А. В. Жуков
С. С. Лебедєв

Затверджено на засіданні кафедри вищої математики й економіко-математичних методів.
Протокол № 9 від 25.04.2018 р.

Самостійне електронне текстове мережеве видання

Е45 **Економетрика** [Електронний ресурс] : методичні рекомендації і завдання до самостійної роботи за темою "Проблеми в побудові лінійних множинних регресійних моделей: гетероскедастичність" для студентів усіх спеціальностей першого (бакалаврського) рівня / уклад. І. Л. Лебедєва, А. В. Жуков, С. С. Лебедєв. – Харків : ХНЕУ ім. С. Кузнеця, 2019. – 33 с.

Подано основні теоретичні відомості за темою "Проблеми в побудові лінійних множинних регресійних моделей: гетероскедастичність", наведено приклад розв'язання типових завдань за допомогою комп'ютера із застосуванням програмного середовища MS Excel, наведено завдання для самостійної роботи та контрольні запитання для перевірки знань студентів.

Рекомендовано для студентів усіх спеціальностей першого (бакалаврського) рівня денної форми навчання.

УДК 519.862(07.034)

Вступ

Навчальна дисципліна "Економетрика" є базовою дисципліною циклу природничо-наукової та загальноекономічної підготовки студентів і вивчається згідно з навчальним планом підготовки фахівців з усіх спеціальностей першого (бакалаврського) рівня всіх форм навчання.

У системі підготовки фахівців з економіки та менеджменту економетрика займає одне з провідних місць. Вивчення цієї навчальної дисципліни дозволяє засвоїти загальні принципи побудови моделей економічних процесів і явищ, навчитися здійснювати перевірку адекватності отриманих моделей, використовувати ці моделі для прогнозування та оцінювання точності прогнозу. Отже, побудова економетричних моделей є фундаментальною основою методології управління економікою. Разом з іншими математичними дисциплінами економетрика озброює сучасного фахівця потужними методами дослідження як теоретичних, так і практичних питань, сприяє формуванню наукового мислення.

Новітньою тенденцією розвитку вищої школи, що спрямована на підвищення якості підготовки фахівців, є зростання ролі самостійної роботи студентів у процесі вивчення навчальної дисципліни. Самостійна робота сприяє розвитку творчої діяльності студента з придбання та засвоєння наукових знань, набуття ними практичних навичок. Оскільки вона здійснюється без безпосередньої участі викладача, виникає необхідність у її чіткій організації, яка б дозволяла здійснювати керівництво та контроль з боку викладача. Саме організація самостійної роботи студентів під час вивчення теми "Проблеми в побудові лінійних множинних регресійних моделей" і є метою цього навчального видання.

Запропоноване навчальне видання містить теоретичні відомості з теми "Проблеми в побудові лінійних множинних регресійних моделей", яка є однією з провідним тем економетрики, зокрема проблеми гетероскедастичності. Досконале вивчення цього питання має велике значення для побудови адекватних економетричних моделей як двофакторної, так і багатофакторної регресії. Запропоноване навчальне видання також містить завдання для самостійного виконання, розв'язання яких дозволяє на практиці закріпити набуті теоретичні знання. Також наведені зразки виконання навчальних прикладів.

Формування професійних компетентностей сучасного фахівця з економіки та менеджменту вимагає набуття знань і вмінь з різнобічних сфер професійної діяльності, що в подальшому стануть у нагоді під час виконання реальних завдань творчого рівня. Однією зі складових таких компетентностей є опанування сучасної обчислювальної техніки, що значно розширює можливості здійснення аналізу реальних економічних проблем, що досліджуються. Отже, обчислення, які необхідно здійснювати в процесі побудови економетричних моделей, доцільно виконувати на комп'ютері зі застосуванням відповідного програмного середовища. Одним із таких програмних середовищ є Microsoft Excel, який створено спеціально для роботи з електронними таблицями. MS Excel входить до складу Microsoft Office і підтримує всі необхідні функції для створення електронних таблиць будь-якої складності. Перевагою цього програмного продукту є те, що він дозволяє не просто реалізовувати алгоритм перевірки емпіричних даних щодо наявності гетероскедастичності, а детально розглядати особливості застосування методів її подолання. Для проведення обчислень доцільно застосовувати як вбудовані функції, так і надбудови MS Excel. У запропонованому вашій увазі навчальному виданні необхідні обчислення виконуються за допомогою програмного середовища MS Excel 2010.

Самостійна контрольна робота виконується за варіантами та подається на перевірку викладачеві у термін, визначений технологічною картою з навчальної дисципліни.

Для самостійної перевірки знань у запропонованому навчальному виданні наведені контрольні запитання з теми. Ці запитання також можуть використовуватись викладачем під час складання тестів з навчальної дисципліни.

1. Теоретичні відомості

1.1. Класична економетрична модель лінійної регресії

Лоуренс Р. Клейн, американський економіст, лауреат Нобелівської премії з економіки 1980 року "За створення економічних моделей і їхнє застосування до аналізу ваг розвитку економіки та економічної політики", зазначав, що завдання економетрики полягає в тому, щоб наповнювати емпіричним вмістом апріорні економічні міркування, тобто надавати кількісні оцінки тим висновкам, які були сформульовані в економічній теорії. Отже, основним завданням економетрики є побудова економіко-математичної моделі, яка на базі узагальнення результатів статистичних спостережень не тільки дає можливість визначити взаємний вплив сукупності факторів, що розглядаються в межах моделі, але і дозволяє здійснювати прогноз за цією моделлю.

Найбільш поширеним класом економетричних моделей є моделі з одним рівнянням регресії. ***Теоретичним рівнянням*** (або просто ***рівнянням***) ***регресії*** Y на X називається рівняння $Y_x = f(x)$. Відповідно, функція $f(x)$ називається ***регресією*** Y на X, а у випадку парної регресії графік цієї функції є ***лінією регресії*** випадкової величини Y на випадкову величину X. За цих умов X є незалежною (***пояснювальною***, або ***екзогенною***) змінною, тоді як Y – залежною (тобто тією, що пояснюється, або ***ендогенною***) змінною.

Множинною регресією називається кореляційна залежність кількох змінних, що визначається співвідношенням:

$$M(Y|x_1, x_2, \dots, x_m) = f(x_1, x_2, \dots, x_m), \tag{1}$$

де $M(Y|x_1, x_2, \dots, x_m)$ – умовне математичне сподівання випадкової величини Y, тобто математичне сподівання випадкової величини Y за умови, що випадкові величини $X_1, X_2, \dots, X_m$ набули, відповідно, такі значення: $x_1, x_2, \dots, x_m$.

Оскільки емпіричні значення внутрішньої змінної не завжди збігаються з її умовними математичними сподіваннями та можуть бути різними за одного і того самого значення зовнішньої змінної (або наборі зовнішніх змінних), фактична залежність повинна враховувати помилку ε, яка також є випадковою величиною.

Отже, кореляційний зв'язок між внутрішньою та зовнішніми змінними можна описати співвідношеннями:

$$Y = M(Y|x_1, x_2, \dots, x_m) + \varepsilon. \tag{2}$$

До того ж вважається, що випадкові помилки моделі розподілені за нормальним законом $\varepsilon \sim N(\mu;\ \sigma^2)$.

Побудова економетричної моделі передбачає визначення параметрів рівняння регресії за результатами статистичного дослідження вибіркових даних. Отримані значення є ***оцінками*** значень відповідних параметрів теоретичного рівняння. Кожна така оцінка є ***точковою***, оскільки вона характеризується одним значенням. Півширина довірчого інтервалу, до якого з певною надійністю належатиме статистична оцінка параметра моделі, визначає ***інтервальну оцінку*** цього параметру.

Статистичні оцінки параметрів мають відповідати таким вимогам:

1) бути ***незсунутими***, тобто математичне сподівання оцінки θ^* дорівнює значенню параметра θ, який підлягає оцінюванню:

$$M(\theta^*) = \theta; \tag{3}$$

2) бути ***ефективними***, тобто така оцінка повинна мати найменшу дисперсію серед усіх можливих оцінок даного параметра за вибірками однакового обсягу:

$$M(\theta^*_{\text{еф}} - \theta)^2 = \sigma^2_{\text{еф}} = \min_k \sigma_k^2, \tag{4}$$

де $\theta^*_{\text{еф}}$ – ефективна оцінка параметра θ;

$\sigma^2_{\text{еф}}$ – дисперсія ефективної оцінки параметра θ;

σ_k^2 – дисперсія оцінки параметра θ, яка була отримана із застосуванням k-го методу;

3) бути ***спроможними***, тобто із збільшенням кількості випробувань оцінка θ^* має прямувати до значення параметра θ у такій генеральній сукупності:

$$\theta^* \xrightarrow{n \to \infty} \theta. \tag{5}$$

Поряд із помилками вимірювань і обмеженістю вибіркових даних найбільш істотними причинами присутності в регресійних моделях випадкових помилок є невірна специфікація моделі, тобто відсутність у моделі всіх пояснювальних факторів, помилка у виборі функції для апроксимації емпіричних даних або наявність зв'язку між факторами, які вважаються незалежними.

Найбільш поширеним методом оцінювання невідомих параметрів регресійних моделей за вибірковими даними є (***звичайний***) ***метод найменших квадратів*** (***МНК***, або ***OLS*** – ***ordinary least squares***). Сутність методу полягає в тому, що параметри функції $f(x_1, x_2, \dots, x_k)$, за допомогою якої здійснюють апроксимацію, визначають відповідно до умови мінімізації суми квадратів похибок моделі, а саме суми квадратів різниць між емпіричними значеннями внутрішнього фактору та тими, що обчислюються за економетричною моделлю:

$$F(\theta^*) = \sum_{i=1}^{n} {\varepsilon_i}^2 = \sum_{i=1}^{n} (y_i - \widehat{y_i})^2 \rightarrow min, \tag{6}$$

де y_i – значення внутрішнього фактору, яке було отримане в i-му вимірюванні;

$\widehat{y_i}$ – розрахункове значення внутрішнього фактору, яке обчислено за економетричною моделлю у разі підстановки в цю модель значень зовнішніх факторів, що відповідають i-му вимірюванню;

n – обсяг вибіркової сукупності;

$F(\theta^*)$ – сума квадратів помилок моделі, яка є функцією від оцінок параметрів моделі.

МНК-оцінки будуть незсунутими, ефективними та спроможними за умови виконання певних вимог, які називаються ***передумовами МНК:***

1) математичне сподівання випадкових помилок моделі має дорівнювати нулю:

$$M(\varepsilon_i) = 0; \tag{7}$$

2) дисперсія випадкових помилок моделі повинна бути сталою для всіх спостережень, тобто $var(\varepsilon_i) = \sigma_\varepsilon^2$. З урахуванням співвідношення (7) це означає, що:

$$M(\varepsilon_i^2) = 0, \tag{8}$$

отже, передбачається, що всі спостереження є рівноточними;

3) систематичний зв'язок між значеннями помилок моделі у будь-яких двох спостереженнях повинен бути відсутнім, тобто $cov(\varepsilon_i;\ \varepsilon_j) = 0$, якщо $i \neq j$. З урахуванням (7) це означає:

$$M(\varepsilon_i \cdot \varepsilon_j) = 0 \text{ для усіх } i \neq j; \tag{9}$$

4) систематичний зв'язок між значеннями помилок моделі та значенням будь-якої зовнішньої змінної повинен бути відсутнім, тобто $cov(\varepsilon_i \cdot x_i) = 0$. З урахуванням (7) це означає:

$$M(\varepsilon_i \cdot x_i) = 0 \text{ для усіх } i. \tag{10}$$

Згідно з ***теоремою Гаусса – Маркова*** у припущенні лінійної форми зв'язку між внутрішньою Y і зовнішніми $X_1, X_2, \ldots, X_m$ змінними, що описується рівнянням регресії:

$$y_i = b_0 + b_1 \cdot x_{i1} + b_2 \cdot x_{i2} + \cdots + b_m \cdot x_{im} + \varepsilon_i, \tag{11}$$

за умов, що всі зовнішні змінні $X_1, X_2, \ldots, X_m$ є попарно незалежними, та відносно помилок моделі справедливими є твердження (7) – (10), оцінки параметрів множинної регресії, що отримані в результаті застосування методу найменших квадратів, є найбільш ефективними, тобто мають найменшу дисперсією у класі лінійних незсунутих оцінок (***Best Linear Unbiased Estimator*** – ***BLUE***).

Співвідношення (11) також можна надати у матричній формі:

$$\mathbf{Y} = \mathbf{X} \cdot \mathbf{B} + \boldsymbol{\varepsilon}, \tag{12}$$

де $\mathbf{Y}$ – матриця-стовпець значень внутрішнього фактору розміром n x 1; $\mathbf{X}$ – матриця значень зовнішніх факторів, елементами x_{ij} якої є значення випадкової величини X_j $(j = \overrightarrow{1,m})$ у вимірюванні під номером i $(i = \overrightarrow{1,n})$, до якої зліва приписано стовпець з самих одиниць.

Матриця значень зовнішніх факторів має такий вигляд:

$$\mathbf{X} = \begin{vmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{vmatrix}. \tag{13}$$

Оскільки матриця $\mathbf{X}$ окрім значень незалежних випадкових величин містить додатковий стовпець, то її розмір дорівнює n x $(m+1)$.

Лінійна модель регресії, для якої виконуються умови теореми Гаусса – Маркова, називається ***класичною*** економетричною моделлю лінійної регресії.

Матриця оцінок $\widehat{\mathbf{B}}$ параметрів такої моделі визначається співвідношенням:

$$\widehat{\mathbf{B}} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot (\mathbf{X}^T \cdot \mathbf{Y}), \tag{14}$$

де індекс $\mathbf{X}^T$ означає, що матриця $\mathbf{X}^T$ є транспонованою, тобто отримана із вихідної матриці $\mathbf{X}$ шляхом заміни її рядків на стовпці з відповідним номером;

індекс $^{-1}$ означає, що матриця $(\mathbf{X}^T \cdot \mathbf{X})^{-1}$ є оберненою до добутку матриць $\mathbf{X}^T \cdot \mathbf{X}$ (зауважимо, що для існування оберненої матриці необхідно, щоб визначник вихідної матриці не дорівнював нулю, однак ця умова порушується, якщо між зовнішніми факторами існують функціональний або кореляційний зв'язки).

1.2. Порушення умов теореми Гаусса – Маркова: гетероскедастичність залишків моделі

Однією з умов, за якою МНК-оцінки параметрів рівняння регресії будуть BLUE, є сталість дисперсії випадкових помилок моделі (8). Така властивість випадкових помилок називається ***гомоскедастичністю***. У протилежному випадку, тобто коли ця умова порушується та дисперсія залишків моделі не є сталою, має місце ***гетероскедастичність*** випадкових помилок моделі.

Будемо розглядати випадок, коли залишки моделі розподілені за нормальним законом, мають математичне сподівання, яке дорівнює нулю, вони некорельовані між собою, але їхня дисперсія змінюється зі зміною одного або кількох зовнішніх факторів. За цих умов зовнішні фактори є попарно незалежними. Тобто порушується лише одна з умов теореми Гаусса – Маркова. Якщо в цьому випадку застосовувати метод найменших квадратів, то отримані МНК-оцінки параметрів моделі будуть неефективними, тобто вони вже не матимуть найменшу дисперсію серед незсунутих оцінок, як це було за умови гомоскедастичності.

Причин гетероскедастичності може бути декілька. Наприклад, гетероскедастичність може виникати, коли значення результатів спостережень суттєво різняться за величиною у різних вимірюванням. Достатньо часто така проблема виникає, коли одним із зовнішніх факторів є час, тобто під час побудови економетричних моделей рядів динаміки. Отже, є необхідність перевірки властивостей дисперсії залишків моделі.

Найбільш простим способом перевірки виконання вимоги гомоскедастичність випадкових залишків є візуальне дослідження графіка залишків. Для його проведення необхідно побудувати графік залежності величин випадкових залишків моделі, параметри якої були визначені за однокроковим методом найменших квадратів, від вирівняного значення зовнішньою змінної. У тому разі, коли має місце гомоскедастичність, "хмара" випадкових залишків має вигляд смуги, яка паралельна осі абсцис. Усі інші випадки розташування цієї "хмари" відповідають гетероскедастичності випадкових залишків моделі. Однак цей метод не дозволяє визначити, на скільки суттєвою є гетероскедастичність і чи треба її враховувати під час побудови економетричної моделі.

Аналітично перевірка на відсутність гетероскедастичності може здійснюватися за спеціальними тестами. Найбільш поширеними серед них є такі тести, як тест рангової кореляції Спірмена (Spearman rank correlation test), тест Голдфелда – Квандта (Goldfeld – Quandt test), тест Уайта (White test), тест Глейзера (Gleizer test), тест Бреуша – Пагапа (Breusch – Pagan test). Під час застосування будь-якого з цих тестів перевірці підлягає основна гіпотеза H_0: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \cdots = \sigma_n^2$ за альтернативної H_1: $\sigma_1^2 \neq \sigma_2^2 \neq \sigma_3^2 \neq \cdots \neq \sigma_n^2$.

Вибір того чи іншого тесту на гетероскедастичність залежить від припущення щодо її причин і, відповідно, щодо змінної, з якою корелює дисперсія помилок. Розглянемо найбільш поширені з цих тестів.

Тест рангової кореляції Спірмена використовується в тому випадку, коли дослідник передбачає, що дисперсія випадкового члена рівняння регресії буде або збільшуватися, або зменшуватися зі зростанням значень певної зовнішньої змінної, тому в рівнянні регресії, параметри якого оцінюється за методом 1МНК, абсолютні величини залишків і значення цієї змінної будуть корельованими. Алгоритм проведення цього тесту передбачає виконання таких кроків:

1) за методом 1МНК здійснюють побудову економетричних моделей парної регресії відносно внутрішньої змінної (Y) і кожної зі зовнішніх змінних ($X_j,\ j = \overrightarrow{1,m}$);

2) для кожної з побудованих моделей обчислюють залишки (ε_j);

3) для кожної моделі здійснюється ранжування окремо значень зовнішньої змінної (X_j) та значень залишків цієї моделі (ε_{ij}) за модулем від меншого до більшого, після чого замінюють значення їхніми рангами;

4) розраховують коефіцієнт рангової кореляції Спірмена за формулою:

$$r_{X_j \varepsilon_j} = 1 - 6\frac{\sum_{i=1}^{n} d_{ij}^2}{n(n^2-1)}, \tag{15}$$

де d_{ij} – різниця між рангом змінною X_j та рангом модуля випадкових залишків моделі ε_j в i-му вимірюванні ($i = \overrightarrow{1,n}$);

5) для перевірки рівня статистичної значущості коефіцієнта кореляції Спірмена обчислюють емпіричне значення статистики Стьюдента за формулою:

$$t = \frac{|r_{X_j \varepsilon_j}| \cdot \sqrt{n-2}}{\sqrt{1-r_{X_j \varepsilon_j}^2}}; \tag{16}$$

6) за довідковою таблицею визначають критичне значення статистики Стьюдента для рівня значущості $\alpha = 0{,}05$ та кількості ступенів вільності $df = n - 2$;

7) порівнюють емпіричне та критичне значення статистики Стьюдента. Якщо $t_{\text{емп.}} < t_{0{,}05}$, то нульову гіпотезу про відсутність гетероскедастичності нема підстав відкинути, тобто оцінки параметрів рівняння регресії, які побудовані за однокроковим методом найменших квадратів, будуть ефективними. Якщо $t_{\text{емп.}} > t_{0{,}05}$, то з надійністю 95 % нульову гіпотезу слід відкинути, тобто гетероскедастичність залишків моделі є статистично значущою.

Слід зазначити, що на відміну від коефіцієнта кореляції Пірсона застосування коефіцієнта рангової кореляції не пов'язане з передумовою нормальності розподілу вихідних даних. Як і для коефіцієнта кореляції Пірсона, значення коефіцієнта рангової кореляції Спірмена належать відрізку [−1; 1]. Якщо коефіцієнт рангової кореляції дорівнює 1, то між елементами двох послідовностей є повна узгодженість, тобто елементи, що відповідають одному вимірюванню, займають однакові місця в обох рядах. Якщо коефіцієнт рангової кореляції дорівнює −1, то елементи двох послідовностей розташовані в зворотному порядку, тобто має місце повна негативна кореляція рангів. Якщо коефіцієнт кореляції Спірмена дорівнює нулю, то кореляції між рангами відсутня.

Тест Голдфелда – Квандта є найбільш популярним формальним критерієм, який застосовується для перевірки гіпотези про відсутність гетероскедастичності. Як і у випадку тесту рангової кореляції Спірмена, тест Голдфелда – Квандта застосовують тоді, коли випадкові помилки моделі змінюються пропорційно значенням незалежної випадкової змінної X_j. Цей тест не вимагає наявності великого обсягу спостережень. Алгоритм його проведення передбачає такі кроки:

1) результати вимірювань упорядковують відповідно до величини компонентів тієї зі зовнішніх змінних ($X_j,\ j = \overline{1,m}$), яка, за припущенням, впливає на зміну дисперсії випадкових залишків;

2) розбивають впорядковану вибіркову сукупність на три підгрупи у такий спосіб, що ті підгрупи, які розташовані на кінцях ряду, є рівними за обсягом (кількість вимірювань у кожній з цих підгруп дорівнює k) і трохи більшими за центральну ($k > {}^{n}/_{3}$). Ті спостереження, що містяться всередині ряду, відкидають. Для вибіркової сукупності, обсяг якої становить 30 – 60 вимірювань, оптимальна кількість спостережень, які треба відкинути, дорівнює 8 – 16. Зауважимо, що якщо вибірка мала за обсягом, то її можна ділити навпіл і в подальших дослідженнях застосовувати обидві її частини;

3) за тими спостереженнями, що були залишені для дослідження, за допомогою однокрокового методу найменших квадратів будують дві економетричні моделі регресії тієї ж структури, що була побудована для всієї сукупності в цілому, і обчислюють суму квадратів помилок (***ESS***, тобто ***Error Sum of Squares***) за цими двома моделями. У тому випадку, коли вибірка була поділена навпіл, маємо:

$$ESS_1 = \sum_{i=1}^{k}(y_i - \hat{\tilde{y}}_i)^2 \quad \text{та} \quad ESS_2 = \sum_{i=k+1}^{n}(y_i - \hat{\tilde{y}}_i)^2; \qquad (17)$$

4) за даними щодо суми квадратів залишків моделей обох підгруп розраховують емпіричне значення F-критерію Фішера за формулою:

$$F = \frac{ESS_2}{ESS_1}, \text{ якщо } \quad ESS_2 > ESS_1, \qquad (18)$$

або

$$F = \frac{ESS_1}{ESS_2}, \text{ якщо } \quad ESS_1 > ESS_2; \qquad (18^{\prime})$$

5) за довідковою таблицею визначають критичне значення статистики Фішера для рівня значущості $\alpha = 0{,}05$ та кількості ступенів вільності чисельника і знаменника $df_1 = df_2 = k - m - 1$;

6) порівнюють емпіричне та критичне значення статистики Фішера. Якщо має місце співвідношення $F_{\text{емп.}} < F_{0{,}05}$, то нульову гіпотезу про відсутність гетероскедастичності нема підстав відкинути, тобто оцінки параметрів рівняння регресії, які побудовані за однокроковим методом найменших квадратів, будуть ефективними. Якщо $F_{\text{емп.}} > F_{0{,}05}$, то з надійністю 95 % нульову гіпотезу відкидаємо, тобто гетероскедастичність залишків моделі є статистично значущою.

У тому випадку, коли гетероскедастичність залишків моделі є статистично значущою, замість однокрокового методу найменших квадратів необхідно застосовувати інший метод, який би забезпечив ефективність оцінок параметрів моделі. Таким методом побудови економетричної моделі є ***узагальнений метод найменших квадратів*** (***УМНК***, або ***GLS*** – ***generalized least squares***).

1.3. Використання узагальненого методу найменших квадратів для оцінювання параметрів регресійної моделі у разі гетероскедастичності

Узагальнений метод найменших квадратів дозволяє з єдиної позиції розглядати деякі важливі класи регресійних моделей, зокрема моделі з гетероскедастичністю. Ефективність оцінок для параметрів регресійної моделі (12) доведена ***теоремою Айткена*** (Aitken's theorem), яка стверджує, що найменшу матрицю коваріацій у класі лінійних незсунутих оцінок параметрів узагальненої регресійної моделі має оцінка:

$$\widehat{\mathbf{B}} = (\mathbf{X}^T \cdot \mathbf{\Omega}^{-1} \cdot \mathbf{X})^{-1} \cdot (\mathbf{X}^T \cdot \mathbf{\Omega}^{-1} \cdot \mathbf{Y}), \quad (19)$$

де $\mathbf{\Omega}$ – ***матриця коваріацій помилок моделі*** (***covariance matrix of the errors of a linear regression***).

Зауважимо, що коли гетероскедастичність відсутня, то коваріаційна матриця помилок моделі набуває вигляду $\sigma^2 \cdot \mathbf{I}$ (за умови, що $\sigma^2 = const$), де $\mathbf{I}$ є одиничною матрицею розміром n x n.

Слід зазначити, що узагальнений метод найменших квадратів має широке застосування і використовується не тільки для подолання гетероскедастичності, але і в тих випадках, коли існує таке порушення умови теореми Гаусса – Маркова, як автокореляція залишків моделі.

Для побудови економетричної моделі за допомогою УМНК необхідно знати матрицю коваріацій помилок моделі, але на практиці це буває дуже рідко. Тому доцільно будь-яким способом оцінити матрицю коваріацій помилок моделі і саме цю оцінку застосовувати у співвідношенні (19) замість матриці $\boldsymbol{\Omega}$. Такий підхід покладено в основу ***доступного узагальненого методу найменших квадратів*** (***Feasible Generalized Least Square***).

Загальною вимогою щодо матриці $\boldsymbol{\Omega}$ є її додатна визначеність. Оскільки матриця $\boldsymbol{\Omega}$ є невиродженою симетричною матрицею, то її можна надати за допомогою деякої невиродженої матриці $\mathbf{P}$ у такий спосіб:

$$\boldsymbol{\Omega} = \mathbf{P} \cdot \mathbf{P}^{\boldsymbol{T}}. \tag{20}$$

Якщо застосувати цю матрицю до зовнішніх і внутрішнього факторів економетричної моделі та перейти до нових змінних:

$$\mathbf{X}^* = \mathbf{P}^{-1} \cdot \mathbf{X}, \quad \mathbf{Y}^* = \mathbf{P}^{-1} \cdot \mathbf{Y}, \tag{21}$$

то відносно цих змінних оцінки параметрів регресійної моделі $\widehat{\mathbf{B}}^*$, що отримані за методом 1МНК:

$$\widehat{\mathbf{B}}^* = (\mathbf{X}^{*T} \cdot \mathbf{X}^*)^{-1} \cdot (\mathbf{X}^{*T} \cdot \mathbf{Y}^*), \tag{22}$$

будуть BLUE. Інакше кажучи, якщо перетворити фактори економетричної моделі за формулами (21), то застосування до них однокрокового МНК приведе до того ж результату, що і застосування УМНК до вихідних змінним.

Якщо залишки моделі є лише гетероскедастичними, а автокореляція залишків відсутня, то матрицю коваріацій помилок моделі доцільно обирати у такому вигляді:

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}. \tag{23}$$

У цій матриці як оцінки дисперсій випадкових залишків моделі (елементів головної діагоналі матриці) можуть бути використані квадрати залишків, що отримані під час застосування однокрокового МНК.

Відповідно:

$$\mathbf{\Omega^{-1}} = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n^2 \end{pmatrix}, \tag{24}$$

тоді

$$\mathbf{P^{-1}} = \begin{pmatrix} 1/\sigma_1 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\sigma_n \end{pmatrix}. \tag{25}$$

Отже, фактично ми змінюємо специфікацію моделі, переходячи до нових змінних, значення яких у кожному вимірюванні обчислюються як $y_i^* = {}^{y_i}/_{\sigma_i}$ та $x_{ij}^* = {}^{x_{ij}}/_{\sigma_i}$ $(i = \overline{1,n})$. У разу застосування цих змінних для побудови моделі випадкова помилка в i-му вимірюванні буде дорівнювати $\varepsilon_i^* = {}^{\varepsilon_i}/_{\sigma_i}$ $(i = \overline{1,n})$, відповідно, дисперсія помилок завдяки застосуванню матриці коваріацій помилок у вигляді (24) є сталою величиною, тобто має місце гомоскедастичність залишків моделі. Отже, отримані МНК-оцінки параметрів моделі стають BLUE.

Існує кілька підходів до побудови матриці $\mathbf{\Omega}$ для подолання гетероскедастичності. Наприклад, вище ми вже зазначали, що у деяких випадках можна припускати, що стандартні відхилення помилок є прямо пропорційними певному зовнішньому фактору X_k.

Процедуру визначення оцінок параметрів моделі за допомогою УМНК можна зробити значно простішою, якщо припустити, що дисперсію помилок моделі можна задати таким співвідношенням:

$$\sigma_i^2 = \sigma^2 \cdot x_{kj}^2, \quad i = \overline{1,n} \tag{26}$$

де σ_i^2 – дисперсія помилки в i-му вимірюванні.

У цьому випадку коваріаційна матриця випадкових помилок моделі набуває вигляду:

$$\mathbf{\Omega} = \sigma^2 \begin{pmatrix} x_{1k}^2 & 0 & \cdots & 0 \\ 0 & x_{2k}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{nk}^2 \end{pmatrix}. \tag{27}$$

Відповідно, для переходу до нових змінних, що визначаються формулою (21), застосовують таку матрицю:

$$\mathbf{P^{-1}} = \begin{pmatrix} 1/x_{1k} & 0 & \cdots & 0 \\ 0 & 1/x_{2k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/x_{nk} \end{pmatrix}. \qquad (28)$$

У такий спосіб ми змінюємо специфікацію моделі, переходячи до нових змінних, значення яких у кожному вимірюванні визначаються співвідношеннями: $y_i^* = {}^{y_i}/_{x_{ik}}$ та $x_{ij}^* = {}^{x_{ij}}/_{x_{ik}}$ $(i = \overrightarrow{1,n})$. Завдяки цьому отримані МНК-оцінки параметрів моделі є BLUE.

Також може мати місце таке співвідношення для дисперсія помилок: $\sigma_i^2 = \sigma^2 \cdot x_{ik}$ $(i = \overrightarrow{1,n})$. У цьому випадку для переходу до нових змінних застосовується ваговий коефіцієнт $1/\sqrt{x_{ik}}$, і у такій специфікації оцінювання параметрів регресійної моделі здійснюється за методом 1МНК.

2. Приклад побудови регресійної моделі за наявності гетероскедастичності залишків

Розглянемо залежність витрат на освіту від обсягу валового внутрішнього продукту різних країн (табл. 1, стовпці 2 – 4).

Таблиця 1

Залежність витрат на освіту від обсягу ВВП країни*

№ п/п	Країна	ВВП (x_i), млрд USD	Витрати на освіту (y_i), млрд USD	$\widetilde{y}_i$	$e_i = \lvert y_i - \widetilde{y}_i \rvert$	Ранг помилок моделі
1	2	3	4	5	6	7
1	Люксембург	5,67	0,34	−1,93	2,27	17
2	Уругвай	10,13	0,22	−1,63	1,85	14
3	Сінгапур	11,34	0,32	−1,55	1,87	15
4	Ірландія	15,88	1,23	−1,25	2,48	20
5	Ізраїль	20,94	1,81	−0,91	2,72	21
6	Угорщина	22,16	1,02	−0,83	1,85	13
7	Нова Зеландія	23,83	1,27	−0,71	1,98	16

Закінчення табл. 1

1	2	3	4	5	6	7
8	Португалія	24,67	1,07	−0,66	1,73	12
9	Гонконг	27,56	0,67	−0,47	1,14	5
10	Чилі	27,57	1,25	−0,46	1,71	11
11	Греція	40,15	0,75	0,38	0,37	1
12	Фінляндія	51,62	2,80	1,14	1,66	10
13	Норвегія	57,71	4,90	1,55	3,35	22
14	Югославія	63,03	3,5	1,91	1,59	9
15	Данія	66,32	4,45	2,13	2,32	18
16	Туреччина	66,97	1,60	2,17	0,57	2
17	Австрія	76,88	4,26	2,83	1,43	6
18	Швейцарія	101,65	5,31	4,49	0,82	3
19	Саудівська Аравія	115,97	6,40	5,45	0,95	4
20	Бельгія	119,49	7,15	5,69	1,46	7
21	Швеція	124,15	11,22	6,00	5,22	27
22	Австралія	140,98	8,66	7,12	1,54	8
23	Аргентина	153,85	5,56	7,98	2,42	19
24	Нідерланди	169,38	13,41	9,02	4,39	25
25	Мексика	186,33	5,46	10,16	4,70	26
26	Іспанія	211,78	4,79	11,86	7,07	30
27	Бразилія	249,75	8,92	14,40	5,48	28
28	Канада	261,41	18,90	15,18	3,72	24
29	Італія	395,52	15,95	24,15	8,20	32
30	Велика Британія	534,97	29,90	33,48	3,58	23
31	Франція	655,29	33,59	41,53	7,94	31
32	Німеччина	815,00	38,92	52,22	13,30	34
33	Японія	1 040,45	61,61	67,30	5,69	29
34	США	2 586,40	181,30	170,74	10,56	33

* *Джерело вихідних даних*: Доугерті.

Розташування країн у табл. 1 наведено у порядку зростання їхнього ВВП, який у цій моделі є зовнішнім (екзогенним) фактором X, а витрати на освіту є внутрішнім (ендогенним) фактором Y. Для проведення розрахунків скористаємося надбудовами та вбудованими функціями MS Excel, тому дані табл. 1 перенесемо на робочий аркуш MS Excel.

Для проведення дослідження необхідно побудувати модель парної регресії. Спочатку за виглядом графіка, побудованого за емпіричними даними (рис. 1), виберемо функцію апроксимації. Для побудови графіка вибираємо шлях: *Вставка* → *Диаграммы* → *Точечная*.

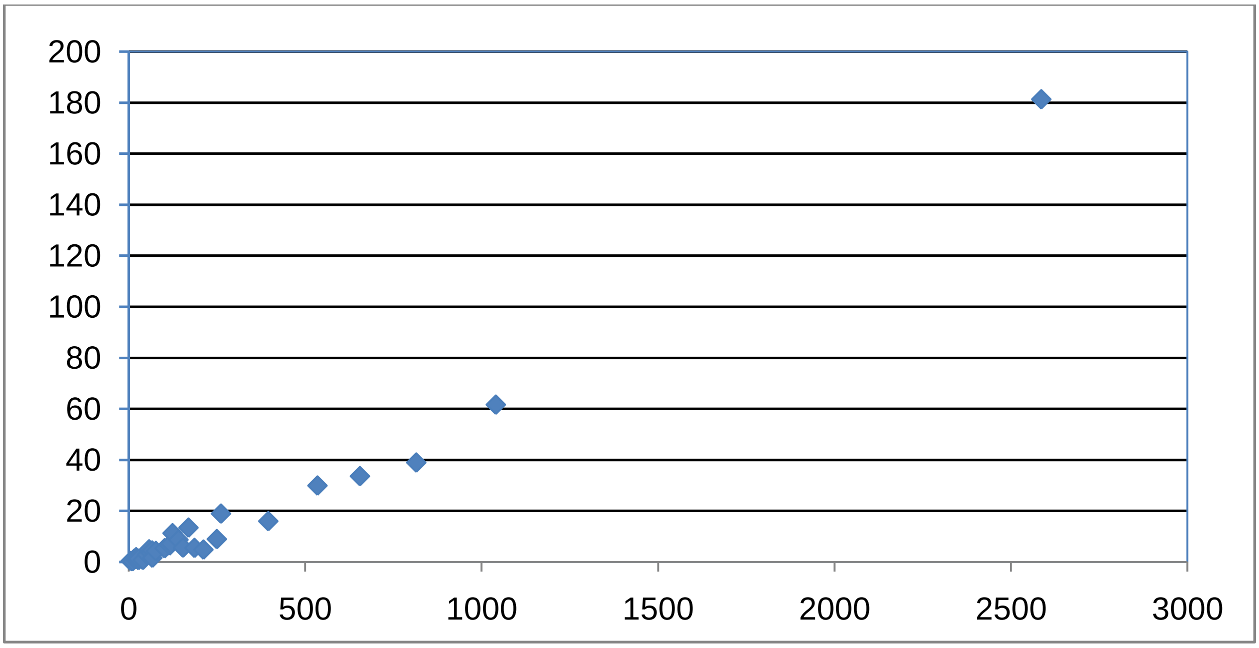

Рис. 1. **Залежність витрат на освіту (млрд USD) від ВВП (млрд USD) країни**

За виглядом графіка можна припустити, що функція апроксимації є лінійної. Побудуємо модель парної регресії за допомогою 1МНК. Для цього скористаємося вбудованою функцією ЛИНЕЙН(). Результати її застосування надані в табл. 2.

Таблиця 2

Результати функції ЛИНЕЙН() для всієї вибірки

b_1 = 0,0669	b_0 = −2,3090
σ_{b_1} = 0,0017	σ_{b_0} = 0,9092
R^2 = 0,9797	σ_e = 4,6872
F = 1 540,83	df_e = 32
RSS = 33 852,49	ESS = 703,0492

Отже, ми отримали таке рівняння регресії:

$$\tilde{y} = -2{,}3090 + 0{,}0669 \cdot x. \quad (29)$$

Модель є значущою в цілому за критерієм Фішера, оскільки маємо $F_{\text{емп.}} = 1540{,}83 > F_{0,05}(2;32) = 3{,}19$. Критичне значення статистики Фішера знаходимо за допомогою вбудованої функції F.ОБР.ПХ(). Також за критерієм Стьюдента обидва параметри моделі є значущими, оскільки емпіричні значення t-критерію дорівнюють, відповідно, $t_{b_1} = 39{,}25$ та $t_{b_0} = 2{,}54$, тоді як критичне значення становить $t_{0,05}(32) = 2{,}04$. Його знаходимо за допомогою вбудованої функції СТЬЮДЕНТ.ОБР.2Х(). За коефіцієнтом детермінації мінливість зовнішнього фактору в межах отриманої моделі зумовлює 97,97 % мінливості внутрішнього фактору.

Перевіримо, чи є оцінки, які ми отримали за однокроковим методом найменших квадратів, ефективними, тобто чи виконуються умови теореми Гаусса – Маркова щодо гомоскедастичності залишків моделі. Для цього за рівнянням регресії (29) знайдемо теоретичні значення внутрішнього фактору (стовпець 5 табл. 1) й обчислимо абсолютні значення залишків моделі, застосувавши для цього вбудовану функцію ABS(). Результати розрахунків виведені у стовпці 6 табл. 1. Графік абсолютних величин залишків моделі наведено на рис. 2.

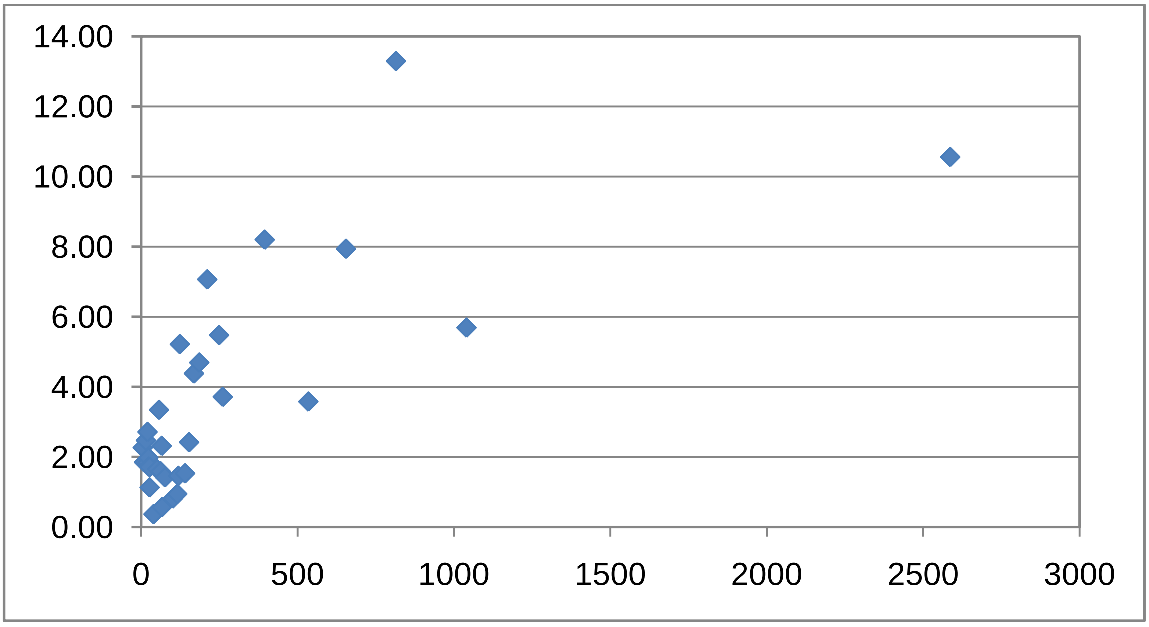

Рис. 2. **Залежність модуля залишків моделі (29) від значення зовнішнього фактору**

З рис. 2 видно, що із збільшенням значення зовнішнього фактору дисперсія залишків моделі, що була побудована за допомогою 1МНК, збільшується, тобто може мати місце гетероскедастичність залишків моделі. Перевіримо це за тестом рангової кореляції Спірмена. Оскільки у табл. 1 країни розташовані у порядку зростання їхнього ВВП (зовнішній фактор), то порядковій номер у цій таблиці відповідає рангу за X. Для визначення рангу модулів залишків моделі застосуємо вбудовану функцію РАНГ.СР(), яка повертає ранг числа у списку чисел (його порядковий номер відносно інших чисел списку). Якщо декілька чисел мають однакові ранги, то функція повертає їхній середній ранг. Функція дозволяє здійснювати ранжування як у порядку зростання, так і спадання значень. Оскільки ми розглядаємо ранги в порядку зростання, це треба вказати у рядку *Порядок* діалогового вікна функції, поставивши там 1. Результати ранжування залишків моделі наведені у стовпці 7 табл. 1.

Обчислюємо за формулою (15) коефіцієнт рангової кореляції Спірмена:

$$r_{X\varepsilon} = 1 - 6 \cdot \frac{2736}{34 \cdot (34^2 - 1)} = 0{,}5820.$$

Тепер за формулою (16) визначаємо емпіричне значення статистики Стьюдента:

$$t = \frac{0{,}5820 \cdot \sqrt{34 - 2}}{\sqrt{1 - 0{,}5820^2}} = 4{,}0.$$

Застосувавши вбудовану функцію СТЬЮДЕНТ.ОБР.2Х(), визначаємо критичне значення статистики Стьюдента. Воно дорівнює $t_{0,05}(32) = 2{,}04$. Оскільки $t_{\text{емп.}} > t_{0,05}$, то нульову гіпотезу про гомоскедастичність залишків моделі відкидаємо на користь альтернативної. Отже, тест рангової кореляції Спірмена свідчить, що гетероскедастичність залишків моделі, яка була побудована за 1МНК, є статистично значущою.

Перевіримо нульову гіпотезу про гомоскедастичність залишків моделі за допомогою тесту Голдфелда – Квандта. Для цього об'єднаємо країни у підгрупи за обсягом ВВП. До першої підгрупи беремо 12 країн, що мають найменший ВВП, до другої – 12 країн, що мають найбільший ВВП. Ці дві підгрупи залишаємо для подальшого дослідження, а решту 10 країн із середини сукупності вилучаємо з розгляду. Для першої та другої підгруп окремо побудуємо лінійні регресійні моделі за допомогою 1МНК. Результати застосування функції ЛИНЕЙН() для обох підгруп виведено в табл. 3.

Таблиця 3

Результати функції ЛИНЕЙН() для підгруп країн, що об'єднані за обсягом ВВП

Підгрупа країн з низьким ВВП (1 – 12 країни рейтингу)		Підгрупа країн з високим ВВП (22 – 34 країни рейтингу)	
0,0402	0,1188	0,0711	−8,1699
0,0124	0,3283	0,0027	2,4244
0,5129	0,5279	0,9856	6,1776
10,53	10	686,36	10
2,93	2,79	26 193,65	381,63

За моделлю для першої підгрупи сума квадратів помилок становить $ESS_1 = 2{,}79$, для другої $ESS_2 = 381{,}63$. Оскільки обсяги обох підгруп однакові, то обчислюємо маємо емпіричне значення критерію Фішера за формулою (18):

$$F = \frac{381{,}63}{2{,}79} = 136{,}96.$$

За допомогою вбудованої функції F.ОБР.ПХ() визначаємо критичне значення статистики Фішера для рівня значущості $\alpha = 0{,}05$ та кількості ступенів вільності чисельника та знаменника $df_1 = df_2 = 12 - 1 - 1 = 10$. Маємо $F_{0{,}05}(10; 10) = 2{,}98$. Оскільки $F_{\text{емп.}} > F_{0{,}05}$, то нульову гіпотезу про гомоскедастичність залишків відкидаємо. Отже, результати обох тестів співпадають.

Оскільки гетероскедастичність залишків моделі виявилась статистично значущою, то оцінки параметрів моделі, які отримані за допомогою 1МНК не є ефективними. Застосуємо для визначення параметрів лінійної моделі узагальнений метод найменших квадратів.

Припустимо, що дисперсія помилок моделі пропорційна квадрату значення зовнішнього фактору, тобто має місце співвідношення (26). Для переходу до нової специфікації моделі слід застосовувати матрицю (28). Отже, вибіркова сукупність, для якої будемо застосовувати метод 1МНК, складається із значень ${}^{y_i}/_{x_i}$ та ${}^{1}/_{x_i}$, $i = \overrightarrow{1;\ 34}$. Пропонуємо побудувати цю модель самостійно. Для цього треба сформувати масив значень нових змінних і застосувати до них функцію ЛИНЕЙН(). Повернувшись до вихідних змінних, отримуємо таку економетричну модель:

$$\tilde{\hat{y}} = -0{,}0511 + 0{,}0532 \cdot x. \tag{30}$$

Перевіримо випадкові залишки моделі, що описується рівнянням регресії (30), на відсутність гетероскедастичності, скориставшись для цього тестом рангової кореляції Спірмена. Необхідні для цього обчислення ми теж пропонуємо виконати самостійно за тим алгоритмом, за яким здійснювалась перевірка моделі (29). Однак перевірка на гомоскедастичність випадкових залишків моделі (30) показала, що гетероскедастичність залишків моделі є статистично значущою.

Також не дало результатів і перехід до нових змінних за допомогою вагового коефіцієнта $1/\sqrt{x_{ik}}$. У чому теж радимо переконатись, провівши самостійно побудову моделі за допомогою 1МНК і перевірку її випадкових залишків на гомоскедастичність за тестом рангової кореляції Спірмена за описаною вище схемою.

Оскільки жоден з формальних методів, що спираються на застосування нових змінних, не допоміг позбутися гетероскедастичності залишків моделі, розглянемо саму суть проблеми.

Вибіркова сукупність, яку ми досліджуємо (див. табл. 1), характеристики об'єктів, що визначаються у натуральних величинах і відповідають певному перерізу часу (це так звані ***cross section*** дані). Гетероскедастичність залишків моделі виникає у зв'язку з тим, що об'єкти дослідження, характеристики яких розглядаються, дуже сильно різняться за значеннями як внутрішнього, так і зовнішнього факторів. Так, до вибіркової сукупності належать одночасно і Люксембург, який має найменший за обсягом ВВП серед країн, що розглядаються, і щорічно витрачає на освіту 0,34 млрд USD, і США, що має найбільший обсяг ВВП серед країн, що розглядаються, і щорічно витрачають на освіту 181,30 млрд USD. Спробуємо скоротити розмах вибіркової сукупності таким чином. Будемо розглядати не абсолютні показники ВВП і витрати на освіту для кожної країни, а питомі величини, що припадають на одну особу. Отже, нам знадобиться додаткова інформація щодо кількості населення країн, що розглядаються. Вихідні дані та результати обчислення питомих характеристик наведені в табл. 4.

Таблиця 4

Вихідні дані для побудови економетричної моделі у питомих показниках

№ п/п	Країна	ВВП, млрд USD	Витрати на освіту, млрд USD	Кількість населення, млн осіб	Питомий ВВП	Питомі витрати на освіту
1	Люксембург	5,67	0,34	0,36	15,75	0,94
2	Уругвай	10,13	0,22	2,90	3,49	0,08
3	Сінгапур	11,34	0,32	2,39	4,74	0,13
4	Ірландія	15,88	1,23	3,4	4,62	0,36
5	Ізраїль	20,94	1,81	3,87	5,41	0,47
6	Угорщина	22,16	1,02	10,71	2,07	0,10
7	Нова Зеландія	23,83	1,27	3,10	7,69	0,41
8	Португалія	24,67	1,07	9,93	2,48	0,11
9	Гонконг	27,56	0,67	5,07	5,44	0,13
10	Чилі	27,57	1,25	11,10	2,48	0,11
11	Греція	40,15	0,75	9,60	4,18	0,08
12	Фінляндія	51,62	2,80	4,75	10,80	0,59
13	Норвегія	57,71	4,90	4,09	14,11	1,20
14	Югославія	63,03	3,5	22,34	2,82	0,16
15	Данія	66,32	4,45	5,12	12,95	0,8/
16	Туреччина	66,97	1,60	11,92	1,49	0,04
17	Австрія	76,88	4,26	7,51	10,24	0,57
18	Швейцарія	101,65	5,31	6,37	15,96	0,83
19	Саудівська Аравія	115,97	6,40	8,37	13,86	0,76
20	Бельгія	119,49	7,15	9,86	12,19	0,73
21	Швеція	124,15	11,22	8,31	14,94	1,35
22	Австралія	140,98	8,66	14,62	9,64	0,59
23	Аргентина	153,85	5,56	27,06	5,69	0,21
24	Нідерланди	169,38	13,41	14,14	11,98	0,95
25	Мексика	186,33	5,46	67,40	2,76	0,08
26	Іспанія	211,78	4,79	37,43	5,66	0,13
27	Бразилія	249,75	8,92	123,03	2,03	0,07
28	Канада	261,41	18,90	23,94	10,92	0,79
29	Італія	395,52	15,95	57,04	6,93	0,28
30	Велика Британія	534,97	29,90	55,95	9,56	0,53
31	Франція	655,29	33,59	53,71	12,20	0,63
32	Німеччина	815,00	38,92	61,59	13,24	0,63
33	Японія	1 040,45	61,61	116,75	8,91	0,53
34	США	2 586,40	181,30	227,64	11,36	0,80

Виберемо функцію апроксимації, за якою будуватимемо модель. За виглядом графіка (рис. 3) можна зробити припущення, що залежність питомих витрат на освіту від питомого ВВП країни є лінійно.

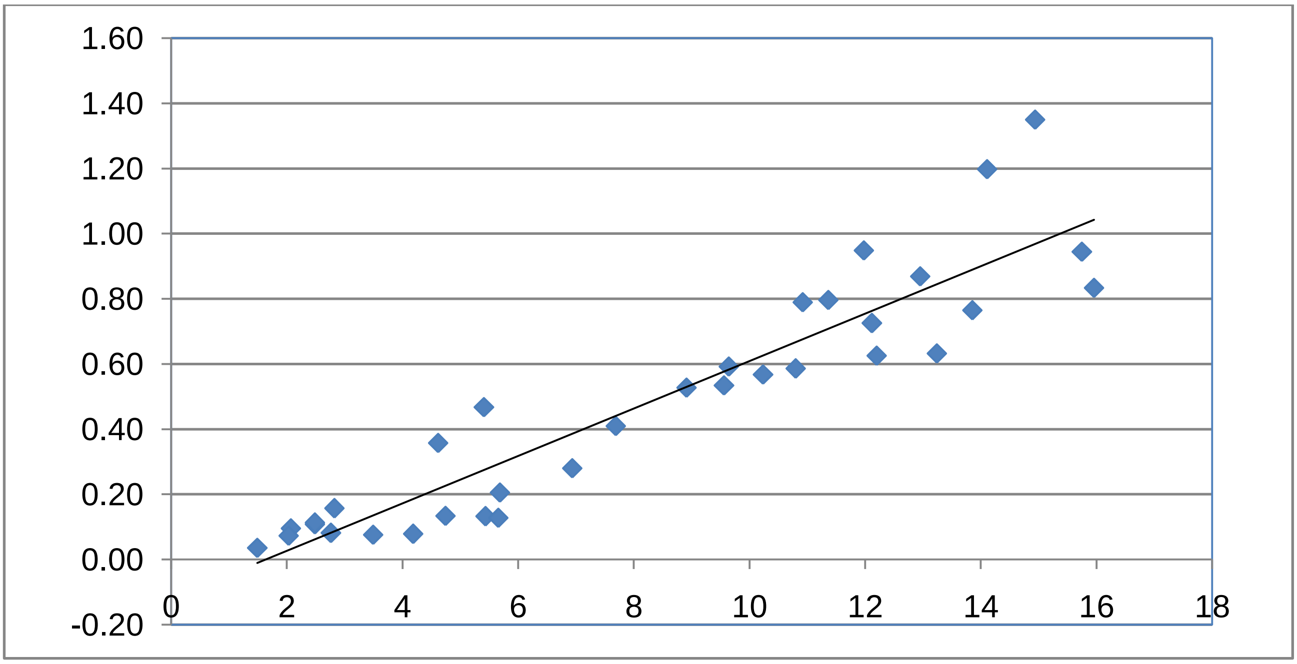

Рис. 3. **Залежність питомих витрат на освіту (тис. USD/особу) від питомого ВВП (тис. USD/особу) країни**

За графіком (рис. 3) також можна зробити висновок, що залишки регресійної моделі, що побудована у питомих показниках, теж мають гетероскедастичність, оскільки розпорошення емпіричних даних відносно лінії тренду зростає зі збільшеннях значення зовнішнього фактору. Перевіримо за допомогою тесту рангової кореляції Спірмена.

За даними табл. 4 (стовпці 6 і 7) побудуємо модель парної лінійної регресії. Результати застосування функції ЛИНЕЙН() наведені в табл. 5.

Таблиця 5

Результати функції ЛИНЕЙН() для моделі у питомих величинах

b_1 = 0,0728	b_0 = −0,1193
σ_{b_1} = 0,0052	σ_{b_0} = 0,0487
R^2 = 0,8591	σ_e = 0,1372
F = 195,05	df_e = 32
RSS = 3,67	ESS = 0,60

Отже, ми отримали таке рівняння регресії:

$$\tilde{y} = -0{,}1193 + 0{,}0728 \cdot x. \quad (31)$$

Модель є значущою в цілому за критерієм Фішера, оскільки маємо $F_{\text{емп.}} = 195{,}05 > F_{0{,}05}(2; 32) = 3{,}19$. Критичне значення критерію Фішера було знайдено за допомогою вбудованої функції F.ОБР.ПХ(). Також за критерієм Стьюдента обидва параметри моделі є значущими, оскільки емпіричні значення t-критерію дорівнюють, відповідно, $t_{b_1} = 13{,}97$ та $t_{b_0} = 2{,}45$, тоді як критичне значення становить $t_{0{,}05}(32) = 2{,}04$. Його визначаємо за допомогою вбудованої функції СТЬЮДЕНТ.ОБР.2Х(). За величиною коефіцієнта детермінації робимо висновок, що мінливість зовнішнього фактора згідно з отриманою моделлю зумовлює 85,91 % мінливості внутрішнього фактора.

Тепер перевіримо, чи є гетероскедастичність залишків моделі (31) статистично значущою. Результати розрахунків, які необхідні для здійснення тестування, наведені у табл. 6.

Таблиця 6

Результати розрахунків за тестом рангової кореляції Спірмена для моделі у питомих показниках

№ п/п	Країна	Питомий ВВП, x_i	Питомі витрати на освіту, y_i	$\tilde{y}_i$	$e_i = \lvert y_i - \tilde{y}_i \rvert$	Ранг помилок моделі
1	2	3	4	5	6	7
1	Люксембург	15,75	0,94	1,03	0,09	17
2	Уругвай	3,49	0,08	0,13	0,05	13
3	Сінгапур	4,74	0,13	0,23	0,10	20
4	Ірландія	4,62	0,36	0,22	0,14	25
5	Ізраїль	5,41	0,47	0,27	0,20	29
6	Угорщина	2,07	0,10	0,03	0,07	14
7	Нова Зеландія	7,69	0,41	0,44	0,03	4
8	Португалія	2,48	0,11	0,06	0,05	9
9	Гонконг	5,44	0,13	0,28	0,15	27
10	Чилі	2,48	0,11	0,06	0,05	11
11	Греція	4,18	0,08	0,19	0,11	22
12	Фінляндія	10,80	0,59	0,67	0,08	16
13	Норвегія	14,11	1,20	0,91	0,29	33
14	Югославія	2,82	0,16	0,09	0,07	15
15	Данія	12,95	0,88	0,82	0,06	8
16	Туреччина	1,49	0,04	-0,01	0,05	10
17	Австрія	10,24	0,57	0,63	0,06	12
18	Швейцарія	15,96	0,83	1,04	0,21	31
19	Саудівська Аравія	13,86	0,76	0,89	0,13	24

Закінчення табл. 6

1	2	3	4	5	6	7
20	Бельгія	12,12	0,73	0,76	0,03	5
21	Швеція	14,94	1,35	0,97	0,38	34
22	Австралія	9,64	0,59	0,58	0,01	3
23	Аргентина	5,69	0,21	0,29	0,08	19
24	Нідерланди	11,98	0,95	0,75	0,20	30
25	Мексика	2,76	0,08	0,08	0,00	1
26	Іспанія	5,66	0,13	0,29	0,16	28
27	Бразилія	2,03	0,07	0,03	0,04	7
28	Канада	10,92	0,79	0,68	0,11	23
29	Італія	6,93	0,28	0,39	0,11	21
30	Велика Британія	9,56	0,53	0,58	0,05	6
31	Франція	12,20	0,63	0,77	0,14	26
32	Німеччина	13,24	0,63	0,84	0,21	32
33	Японія	8,91	0,53	0,53	0,00	2
34	США	11,36	0,80	0,71	0,09	18

Обчислюємо за формулою (15) коефіцієнт рангової кореляції Спірмена:

$$r_{X\varepsilon} = 1 - 6 \cdot \frac{6618}{34 \cdot (34^2 - 1)} = -0{,}011.$$

Тепер за формулою (16) визначаємо емпіричне значення статистики Стьюдента:

$$t = \frac{0{,}011 \cdot \sqrt{34 - 2}}{\sqrt{1 - 0{,}011^2}} = 0{,}06.$$

Критичне значення статистики Стьюдента дорівнює $t_{0,05}(32) = 2{,}04$. Оскільки $t_{\text{емп.}} < t_{0,05}$, то нульову гіпотезу про гомоскедастичність залишків моделі нема причин відкидати. Отже, тест рангової кореляції Спірмена свідчить, що з надійністю 95 % гетероскедастичність залишків моделі, яка була побудована за 1МНК, не є статистично значущою. Ми отримали регресійну модель (31), за якою МНК-оцінки є BLUE. Таким чином, завдяки зменшенню розмаху вихідних даних ми побудували модель, яка відповідає завданню дослідження і МНК-оцінки за цією моделлю є не тільки незсунутими та спроможними, але й ефективними.

3. Завдання для самостійного виконання

За вихідними даними, що наведені в табл. 7 (значення зовнішнього фактору надані у першому стовпці таблиці, вони є спільним для всіх варіантів, а значення внутрішнього фактора треба вибирати за номером стовпця, що відповідає номеру варіанта) необхідно:

побудувати економетричну модель парної лінійної регресії за допомогою 1МНК;

перевірити якість моделі в цілому за критерієм Фішера і значущість її параметрів за критерієм Стьюдента;

обчислити випадкові залишки моделі та за тестом рангової кореляції Спірмена і тестом Голдфелда – Квандта перевірити, чи справедливою є нульова гіпотеза про їхню гомоскедастичність;

побудувати регресійну модель за методом УМНК у припущенні, що дисперсія залишків моделі пропорційна квадрату значення зовнішнього фактора та за тестом рангової кореляції Спірмена перевірити, чи справедливим є припущення щодо гомоскедастичності залишків цієї моделі.

Таблиця 7

Вихідні дані завдання для самостійного виконання

X	Номери варіантів									
	1	2	3	4	5	6	7	8	9	10
	Y	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
0,5	18	36	94	29	82	402	484	98	770	64
1	16	48	108	28	96	404	472	116	746	60
1,5	24	48	102	33	90	396	492	102	746	76
2	28	48	105	29	93	392	472	128	746	84
2,5	26	51	111	31	99	394	469	126	740	80
3	31	68	128	35	116	389	452	131	706	90
3,5	52	72	132	33	120	368	448	152	698	132
4	46	66	128	29	116	374	454	132	710	120
4,5	48	68	126	39	114	372	452	148	706	124
5	62	82	142	37	130	358	420	162	678	152
5,5	59	79	137	39	110	361	441	159	684	146
6	45	65	125	42	113	375	455	145	712	118
6,5	64	84	144	50	132	356	436	164	674	156

Закінчення табл. 7

1	2	3	4	5	6	7	8	9	10	11
7	74	94	154	49	142	346	426	174	654	176
7,5	71	91	151	46	139	349	440	171	660	170
8	82	102	162	57	150	361	418	159	612	192
8,5	74	94	154	42	142	346	426	174	654	176
9	80	112	172	55	176	325	394	195	618	205
9,5	87	107	167	62	155	333	413	187	628	210
10	74	82	142	36	130	346	438	174	678	157
10,5	95	115	175	70	163	310	405	210	612	218
11	79	99	159	76	130	341	421	179	675	152
11,5	89	109	169	52	157	315	402	205	624	206
12	88	108	168	63	156	332	412	188	626	237
12,5	87	90	150	62	138	333	430	210	662	178
13	94	114	174	82	162	326	406	194	598	216
13,5	112	132	192	92	180	308	388	212	578	231
14	125	145	210	100	198	295	375	234	552	278
14,5	81	101	161	56	149	339	419	181	640	190
15	118	138	198	93	186	276	382	258	566	264
15,5	112	132	194	99	182	308	367	212	578	252
16	110	130	190	107	178	310	390	210	547	248
16,5	95	106	166	86	154	265	414	268	630	173
17	147	167	224	122	212	273	332	231	508	367
17,5	155	175	235	130	256	265	345	255	492	338
18	122	142	202	97	190	298	378	176	558	272
18,5	190	210	270	165	258	230	310	290	422	408
19	94	102	162	86	150	326	418	194	638	216
19,5	160	180	240	135	228	260	340	260	482	348
20	220	240	300	195	324	178	280	342	362	468
20,5	97	117	177	117	152	323	403	197	608	186
21	230	250	310	205	298	190	270	330	342	488
21,5	110	130	190	103	178	310	390	210	582	248
22	300	320	380	275	368	120	200	400	202	628
22,5	108	110	170	130	158	312	410	208	622	244
23	360	380	440	335	428	60	140	438	204	410
23,5	340	360	420	315	408	80	160	459	122	647
24	120	140	200	122	188	300	380	220	562	268
24,5	360	380	452	380	380	148	140	372	182	645
25	210	230	248	205	452	296	325	224	537	448

4. Запитання для самодіагностики

1. Наведіть умови, яким мають відповідати випадкові помилки моделі для того, щоб МНК-оцінки були незсунутими, ефективними та спроможними.

2. Поясніть, що означає, що МНК-оцінки є BLUE.

3. Наведіть формулу для оцінювання параметрів регресійної моделі за однокроковим методом найменших квадратів.

4. Наведіть розмір матриці зовнішніх змінних, який вона повинна мати у випадку, коли обчислення МНК-оцінок здійснюється матричним методом, й охарактеризуйте структуру цієї матриці.

5. Дайте означення гетероскедастичності.

6. Які проблеми щодо властивостей МНК-оцінок пов'язані з гетероскедастичністю залишків регресійної моделі?

7. Що може бути причиною гетероскедастичності?

8. Що таке "специфікація моделі"?

9. Наведіть формулу для оцінювання параметрів регресійної моделі за узагальненим методом найменших квадратів.

10. Який вигляд має коваріаційна матриця випадкових залишків моделі, якщо виконуються умови теореми Гаусса – Маркова?

11. Який вигляд має коваріаційна матриця випадкових залишків моделі, якщо відхилення від умов теореми Гаусса – Маркова проявляється тільки в наявності гетероскедастичності?

12. Які існують методи, що дозволяють визначити, чи є гетероскедастичність статистично значущою?

13. Сформулюйте основну гіпотезу, яка підлягає перевірці за тестами щодо статистичної значущості гетероскедастичності.

14. Наведіть алгоритм теста рангової кореляції Спірмена.

15. За яким критерієм здійснюється перевірка основної гіпотези за тестом рангової кореляції Спірмена?

16. У яких границях може змінюватися значення коефіцієнта рангової кореляції?

17. Про що свідчить від'ємне значення коефіцієнта рангової кореляції Спірмена?

18. Наведіть алгоритм тесту Голдфелда – Квандта.

19. За яким критерієм здійснюється перевірка основної гіпотези за тестом Голдфелда – Квандта?

20. За яким алгоритмом треба будувати статистичний критерій, якщо для перевірки залишків моделі на гомоскедастичність застосовується тест Голдфелда – Квандта? Визначить границі, у яких може змінюватись значення цього критерію.

5. Рекомендована література

1. Айвазян С. А. Прикладная статистика и основы економетрики : учебник для вузов / С. А. Айвазян, В. С. Мхитарян. – Москва : ЮНИТИ, 1998. – 1022 с.

2. Айвазян С. А. Прикладная статистика : учебник для вузов. В 2 т. Т. 2. Основы эконометрики / С. А. Айвазян. – 2-е изд., испр. – Москва : ЮНИТИ-ДАНА, 2001. – 432 с.

3. Доугерти К. Д. Введение в эконометрику / К. Д. Доугерти; пер. с англ. – Москва : ИНФРА-М, 1999. – XIV, 402 с.

4. Єгоршин А. А. Лабораторний практикум з економетрики в Excel : навч.-практ. посіб. / О. О. Єгоршин, Л. М. Малярець. – Харків : Вид. ХНЕУ, 2011. – 140 с.

5. Лугінін О. Є. Економетрія : навч. посіб. / О. Є. Лугінін, С. В. Білоусова, О. М. Білоусов. – Київ : Центр навчальної літератури, 2005. – 252 с.

6. Магнус Я. Р. Эконометрика. Начальный курс : учебник / Я. Р. Мангус, П. К. Катышев, А. А. Пересецкий. – 6-е изд., перераб. доп. – Москва : Дело, 2004. – 576 с.

7. Малярець Л. М. Економіко-математичне моделювання / Л. М. Малярець. – Харків : ХНЕУ, 2010. – 320 с.

8. Марченко В. М. Эконометрика и экономико-математические методы и модели : учеб. пособ. для студентов учреждений высшего образования по экономическим специальностям. В 2 ч. Ч. 1. Эконометрика / В. М. Марченко, Н. П. Можей, Е. А. Шинкевич. – Минск : БГТУ, 2011. – 157 с.

9. Наконечний С. І. Економетрія : підручник / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – Вид. 3-тє., доп. та перероб. – Київ : КНЕУ, 2005. – 520 с.

10. Овсянникова М. М. Компьютерный практикум по эконометрике : для студентов специальности 080109 "Бухгалтерский учет, анализ и аудит", 080105 "Финансы и кредит". – Глазов : Глазовский инженерно-экономический институт, 2011. – 64 с.

11. Эконометрика : учебник для вузов / под ред. проф. Н. Ш. Кремера. – Москва : ЮНИТИ-ДАНА, 2002. – 311 с.

12. Эконометрика : учебник / под ред. И. И. Елисеевой. – Москва : Финансы и статистика, 2004. – 344 с.

13. Экономико-математические методы и модели : учеб.-метод. пособ. по выполнению расчетных заданий с использованием табличного процессора Excel для студентов экономических специальностей / авт.-сост. Е. А. Шинкевич. – Минск : БГТУ, 2005. – 72 с.

14. Проверка на наличие гетероскедастичности (гомоскедастичности) [Электронный ресурс]. – Режим доступа : https://www.youtube.com/watch?v=3AEX1r3dFv4.

15. Что такое Гомоскедастичность и Гетероскедастичность [Электронный ресурс]. – Режим доступа : https://www.youtube.com/watch?v=oP8TxQyFWsY.

16. Econometrics – Heteroscedasticity [Electronic resource]. – Access mode : https://www.youtube.com/watch?v=4uFKL8MFFoM.

17. Heteroscedasticity in Regression Analysis [Electronic resource]. – Access mode : http://statisticsbyjim.com/regression/heteroscedasticity-regression.

18. Heteroscedasticity summary [Electronic resource]. – Access mode : https://www.youtube.com/watch?v=zRklTsY9w9c.

Зміст

НАВЧАЛЬНЕ ВИДАННЯ

ЕКОНОМЕТРИКА

Методичні рекомендації і завдання до самостійної роботи за темою "Проблеми в побудові лінійних множинних регресійних моделей: гетероскедастичність" для студентів усіх спеціальностей першого (бакалаврського) рівня

Самостійне електронне текстове мережеве видання

Укладачі: **Лебедєва** Ірина Леонідівна
Жуков Андрій В'ячеславович
Лебедєв Степан Сергійович

Відповідальний за видання *Л. М. Малярець*

Редактор *А. С. Ширініна*

Коректор *А. С. Ширініна*

План 2019 р. Поз. № 163 ЕВ. Обсяг 33 с.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Науки, 9-А

Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.