

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Методичні рекомендації

до виконання лабораторних робіт
з дисципліни «Економетрика»

для студентів освітньо-кваліфікаційного рівня бакалавр
спеціальностей 051 «Економіка» та 075 «Маркетинг»

Укладачі:

к.е.н. Харченко А.О.

к.е.н. Конохова З.П.

Затверджено
Редакційно-видавничою
радою університету,
протокол № 2
від 25.06.20 р.

Харків - 2019

Методичні вказівки до виконання лабораторних робіт з дисципліни «Економетрика» для студентів освітньо-кваліфікаційного рівня бакалавр спеціальностей 051 «Економіка» та 075 «Маркетинг» / Укл.: А.А. Харченко, З.П. Конохова – Харків: НТУ «ХП», 2020 - 105с.

Методичні вказівки призначені для студентів освітньо-кваліфікаційного рівня бакалавр спеціальностей 051 «Економіка» та 075 «Маркетинг» НТУ «ХП», що вивчають дисципліну «Економетрика».

ЗМІСТ

Вступ	4
1 Загальні вимоги до підготовки і виконання лабораторних робіт	5
2 Лінійна економетрична модель.....	6
3 Загальна лінійна економетрична модель	33
4 Нелінійні економетричні моделі	42
5 Мультиколінеарність	57
6 Гетероскедастичність	68
7 Автокореляція залишків	82
8 Аналіз часового ряду із використанням рядів Фур'є	90
Список рекомендованої літератури.....	101
Додатки.....	103

ВСТУП

Сучасний стан економіки України, розвиток її наукової бази потребує вирішення складних теоретичних і практичних завдань, які вимагають створення і застосування якісно нових напрямків економічної теорії та суміжних наукових дисциплін, тісно пов'язаних з нею.

При аналізі соціально-економічних явищ і процесів, які характеризують той чи інший етап розвитку ринкової економіки, доводиться мати справу з різноманітними масовими явищами, виявляти наявні там закономірності та встановлювати подальший хід їх розвитку. Задачу вивчення кількісних сторін масових явищ і процесів в економіці у нерозривному зв'язку з їх якісною стороною якраз і вирішує економетрія (економетрика). Економетрика на основі свого інструментально-теоретичного апарату встановлює (виявляє) причинно-наслідкові зв'язки в досліджуваних економічних системах з метою проведення їх аналізу, діагностування і прогнозування.

Економетрика є однією з фундаментальних дисциплін у підготовці бакалаврів з економіки для всіх спеціальностей, побудована на основі математичних та економічних знань, і разом з такими дисциплінами як мікро- та макроекономіка утворює базис сучасної економічної освіти. Конструювання та дослідження економетричних моделей сприяє формуванню у студентів нового економіко-математичного мислення, спрямованого на підготовку фахівців-економістів нової формації.

Метою навчальної дисципліни «Економетрика» є одержання теоретичних знань і практичних навичок з оцінювання параметрів залежностей, які характеризують кількісні причинно-наслідкові взаємозв'язки між економічними показниками, а також використання економетричних моделей в економічних дослідженнях, у практиці управління економічними процесами на різних ієрархічних рівнях національної економіки.

1 ЗАГАЛЬНІ ВИМОГИ ДО ПІДГОТОВКИ І ВИКОНАННЯ ЛАБОРАТОРНИХ РОБІТ

Цикл лабораторних робіт з дисципліни «Економетрика» включає 10 лабораторних робіт, які охоплюють основні теми навчальної програми дисципліни. Основною метою цих робіт є закріплення і перевірка теоретичних знань, отриманих студентами на лекціях і в результаті самостійного вивчення курсу, а також отримання практичних навичок побудови та практичного використання економетричних моделей.

На етапі підготовки до кожної лабораторної роботи студент повинен уважно ознайомитися з метою, завданням та порядком виконання роботи, а також вивчити необхідний теоретичний матеріал і бути в змозі дати відповіді на контрольні питання, які наведені в кінці кожної роботи.

Результат виконання кожної лабораторної роботи оформляється студентом у вигляді презентації виконаної у пакеті PowerPoint. Презентація повинна містити основні етапи та результати виконання роботи. Захист лабораторної роботи є завершальним етапом роботи над нею.

ЗМІСТОВИЙ МОДУЛЬ І

КЛАСИЧНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

2 ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ

2.1 Загальна лінійна економетрична модель

Загальна лінійна економетрична модель - це регресійна модель, яка встановлює лінійну залежність між економічними показниками, один з яких є залежною (пояснюваною) змінною, а всі інші – незалежними (пояснюючими) змінними моделі. Залежна змінна для такої моделі називається ендогенна змінна, а незалежні змінні – екзогенні.

Вибіркова (емпірична) загальна лінійна економетрична модель має наступний вигляд :

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + e,$$

де: y – залежна (пояснювана) змінна моделі, x_1, x_2, \dots, x_m – незалежні (пояснюючі) змінні моделі (фактори), a_0, a_1, a_m – параметри вибіркової моделі (або коефіцієнти регресії), e – залишки моделі.

Вибіркова модель є реальною конструкцією і будується на основі певної статистичної вибірки з генеральної сукупності. Параметри вибіркової моделі a_0, a_1, \dots, a_m є випадковими величинами, а залишки e можна оцінити (розрахувати) на основі статистичних даних.

Вибіркова (емпірична) функція регресії для загальної лінійної економетричної моделі має наступний вигляд :

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m,$$

де: \hat{y} – оцінка математичного сподівання залежної (пояснюваної) змінної моделі, x_1, x_2, \dots, x_m – незалежні (пояснюючі) змінні моделі (фактори), a_0, a_1, a_m – параметри вибіркової регресії.

Для побудови і аналізу загальної лінійної економетричної моделі широко застосовується апарат матричної алгебри. Тому для подальших викладок подамо загальну лінійну економетричну модель у матричній формі. Оскільки теоретична модель використовується для канонічного подання деякого економічного явища або процесу, а реально ми оперуємо у процесі дослідження цього явища (процесу) тільки вибірковою моделлю, саме вибірккову модель подамо у матричному вигляді :

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

де

– вектор спостережень за залежною змінною моделі:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \dim Y = n \times 1$$

– матриця спостережень за пояснюючими змінними моделі, яку ще називають **регресійною матрицею**:

$$H = \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{m1} \\ 1 & x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{mn} \end{pmatrix}, \dim H = n \times k$$

- вектор оцінок параметрів моделі (вектор параметрів вибіркової моделі):

$$A = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_k \end{pmatrix}, \dim A = k \times 1$$

- вектор залишків моделі:

$$e = \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix}, \dim e = n \times 1$$

Для всіх наведених вище форм представлення загальної лінійної моделі прийняті наступні загальні позначення, які будуть у подальшому постійно використовуватися :

n – розмір статистичної вибірки (кількість спостережень в статистичній вибірці);

m – число незалежних (пояснюючих) змінних моделі ;

k = m + 1 – число параметрів моделі.

Найпростішою серед лінійних економетричних моделей є модель парної лінійної регресії (або проста лінійна модель), яка описує зв'язок всього між двом економічними змінними - показниками.

Виходячи з вище розглянутих позначень для простої лінійної регресії маємо

:

• **вибіркова (емпірична) модель парної лінійної регресії**

$$y = a_0 + a_1x + e,$$

• **вибіркова функція парної лінійної регресії**

$$\hat{y} = a_0 + a_1x. \quad (2)$$

Рівняння (2) представляє собою параметричне рівняння прямої, тому на площині xOy вибірковій функції парної лінійної регресії відповідає вибіркова (емпірична) пряма регресії. Графічно вибіркова функція регресії і пряма регресії для деякої вибірки представлені на рис. 1.

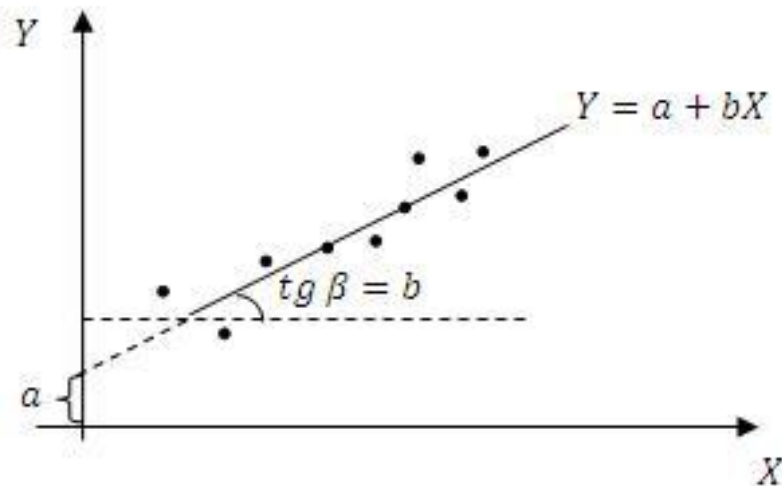


Рисунок 1 - Парна лінійна регресія

2.2 Оцінювання параметрів моделі

Для оцінювання параметрів вибіркової моделі використовуються різні методи, такі як: метод моментів, метод найменших квадратів, метод максимальної правдоподібності та інші. До найбільш простих і розповсюджених методів відноситься широко відомий метод найменших квадратів, який становить основу регресійного аналізу. В принципі цей метод може застосовуватися для оцінювання параметрів будь-яких лінійних економетричних моделей. Але результат застосування цього методу залежить від того, чи задовольняє економетрична модель усім припущенням класичного лінійного регресійного аналізу, чи ні. Якщо модель задовольняє усім припущенням класичного лінійного регресійного аналізу, то вона є по суті багатofакторною (або однофакторною) **класичною регресійною моделлю**. У цьому випадку для оцінювання її параметрів може коректно застосовуватися однокрокова

процедура методу найменших квадратів (МНК) і отримані при цьому оцінки параметрів моделі будуть найкращими серед усіх можливих. Якщо ж для моделі порушується хоча б одне з положень класичної лінійної регресії така модель називається **узагальненою моделлю** і для оцінювання її параметрів застосовуються інші більш ефективні методи.

До основних припущень класичного лінійного регресійного аналізу належать наступні припущення .

Припущення 1. Математичне сподівання стохастичної складової моделі дорівнює нулю для всіх спостережень: $M(\varepsilon_i) = 0, i = 1, 2, \dots, n,$

Припущення 2. Гомоскедастичність.

Це припущення означає, що дисперсія стохастичної складової у всіх спостереженнях повинна бути однаковою $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2 = \text{const}, i = 1, 2, \dots, n.$

Припущення 3. Відсутність автокореляції залишків.

Це означає, що значення $\varepsilon_i, (i = \overline{1, n})$ вектора стохастичної складової моделі повинні бути незалежні між собою $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ для $i \neq j.$

Припущення 4. Незалежні (пояснюючі) змінні не пов'язані із стохастичною складовою моделі : $\text{cov}(\varepsilon_i, x_{ji}) = 0$ для $i = 1, 2, \dots, n; j = 1, 2, \dots, m$

Припущення 5. Матриця спостережень N не є стохастичною.

Припущення 6. Відсутність мультиколіарності .

Незалежні змінні моделі не повинні бути мультиколінеарними, тобто між ними не повинно існувати лінійного функціонального або тісного кореляційного зв'язку. Слід зазначити, що дане припущення має місце тільки для багатofакторної регресії.

Припущення 7. Випадкова величина $\varepsilon_i, i = 1, 2, \dots, n$ має нормальний закон розподілу з математичним сподіванням 0 і сталою дисперсією:

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2).$$

Перше припущення стверджує що, невраховані у моделі і тому віднесені до стохастичної складової ε фактори не впливають систематично на математичне сподівання залежної змінної y . У протилежному ж випадку існує систематичний вплив на залежну змінну і до моделі не введено всі основні пояснюючі змінні.

Друге і третє припущення передбачають відсутність **автокореляції залишків** (залежність між залишками у вибірці) і **гомоскедастичність** моделі (сталість дисперсії залишків). Відсутність автокореляцій залишків дає можливість вивчати тільки систематичний вплив незалежних змінних на залежну. Сталість дисперсії залишків (гомоскедастичність) дає підстави вважати всі значення залежної змінної y_i , які відносяться до різних спостережень однаково важливими при оцінюванні параметрів моделі.

П'яте припущення означає, що пояснюючі змінні моделі утворюють лінійно незалежну систему векторів, внаслідок чого матриця спостережень за незалежними змінними моделі має повний ранг і $\det(H^T H) \neq 0$.

В основу методу найменших квадратів покладено критерій, згідно якого, „найкращою” серед усіх можливих вважається функція регресії з такими параметрами, для якої сума квадратів залишків є мінімальною. У математичному вигляді цей критерій має наступний вигляд :

$$\sum_{i=1}^n e_i^2 \rightarrow \min. \quad (3)$$

Використовуючи цей критерій і визначаються параметри вибіркової лінійної моделі.

Подамо рівняння (1) у вигляді : $e = Y - HA$. Тоді суму квадратів залишків e можна записати таким чином :

$$\sum_{i=1}^n e_i^2 = e'e = (Y - HA)^T (Y - HA) = Y^T Y - 2A^T H^T Y + A^T H^T HA. \quad (4)$$

Для мінімізації суми квадратів залишків візьмемо похідну від виразу (4) за вектором оцінок параметрів A і прирівняємо похідні до нуля :

$$\frac{\partial \sum_{i=1}^n e_i^2}{\partial A} = \frac{\partial (e'e)}{\partial A} = -2H^T Y + 2H^T HA = 0. \quad (5)$$

З виразу (5) маємо :

$$H^T HA = H^T Y. \quad (6)$$

Вираз (6) представляє собою матричну форму запису так званої **системи нормальних рівнянь**. Розв'язавши її відносно вектора оцінок A отримаємо оцінку параметрів моделі. З цією метою помножимо обидві частини виразу (6) на обернену матрицю $(H^T H)^{-1}$:

$$(H^T H)^{-1} H^T HA = (H^T H)^{-1} H^T Y.$$

Оскільки $(H^T H)^{-1} H^T H = E$, де E – одинична матриця остаточно отримуємо:

$$A = (H^T H)^{-1} H^T H Y. \quad (7)$$

Таким чином, сформувавши на основі статистичної вибірки вектор спостережень за залежною змінною моделі Y і матрицю спостережень за незалежними змінними моделі X , за формулою (7) можна обчислити вектор оцінок параметрів моделі A .

Слід зазначити, що рівняння регресії, параметри якого оцінені методом найменших квадратів, має декілька корисних властивостей, які використовуються у практичних дослідженнях.

1. Пряма регресії (поверхня або гіперповерхня регресії) проходить через середню точку з координатами $\bar{y}, \bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$.

2. Середнє значення оцінки залежної змінної дорівнює фактичному середньому значенню, тобто $\bar{\hat{y}} = \bar{y}$.

3. Сума залишків моделі дорівнює нулю, тобто $\sum_{i=1}^n e_i = 0$.

Так наприклад, третя властивість широко використовується для контролю за тим, чи правильно обчислені параметри моделі і її залишки.

1.3 Властивості оцінок параметрів моделі, отриманих за МНК

Теорема Гауса – Маркова показує, що якщо виконуються усі основні припущення класичного лінійного регресійного аналізу, розглянуті вище, то оцінки параметрів моделі, отримані МНК є **BLUE** – оцінками, тобто **найкращими незміщеними лінійними оцінками (best linear unbiased estimators)** і мають наступні властивості.

1. Вони є **незміщеними оцінками**, тобто оцінками, які задовольняють умові

$$M(a_j) = a_j, \quad (j = \overline{0, m}).$$

Незміщеність оцінки означає, що при багаторазовому повторенні випадкової вибірки попри те, що для окремих вибірок, можливо будуть помилки оцінювання, середнє значення цих помилок дорівнює нулю.

2. Вони є **ефективними**, тобто ці оцінки мають найменшу дисперсію серед усіх незміщених оцінок, отриманих іншими методами.

3. Вони є **обґрунтованими** оцінками, тобто оцінками, які при довільному $\delta > 0$ задовольнить умові

$$\lim P\{|a_j - \alpha_j| < \delta\} = 1.$$

Обґрунтованість оцінок означає, що чим більше вибірка, тим більше ймовірність того, що помилка оцінки не перевищуватиме достатньо малої попередньо заданої величини δ . Іншими словами, при збільшенні розміру вибірки зростає надійність обчислених оцінок.

1.4 Верифікація моделі

Етап верифікації економетричної моделі є достатньо об'ємним і його можна умовно поділити на дві частини :

- 1) визначення за даними вибірки показників якості моделі і їх інтерпретація ;
- 2) перевірка статистичної значимості моделі.

Показники якості моделі

До основних показників якості побудованої вибіркової моделі відносяться наступні показники :

- стандартна похибка рівняння регресії ;
- стандартні похибки параметрів моделі ;
- коефіцієнт кореляції ;
- коефіцієнт детермінації .

Зазначені показники використовуються :

- 1) для оцінювання точності вибіркової моделі, тісноти лінійного кореляційного зв'язку між змінними моделі і сили впливу пояснюючих змінних моделі на залежну ;
- 2) при виконанні ряду статистичних тестів, пов'язаних з перевіркою статистичної значимості побудованої моделі ;
- 3) при прогнозуванні.

А) Стандартна похибка рівняння регресії $\hat{\sigma}_\varepsilon$ визначається за наступною залежністю :

$$\hat{\sigma}_\varepsilon = \sqrt{\hat{\sigma}_\varepsilon^2} ,$$

де $\hat{\sigma}_\varepsilon^2$ - незміщена оцінка дисперсії залишків, яка в свою чергу ,визначається наступним чином :

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^n e_i^2}{n - k} . \quad (8)$$

Сума квадратів залишків у виразі (8) обчислюється на основі попередньо обчислених залишків. Значення залишку у кожному спостереженні статистичної вибірки при цьому обчислюється за наступною формулою :

$$e_i = y_i - \hat{y}_i, \quad (i = \overline{1, n}), \quad (9)$$

де y_i - фактичні значення залежної змінної моделі, що спостерігаються у статистичній вибірці, \hat{y}_i - розрахункові значення залежної змінної моделі, що обчислені на основі оціненого рівняння регресії за залежністю :

$$\hat{y}_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_m x_{mi}, \quad (i = \overline{1, n}).$$

Б) Стандартні похибки параметрів моделі $\hat{\sigma}_{a_j}, (j = \overline{1, m})$, як і у попередньому випадку обчислюються на основі попередньо обчислених оцінок дисперсії параметрів моделі, тобто :

$$\hat{\sigma}_{a_j} = \sqrt{\hat{\sigma}_{a_{jj}}^2}, \quad (j = \overline{1, m}),$$

де $\hat{\sigma}_{a_j}^2$ - оцінка дисперсії j - го параметра моделі, яка може бути визначена на основі так званої **дисперсійно – коваріаційної матриці** оцінок параметрів моделі , яка має наступний вигляд і структуру :

$$\text{var} - \text{cov}(a) = \begin{pmatrix} \hat{\sigma}_{a_0}^2 & \text{cov}(a_0, a_1) & \dots & \text{cov}(a_0, a_m) \\ \text{cov}(a_1, a_0) & \hat{\sigma}_{a_1}^2 & \dots & \text{cov}(a_1, a_m) \\ \dots & \dots & \dots & \dots \\ \text{cov}(a_m, a_0) & \text{cov}(a_m, a_1) & \dots & \hat{\sigma}_{a_m}^2 \end{pmatrix}.$$

Як видно, на головній діагоналі цієї матриці знаходяться необхідні дисперсії оцінок параметрів моделі, що дає можливість на основі попередньо розрахованої дисперсійно-коваріаційної матриці обчислити дисперсії оцінок параметрів $\hat{\sigma}_{a_0}^2, \hat{\sigma}_{a_1}^2, \dots, \hat{\sigma}_{a_m}^2$, а також їхні стандартні відхилення $\hat{\sigma}_{a_0}, \hat{\sigma}_{a_1}, \dots, \hat{\sigma}_{a_m}$.

Сама ж дисперсійно-коваріаційна матриця оцінок параметрів моделі обчислюється за наступною залежністю :

$$\text{var} - \text{cov}(A) = \hat{\sigma}_\varepsilon^2 (H^T H)^{-1}. \quad (10)$$

У випадку парної лінійної регресії оцінки дисперсії параметрів a_0 і a_1 можна визначити за наступними залежностями:

$$\hat{\sigma}_{a_0}^2 = \hat{\sigma}_\varepsilon^2 \frac{\sum_{i=1}^n x_i^2}{n \cdot \sum_{i=1}^n (x_i - \bar{x})^2},$$

$$\hat{\sigma}_{a_1}^2 = \hat{\sigma}_\varepsilon^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

В) Коефіцієнт кореляції є кількісною мірою тісноти лінійного кореляційного зв'язку між змінними моделі і у загальному випадку представляє собою **коефіцієнт множинної кореляції**, який обчислюється за наступною залежністю :

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}},$$

де: y_i – фактичні (статистичні) значення залежної змінної, \hat{y}_i – розрахункові значення.

У випадку парної лінійної регресії щільність лінійного кореляційного зв'язку між залежною змінною y і незалежною змінною x оцінюється за допомогою відомого **коефіцієнта парної кореляції** :

$$r_{yx} = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x)} \sqrt{\text{var}(y)}},$$

де

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \text{ - вибірковий коефіцієнт коваріації,}$$

$$\text{var}(x) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \text{ - вибіркова дисперсія пояснюючої змінної моделі,}$$

$$\text{var}(y) = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \text{ - вибіркова дисперсія залежної змінної моделі,}$$

Г) Коефіцієнт детермінації використовується як **критерій адекватності** (відповідності) моделі статистичним даним, оскільки він є мірою пояснювальної сили незалежних змінних і показує, яка частина варіації залежної змінної пояснюється саме варіацією (змінною) незалежних змінних, а не іншими випадковими факторами, які акумулюються у стохастичній складовій моделі. Іншими словами, коефіцієнт детермінації показує наскільки значним є вплив пояснюючих змінних моделі на залежну. Якщо цей вплив є значним, то побудована модель дійсно описує лінійну залежність між відповідними економічними показниками і ця залежність є суттєвою. Якщо ж цей вплив є незначним модель є неадекватною статистичним даним і лінійна регресійна залежність між економічними показниками у ній є достатньо сумнівною і неякісною.

Позначається коефіцієнт детермінації через R^2 і обчислюється за наступною залежністю :

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} . \quad (11)$$

Як видно з виразу (11) коефіцієнт детермінації завжди є **додатною величиною** і може змінювати своє значення у **межах від 0 до 1**, тобто

$$0 \leq R^2 \leq 1 .$$

Чим більше значення коефіцієнта детермінації (чим ближче воно до 1) тим більш вагомим і систематичним є вплив пояснюючих змінних на залежну і тим більше підстав стверджувати, що саме змінною значень пояснюючих змінних пояснюється змінна значення залежної змінної моделі, а не іншими випадковими і неврахованими у моделі випадковими факторами. Іншими словами високе, близьке до 1 значення коефіцієнта детермінації свідчить про високий рівень адекватності оціненої моделі статистичним даним.

І навпаки, чим менше значення коефіцієнта детермінації (чим ближче воно до 0) тим менш вагомим є вплив пояснюючих змінних на залежну і тим менше підстав стверджувати, що саме змінною значень пояснюючих змінних пояснюється змінна значення залежної змінної моделі, а не іншими випадковими і неврахованими у моделі випадковими факторами. Іншими словами низьке, близьке до 0 значення коефіцієнта детермінації свідчить про низький рівень адекватності оціненої моделі статистичним даним.

У граничному випадку, коли $R^2=1$ статистична лінійна залежність між економічними показниками перетворюється на однозначну, функціональну лінійну залежність. При цьому, для парної лінійної регресії всі точки діаграми розсіювання „лягають” на пряму регресії. При $R^2=0$ варіація значень залежної змінної взагалі не залежить від варіації значень пояснюючих змінних моделі, а пояснюється тільки впливом випадкових факторів. У цьому випадку пояснюючі змінні не мають систематичного впливу на залежну змінну моделі і між економічними показниками відсутній лінійний кореляційно-регресійний зв'язок – тобто побудована модель взагалі не відповідає статистичним даним і слід переглянути її специфікацію.

Як інтерпретується значення коефіцієнта детермінації? Нехай для деякої вибіркової моделі коефіцієнт детермінації дорівнює $R^2=0,89$. Це означає, у даному випадку, що зміна значення залежної змінної моделі на 89% пояснюється зміною значень пояснюючих змінних моделі, а на 11% - іншими випадковими факторами. Таким чином вплив пояснюючих змінних на залежну є достатньо суттєвим, що свідчить про високий рівень адекватності моделі статистичним даним.

Перевірка статистичної значимості і інтервали довіри

Оскільки значення параметрів $a_j, (j = \overline{0, m})$, коефіцієнта кореляції $R (r_{yx})$ і коефіцієнта детермінації R^2 обчислюються за даними деякої статистичної вибірки з точки зору математичної статистики вони є **вибірковими характеристиками (статистиками)** відповідних характеристик генеральної сукупності. Значення цих вибірових характеристик є **оцінками** значень відповідних характеристик генеральної сукупності.

Особливостями зазначених вибірових характеристик, як відомо, є те, що вони є **випадковими величинами**, оскільки є функціями вибірки і можуть змінювати свої значення від однієї вибірки до іншої приймаючи будь-які значення. Внаслідок цього виникає мабуть основне питання етапу верифікації, чи значення оцінок $a_j, (j = \overline{0, m})$, $R (r_{yx})$ і R^2 є результатом дії тільки випадку, чи вони відображують суттєву (значиму) інформацію і реальний зв'язок між змінними моделі.

Для відповіді на це питання необхідно перевірити зазначені вибірові характеристики на значимість. Поки цього не зроблено, отримані статистичні результати є лише деякою гіпотезою, яка може бути як прийнята так і відхилена. Перевірка розглянутих вибірових характеристик на значимість виконується за відомою схемою перевірки статистичних гіпотез.

А) Перевірка статистичної значимості моделі у цілому.

Регресійна модель, а точніше рівняння регресії, можна вважати статистично значимим („правильним”), якщо воно дійсно відображує лінійну залежність між економічними показниками у цьому рівнянні. Це можливо, якщо коефіцієнт детермінації R^2 є дуже близьким до одиниці. Якщо ж він дорівнює нулю або несуттєво відрізняється від нуля залежність між змінними моделі відсутня і отримане вибіркоче рівняння регресії не відтворює реальної ситуації.

Таким чином перевірка статистичної значимості моделі у цілому зводиться до перевірки статистичної значимості коефіцієнта детермінації, тобто перевірки того, чи суттєво його реальне значення відрізняється від нуля.

З цією метою висуваються наступні дві гіпотези :

$$H_0 : R^2 = 0,$$

$$H_1 : R^2 \neq 0.$$

Перша, **нульова гіпотеза** стверджує, що оцінене вибіркоче рівняння регресії не пояснює зміну залежної змінної під впливом пояснюючих змінних. Друга, **альтернативна гіпотеза** навпаки стверджує, що варіація залежною зміною пояснюється впливом пояснюючих змінних.

Слід зазначити, що наведена вище нуль-гіпотеза є справедливою тільки у тому випадку, якщо всі коефіцієнти регресії одночасно дорівнюють нулю. Виходячи із цього гіпотези можна подати в наступному еквівалентному вигляді :

$$H_0 : a_1 = a_2 = \dots = a_m = 0,$$

H₁ : значення хоча б одного параметра відмінне від нуля.

Якщо нуль-гіпотеза неправильна, то не всі параметри моделі незначною мірою відрізняються від нуля. Це дає підставу вважати, що відібрані фактори пояснюють варіацію залежної змінної y , тобто побудована модель адекватна фактичним статистичним даним.

Для перевірки гіпотез використовується наступна F – статистика (F – критерій Фішера) :

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-k}{m}, \quad (12)$$

яка, якщо нуль-гіпотеза є вірною, має розподіл Фішера із ступенями вільності $v_1 = m$ і $v_2 = n - k$. Тоді перевірка статистичної значимості моделі у цілому здійснюється наступним чином .

1. На основі значення вибіркового коефіцієнта детермінації R^2 за формулою (12) обчислюється розрахункове значення критерію Фішера F^* .

2. Задається рівень значимості α (як правило $\alpha = 0,05$ або $\alpha = 0,01$).

3. За статистичними таблицями F - розподілу Фішера для прийнятого рівня значимості α і ступенів вільності $\nu_1 = m$ і $\nu_2 = n - k$ визначається критичне (табличне) значення критерію Фішера $F_{кр}$.

4. Якщо виконується умова $F^* > F_{кр}$, нуль-гіпотеза відкидається на користь альтернативної, що свідчить про статистичну значимість побудованої моделі у цілому і її адекватність. У протилежному випадку нуль-гіпотеза приймається і модель вважається статистично не значимою у цілому, тобто неадекватною статистичним даним.

Якщо модель не є статистично значимою, необхідно припинити процес верифікації і повернутися на етап параметризації або етап специфікації економетричної моделі.

Б) Перевірка статистичної значимості параметрів моделі . Інтервали довіри для параметрів моделі.

Якщо модель є статистично значимою у цілому, необхідно визначити за рахунок значимості яких параметрів забезпечується її загальна статистична значимість. Для цього необхідно перевірити статистичну значимість кожного параметра моделі. При цьому для кожного параметра моделі формулюються наступні статистичні гіпотези, які потрібно перевірити з деяким рівнем значимості α :

нульова гіпотеза $H_0 : a_j = 0, (j = \overline{0, m}) ,$

альтернативна гіпотеза $H_1 : a_j \neq 0, (j = \overline{0, m}) .$

Для перевірки наведених статистичних гіпотез використовується наступна t – статистика (критерій Ст'юдента) :

$$t_{a_j} = \frac{a_j}{\hat{\sigma}_{a_j}}, (j = \overline{0, m}) , \quad (13)$$

де a_j - оцінка параметра теоретичної регресії,

$\hat{\sigma}_{a_j}$ - стандартна похибка j -го параметра моделі;

m - кількість оцінених параметрів моделі.

Статистика (13) у разі вірності нуль – гіпотези має t – розподіл Ст’юдента з ступенем вільності $\nu = n - k$. Тоді перевірка статистичної значимості кожного параметра моделі здійснюється за наступною схемою :

1. Для кожного параметру за формулою (13) розраховується t -статистика (критерій Ст’юдента) $t_{a_j}^*$.

2. Вибирається рівень значимості α (як правило $\alpha = 0,05$ або $\alpha = 0,01$).

Для вибраного рівня значимості α і ступеня вільності $\nu = n - k$ за статистичними таблицями t -розподілу Ст’юдента знаходиться критичне значення критерію Ст’юдента $t_{кр} = t_{\alpha/2}$.

Таке позначення $t_{кр}$ як $t_{\alpha/2}$ пояснюється тим, що t -розподіл Ст’юдента є симетричним розподілом і критична область для нього складається з двох фрагментів, межі яких задаються квантилем $\alpha/2$. У самій же таблиці t -розподілу Ст’юдента критичне значення $t_{кр}$ слід вибирати для значення α при двосторонньому тесті і для значення $\alpha/2$ - для одностороннього. Оскільки розрахункові значення значення $t_{b_j}^$ можуть бути як додатними так і від’ємними, слід орієнтуватися на двосторонній тест на значимість.*

3. Якщо виконується умова $t_{a_j}^* > t_{кр}$, нуль-гіпотеза відкидається на користь альтернативної, що свідчить про суттєву відмінність від нуля відповідної оцінки a_j і її статистичну значимість. У протилежному випадку приймається нуль-гіпотеза і оцінка a_j вважається статистично не значимою (статистично близькою до нуля).

Якщо деякий параметр моделі a_j є статистично незначимим, його потрібно вилучити з рівняння регресії разом із відповідною пояснюючою змінною x_j , оскільки статистична не значимість цього параметра свідчить, що вплив відповідної пояснюючої змінної на залежну є несуттєвим і ним можна знехтувати.

Для статистично значимих параметрів моделі можна побудувати їх інтервали довіри. **Інтервал довіри** – це інтервал, у який з наперед заданою ймовірністю потрапляють фактичні, дійсні значення параметрів моделі – тобто параметри теоретичної регресії. Інтервали довіри параметрів моделі представляють по суті інтервальні оцінки дійсних параметрів на відміну від точкових $a_j, (j = \overline{0, m})$. Ймовірність p з якою визначаються довірчі інтервали називається **довірчою** (або **рівнем довіри**) і пов’язана з рівнем значимості α наступною залежністю :

$$p = 1 - \alpha .$$

Самі довірчі інтервали обчислюються за наступною залежністю :

$$a_j - t_{\alpha/2} \cdot \hat{\sigma}_{a_j} < \alpha_j < a_j + t_{\alpha/2} \cdot \hat{\sigma}_{a_j}, \quad (j = \overline{0, m}),$$

де a_j - оцінка параметра моделі (вибірковий параметр моделі),

$\hat{\sigma}_{a_j}$ - стандартна похибка цього параметра,

$t_{\alpha/2}$ - критичне значення критерію Ст'юдента для рівня значимості α і ступеня вільності $\nu = n - k$.

1.5 Прогнозування із застосуванням економетричної моделі

Якщо побудована модель є адекватною і статистично значимою, її можна застосовувати для **прогнозування** або **передбачення** значень залежної змінної моделі.

Про **прогнозування** кажуть, коли в часових рядах (динамічній вибірці) прогнозний період настає пізніше, ніж базовий. Якщо ж модель побудована за просторовою вибіркою, прогноз здійснюється на основі очікуваних значень пояснюючих змінних, які перебувають за межами застосованої вибірки. Про **передбачення** кажуть, коли в часових рядах (динамічній вибірці) прогнозний період не виходить за базовий. Якщо ж модель побудована за просторовою вибіркою, прогноз здійснюється на основі очікуваних значень пояснюючих змінних, які також не виходять за межі застосованої вибірки. Внаслідок цього прогнозні значення залежної змінної завжди є менш точними, ніж передбачувані, але вони є більш інформативними і бажаними в економічному аналізі. Якість прогнозу тим краща, чим повніше виконуються передумови моделі у прогнозний часовий період чи у прогнозному просторі, чим надійніше обчислено параметри моделі і більш точно визначено прогнозні, очікувані значення пояснюючих змінних.

Нехай відомо вектор прогнозних значень пояснюючих змінних

$$X_{pr} = \begin{pmatrix} 1 \\ x_{1,pr} \\ x_{2,pr} \\ \dots \\ x_{m,pr} \end{pmatrix},$$

де $x_{1,pr}, x_{2,pr}, \dots, x_{m,pr}$ - прогнознi значення пояснюючих змiнних. Тодi можна отримати два види прогнозу стосовно залежної змiнної економетричної моделi: **точковий та iнтервальний**.

Точковий прогноз представляє собою **точкову оцiнку математичного сподiвання** залежної змiнної моделi i дає можливість обчислити **наближене середнє прогнозне значення** залежної змiнної. Точкове прогнозне значення обчислюється на основi оцiненого рiвняння регресiї за наступною залежнiстю :

$$\hat{y}_{pr} = X'_{pr} \cdot A, \quad (14)$$

де X'_{pr} - транспонований вектор прогнозних значень пояснюючих змiнних моделi, B – вектор оцiнок параметрiв моделi.

1.6 Економiко-математичний аналіз

Показники середньої ефективностi впливу пояснюючих змiнних на залежну змiнну

Середня ефективнiсть впливу деякої пояснюючої змiнної x_j на залежну змiнну y визначається за наступною залежнiстю :

$$Es_j = \frac{\bar{y}}{\bar{x}_j}, \quad (j = \overline{1, m}), \quad (15)$$

де \bar{y} - середнє значення залежної змiнної, обчислене для масиву розрахункових значень залежної змiнної, \bar{x}_j - середнє значення j – iї пояснюючої змiнної. Цей показник показує на скiльки y середньому змiнить своє значення залежна змiнна моделi, якщо деяка пояснююча змiнна x_j змiнить своє значення на одиницю при умовi, що iншi пояснюючi змiннi моделi будуть незмiнними. За економiчним змiстом показники середньої ефективностi впливу можуть розглядатися як середня продуктивнiсть працi, середня продуктивнiсть основного капiталу i т.і.

Показники граничної ефективностi впливу пояснюючих змiнних на залежну змiнну

Гранична ефективнiсть впливу деякої пояснюючої змiнної x_j на залежну змiнну y визначається за наступною залежнiстю :

$$M_j = \frac{\partial \hat{y}}{\partial x_j} = a_j, \quad (j = \overline{1, m}), \quad (16)$$

де a_j - параметр вибіркового рівняння регресії при пояснючій змінній \bar{x}_j .

Цей показник показує на скільки змінить своє значення залежна змінна моделі, якщо деяка пояснююча змінна x_j змінить своє значення на одиницю при умові, що інші пояснюючі змінні моделі будуть незмінними.

Як видно з виразу (16) гранична ефективність впливу деякої пояснюючої змінної x_j на залежну змінну y визначається значенням відповідного коефіцієнта регресії a_j , що дає можливість надавати кожному з цих коефіцієнтів відповідного економічного змісту. Так наприклад, за економічним змістом параметри лінійного рівняння регресії можуть представляти собою граничну схильність до споживання (MPC), граничну продуктивність праці (MP_L) або основного капіталу (MP_K) і т. ін.

Показники відносного впливу пояснюючих змінних на залежну змінну

Параметри вибіркового рівняння регресії характеризують **абсолютний вплив** кожної пояснюючої змінної на залежну. Крім цього на основі загальної лінійної економетричної моделі можна оцінити **відносний вплив** кожної з пояснюючих змінних на залежну. Такий вплив характеризує **частковий середній коефіцієнт еластичності**, який визначається для будь-якої пояснюючої змінної за наступною залежністю:

$$\bar{E}_{y/x_j} = \frac{M_j}{Es_j} = a_j \frac{\bar{x}_j}{\bar{y}}, (j = 1, m). \quad (17)$$

Як відомо цей коефіцієнт характеризує еластичність поведінки залежної змінної моделі при змінні значення змінної x_j і показує на скільки відсотків у середньому змінить своє значення залежна змінна (залежний економічний показник) при зміні значення змінної x_j на 1% при незмінних значеннях інших пояснюючих змінних моделі.

Крім часткових коефіцієнтів еластичності також можна визначити і **загальний коефіцієнт еластичності**. Він показує на скільки відсотків у середньому змінить своє значення залежна змінна (залежний економічний показник) при одночасній зміні значень усіх пояснюючих змінних моделі на 1%. Загальний коефіцієнт еластичності обчислюється за наступною залежністю :

$$p = \sum_{j=1}^m \bar{E}_{y/x_j}. \quad (18)$$

Показники сили впливу пояснюючих змінних на залежну змінну

При дослідження економічних явищ і процесів дуже часто виникає питання – які з пояснюючих змінних впливають більше на залежну змінну, а які менше. Це дає можливість визначити вагомі і ефективні важелі впливу на залежний економічний показник у необхідному напрямку.

Який показник потрібно використати у цьому випадку? Можливо параметри моделі, які як зазначалося вище, дозволяють оцінити абсолютний вплив кожної пояснюючої змінної на залежну. Але оскільки пояснюючі змінні моделі можуть бути виражені у різних одиницях вимірювання, то і відповідні коефіцієнти регресії a_1, a_2, \dots, a_m також можуть мати різні одиниці вимірювання. Це не дає можливість застосувати їх безпосередньо для порівняння сили впливу кожної пояснюючої змінної на залежну.

Для цієї мети використовуються **стандартизовані коефіцієнти регресії**, які обчислюються за наступною залежністю :

$$a_j^* = a_j \frac{\hat{\sigma}_{x_j}}{\hat{\sigma}_y}, (j = \overline{1, m}), \quad (19)$$

де a_j - коефіцієнт регресії при пояснюючій змінній x_j ,

$\hat{\sigma}_{x_j}$ - стандартна похибка пояснюючої змінної x_j ,

$\hat{\sigma}_y$ - стандартна похибка залежної змінної моделі.

Стандартизований коефіцієнт регресії представляє собою очікувану абсолютну зміну залежної змінної y (у стандартизованих одиницях вимірювання $\hat{\sigma}_y$), викликану зміною x_j на одну відповідну стандартизовану одиницю (тобто на $\hat{\sigma}_{x_j}$) при незмінних значеннях інших пояснюючих змінних. Кожний стандартизований коефіцієнт регресії вимірюється в одиницях стандартних відхилень y на одне стандартне відхилення змінної x_j . Абсолютні значення стандартизованих коефіцієнтів регресії можна порівнювати і виконувати ранжування пояснюючих змінних за ступенем впливу їх на залежну змінну моделі.

Лабораторна робота № 1 «Економетрична модель парної лінійної регресії»

Мета роботи: Набуття практичних навичок побудови економетричної моделі у вигляді парної класичної лінійної регресії, її верифікації і практичного використання в економічних дослідженнях.

Завдання роботи і вихідні дані

Громадськість промислового центру N не на жарт стривожена результатами спостережень місцевої станції екологічного контролю за останні 45 років. В результаті впливу антропогенних факторів (індустріального зростання, урбанізації, забруднення природного середовища) в річці з року в рік зменшується рівень води (таблиця 1).

Таблиця 1 - Залежність середньорічного рівня води в річці від річного споживання містом водних ресурсів

Показники	Спостережувані значення				
Роки	1965	1970	1975	1980	1985
Споживання води (тис. м ³ /рік)	1200	1900	2500	4500	7800
Рівень води (см)	315	302	290	284	271
Роки	1990	1995	2000	2005	2010
Споживання води (тис. м ³ /рік)	12400	24000	32000	39000	50000
Рівень води (см)	258	237	211	202	174

Через те, що берега водойми - улюблене місце відпочинку городян, місцевим вченим було дано завдання: дослідити, до чого призведе збереження сформованих небезпечних тенденцій.

Грунтуючись на наведених статистичних даних необхідно:

1. Побудувати модель регресії, яка описує залежність глибини річки від середньорічного споживання води
2. Оцінити адекватність і точність побудованої моделі
3. Розрахувати коефіцієнт еластичності.
4. Продовжити сформовані тенденції на майбутнє
5. Дати відповідь на питання: в якому році річка як така, припинить своє існування?

Порядок виконання роботи:

Вказівки до виконання завдання в пакеті MathCad

1. Поставити вихідні дані у векторній формі Y (10 * 1) і X (10 * 1).

2. Побудувати кореляційне поле. На основі результатів побудови висловити припущення про наявність або відсутність якої-небудь закономірності.

3. Оцінити тісноту зв'язку між обсягами споживання води і глибиною водойми, розрахувавши для цього коефіцієнт кореляції Пірсона із використанням вбудованої в пакеті MathCad функції $corr(x,y)$.

4. Задати матрицю H ($10 * 2$):

$$i := 0..9$$

$$H_{i,0} := 1 \quad H_{i,1} := X_i$$

5. Оцінити вектор невідомих параметрів за методом найменших квадратів формула (7). Здійснити висновки про ступінь впливу середньорічного споживання води на глибину водойми.

6. Розрахувати залишки в кожному спостереженні за формулою (9).

7. Розрахувати коефіцієнт детермінації за формулою (11). Зробити висновок про ступінь впливу пояснюватиме змінної моделі на залежну змінну. Перевірити значущість моделі у цілому із використанням критерію Фішера (формула (12)).

8. оцінити їх статистичну значущість параметрів моделі за критерієм Ст'юдента. Для цього спочатку оцінити незміщену оцінку дисперсії залишків за формулою (8). Далі розрахувати дисперсійно-коваріаційну матрицю за формулою (10). Потім оцінити стандартну помилку параметрів моделі $\hat{\sigma}_{a_j}, (j = \overline{1,m})$ та оцінити їх статистичну значущість за критерієм Ст'юдента - формула (13).

9. Оцінити тенденцію зміни середньорічного споживання води наступним чином. Спочатку розрахувати

$$\Delta x = (x_{\max} - x_{\min}) / (n - 1)$$

А потім безпосередньо самі прогнозні значення фактору

$$x_{n+k} = x_n + \Delta x * k, \quad k = 1, 2, \dots$$

де k – періоди прогнозу.

10. Продовжити сформовану тенденцію на майбутнє і оцінити рівень води у водоймі наступним чином:

$$y_{n+k} = a_0 + a_1 * k, \quad k = 10, 11, 12, \dots$$

13. Оцінити через скільки років водоймище пересохне. Проаналізувати отримані результати.

Завдання та запитання для самоконтролю.

1. Який загальний вигляд має економетрична модель парної лінійної регресії і її структура?

2. Математична і економічна інтерпретація параметрів парної лінійної регресії.

3. Як визначається вибірковий коефіцієнт парної кореляції, його властивості і застосування?

4. Як визначається вибірковий коефіцієнт детермінації для парної лінійної регресії, його властивості і застосування?

5. За яким критерієм і як здійснюється перевірка загальної статистичної значимості моделі парної лінійної регресії?

6. За яким критерієм і як здійснюється перевірка статистичної значимості оцінок параметрів моделі парної лінійної регресії?

7. За яким критерієм і як здійснюється перевірка статистичної значимості вибіркового коефіцієнта парної кореляції для лінійної парної регресії?

8. Для чого будуються інтервали довіри для параметрів парної лінійної регресії?

9. Призначення та застосування точкового прогнозу.

10. Як оцінюється граничний вплив пояснюючої змінної на залежну у випадку парної регресії?

11. Як оцінюється відносний вплив пояснюючої змінної на залежну у випадку парної регресії?

Лабораторна робота №2 «Побудова двофакторного рівняння регресії»

Мета роботи: Набуття практичних навичок побудови економетричної моделі у вигляді двофакторної класичної лінійної регресії, її верифікації і практичного використання в економічних дослідженнях

Завдання роботи і вихідні дані

Чудова корова кота Матроскіна радувала чудовим надоем, і тому він намірився надлишки молока продавати. При цьому кіт Матроскін вирішив з'ясувати, яким чином обсяг щоденної продажу молока у (літрів в день) залежить від а) присутності серед покупців бабусь з онуками (їх частка від загального числа покупців x_1 ,%) і б) участі в комерції пса Шарика (відносне час x_2 , коли він допомагав працювати за прилавком,%). Ретельні спостереження кіт Матроскін вів протягом 20 робочих днів, результати яких представив у табличній формі (табл. 2).

Таблиця 2 - Вихідні дані про ефективність продажу молока

№ дня продажу	1	2	3	4	5	6	7	8	9	10
у, л/день	6	4,6	4,4	4,5	5,5	4,8	5,1	5,2	7	5,3
x_1 , %	40	20	31	32	34	35	37	32	39	35
x_2 , %	30	33	20	25	29	20	21	20	35	30
№ дня продажу	11	12	13	14	15	16	17	18	19	20
у, л/день	7,5	7,7	7,3	7	6,7	5,7	6	6,4	7,1	6,3
x_1 , %	50	37	50	38	50	35	46	49	51	45
x_2 , %	35	30	40	42	39	35	36	38	41	34

Необхідно з'ясувати, чи впливають зазначені фактори (x_1 , x_2) на комерційну діяльність кота Матроскіна в області молочного бізнесу, а якщо це так, то наскільки відчутно.

Грунтуючись на наведених статистичних даних:

1. Побудувати рівняння множинної регресії
2. Провести верифікацію побудованої моделі
3. Пояснити характер впливу факторів на досліджуваний показник

Порядок виконання роботи

1. Ввести вихідні дані, наведені в таблиці.
2. Оцінити наявність впливу факторів (x_1 , x_2) на обсяг продажів молока, розрахувавши відповідні коефіцієнт кореляції.

3. Оцінити параметри багатофакторного рівняння, використовуючи МНК. Задати для цього матрицю H і оцінити невідомі параметри A за формулою (7).

4. Оцінити якість побудованої моделі, використовуючи коефіцієнт детермінації формула (11) і перевірити його статистичну значущість за формулою (12).

5. Використовуючи t -критерій, перевірити значущість коефіцієнтів рівняння регресії формула (13).

6. Оцінити ступінь впливу факторів, розрахувавши для цього стандартизовані коефіцієнти рівняння регресії формула (19).

7. Оцінити середню помилку апроксимації моделі

$$\bar{A} = \frac{\sum_{i=1}^n |y_i - \hat{y}| / y_i}{n} \cdot 100\%$$

8. Допоможіть коту Матроскіну оцінити обсяг продажів молока за умови, що частка присутності серед покупців бабусь з онуками склади 40%, а участь у комерції пса Шарика складе 50%.

9. Оцініть, як зміниться обсяг продажів молока, якщо частка присутності серед покупців бабусь з онуками зросте до 60%, а участь у комерції пса Шарика скоротиться до 30%.

Завдання та запитання для самоконтролю

1. Як специфікується економетрична модель багатофакторної лінійної регресії, її структура і математичний зміст її параметрів?

2. За яким критерієм і як здійснюється перевірка загальної статистичної значимості моделі багатофакторної лінійної регресії?

3. За яким критерієм і як здійснюється перевірка статистичної значимості параметрів моделі багатофакторної лінійної регресії?

4. Для чого і як будуються інтервали довіри параметрів моделі багатофакторної лінійної регресії?

5. Для чого і як будуються прогнози для моделі багатофакторної лінійної регресії?

6. Як оцінити абсолютний граничний вплив кожної пояснюючої змінної багатофакторної лінійної економетричної моделі на залежну змінну?

7. Як оцінити відносний вплив кожної пояснюючої змінної багатофакторної лінійної економетричної моделі на залежну змінну?

8. Як оцінити силу впливу кожної пояснюючої змінної багатофакторної лінійної економетричної моделі на залежну змінну?

Лабораторна робота №3 «Багатофакторне рівняння регресії»

Мета роботи: Набуття практичних навичок побудови економетричної моделі у вигляді багатофакторної класичної лінійної регресії та отримання з її допомогою прогнозу економічних показників

Завдання роботи і вихідні дані

Зміну індексу ПФТС (Позабіржова Фондова Торгова Система) наведено у таблиці 3. Необхідно побудувати модель залежності фондового індексу від макроекономічних показників (темпи інфляції, курс долара, дисконтна ставка Національного банку України). У таблицях 4 - 6 наведено зміну макроекономічних показників з часом. Із-за допомогою отриманого ієрархічного багатофакторного рівняння регресії побудувати прогноз значень індексу ПФТС на наступні два моменти часу.

Таблиця 3- Динаміка індексу ПФТС

Моменти часу	1	2	3	4	5	6	7	8	9	10	11	12
Значення фондового індексу	60,75	57,39	56,62	57,05	55,45	56,1	58,28	56,13	55,12	56,13	56,82	56,74
Моменти часу	13	14	15	16	17	18	19	20	21	22	23	24
Значення фондового індексу	54,84	54,88	52,84	53,88	52,98	52,67	51,27	53,71	53,14	51,04	53,04	52,2

Таблиця 4 – Динаміка темпів інфляції

Моменти часу	1	2	3	4	5	6	7	8	9	10	11	12
Темпи інфляції	98,5	98,5	99,8	99,8	99,8	100,2	100,2	100,2	100,7	100,7	100,7	100,7
Моменти часу	13	14	15	16	17	18	19	20	21	22	23	24
Темпи інфляції	100,7	100,7	100,7	100,7	100,7	101,4	101,4	101,5	101,5	101,1	101,1	100,7

Таблиця 5 – Динаміка курсу долара

Моменти часу	1	2	3	4	5	6	7	8	9	10	11	12
Значення курсу долара	25,32	25,38	25,33	25,39	25,33	25,39	25,33	25,39	25,31	25,33	25,32	25,34
Моменти часу	13	14	15	16	17	18	19	20	21	22	23	24
Значення курсу долара	25,33	25,33	25,33	25,33	25,34	25,36	25,39	25,38	25,39	25,37	25,38	25,37

Таблиця 6 – Динаміка дисконтної ставки Національного банку України

Моменти часу	1	2	3	4	5	6	7	8	9	10	11	12
Дисконтна ставка НБУ	8,0	8,0	8,0	8,0	8,0	8,0	8,0	8,0	8,0	8,0	8,0	8,0
Моменти часу	13	14	15	16	17	18	19	20	21	22	23	24
Дисконтна ставка НБУ	8,0	8,0	8,0	8,0	8,0	8,0	7,0	7,0	7,0	7,0	7,0	7,0

Грунтуючись на наведених статистичних даних:

1. Побудувати багатofакторну модель залежності фондового індексу від макроекономічних показників (темпи інфляції, курс долара, дисконтна ставка Національного банку України).

2. Для кожного макроекономічного показника отримати прогнозні значення

3. Застосовуючи отримане багатofакторне рівняння регресії та прогнозні оцінки макроекономічних показників побудувати прогноз значень індексу ПФТС на наступні два моменти часу.

4. Здійснити аналіз впливу факторів на індекс ПФТС

Порядок виконання роботи

1. Ввести вихідні дані, які наведено у таблицях.

2. Побудувати модель залежності значень індексу ПФТС від макроекономічних показників виду:

$$Y = a_0 + a_1F_1 + a_2F_2 + a_3F_3,$$

де Y - значення індексу ПФТС;

F_1, F_2, F_3 – макроекономічні показники, які впливають на значення фондового індексу (темпи інфляції, курс долара, дисконтна ставка Національного банку України).

Оцінити параметри багатofакторного рівняння, використовуючи МНК. Задати для цього матрицю N і оцінити невідомі параметри A за формулою (7).

3. Здійснити оцінку якості побудованої моделі з використанням критерію Фішера за формулами (11) – (13).

4. Для кожного макроекономічного показника отримати прогнозні значення наступним чином:

4.а.

$$F_i^{пр} = \frac{(F_i^{\max} - F_i^{\min})}{(n - 1)} * (n + k), \quad k = 1, 2, \dots$$

4.б. (для всіх факторів за виключенням дисконтної ставки НБУ)

$$F_i^{IP} = (F_i^n - F_i^{n-1}) * (n + k), \quad k = 1, 2, \dots$$

6. Здійснити прогноз значень індексу ПФТС на наступні два періоди:

$$Y^{IP} = a_0 + a_1 F_1^{IP} + a_2 F_2^{IP} + a_3 F_3^{IP} .$$

7. Зробити висновки щодо поведінки індексу ПФТС.

8. Здійснити аналіз впливу факторів на індекс ПФТС формули (15) – (19).

Завдання та запитання для самоконтролю: див. до лабораторної роботи №2.

3 МЕТОДИ ПОБУДОВИ ЗАГАЛЬНОЇ ЛІНІЙНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ

3.1 Характеристика основних методів побудови загальної лінійної економетричної моделі

На кожний економічний показник впливає безліч різних факторів. При ідентифікації загальної лінійної економетричної моделі виникає питання, які саме з них потрібно увести до моделі у якості пояснюючих змінних. Для вирішення цього питання в принципі існує два наступні підходи.

1-й підхід. З точки зору надійності прогнозування, необхідно включати до моделі якомога більше факторів. Але при цьому, кожна „зайва” пояснююча змінна погіршує ситуацію, оскільки вона зменшує надійність F-тесту на загальну статистичну значимість моделі і може призвести до невірних статистичних висновків.

2-й підхід. З точки зору отримання надійної статистичної інформації по кожному фактору слід прагнути, щоб модель мала якомога менше факторів, оскільки збір і обробка великих статичних масивів потребує великих затрат при недостатній надійності статистичних даних.

Компромідом між цими крайніми підходами є те, що називають вибором "найкращого рівняння" регресії. Для реалізації такого вибору немає єдиної статистичної процедури, єдиного методу. Існує декілька методів побудови "найкращої" лінійної регресії, найбільш поширеними серед яких є:

- 1) метод усіх можливих регресій ;
- 2) метод покрокової регресії ;
- 3) метод виключень .

Метод усіх можливих регресій – історично перший метод побудови лінійних регресійних моделей і найбільш громіздкий серед усіх методів. Ідея методу полягає у побудові множини регресійних рівнянь, які містять усі можливі комбінації попередньо відібраних факторів, і у порівнянні цих рівнянь за трьома критеріями : коефіцієнтом детермінації R^2 , стандартною похибкою $\hat{\sigma}_\varepsilon$ і критерієм Меллоуза C_p . У загальному випадку для m відібраних факторів (пояснюючих змінних) можна побудувати 2^m рівнянь регресії і виконати їх порівняння.

Побудова і аналіз усіх можливих регресійних рівнянь є доволі громіздка і ненадійна процедура, тому цей метод рекомендується використовувати при невеликій кількості відібраних факторів.

Метод покрокової регресії є найпоширенішим на практиці і більш економічним у порівнянні з попереднім. Ідея методу полягає у послідовному

включенні до моделі факторів (пояснюючих змінних) до тих пір, поки модель не стане задовільною. Порядок включення факторів до моделі вибирається на основі значень коефіцієнтів парної кореляції між пояснюючими і залежною змінною моделі. Алгоритм методу покрокової кореляції можна подати у наступному вигляді.

Алгоритм методу:

1. Розраховується кореляційна матриця \mathbf{r} для усіх змінних моделі, які планується включити до моделі.

2. Спочатку з кореляційної матриці вибирається і включається до моделі той фактор $x_j, (j=\overline{1,m})$, якому у кореляційній матриці відповідає найбільший за модулем коефіцієнт парної кореляції з залежною змінною моделі y (нехай це буде змінна x_1). Будується регресійне рівняння з однією незалежною змінною $\hat{y} = b_0 + b_1x_1$ і для нього обчислюється коефіцієнт детермінації. Після цього перевіряється чи значима ця змінна за коефіцієнтом детермінації і за частковим F- критерієм. Якщо ні, то приймаємо $y = \bar{y}$ і процес побудови моделі припиняється. Якщо так, то переходимо до наступного кроку 3.

3. На основі аналізу кореляційної матриці серед тих пояснюючих змінних, що залишились, шукаємо нову змінну, яка має найбільший за модулем коефіцієнт кореляції з y і включаємо її до моделі (нехай це буде змінна x_2).

4. Будується нове рівняння регресії :

$$\hat{y} = b_0 + b_1x_1 + b_2x_2$$

і для нього розраховується звичайний R^2 і оцінений \bar{R}^2 коефіцієнт детермінації. Аналізується зміна цих показників у порівнянні з попередньою моделлю. Потім розраховуються часткові F- критерії для кожного фактору. Серед них обирається найменше значення і порівнюється із заздалегідь обраним критичним значенням F - критерію. В залежності від результатів перевірки додана на цьому кроці змінна або залишається у моделі, або відкидається.

5. Після цього модель перераховується в залежності від факторів, які залишились і здійснюється перехід до кроку 3.

Процес побудови моделі за наведеним алгоритмом припиняється, якщо жодний фактор, що знаходиться у поточному рівнянні, не вдається виключити, а новий претендент на включення не відповідає частковому F - критерію.

Метод виключень діє у зворотному порядку порівняно з методом покрокової регресії і є також досить поширеним. Загальний алгоритм методу складається з 5 кроків.

1. Будується рівняння регресії, яке включає всі відібрані фактори. Якщо попередньо було відібрано m факторів, то вихідне базове рівняння має вигляд :

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m .$$

2. Для кожного фактору (пояснюючої змінної) $x_j, (j = \overline{1, m})$ обчислюється значення часткового F- критерію.

3. Серед розрахованих значень часткового F- критерію вибирається найменше F_{\min} і порівнюється із заздалегідь обраним критичним значенням розподілу Фішера $F_{кр}$.

4. Якщо $F_{\min} < F_{кр}$, то відповідний фактор виключається з рівняння. Проводиться новий розрахунок регресійного рівняння вже без виключеного фактора і здійснюється перехід знову до кроку 2.

5. Якщо $F_{\min} > F_{кр}$, то регресійне рівняння залишається без змін.

3.2 Статистичні показники, які використовуються при побудові загальної лінійної економетричної моделі

При розгляді методів побудови загальної лінійної економетричної моделі були використані такі нові поняття і статистичні показники, як **частковий F – критерій, оцінений коефіцієнт детермінації, кореляційна матриця**. Розглянемо ці показники більш детально.

Частковий F- критерій.

Одним з головних питань будь-якого методу побудови багатofакторної регресійної моделі є питання визначення суттєвості впливу на залежну змінну у окремих факторів. Таку оцінку можна зробити з використанням F- статистики на основі часткового F- критерію Фішера. Зміст часткового F- критерію розглянемо на наступному прикладі.

Нехай є економетрична модель, яка враховує вплив k факторів на залежну змінну y , тобто :

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + e .$$

Припустимо, що з k факторів p факторів несуттєво впливають на показник y . Тоді побудуємо другу регресійну модель, в яку не включаємо ці p факторів.

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_qx_q + e, \quad q < k; \quad p = k - q .$$

Позначимо суму квадратів залишків 1-ї моделі через SSE_1 , а 2-ї моделі - через SSE_2 . Тоді різниця $SSE_1 - SSE_2$ дорівнює додатковій сумі квадратів залишків, яка пов'язана з включенням (або вилученням) до 1-ї моделі p

додаткових факторів. Зазначимо, що ця додаткова сума квадратів буде мати ступінь вільності $p = k - q$.

Знайдемо наступне розрахункове значення F- статистики :

$$F_{p,n-k} = \frac{SSE_1 - SSE_2}{SSE_1} \cdot \frac{n-k}{p} . \quad (20)$$

Для заданого рівня значимості α і ступенів вільності $v_1 = p$ і $v_2 = n - k$ за статистичними таблицями F- розподілу знаходимо критичне значення критерію Фішера $F_{кр}$. Якщо $F_{p,n-k} > F_{кр}$, то із надійністю $1 - \alpha$ можна вважати, що вилучені фактори суттєво впливають на результуючий показник y і їх потрібно залишити у складі пояснюючих змінних моделі. У протилежному випадку ($F_{p,n-k} < F_{кр}$) з надійністю $1 - \alpha$ можна вважати, що виключення із моделі p факторів несуттєво впливає на показник y і для моделювання можна вибрати другу модель з q пояснюючими змінними.

Якщо розглядати процес поступово включення (або вилучення) факторів до моделі, коли на кожному етапі до моделі включається (або вилучається) **тільки один фактор** за допомогою наведеного F- відношення (20) будемо мати критерій, який і визначатиме суттєвість впливу цього окремого фактору на залежну змінну y . Такий варіант F- критерію називається **частковим F-критерієм** і визначається за формулою (20) для $p=1$.

Оцінений коефіцієнт детермінації

Важливою властивістю коефіцієнта детермінації R^2 є те, що він – неспадна функція від кількості факторів, які входять до моделі. Якщо кількість факторів зростає, то R^2 також зростає і ніколи не зменшується. Це ускладнює порівняння економетричних моделей і вибір серед них найкращої. Так наприклад, якщо ми порівнюємо дві економетричні моделі з однаковою залежною змінною, але різною кількістю пояснюючих змінних, ми звичайно віддаємо перевагу тій, яка має більше значення R^2 , хоча це може і не відповідати дійсності.

Тому щоб запобігти невиправданому розширенню моделі і **мати можливість порівнювати моделі** з різною кількістю факторів вводять так званий **оцінений коефіцієнт детермінації** \bar{R}^2 , який зменшує вплив зростання кількості факторів на коефіцієнт детермінації за рахунок поправки на ступені вільності. У практиці економетричного дослідження використовуються два різновиди оціненого коефіцієнта детермінації :

- коефіцієнт детермінації, скоригований за Тейлом - \bar{R}_T^2 ;
- коефіцієнт детермінації, скоригований за Аемією - \bar{R}_A^2 .

Перший з них обчислюється за наступною залежністю :

$$\bar{R}_T^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \frac{n-1}{n-k},$$

а другий – за залежністю :

$$\bar{R}_A^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \frac{n+m}{n-k}.$$

Коефіцієнт детермінації R^2 і оцінені коефіцієнти детермінації \bar{R}_T^2 та \bar{R}_A^2 пов'язані між собою наступними співвідношеннями :

$$\bar{R}_T^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k},$$

$$\bar{R}_A^2 = 1 - (1 - R^2) \cdot \frac{n+m}{n-k}.$$

Вочевидь, для кожного оціненого коефіцієнта детермінації виконується нерівність $\bar{R}^2 \leq R^2$, тобто зі збільшенням числа пояснюючих змінних моделі оцінені коефіцієнти детермінації зростають повільніше, ніж R^2 , зменшуючи таким чином вплив числа факторів на величину коефіцієнта детермінації. Крім того, якщо $R^2 = 1$, то і $\bar{R}^2 = 1$. Якщо R^2 прямує до нуля, оцінені коефіцієнти кореляції стають від'ємними. Така властивість скоригованих коефіцієнтів детермінації дає змогу більш об'єктивно оцінювати якість моделей з різним числом факторів.

Кореляційна матриця

Кореляційна матриця дозволяє оцінити щільність лінійного кореляційного зв'язку між залежною змінною моделі і окремими факторами, а також між окремими незалежними змінними. У загальному випадку вона представляє собою квадратну симетричну матрицю, елементами якої є коефіцієнти парної кореляції між залежною змінною моделі і кожною пояснюючою змінною, а також коефіцієнти парної кореляції між самими пояснюючими змінними моделі.

Для випадку m пояснюючих змінних кореляційна матриця має наступний вигляд і структуру :

$$r = \begin{pmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_m} \\ r_{x_1y} & r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_m} \\ r_{x_2y} & r_{x_2x_1} & r_{x_2x_2} & \cdots & r_{x_2x_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_my} & r_{x_mx_1} & r_{x_mx_2} & \cdots & r_{x_mx_m} \end{pmatrix}, \dim r = (m+1) \times (m+1).$$

Діагональні елементи матриці r дорівнюють 1. Коефіцієнти парної кореляції r_{yx_j} та $r_{x_jx_k}$ обчислюються за формулою визначення коефіцієнта парної кореляції.

Будь-який коефіцієнт парної кореляції з наведеної вище кореляційної матриці характеризує тісноту зв'язку між відповідними змінними за умови, що інші змінні ведуть себе „природним чином” – тобто також змінюють свої значення разом з тими, для яких обчислюється коефіцієнт парної кореляції. Це не дає можливість оцінити тісноту кореляційного зв'язку між двома змінними моделі так би мовити у „чистому вигляді”.

Тому при вивченні зв'язку між змінними економетричної моделі недостатньо спиратися тільки на кореляційну матрицю. Необхідно також проаналізувати **часткові коефіцієнти кореляції**. На відміну від парних часткові коефіцієнти кореляції характеризують тісноту зв'язку між двома змінними за умови, що інші змінні сталі. Такі коефіцієнти кореляції більш коректно вимірюють силу лінійного кореляційного зв'язку між змінними моделі і дають більш точну інформацію, необхідну при побудові загальної лінійної економетричної моделі. Формула для визначення часткового коефіцієнта кореляції між двома змінними моделі має вигляд :

$$r_{jk,12\dots m+1} = \frac{c_{jk}}{\sqrt{c_{jj} \cdot c_{kk}}}, (j = \overline{1, m+1}), (k = \overline{1, m+1}),$$

де c_{jk}, c_{jj}, c_{kk} – елементи матриці C , оберненої до кореляційної матриці r .

За цією формулою визначається частковий коефіцієнт кореляції між економічним показником j і показником k при умові, що усі інші економічні показники, які фігурують в економетричній моделі є сталими. Слід зазначити, що у якості показника j може виступати як залежна змінна моделі так і незалежна, а у якості показника k – незалежна, пояснююча змінна.

Лабораторна робота №4 «Метод виключень»

Мета роботи: Набуття практичних навичок побудови загальної економетричної моделі із використанням методу виключень

Завдання роботи і вихідні дані

Маємо 12 транспортних підприємств для яких досліджується залежність чистого річного прибутку (Y , млн. грн.) від кількості вантажних автомобілів із дозволеною максимальною вагою: а) не більш 3,5 т (X_1 , од.), б) понад 3,5 т (X_2 , од.). У якості фактору також враховується форма власності (фіктивна змінна X_3 : 0 - муніципальне підприємство, 1 - приватне):

Таблиця 7 – Вихідні дані

№ підприємства	Y	X_1	X_2	X_3
1	13	23	9	0
2	22	28	8	1
3	17	20	12	0
4	19	28	13	0
5	24	25	7	1
6	33	27	20	1
7	8	13	10	0
8	17	25	10	0
9	32	36	10	1
10	21	23	17	1
11	27	28	8	1
12	20	23	10	0
Середнє значення	21,1	24,9	11,2	0,50

Грунтуючись на наведених статистичних даних:

1. Побудувати трьох факторну модель.
2. Оцінити якість моделі та здійснити інтерпретацію отриманих результатів
3. Застосувати метод виключень.
4. Обрати найбільш прийнятну модель

Порядок виконання роботи

1. Ввести вихідні дані, наведені в таблиці.
2. Побудувати трьох факторну модель, оцінивши її параметри з використанням МНК. Задати для цього матрицю N і оцінити невідомі параметри A за формулою (7).
3. Оцінити якість побудованої трьох факторної моделі, використовуючи коефіцієнт детермінації формула (11) і перевірити її статистичну значущість за формулою (12).

4. Використовуючи t-критерій, перевірити значущість коефіцієнтів рівняння регресії формула (13).

5. Оцінити середню помилку апроксимації моделі

$$\bar{A} = \frac{\sum_{i=1}^n |y_i - \hat{y}| / y_i}{n} \cdot 100\%$$

6. Розрахувати коефіцієнти кореляції для усіх змінних моделі.

7. Обрати змінну, яка має найменший вплив на y (має найменше за модулем значення коефіцієнту кореляції) та побудувати двох модель, виключивши з неї обрану змінну.

8. Здійснити перевірку чи значима виключена з моделі змінна за частковим F- критерієм формула (20). Якщо так, то припинити процес і обрати трьох факторну модель. Якщо змінна не значима, то перейти до наступного кроку.

9. Обрати наступну змінну, яка має найменший вплив на y (має найменше за модулем значення коефіцієнту кореляції) та побудувати модель парної регресії, виключивши з неї обрану змінну.

8. Здійснити перевірку чи значима виключена з моделі змінна за частковим F- критерієм формула (20). Якщо так, то припинити процес і обрати двох факторну модель. Якщо змінна не значима, то обрати модель парної регресії.

Завдання та запитання для самоконтролю

1. Ідея методу усіх можливих регресій
2. Недоліки методу усіх можливих регресій
3. Алгоритм методу покрокової регресії
4. Алгоритм методу виключень
5. Для чого використовується частковий F- критерій
6. Мета застосування оціненого коефіцієнту детермінації

4 НЕЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

4.1 Загальні поняття і визначення

Багато економічних залежностей, як свідчить економічна теорія, не є лінійними по суті, і тому їх моделювання лінійними залежностями, безумовно, не дасть позитивний результат. У цьому випадку необхідно використовувати **нелінійні економетричні моделі**.

Нелінійна економетрична модель – це регресійна модель, яка встановлює нелінійну залежність між економічними показниками, один з яких є залежною (пояснюваною) змінною, а інші – незалежними (пояснюючими) змінними.

За методами оцінювання параметрів усі нелінійні економетричні моделі поділяються на два типи:

- 1) нелінійні за факторами, але лінійні за параметрами;
- 2) нелінійні за факторами і за параметрами.

Економетричні моделі, які є нелійними за факторами, але лінійними за параметрами називаються **квазілінійними**.

Будь-яка квазілінійна модель у загальному випадку може бути представлена у наступному вигляді:

$$y = a_0 + a_1 f_1(x_1) + a_2 f_2(x_2) + \dots + a_m f_m(x_m) + e \quad (21)$$

де $f_j(x_j)$, ($j = \overline{1, m}$) – може бути різними нелінійними функціями x_j .

Квазілінійні моделі є більш простим і привабливим варіантом нелінійних моделей. Їх привабливість пояснюється тим, що оцінки параметрів таких моделей можуть бути отримані методами лінійного регресійного аналізу. Прикладом такої моделі може бути модель попиту, для якої функція попиту має наступний вигляд :

$$Q = a_0 + a_1 P + a_2 P^2$$

де **Q** – попит; **P** - ціна; **a₀**, **a₁** і **a₂** - параметри моделі.

Нелінійні за факторами і за параметрами економетричні моделі є більш складними і оцінювання їх параметрів, як правило, виконується методами нелінійного регресійного аналізу. Прикладом такої моделі може бути модель, яка описує залежність між об'ємом надходжень до бюджету і податковою ставкою на основі відомої **кривої Лаффера** :

$$Y = a \cdot e^{b(x-c)^2},$$

де Y – податкові надходження; x – податкова ставка; a , b і c – параметри моделі.

Порівнюючи лінійні і нелінійні економетричні моделі можна зробити наступні висновки.

1) Лінійна класична регресійна модель і відповідні методи оцінювання її параметрів, тестування і прогнозування в цілому теоретично краще обгрунтовані, ніж нелінійні.

2) Відносно простоти обчислювального апарату, лінійна модель регресії також не перевершена ніякими іншими типами моделей.

Виходячи із сказаного у сучасній практиці економетричного моделювання при необхідності застосування нелінійних моделей практикуються наступні два підходи:

1) замість складної нелінійної функціональної залежності навіть з невеликою кількістю факторів намагаються застосувати лінійну регресійну функцію з великою кількістю факторів ;

2) за допомогою математичних перетворень намагаються звести (перетворити) нелінійну функцію на лінійну.

Процес приведення нелінійної регресійної моделі до лінійної називається **лінеаризацією**, а сама модель **лінеаризованою** або **лінійною формою**.

Для лінеаризованих моделей повністю зберігається вся методологія економетричного дослідження. Можливість лінеаризації моделі залежить від типу нелінійної моделі.

4.2 Нелінійні економетричні моделі, які зводяться до лінійних Степенева модель

Степенева модель характеризується наступною функцією регресії:

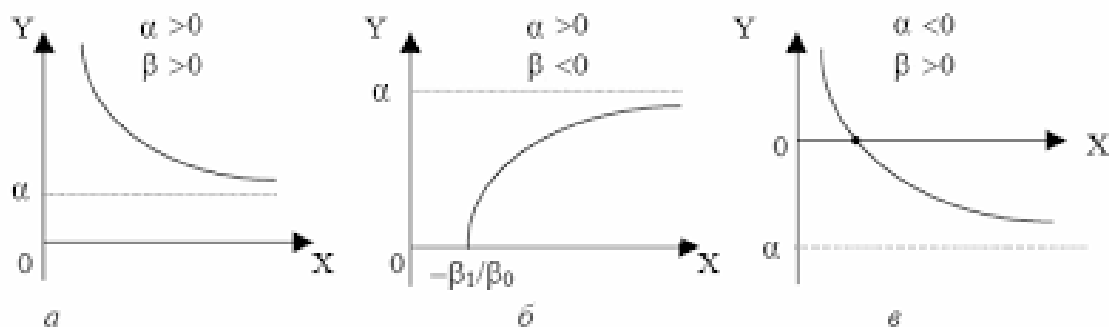
$$y = \alpha \cdot x^\beta, \quad (22)$$

де α , β – параметри моделі.

Нелінійна модель на основі степеневі функції регресії є однією з найпоширеніших у практиці моделей і описує достатньо широкий спектр економічних явищ і процесів, таких як: процес виробництва (виробничі функції), попит на товари різних категорій (криві Енгеля), для опису кривих байдужості та інших.

При моделювання економічних явищ і процесів розглядаються і мають зміст тільки випадки коли $\alpha \geq 0$, що є типовим для економічних процесів. Якщо значення параметру β – не ціле число, то розглядається лише випадок, коли $x \geq 0$. При цьому в залежності від позначки і значення параметра β модель може

описувати різні економічні процеси: **прискорене зростання, уповільнене зростання і спад.**



а) прискорене зростання б) уповільнене зростання в) спад

Рисунок 2 - Графіки степеневі функції

Лінеаризація функції (22) виконується у два кроки. Спочатку виконується логарифмування лівої і правої частини виразу (22):

$$\ln y = \ln \alpha + \beta \cdot \ln x$$

а потім вводиться наступна заміна змінних:

$$y_1 = \ln y, \beta_0 = \ln \alpha, \beta_1 = \beta, x_1 = \ln x,$$

і нелінійна функція (22) зводиться до наступної лінійної форми:

$$y_1 = \beta_0 + \beta_1 x_1.$$

Крім однофакторної степеневі функції (22) широке розповсюдження, особливо при побудові виробничих функцій (виробнича функція Кобба-Дугласа), набули багатфакторні мультиплікативні моделі виду:

$$y = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m}, \quad (23)$$

лінеаризація яких виконується так само, як і для розглянутого випадку однієї пояснюючої змінної.

Слід зазначити, що параметр β у степеневій моделі характеризує еластичність змінної y за змінною x , тобто цей параметр фактично дорівнює коефіцієнту еластичності y за x . Тому часто степеневу модель ще називають

моделлю постійної еластичності, що вказує на можливі напрямки її застосування.

Показникова (експоненційна) модель

Для показникової моделі функція регресії має декілька форм представлення:

Основна:

$$y = \alpha \cdot \beta^x, \quad (24)$$

Еквівалентні:

$$\begin{aligned} y &= \alpha \cdot e^{\beta x}, \\ y &= \alpha(1-r)^x, \\ y &= e^{\beta_0 + \beta_1 x}. \end{aligned} \quad (25)$$

де α , β , β_0 , β_1 , r , – параметри моделей.

Використовується експоненціальна модель для опису **швидкозростаючих** або **швидкоспадних** економічних процесів. Найбільш типовим її застосуванням є ситуація, коли аналізується змінна залежної змінної y з постійним темпом приросту у часі. Наприклад, форма (25) найчастіше використовується у фінансах (r – норма річного відсотка).

Лінеаризація показникової функції виконується, як і для степеневій, шляхом застосування операції логарифмування і подальшої заміни змінних.

Зворотна модель

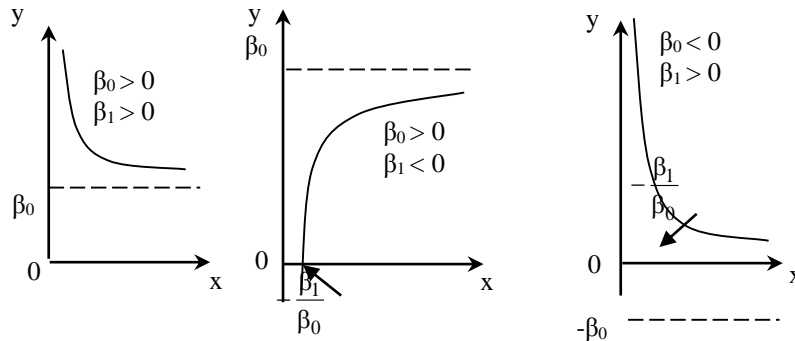
Зворотна функція регресії має вигляд :

$$y = \beta_0 + \beta_1 \frac{1}{x}, \quad (26)$$

де β_0 і β_1 – параметри моделі.

Особливістю функції (26) є те, що при $x \rightarrow \infty$ залежна змінна моделі y прямує до свого граничного значення β_0 , тобто функція (26) має асимптоту β_0 . Внаслідок цього функція (26) застосовується для моделювання економічних явищ і процесів, **обмежених зверху** або **знизу**. Найбільш поширеним застосуванням таких моделей є кількісний опис залежності попиту на блага y від доходу x (криві Торнквіста), а також залежності зміни заробітної плати y від рівня безробіття x (крива Філіпса).

Графік зворотної функції значною мірою залежить від **значень і знаків** параметрів β_0 і β_1 . Вигляд графіків зворотної функції при різних значеннях і знаках параметрів β_0 і β_1 наведено на рис. 3.



0

Рисунок 3 - Графіки зворотної функції

Як видно з (26) зворотна модель відноситься до квазілінійних моделей, оскільки може бути поданою у вигляді (21). У зв'язку з цим лінеаризація зворотної моделі здійснюється простою заміною змінної:

$$z = \frac{1}{x} .$$

Після підстановки в (26) нової змінної отримуємо наступну модель парної лінійної регресії :

$$y = \beta_0 + \beta_1 z + \varepsilon .$$

Квадратичні моделі

Квадратичні моделі використовуються для опису дуже широкого спектру економічних процесів завдяки їхнім універсальним можливостям. У загальному випадку квадратична модель має вигляд:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon , \quad (27)$$

де β_0 , β_1 і β_2 – параметри моделі.

В залежності від знаків і значень параметрів моделі вона може відображати **еволюцію** – дуже різну на різних проміжках інтервалу зміни незалежної змінної

x : зростання, спад, зростання з наступним спадом, спад з наступним зростанням. Так наприклад, квадратична модель може описувати кількісну залежність між обсягом випуску і середніми або граничними видатками, чи між витратами на рекламу і прибутками і т. ін. Можливі варіанти графіків квадратичної функції мають вигляд на рис. 4.

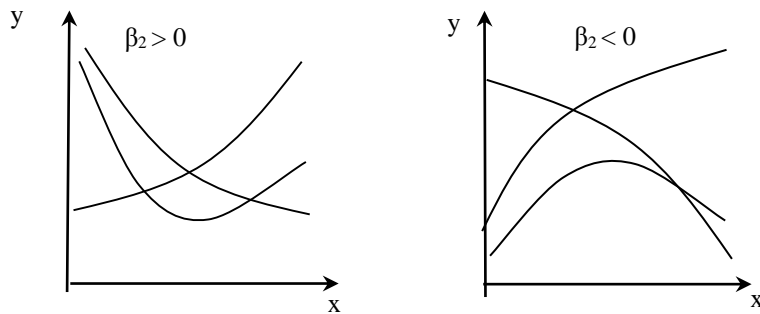


Рисунок 4 - Графіки квадратичної функції

Квадратична модель, як і зворотна, також відноситься до квазілінійних і її лінеаризація виконується шляхом елементарної заміни змінних

$$z_1 = x, z_2 = x^2$$

і зведенням її до двохфакторної лінійної регресії :

$$y = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \varepsilon,$$

параметри якої розраховуються за відомою методикою для лінійних економетричних моделей.

4.3 Прогнозування та економіко-математичний аналіз на основі нелінійних економетричних моделей

Прогнозування на основі нелінійних економетричних моделей

Для побудови **точкового прогнозу** на основі нелінійної економетричної моделі може бути використана як **нелінійна** (звичайна) форма моделі так і **лінійна** форма, отримана у процесі лінеаризації.

При використанні **нелінійної форми** моделі прогнозні значення пояснюючих змінних моделі безпосередньо підставляються у нелінійне рівняння регресії.

При використанні **лінійної форми** моделі для визначення точкових прогнозних значень залежної змінної у випадку **степеневої** або **показникової**

моделі слід пам'ятати ,що до рівняння регресії потрібно підставляти не безпосередньо прогнознi значення пояснюючих змінних ,а перетворені - тобто **логарифми натуральні** прогнозних значень пояснюючих змінних. Слід також пам'ятати ,що отримане точкове прогнозне значення залежної змінної для цих функції потрібно також шляхом зворотних перетворень (експонування) привести до нормального виду.

При використані **квазілінійних моделей** (зворотної і квадратичної) потрібно робити перехід тільки від прогнозних значень пояснюючих до перетворених у відповідності до виду перетворень, які були застосовані при лінеаризації цих моделей. Після обчислення точкового прогнозу зворотне перетворення , як у випадку степеневої і показникової моделей, не потрібне.

Економіко - математичний аналіз на основі нелінійних моделей

Як і у випадку лінійних економетричних моделей на основі нелінійних економетричних моделей можна обчислити наступні показники:

- показники середньої ефективності впливу пояснюючих змінних на залежну ;
- показники граничної ефективності впливу пояснюючих змінних на залежну ;
- показники відносного впливу пояснюючих змінних на залежну – коефіцієнти еластичності.

Ці показники можна визначити через оцінки **параметрів лінійної форми** за наступними відомими співвідношеннями:

$$A_j = \frac{\bar{y}}{\bar{x}_j}, (j = \overline{1, m}), \quad (28)$$

$$M_j = \frac{\partial \hat{y}}{\partial x_j}, (j = \overline{1, m}), \quad (29)$$

$$\bar{E}_{y/x_j} = \frac{M_j}{A_j}, (j = \overline{1, m}), \quad (30)$$

де A_j – показник середньої ефективності впливу пояснюючої змінної x_j на залежну змінну,

M_j – показник граничної ефективності впливу пояснюючої змінної x_j на залежну змінну,

\bar{E} – середній коефіцієнт еластичності, який характеризує відносний вплив пояснюючої змінної x_j на залежну змінну.

Конкретний вираз наведених показників залежить від конкретного вигляду моделі, при цьому наведені формули слід застосовувати саме до **лінійної форми** відповідної нелінійної моделі. Для основних типів нелінійних економетричних моделей, розглянутих вище, формули для визначення цих показників **на основі вибіркової моделі**, подано у таблиці 8.

Таблиця 8 - Основні показники для нелінійних економетричних моделей

Тип моделі	Загальний вигляд вибіркової лінійної форми	M	E
Степенева	$\ln y = b_0 + b_1 \cdot \ln x$	$b_1 \cdot \left(\frac{\bar{y}}{\bar{x}} \right)$	b_1
Показникова	$\ln y = b_0 + b_1 x$	$b_1 \cdot \bar{y}$	$b_1 \cdot \bar{x}$
Зворотна	$y = b_0 + b_1 \frac{1}{x}$	$-b_1 \cdot \left(\frac{1}{\bar{x}^2} \right)$	$-b_1 \cdot \left(\frac{1}{\bar{x} \bar{y}} \right)$
Квадратична	$y = b_0 + b_1 x + b_2 x^2$	$b_1 + 2b_2 \bar{x}$	$\frac{b_1 \bar{x} + 2b_2 \bar{x}^2}{\bar{y}}$

У таблиці: \bar{y} - середнє значення залежної змінної, обчислене для масиву розрахункових значень залежної змінної, \bar{x} - середнє значення пояснюючої змінної, b_0, b_1, b_2 - параметри вибіркової функції регресії.

Лабораторна робота №5 «Нелінійна економетрична модель»

Мета роботи: Набуття практичних навичок побудови економетричної моделі у вигляді нелінійної регресії

Завдання роботи і вихідні дані

Швидкість гоночного автомобілю у перші секунди руху представлена у таблиці 9.

Таблиця 9 – Швидкість гоночного автомобілю в перші секунди руху

t	1	2	3	4	5	6	7	8
$v(t)$	3	6	12	24	48	96	192	384

Грунтуючись на наведених статистичних даних:

1. Побудувати різні нелінійні моделі, які описують швидкість руху гоночного автомобілю.
2. Оцінити якість побудованих моделей.
3. Обрати найбільш прийнятну модель.

Порядок виконання роботи

1. Ввести вихідні дані, наведені в таблиці.
2. Побудувати різні моделі, описуючі залежність швидкості гоночного автомобіля від часу. Як моделі можна вибрати функції наступного виду:

$$a) y(t) = a_0 t^{a_1}, \quad b) y(t) = a_0 a_1^t, \quad c) y(t) = a_0 + a_1 t + a_2 t^2.$$

Привести запропоновані моделі до лінійного вигляду.

3. Здійснити оцінку невідомих параметрів для отриманих лінійних моделей із використанням методу МНК. Задати для цього матрицю N і оцінити невідомі параметри A за формулою (7).

4. Оцінити якість кожної із трьох побудованих лінійних моделей, використовуючи коефіцієнт детермінації формула (11) і перевірити їх статистичну значущість за формулою (12).

5. Шляхом зворотніх перетворень оцінити параметри вихідних нелінійних моделей.

6. Оцінити якість кожної із трьох нелінійних моделей, використовуючи коефіцієнт детермінації формула (11) і перевірити їх статистичну значущість за формулою (12).

7. Для кожної нелінійної моделі оцінити середню помилку апроксимації.

$$\bar{A} = \frac{\sum_{i=1}^n |y_i - \hat{y}| / y_i}{n} \cdot 100\%$$

8. Здійснити висновок о якості та адекватності кожної нелінійної моделі та обрати дві найкращі.

9. Визначити прогнозну швидкість автомобіля на 12-ї секунді руху з використанням обох обраних моделей та порівняти отримані результати.

Завдання та запитання для самоконтролю

1. Які нелінійні економетричні моделі існують
2. Що таке квазілінійна модель
3. Що таке лінеаризація моделі
4. Які методи використовують для оцінки параметрів нелінійної моделі
5. Степенева модель та її лінеаризація
6. Показникова модель та її лінеаризація
7. Квадратична модель та підходи до оцінки її параметрів

Лабораторна робота №6 «Нелінійна економетрична модель»

Мета роботи: Набуття практичних навичок здійснення економіко-математичного аналізу за результатами побудови моделі нелінійної регресії

Завдання роботи і вихідні дані.

На основі вибіркових статистичних спостережень на протязі 12 років за деякою галуззю отримані статистичні дані щодо річного випуску продукції галузі **Y** (млн. гр. од.), вартості основного капіталу **K** (млн. гр. од.) і чисельності зайнятих у галузі **L** (тис. чоловік). Дані наведені у таблиці 10.

Таблиця 10 – Вихідні дані

Рік	Номер варіанту														
	1			2			3			4			5		
	Y	K	L	Y	K	L	Y	K	L	Y	K	L	Y	K	L
1	34,09	21,6	50,6	67,38	31,6	62,6	36,51	22,6	55,6	44,42	25,6	52,6	45,42	26,6	57,6
2	41,92	23,9	73,3	72,32	28,9	67,3	39,31	29,9	54,3	47,93	27,9	58,3	48,92	28,9	54,3
3	37,77	20,7	64,0	74,49	33,7	62,7	40,16	26,7	63,0	49,18	24,7	62,5	50,17	25,7	65,0
4	41,06	27,2	66,1	79,39	35,2	71,1	43,47	25,2	69,1	52,96	32,2	60,1	53,96	30,2	71,1
5	43,64	29,8	65,7	83,52	34,8	73,7	44,02	30,8	80,7	56,05	28,8	69,7	57,03	28,8	68,7
6	45,77	26,6	74,6	87,78	38,6	74,6	48,14	31,6	77,6	58,84	30,6	78,6	59,83	31,6	79,6
7	48,09	31,1	87,9	91,74	41,1	82,9	50,43	32,1	80,9	61,70	35,1	70,9	56,68	26,1	82,9
8	49,37	33,6	78,5	93,66	37,6	78,5	51,68	36,6	81,5	63,21	32,6	77,5	64,18	32,6	83,5
9	50,71	30,3	82,6	96,68	40,3	92,6	58,03	36,3	85,6	65,06	35,3	86,6	66,04	38,3	87,6
10	52,22	34,0	92,0	101,7	44,0	86,9	54,54	37,0	89,0	67,05	40,8	80,2	64,02	36,0	91,0
11	54,14	33,2	97,2	106,4	45,2	99,2	56,46	38,2	92,2	69,21	39,2	93,2	70,19	40,2	94,2
12	57,59	37,3	95,4	104,7	47,3	95,4	59,91	40,3	98,4	73,25	41,3	99,4	74,23	42,3	100,4

Грунтуючись на наведених статистичних даних:

1. Побудувати неокласичну виробничу функцію Кобба–Дугласа:

$$Y = a_0 K^\alpha L^\beta, \quad (6.1)$$

де Y – річний випуск продукції у галузі; K – вартість основного капіталу; L – чисельність зайнятих у галузі; a_0 , α , β – параметри моделі.

2. Оцінити якість, адекватність і статистичну значимість побудованої виробничої функції для рівня значимості $\alpha = 0,05$.

3. На основі побудованої виробничої функції:

- оцінити вплив виробничих ресурсів на річний випуск продукції;
- оцінити вплив зростання масштабів виробництва на темпи росту випуску продукції і ефективність виробництва;

– для планового випуску продукції $Y = Y^*$ (Табл. 11) обчислити необхідну чисельність зайнятих у галузі L^* у припущенні, що вартість основного капіталу залишиться на рівні останнього року у вибірці;

– для планового випуску продукції $Y = Y^*$ (Табл. 11) обчислити необхідну вартість основного капіталу K^* у припущенні, що чисельність зайнятих у галузі залишиться на рівні останнього року у вибірці;

– для прогнозних значень основного капіталу K_{pr} і кількості зайнятих у галузі L_{pr} (Табл. 11) обчислити середню і граничну продуктивність праці та основного капіталу;

– для прогнозних значень основного капіталу K_{pr} і кількості зайнятих у галузі L_{pr} (Табл. 11) розрахувати точковий прогноз випуску продукції.

Таблиця 11 – Дані для прогнозування

Показники	Номер варіанту				
	1	2	3	4	5
Y^*	60	60	62	75	75
K_{pr}	40	40	42	42	44
L_{pr}	97	99	99	100	102

Порядок виконання роботи.

1. Ввести вихідні дані.
2. Виконати лінеаризацію виробничої функції у два кроки наступним чином.

Спочатку виконати логарифмування обох частин виразу (6.1):

$$\ln Y = \ln a_0 + \alpha \ln K + \beta \ln L .$$

Наступним кроком виконати наступну заміну змінних:

$$y = \ln Y; \quad x_1 = \ln K; \quad x_2 = \ln L .$$

В результаті цього нелінійна мультиплікативна виробнича функція (6.1) зводиться до наступної лінійної :

$$y = b_0 + b_1 x_1 + b_2 x_2 , \tag{6.2}$$

де параметри лінійної і нелінійної форм пов'язані наступним співвідношеннями:

$$b_0 = \ln a_0, \quad b_1 = \alpha, \quad b_2 = \beta . \tag{6.3}$$

3. Сформувати нові дані для подальшого оцінювання параметрів лінійної форми (6.2).

4. Побудувати лінійну економетричну модель. Для цього оцінити параметри b_0, b_1, b_2 лінійної форми (6.2) виробничої функції з використанням МНК. Задати для цього матрицю H і оцінити невідомі параметри A за формулою (7).

5. Оцінити якість побудованої моделі, використовуючи коефіцієнт детермінації формула (11) і перевірити її статистичну значущість за формулою (12).

6. Використовуючи t -критерій, перевірити значущість коефіцієнтів рівняння регресії формула (13).

7. Здійснити загальний висновок щодо якості, адекватності і статистичної значимості побудованої виробничої функції.

8. Шляхом зворотних перетворень виробничу функцію представити для її подальшого використання у традиційній нелінійній формі $Y = a_0 K^\alpha L^\beta$. Параметри a_0, α, β визначити на основі зворотних перетворень виразу (6.3).

9. Визначити часткові коефіцієнти еластичності випуску продукції за виробничими ресурсами за наступними співвідношеннями:

$$E_K = \alpha; E_L = \beta,$$

де E_K – коефіцієнт еластичності випуску продукції за основним капіталом, E_L – коефіцієнт еластичності випуску продукції за працею. На основі цих коефіцієнтів еластичності оцінюється вплив кожного з зазначених ресурсів на річний випуск продукції галузі.

10. Визначити загальний коефіцієнт еластичності p :

$$p = E_K + E_L = \alpha + \beta.$$

Оцінити вплив зростання виробничих ресурсів (зростання масштабів виробництва) на темпи росту випуску продукції і ефективність виробництва:

а) якщо $p > 1$, то темпи росту випуску продукції вищі за темпи росту виробничих ресурсів і ми маємо зростання ефективності виробництва при зростанні масштабів виробництва і економію виробничих ресурсів.

б) якщо $p < 1$, то темпи росту випуску продукції нижчі за темпи росту ресурсів і ми маємо падіння ефективності виробництва при зростанні масштабів виробництва і зростання витрат ресурсів на одиницю продукції.

в) якщо $p = 1$ – маємо постійну ефективність виробництва, тобто темпи росту випуску продукції дорівнюють темпу росту виробничих ресурсів.

11. Для планового випуску продукції $Y = Y^*$ розрахувати необхідну чисельність зайнятих у галузі L^* :

$$L^* = \left(\frac{Y^*}{a_0 K^\alpha} \right)^{\frac{1}{\beta}}.$$

12. Для планового випуску продукції $Y = Y^*$ розрахувати необхідну вартість основного капіталу K^* :

$$K^* = \left(\frac{Y^*}{a_0 L^\beta} \right)^{\frac{1}{\alpha}}.$$

13. Для прогнозних значень основного капіталу K_{pr} і кількості зайнятих у галузі L_{pr} обчислити середню і граничну продуктивність праці:

$$AP_L = \frac{Y}{L} = a_0 K^\alpha L^{\beta-1}, \quad MP_L = \frac{\partial Y}{\partial L} = \beta \cdot AP_L,$$

де AP_L – середня продуктивність праці;

MP_L – гранична продуктивність праці.

14. Для прогнозних значень вартості основного капіталу K_{pr} і кількості зайнятих у галузі L_{pr} обчислити середню і граничну продуктивність основного капіталу (фондовіддачу):

$$AP_K = \frac{Y}{K} = a_0 K^{\alpha-1} L^\beta, \quad MP_K = \frac{\partial Y}{\partial K} = \alpha \cdot AP_K,$$

де AP_K – середня продуктивність основного капіталу,

MP_K – гранична продуктивність основного капіталу.

15. Для прогнозних значень K_{pr} і L_{pr} розрахувати точковий прогноз обсягу випуску продукції по галузі:

$$\hat{Y}_{pr} = a_0 K_{pr}^\alpha L_{pr}^\beta.$$

16. Здійснити відповідні висновки.

Завдання та запитання для самоконтролю

1. Який економічний зміст мають параметри неокласичної виробничої функції Кобба–Дугласа?
2. Що таке повний (сумарний) коефіцієнт еластичності неокласичної виробничої функції Кобба–Дугласа і для чого він використовується?
3. Як визначається середня продуктивність праці і основного капіталу на основі неокласичної виробничої функції Кобба–Дугласа?
4. Як визначається гранична продуктивність праці і основного капіталу на основі неокласичної виробничої функції Кобба–Дугласа?
5. Як обчислити необхідні виробничі ресурси при заданому рівні випуску на основі виробничої функції Кобба-Дугласа?
6. Як оцінюється вплив зростання масштабів виробництва на темпи росту випуску продукції і ефективність виробництва.

ЗМІСТОВИЙ МОДУЛЬ II УЗАГАЛЬНЕНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

5 МУЛЬТИКОЛІНЕАРНІСТЬ

5.1 Визначення мультиколінеарності, її природа, причини виникнення і наслідки

У попередніх темах розглядалися економетричні моделі, які будувалися на основі припущень класичної лінійної регресії. Параметри цих моделей, які оцінювалися на основі МНК, внаслідок виконання цих припущень мали властивості BLUE – оцінок, а самі моделі по суті були класичними регресійними моделями.

Але при дослідженні деяких економічних явищ і процесів економетричними методами приходиться мати справу з випадками, коли одне або декілька з цих припущень порушується. Одним з таких порушень основних положень класичного лінійного регресійного аналізу, яке може мати місце **тільки для багатofакторних** регресійних моделей, є **мультиколінеарність**.

Мультиколінеарністю називається існування у багатofакторній лінійній регресійній моделі лінійної функціональної залежності, або сильної кореляції між двома чи більше пояснюючими (незалежними) змінними.

Наприклад, мультиколінеарність може мати місце для такої економетричної моделі:

$$y = b_0 + b_1x_1 + b_2x_2 + e ,$$

де y – ціна акції, x_1 - дивіденди на акцію, x_2 - зароблений прибуток на акцію, оскільки в даному випадку дивіденди та зароблений прибуток мають високий рівень кореляції.

Таким чином, наявність мультиколінеарності означає порушення припущення класичного лінійного регресійного аналізу про незалежність між пояснюючими змінними моделі.

У практиці економетричного моделювання розрізняють **повну(довершену)** і **неповну (недовершену) мультиколінеарність**.

Повною мультиколінеарністю називається існування у багатofакторній лінійній регресійній моделі лінійної функціональної залежності між двома чи більше пояснюючими (незалежними) змінними.

Неповною мультиколінеарністю називається існування у багатofакторній лінійній регресійній моделі тісного кореляційного зв'язку між двома чи більше пояснюючими (незалежними) змінними.

Природу мультиколінеарності для випадку тільки двох пояснюючих змінних можна достатньо наочно проілюструвати наступною діаграмою Вена, наведеною на рис. 5.



Рисунок 5 - Природа мультиколінеарності

На рис. 1,а залежність між пояснюючими змінними моделі x_1 і x_2 , як функціональна так і кореляційна, відсутня і чітко можна виділити (розрізнити) вплив кожної змінної на залежну змінну y . На рис. 1,б, який ілюструє випадок неповної мультиколінеарності, видно, що внаслідок лінійної залежності між пояснюючими змінними x_1 і x_2 стає важчим розрізнити окремо вплив кожної пояснюючої змінної на залежну, оскільки починає проявлятися і одночасний вплив обох цих змінних на залежну. На рис. 1,в при повній мультиколінеарності взагалі неможливо розрізнити ступінь індивідуального впливу пояснюючих змінних x_1 і x_2 на залежну змінну y . Тобто, в останньому випадку залежна змінна y „не розрізняє” пояснюючі змінні x_1 і x_2 і сприймає їх як одну.

Таким чином природа мультиколінеарності полягає у неможливості статистично оцінити і обґрунтувати вплив кожної пояснюючої змінної на залежну зміну моделі, що, в свою чергу, робить ненадійною економічну інтерпретацію оціненого рівняння регресії.

З математичної точки зору мультиколінеарність означає, що у матриці незалежних змінних

$$X = \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix}$$

між різними стовпцями (векторами спостережень за окремим пояснюючими змінними) може існувати тісна кореляція, або ж елементи деякого стовпця отримані з відповідних елементів іншого шляхом лінійних перетворень останніх. В першому випадку маємо неповну мультиколінеарність, у другому – повну. У

випадку повної мультиколінеарності матриця X має неповний ранг, тобто вона містить менше ніж m незалежних стовпців. Внаслідок цього матриця $X'X$ є виродженою (сингулярною) і її визначник дорівнює нулю - $\det(X'X) = 0$.

Мультиколінеарність може виникнути за різних умов. Основними причинами такого явища є дві причини.

По-перше, існує глобальна тенденція **одночасної зміни** економічних показників. Такі економічні показники, як доход, споживання, накопичення, інвестиції, ціни, зайнятість мають тенденцію до одночасного зростання у період економічної експансії і до спаду у період рецесії. Наявність тренду у зміні цих показників у часі і є причиною мультиколінеарності.

По-друге, до мультиколінеарності приводить широке використання в економетричних моделях **лагових змінних**, тобто змінних, значення якої в економетричній моделі фігурують з деяким часовим запізненням - **лагом** (наприклад, із запізненням на місяць, квартал, рік).

Що стосується наслідків мультиколінеарності, то вони залежать від типу мультиколінеарності.

У випадку **повної мультиколінеарності** взагалі неможливо оцінити вплив незалежних змінних на залежну і побудувати регресійну модель, оскільки неможливо оцінити параметри моделі МНК. Це пов'язано з тим фактом, що, як зазначалося вище, матриця $X'X$ у цьому випадку буде виродженою (сингулярною), визначник якої буде дорівнювати 0. Оскільки для такої матриці неможливо знайти обернену $(X'X)^{-1}$, оцінювання параметрів моделі стає неможливим з чисто математичної точки зору.

У випадку **неповної мультиколінеарності** теоретично для оцінювання параметрів моделі можна застосовувати МНК, але це може призвести до наступних **теоретичних** наслідків:

1. Зміщення оцінок параметрів моделі, що не дає можливість зробити коректні висновки стосовно зв'язку між змінними моделі і економічну інтерпретацію цих параметрів.

2. Різке суттєве збільшенні дисперсії оцінок параметрів $\sigma_{b_j}^2, (j = \overline{0, m})$.

Це в свою чергу призводить до наступних **негативних практичних** наслідків:

1. Збільшення інтервалів довіри параметрів моделі.

2. Статистична незначимість оцінок деяких параметрів моделі. Це пов'язано із зменшенням t – статистики для деяких параметрів. Внаслідок цього із моделі можуть бути вилучені змінні, які за економічним змістом як раз суттєво впливають на залежну змінну.

3. Оцінки параметрів стають чутливими до розміру статистичної вибірки. Збільшення сукупності спостережень внаслідок цього іноді може привести до істотних змін в оцінках параметрів.

Таким чином, мультиколінеарність загалом **негативно** впливає на кількісні характеристики економетричної моделі, або робить її побудову взагалі неможливою. Внаслідок цього важливим стає питання тестування наявності мультиколінеарності у моделі і вилучення її.

Мультиколінеарність не завжди є такою поважною проблемою, щоб прикладати суттєві зусилля щодо її виявлення і усунення. Все залежить від мети економетричного дослідження.

Якщо єдиною метою економетричного дослідження є прогнозування, то при достатньо великому значення коефіцієнта детермінації $R^2 \geq 0,9$ наявність мультиколінеарності не впливає на якість і точність прогнозу. Хоча таке твердження має підстави тільки у тому випадку, якщо між прогнозними значеннями корельованих пояснюючих змінних будуть зберігатися ті ж самі відношення, що і раніше у вибірці.

Якщо ж метою економетричного аналізу є не прогноз, а оцінювання впливу кожної пояснюючої змінної на залежну змінну, то наявність мультиколінеарності скоріш за все спотворить дійсні залежності між змінними моделі. У цій ситуації мультиколінеарність є поважною проблемою.

5.2 Тестування наявності мультиколінеарності

Слід зазначити, що на даний момент не існує єдиного універсального методу виявлення мультиколінеарності, тому на практиці, як правило, використовують декілька методів і підходів.

Серед сучасних методів і підходів до тестування мультиколінеарності можна виділити 2 групи методів:

1. Методи і підходи, які базуються на деяких **зовнішніх ознаках**, які дозволяють **тільки встановити** наявність мультиколінеарності.
2. Методи, які базуються на спеціальних статистичних тестах.

Ознаки мультиколінеарності

1. **Високе значення парних коефіцієнтів кореляції між незалежними змінними.**

Якщо у кореляційній матриці

$$r = \begin{pmatrix} \Gamma_{x_1x_1} & \Gamma_{x_1x_2} & \cdots & \Gamma_{x_1x_m} \\ \Gamma_{x_2x_1} & \Gamma_{x_2x_2} & \cdots & \Gamma_{x_2x_m} \\ \cdots & \cdots & \cdots & \cdots \\ \Gamma_{x_mx_1} & \Gamma_{x_mx_2} & \cdots & \Gamma_{x_mx_m} \end{pmatrix}$$

значення хоча б одного коефіцієнта парної кореляції більше **0,8**, то мультиколінеарність є серйозною проблемою. Але високе значення парних коефіцієнтів кореляції є **достатньою**, але **не необхідною** умовою мультиколінеарності. Тому мультиколінеарність може бути і при відносно невеликих значеннях коефіцієнтів парної кореляції.

Слід також зазначити, що дана ознака буде надійною тільки у випадку невеликої кількості пояснюючих змінних (2 – 3). При великій кількості пояснюючих змінних більш надійним є використання частинних коефіцієнтів кореляції у формі **матриці часткових коефіцієнтів кореляції**.

2. **Мале значення визначника кореляційної матриці.**

Якщо $\det r = 0$ то існує повна (довершена) мультиколінеарність і між незалежними змінними моделі фактично існує функціональний зв'язок. Якщо $\det r = 1$, мультиколінеарність відсутня. Чим ближче визначник кореляційної матриці до нуля, тим більша ймовірність того, що між пояснюючими змінними існує мультиколінеарність.

3. **Високе значення коефіцієнта детермінації R^2 і не значимість t -статистики.**

Якщо в економетричній моделі при високому значенні R^2 знайдені параметри, які є статистично незначимими за t -статистикою, це свідчить про наявність мультиколінеарності.

Основним спільним недоліком наведених ознак мультиколінеарності є те, що ні одна з них чітко не розмежовує випадки, коли мультиколінеарність істотна, а коли нею можна знехтувати. На це питання дає відповідь тест Фаррара-Глобера, який належить до 2-ї групи методів тестування на мультиколінеарність.

Тест Фаррара-Глобера

Тест Фаррара-Глобера складається з трьох етапів:

- 1) тестування на мультиколінеарність усього масиву незалежних змінних - за статистичним критерієм χ^2 ;
- 2) перевірка на мультиколінеарність кожної незалежної змінної з рештою – за F- критерієм;
- 3) перевірка на кореляцію кожної пари незалежних змінних – за t -критерієм.

Алгоритм тесту Фаррара-Глобера.

Крок 1. Стандартизація (нормалізація) незалежних змінних.

Елементи векторів стандартизованих незалежних змінних визначаються за однією із наступних формул:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sigma_{xk}}, \quad (i = \overline{1, n}; k = \overline{1, m}) \quad (31)$$

$$\text{або} \quad x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n\sigma_{xk}^2}}, \quad (i = \overline{1, n}; k = \overline{1, m}), \quad (32)$$

де n - число спостережень у вибірці ; m - число пояснюючих змінних моделі;
 \bar{x}_{x_k} - середнє арифметичне k -ї пояснюючої змінної;

$\sigma_{x_k}^2$ і σ_{x_k} - відповідно дисперсія і стандартна похибка k -ї пояснюючої змінної.

Крок 2. Визначення кореляційної матриці нормалізованих змінних.

$$r = \frac{1}{n} X^* X \quad (33)$$

$$\text{або} \quad r = X^* X, \quad (34)$$

де X^* - матриця стандартизованих незалежних (пояснюючих) змінних,
 X^* - матриця, транспонована до матриці X^* .

Вираз (33) використовується, якщо стандартизація пояснюючих змінних виконується за виразом (31), а вираз (34), якщо стандартизація пояснюючих змінних виконується за виразом (32).

Крок 3. Визначення статистичного критерію χ^2

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m + 5) \right] \cdot \ln|r| \quad (35)$$

де $|r|$ - визначник матриці r .

Розраховане значення χ^2 - критерію порівнюється з табличним $\chi_{\text{табл}}^2$, який знаходиться за статистичними таблицями χ^2 - розподілу для прийнятого рівня значимості α (або рівня довіри $p = 1 - \alpha$) і ступеня вільності $df = \frac{1}{2}m(m-1)$.

Якщо розрахункове, фактичне значення критерію більше за табличне, тобто якщо виконується умова $\chi^2 > \chi_{\text{табл}}^2$, то в масиві незалежних змінних існує мультиколінеарність. У протилежному випадку мультиколінеарність відсутня і тест припиняється.

Крок 4. Визначення матриці похибок .

$$C = r^{-1}. \quad (36)$$

Крок 5. Обчислення F-критеріїв

$$F_k = (C_{kk} - 1) \frac{n - m}{m - 1}, (k = \overline{1, m}) \quad (37)$$

де C_{kk} - діагональні елементи матриці C .

Фактичні значення критеріїв порівнюються з табличними при ступенях вільності $v_1 = n - m$ і $v_2 = m - 1$ і заданому рівні значимості α . Якщо для деякої пояснюючої змінної k виконується умова $F_k > F_{\text{табл}}$, то відповідна k -та незалежна змінна мультиколінеарна з іншими змінними, тобто всі інші пояснюючі змінні моделі впливають на цю змінну внаслідок суттєвої кореляції між ними.

Крок 6. Визначення часткових коефіцієнтів кореляції

$$r_{kj} = \frac{-C_{kj}}{\sqrt{C_{kk} \cdot C_{jj}}}, (k = \overline{1, m}; j = \overline{1, m}), \quad (38)$$

де C_{kj} - елемент матриці C , що міститься у k -му рядку і j -му стовпці;
 C_{kk} і C_{jj} - діагональні елементи матриці C .

Крок 7. Обчислення t-критеріїв

$$t_{kj} = \frac{r_{kj} \sqrt{n - m}}{\sqrt{1 - r_{kj}^2}}, (k = \overline{1, m}; j = \overline{1, m}). \quad (39)$$

Розрахункові значення t - критеріїв порівнюються з табличними $t_{\text{табл}}$, визначеними для рівня значимості α і ступеня вільності $v = n - m$. Якщо для деякого розрахункового значення t – критерію виконується умова $t_{kj} > t_{\text{табл}}$, то між незалежними змінними x_k і x_j існує кореляція, тобто змінні x_k і x_j утворюють колінеарну пару.

5.3 Шляхи і засоби усунення мультиколінеарності

У випадку виявлення наявності мультиколінеарності існує декілька простих шляхів її усунення. Основними серед них є наступні.

1. Вилучення змінної (або змінних) з моделі. При цьому з моделі вилучається одна із змінних колінеарної пари. Слід зазначити, що таке вилучення

змінних можливе тільки у випадку коли це не суперечить логіці економічних зв'язків. У протилежному випадку це може призвести до помилки специфікації.

2. Зміна аналітичної форми економетричної моделі. Іноді заміна однієї функції регресії іншою (наприклад лінійної нелінійною) , якщо це не суперечить апріорній інформації, дає змогу уникнути явища мультиколінеарності.

3. Збільшення спостережень. З точки зору теорії, мультиколінеарність та невелика кількість спостережень у вибірці – це одна і та ж проблема. Тому збільшення спостережень у статистичній вибірці або використання іншої статистичної вибірки може усунути, або принаймні зменшити вплив мультиколінеарності.

4. Перетворення статистичних даних. Позбутися мультиколінеарності можна і шляхом наступних перетворень вихідних даних стосовно пояснюючих змінних :

- а) замість самих даних узяти їхні відхилення від середніх;
- б) замість абсолютних значень даних взяти відносні значення ;
- в) стандартизувати змінні.

5. Використання додаткової первинної інформації. Аналіз і використання первинної додаткової інформації інколи дозволяє зняти проблему мультиколінеарності.

Якщо ж жодний з розглянутих способів не дає змоги позбутися мультиколінеарності, для оцінювання параметрів моделі застосовують такі методи, як **метод головних компонентів, факторний аналіз, гребенева регресія.**

Лабораторна робота № 7 «Мультиколінеарність»

Мета роботи: Набуття практичних навичок тестування наявності мультиколінеарності в економетричних моделях і її усунення.

Завдання роботи і вихідні данні

Для деякого регіону виконується економетричне дослідження, метою якого є аналіз реального споживання населення y (в млн. грошових одиниць) в залежності від наступних трьох факторів: x_1 – купівлі та оплати товарів і послуг (в млн. грошових одиниць), x_2 – заощаджень (в % від загального доходу) і x_3 – заробітної плати (в млн. грошових одиниць). Дані вибіркового статистичного спостереження наведені нижче у таблиці 12.

Таблиця 12 – Вихідні данні

i	y	x₁	x₂	x₃
1	14	9,1	7,90	16,78
2	16	10,1	9,04	19,68
3	15	11,1	9,95	21,56
4	14	13,1	9,22	22,46
5	20	13,1	11,12	22,50
6	19	15,1	13,47	27,20
7	22	14,1	13,46	28,52
8	27	16,1	12,57	30,00
9	29	18,1	12,40	29,56
10	29	16,1	13,20	24,23
11	30	14,1	13,50	25,00
12	30	20,1	14,52	30,00
13	31	21,1	14,00	32,15
14	28	23,1	15,00	32,00
15	31	20,1	14,50	33,00

Грунтуючись на наведених статистичних даних:

- 1) за допомогою тесту Фаррара-Глобера перевірити модель на мультиколінеарність;
- 2) при наявності мультиколінеарності запропонувати шляхи її вилучення.

Порядок виконання роботи

1. Ввести вихідні данні.
2. Визначити середні значення і стандартні відхилення всіх пояснюючих змінних моделі.
3. Виконати стандартизацію пояснюючих змінних. Елементи стандартизованих векторів пояснюючих змінних визначити за формулою (31).

4. Сформуувати матрицю стандартизованих пояснюючих змінних X^* .
5. Обчислити кореляційну матрицю пояснюючих змінних моделі \mathbf{r} за формулою (33).
6. Обчислити розрахункове значення критерію χ^2 за формулою (35).
7. Для рівня значимості $\alpha=0,05$ і ступеня вільності $v = \frac{1}{2}m(m-1)$ за статистичними таблицями χ^2 -розподілу знайти табличне значення $\chi^2_{\text{табл}}$ і порівняти з фактичним розрахунковим. Зробити відповідний висновок.
8. Обчислити матрицю \mathbf{C} за формулою (36).
9. Для кожної пояснюючої змінної моделі розрахувати F-критерій Фішера за формулою (37).
10. Для рівня значимості $\alpha = 0,05$ і ступенів вільності $v_1 = m-1$ та $v_2 = n-m$ за статистичними таблицями F-розподілу знайти критичне значення критерію Фішера $F_{\text{кр}}$. Табличне значення $F_{\text{кр}}$ порівняти з розрахунковими значеннями F_k і зробити відповідні висновки.
11. Використовуючи матрицю \mathbf{C} обчислити часткові коефіцієнти кореляції між пояснюючими змінними моделі за формулою (38).
12. На основі знайдених часткових коефіцієнтів кореляції обчислюються розрахункові значення t-критерію Ст'юдента за формулою (39).
13. Для рівня значимості $\alpha = 0,05$ і для ступеню вільності $v = n-m$ за статистичними таблицями t-розподілу Ст'юдента знайти критичне значення t-критерію Ст'юдента – $t_{\text{кр}}$. Порівняти розрахункові значення t_{kj} з критичним $t_{\text{кр}}$ і зробити відповідні висновки.
14. У разі виявлення наявності мультиколінеарності запропонувати шляхи її усунення.

Завдання та запитання для самоконтролю.

1. Що означає мультиколінеарність пояснюючих змінних економетричної моделі?
2. При моделюванні яких економічних явищ і процесів найчастіше можливо очікувати на мультиколінеарність?
3. Чим відрізняється повна мультиколінеарність від неповної?
4. Як впливає наявність мультиколінеарності на статистичні показники і оцінки параметрів моделі?
5. Основні ознаки мультиколінеарності.
6. Які статистичні критерії включає алгоритм Феррара-Глобера?
7. З якою метою використовується критерій χ^2 ?
8. З якою метою використовується F-критерій Фішера?
9. З якою метою використовується t-критерій Ст'юдента?

10. Сформулюйте нульову гіпотезу для критерію χ^2 .
11. Сформулюйте нульову гіпотезу для F -критерію Фішера.
12. Сформулюйте нульову гіпотезу для t -критерію Стьюдента.
13. Визначник кореляційної матриці дорівнює 0,015. Ваш висновок?
14. Визначник кореляційної матриці дорівнює 0,99. Ваш висновок?
15. Коефіцієнт парної кореляції між двома чинниками дорівнює 0,996. Ваш висновок?
16. Перерахувати способи усунення мультиколінеарності в масиві чинників?

6 ГЕТЕРОСКЕДАСТИЧНІСТЬ

6.1 Визначення гетероскедастичності, її природа та наслідки

Одним з основних припущень класичної лінійної регресії, яке дозволяє коректно застосувати для оцінювання параметрів моделі 1МНК, є припущення про сталість дисперсії стохастичної складової ε_i , тобто припущення про **гомоскедастичність** стохастичної складової економетричної моделі.

Гомоскедастичністю називається явище, при якому дисперсія стохастичної складової економетричної моделі є сталою (незмінною) для кожного окремого спостереження або групи спостережень.

Слід зазначити, що гомоскедастичність слід розглядати не тільки як явище, а і як **властивість** стохастичної складової моделі. Суть гомоскедастичності полягає в тому, що варіація кожної випадкової величини ε_i навколо її математичного сподівання не залежить від значення незалежних змінних x . Таким чином дисперсія випадкової величини ε_i залишається сталою незалежно від малих чи великих значень факторів, тобто :

$$\sigma_{\varepsilon}^2 \neq f(x_{1i}, x_{2i}, \dots, x_{mi}), i = 1, 2, \dots, n .$$

Графічно випадок гомоскедастичності для простої лінійної регресії можна представити наступним чином (рис. 6).

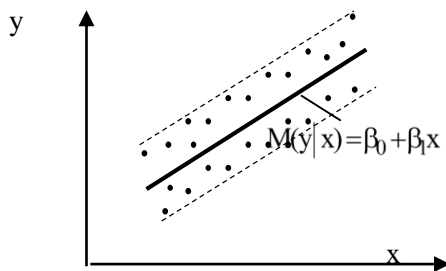


Рисунок 6 - Випадок гомоскедастичності

Як видно з цього рисунку гомоскедастичність характеризується тим, що випадкові відхилення залежної змінної моделі y від прямої регресії, які характеризують дисперсію величини ε_i , розташовані в межах деякого шару **сталого ширини**. Таким чином дисперсія стохастичної складової моделі ε не змінює свого значення при переході від малих значень пояснюючої змінної x до великих і залишається сталою в усьому діапазоні зміни значень пояснюючої змінної. Такий же ефект спостерігається і у випадку множинної регресії.

Формалізовано припущення про гомоскедастичність (сталість) стохастичної складової має вигляд :

$$\text{var}(\varepsilon_i) = M(\varepsilon_i - M(\varepsilon_i))^2 = \sigma_\varepsilon^2 = \text{const}, i = 1, 2, \dots, n,$$

або у матричній формі

$$M(\varepsilon \varepsilon') = \sigma_\varepsilon^2 E, \quad (40)$$

де E – одинична матриця розміру $n \times n$, а вираз $M(\varepsilon \varepsilon')$ визначає коваріаційну, а точніше дисперсійно-коваріаційну матрицю випадкової величини ε .

Дійсно маємо :

$$M(\varepsilon \varepsilon') = M \begin{pmatrix} \varepsilon_1^2 & \varepsilon_1 \varepsilon_2 & \dots & \varepsilon_1 \varepsilon_n \\ \varepsilon_2 \varepsilon_1 & \varepsilon_2^2 & \dots & \varepsilon_2 \varepsilon_n \\ \dots & \dots & \dots & \dots \\ \varepsilon_n \varepsilon_1 & \varepsilon_n \varepsilon_2 & \dots & \varepsilon_n^2 \end{pmatrix},$$

або застосувавши до матриці оператор математичного сподівання отримаємо :

$$M(\varepsilon \varepsilon') = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \text{cov}(\varepsilon_1, \varepsilon_2) & \dots & \text{cov}(\varepsilon_1, \varepsilon_n) \\ \text{cov}(\varepsilon_2, \varepsilon_1) & \sigma_{\varepsilon_2}^2 & \dots & \text{cov}(\varepsilon_2, \varepsilon_n) \\ \dots & \dots & \dots & \dots \\ \text{cov}(\varepsilon_n, \varepsilon_1) & \text{cov}(\varepsilon_n, \varepsilon_2) & \dots & \sigma_{\varepsilon_n}^2 \end{pmatrix}. \quad (41)$$

Матриця, наведена у виразі (41) називається **дисперсійно-коваріаційною матрицею випадкової величини ε (стохастичної складової моделі)**. Діагональними елементами цієї матриці є дисперсії випадкової величини ε у кожному спостереженні, а всі інші – коваріаціями. З огляду на сталість дисперсії стохастичної складової моделі у випадку гомоскедастичності і відсутність автокореляції залишків (нульова коваріації залишків) вираз (41) можна подати у вигляді

$$M(\varepsilon \varepsilon') = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & \dots & 0 \\ 0 & \sigma_\varepsilon^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_\varepsilon^2 \end{pmatrix} = \sigma_\varepsilon^2 \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix} = \sigma_\varepsilon^2 E,$$

що відповідає виразу (40).

Якщо припущення про гомоскедастичність не виконується, то має місце **гетероскедастичність**.

Гетероскедастичністю називається явище, при якому дисперсія стохастичної складової моделі змінює своє значення від одного спостереження до іншого, або від однієї групи спостережень від другої

Суть гетероскедастичності полягає в тому, що значення дисперсії випадкової величини ε_i залежить від значень незалежної змінної x , тобто у цьому випадку можна записати:

$$\sigma_{\varepsilon}^2 = f(x_{1i}, x_{2i}, \dots, x_{mi}), \quad i = 1, 2, \dots, n$$

Графічно випадки гетероскедастичності для випадку простої лінійної регресії можна представити таким чином (рис. 7):

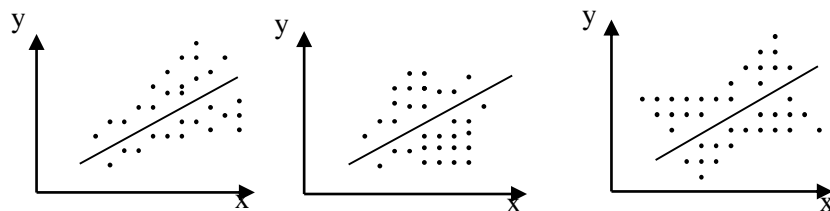


Рисунок 7 - Випадки гетероскедастичності

Оскільки у цьому випадку, як і випадку гомоскедастичності, коваріації випадкових величин ε_i дорівнюють нулю (внаслідок відсутності автокореляції залишків), а дисперсії змінюються від одного спостереження до іншого вираз (41) можна представити наступним чином:

$$M(\varepsilon \varepsilon') = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon_2}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{\varepsilon_n}^2 \end{pmatrix},$$

або

$$M(\varepsilon \varepsilon') = \sigma^2 S \quad (42)$$

де σ^2 - деяка невідома константа, S – відома квадратна діагональна додатньо визначена матриця розмірністю $n \times n$.

З явищем гетероскедастичності приходиться часто зустрічатися у багатьох економетричних дослідженнях. Наявність гетероскедастичності можна прогнозувати при відповідному досвіді і виходячи з аналізу економічних показників, які включаються до економетричної моделі. Прикладом економетричної моделі, для якої скоріш за все буде існувати проблема гетероскедастичності є наступна модель:

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

де

y – заощадження домогосподарства,

x – дохід домогосподарства.

У цьому випадку можна очікувати, що сім'ї з більшим доходом покажуть більшу варіацію у своїй поведінці заощаджень, ніж сім'ї з низьким доходом.

У випадку гетероскедастичності у принципі неможливо використовувати звичайні формули для знаходження оцінок $\hat{\sigma}_{b_j}^2, (j = \overline{0, m})$ дисперсії параметрів моделі, оскільки дисперсія залишків $\hat{\sigma}_{\varepsilon}^2$ в умовах гетероскедастичності не є сталим числом, а змінюється із зростанням значень незалежних змінних x . Внаслідок цього разом із зміною значення незалежних змінних x буде змінюватися і дисперсія оцінок параметрів $\hat{\sigma}_{b_j}^2$.

Оцінки параметрів моделі, отримані МНК в умовах гетероскедастичності будуть незміщеними, обґрунтованими, але **неефективними**, тобто вони будуть мати велику дисперсію, внаслідок чого вони **не-будуть BLUE – оцінками**. Використання таких оцінок призводить до наступних негативних наслідків:

- збільшення інтервалів довіри параметрів;
- помилки при використанні t-тестів і F-тестів;
- неефективність прогнозів, тобто отримання прогнозів з дуже великим інтервалом довіри.

Зрозуміло, що гетероскедастичність є серйозною проблемою, тому необхідно вміти її виявляти і робити оцінювання параметрів іншими методами.

Гетероскедастичність, як і мультиколінеарність не завжди є такою поважною проблемою, щоб прикладати суттєві зусилля щодо її виявлення і усунення. Все залежить від мети економетричного дослідження.

Якщо єдиною метою економетричного аналізу є оцінювання параметрів моделі і їх економічна інтерпретація то наявність гетероскедастичності не

створить проблем, оскільки МНК-оцінки у цьому випадку, як вже відмічалось, будуть незміщеними.

Якщо ж метою економетричного дослідження є прогнозування, то у цій ситуації гетероскедастичність є поважною проблемою, оскільки суттєво збільшить прогнозні інтервали залежної змінної моделі.

6.2 Тестування наявності гетероскедастичності

Як і у випадку мультиколінеарності немає єдиних правил і методів для виявлення гетероскедастичності, а є різноманітні тести. До основних з них належать наступні тести:

- 1) тест на основі графічного аналізу залишків;
- 2) тест на основі М-критерію;
- 3) тест Глейсера;
- 4) тест на основі коефіцієнта рангової кореляції Спірмена;
- 5) параметричний тест Голдфелда – Квондта;
- 6) непараметричний тест Голдфелда – Квондта;

Тестування гетероскедастичності на основі графічного аналізу залишків

Цей тест є найпростішим з усіх і достатньо наочним, оскільки дає можливість візуально визначити наявність гетероскедастичності. Тест умовно можна розбити на два етапи.

На першому етапі на основі статистичної вибірки і припущень про відсутність гетероскедастичності будується класична економетрична модель і обчислюються залишки e_i , ($i = \overline{1, n}$).

На другому етапі виконуються дослідження квадратів залишків e_i^2 , ($i = \overline{1, n}$) і робиться висновок про наявність або відсутність гетероскедастичності. Для цього будуються графіки різних типів. Для парної лінійної моделі будується графік $e^2 = f(x)$. Для моделі багатofакторної лінійної регресії найбільш розповсюдженими є графіки залежності $e^2 = f(y)$ або графіки $e^2 = f(x_j)$, де x_j – пояснююча змінна, яка гіпотетично може впливати на дисперсію залишків. Якщо неможливо однозначно визначитися з такою змінною графіки $e^2 = f(x_j)$ будуються для всіх пояснюючих змінних моделі. Метою побудови таких графіків є встановлення наявності або відсутності систематичності у зміні квадратів залишків e^2 при зміні значення залежної змінної моделі y , або пояснюючої змінної x_j . Звичайно e_i^2 - це тільки оцінки невідомих ε_i^2 , але вони можуть успішно використовуватися, особливо при великих вибірках.

Досліджуючи e_i^2 можна отримати наступні види графіків, які наведені на рис. 8.

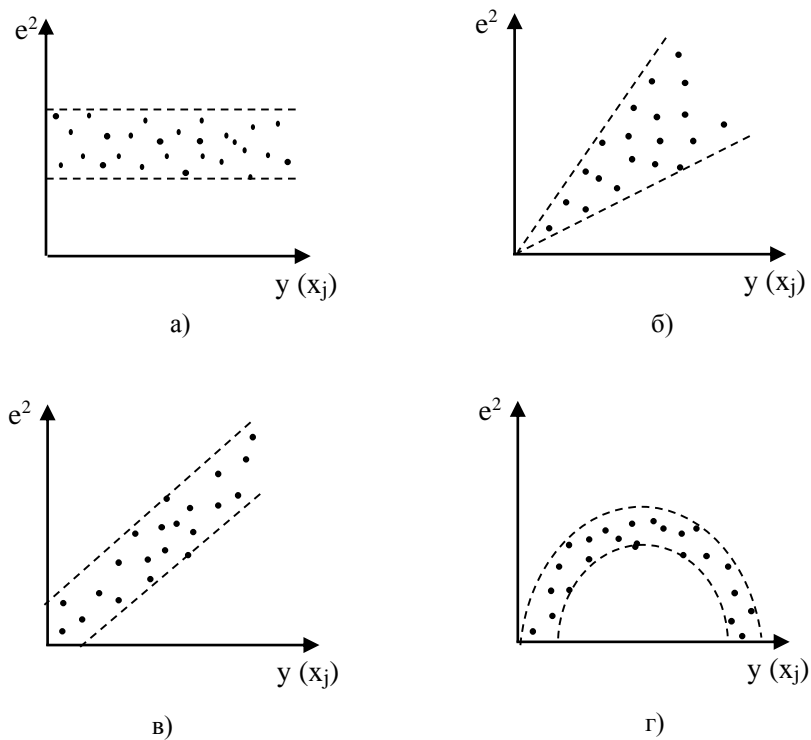


Рисунок 8 - Типи графіків квадратів залишків

На рис. 8,а всі квадрати залишків знаходяться всередині шару постійної ширини, яка паралельна осі абсцис. Це свідчить про незалежність e_i^2 від залежної змінної y або пояснюючої змінної x_j і їх сталості, тобто у цьому випадку виконуються умови гомоскедастичності.

На рис. 8,б – г спостерігаються систематичні зміни у співвідношенні між значеннями y (або x_j) і квадратами залишків e_i^2 . На рис. 8,б і 8,в відображена лінійна, а на рис. 8,г – квадратична залежності між квадратами залишків і значеннями залежної або пояснюючої змінної моделі. Таким чином, ситуації , представлені на 8,б – г, свідчать про наявність у цих випадках гетероскедастичності.

Графічний метод дає можливість не тільки виявити гетероскедастичність, але й зробити висновок щодо самої форми зв'язку між дисперсією залишків і пояснюючими змінними моделі, що особливо важливо для побудови моделі при наявності гетероскедастичності.

Параметричний тест Голдфелда-Квондта

Цей тест застосовується в основному для невеликих вибірок. В основу цього тесту покладено припущення, що дисперсія залишків зростає пропорційно до квадрату однієї з пояснюючих змінних x_j , тобто розглядається випадок коли $\sigma_{\varepsilon,i}^2 = \sigma^2 \cdot x_{ij}^2$, ($i = 1, 2, \dots, n$), де σ^2 - невідомий коефіцієнт пропорційності (константа). Залишки, при цьому розподілені за нормальним законом і некорельовані.

Алгоритм тесту

Крок 1. Виконується впорядкування (ранжування) спостережень у статистичній вибірці в порядку зростання (або спаду) значень пояснюючої змінної x_j .

Крок 2. З усіх спостережень впорядкованої вибірки відкидається c спостережень, які містяться у центрі вибірки. Ця кількість згідно рекомендацій авторів тесту визначається із співвідношення

$$\frac{c}{n} = \frac{4}{15}. \quad (43)$$

У результаті цього утворюються дві підвибірки розміром $n_1 = n_2 = \frac{n-c}{2}$

Крок 3. Для кожної підвибірки на основі МНК будується окрема регресійна модель.

Крок 4. Для кожної підвибірки визначається сума квадратів залишків SSE_1 і SSE_2 :

$$SSE_1 = \sum_{i=1}^{n_1} e_{1,i}^2, \quad (44)$$

$$SSE_2 = \sum_{i=1}^{n_2} e_{2,i}^2, \quad (45)$$

де $e_{1,i}$ – залишки для першої моделі (побудованої на основі першої підвибірки), $e_{2,i}$ – залишки для другої моделі (побудованої на основі другої підвибірки).

Крок 5. Для порівняння зазначених дисперсій обчислюється наступна F – статистика (критерій Фішера):

$$F^* = \frac{SSE_2}{SSE_1}, \quad (46)$$

яка порівнюється з табличним значення F – критерію $F_{\text{табл}}$, що визначається за статистичними таблицями F – розподілу Фішера для заданого рівня значимості α і ступенів вільності $v_1 = v_2 = \frac{n-c}{2} - k$, де n – загальна кількість спостережень, k – кількість параметрів моделі, c – кількість відкинутих спостережень.

Якщо виконується умова $F^* > F_{\text{табл}}$, то гетероскедастичність присутня. У протилежному випадку маємо випадок гомоскедастичності. Слід зазначити, чим більше значення критерію F , обчисленому за виразом (45), тим більше ефект гетероскедастичності стохастичної складової моделі.

Якщо важко априорі визначити пояснюючу змінну x_i , яка впливає на залишки, тест Голдфелда-Квондта потрібно застосувати по черзі до кожної незалежної змінної моделі окремо.

6.3 Оцінювання параметрів моделі у разі гетероскедастичності

Економетрична модель, якій притаманна гетероскедастичність є узагальненою моделлю, і для оцінювання її параметрів використовується так званий узагальнений метод найменших квадратів (УМНК), або метод Ейткена. Свою назву метод отримав внаслідок його застосування для оцінювання параметрів моделі, для якої дисперсійно-коваріаційна матриця стохастичної складової моделі приймається у найбільш загальному вигляді (41), тобто допускається одночасно і гетероскедастичність і автокореляція залишків.

В основу методу Ейткена покладено ідею трансформації економетричної моделі, якій притаманна гетероскедастичність у класичну гомоскедастичну з подальшим застосуванням до такої трансформованої моделі процедури МНК для оцінювання параметрів узагальненої моделі, якій притаманна гетероскедастичність. Трансформація вихідної моделі у гомоскедастичну відбувається шляхом корегування вихідної статистичної інформації стосовно змінних моделі. Спосіб і форма корегування вихідних даних визначаються, при цьому, формою залежності дисперсії стохастичної складової ϵ від тієї чи іншої пояснюючої змінної моделі.

Розглянемо більш докладно цей метод. Нехай ϵ економетрична модель

$$Y = XB + e, \quad (47)$$

для якої $M(\epsilon\epsilon') = \sigma^2 \cdot S$, де σ^2 – як і раніше, деяка невідома константа, S – відома квадратна додатньо визначена матриця розмірністю $n \times n$, яка у випадку гетероскедастичності, як показано раніше, є діагональною матрицею і має наступний вигляд:

$$S = \begin{pmatrix} 1/\lambda_1 & 0 & \dots & 0 \\ 0 & 1/\lambda_2 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1/\lambda_n \end{pmatrix}, \quad \dim P = n \times n, \quad (48)$$

де $\lambda_i, i=1,2,\dots,n$ - власні значення цієї матриці.

Оскільки матриця S симетрична і додатньо визначена, то використовуючи теорію матриць її можна подати у наступному вигляді:

$$S = P' \cdot P,$$

де матриця P є не виродженою і має вигляд

$$P = \begin{pmatrix} 1/\sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/\sqrt{\lambda_n} \end{pmatrix},$$

а обернена до неї відповідно:

$$P^{-1} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{\lambda_n} \end{pmatrix}.$$

Базуючись на особливостях матриць S і P запишемо деякі співвідношення між ними і оберненими до них:

$$P^{-1} \cdot S \cdot P^{-1} = E$$

і

$$P^{-1} P^{-1} = S^{-1}. \quad (49)$$

Помноживши рівняння (47) на матрицю P^{-1} , дістанемо:

$$P^{-1}Y = P^{-1}XB + P^{-1}e \quad (50)$$

Введемо наступні позначення : $Y^* = P^{-1}Y$; $X^* = P^{-1}X$; $e^* = P^{-1}e$.

Тоді модель (50) матиме вигляд:

$$Y^* = X^*B + e^* \quad (51)$$

Використовуючи (49) можна показати, що для цієї перетвореної моделі гетероскедастичність відсутня, оскільки

$$M(\varepsilon^* \varepsilon^{*\prime}) = \sigma_\varepsilon^2 \cdot E ,$$

що дає змогу застосувати до трансформаційної моделі (51) МНК . Тоді отримаємо:

$$B = (X^{*\prime} X^*)^{-1} X^{*\prime} Y^* = (X' P^{-1\prime} P^{-1} X)^{-1} (X' P^{-1\prime} P^{-1} Y) ,$$

або з врахуванням (49) остаточно

$$B = (X'S^{-1}X)^{-1} X'S^{-1}Y. \quad (52)$$

Таким чином, якщо матриця S відома, за формулою (52) можна завжди обчислити оцінки параметрів моделі у разі гетероскедастичності. Проблема полягає у визначенні власних Оскільки дійсні значення випадкової величини ε , як правило невідомі, значення λ_i у матриці S можна обчислити користуючись різними гіпотезами відносно зв'язку дисперсії $\sigma_{\varepsilon,i}^2$ і деякої пояснюючої змінної x_j . В основному при цьому використовуються наступні 2 гіпотези.

Якщо дисперсія залишків пропорційна до зміни пояснюючої змінної x_j , тобто $\sigma_{\varepsilon,i}^2 = \sigma^2 \cdot x_{ji}$, ($i = \overline{1, n}$). Тоді величини λ_i визначається:

$$\lambda_i = \frac{1}{x_{ij}}. \quad (53)$$

Якщо дисперсія залишків пропорційна до зміни квадрату пояснюючої змінної x_j , тобто $\sigma_{\varepsilon,i}^2 = \sigma^2 \cdot x_{ji}^2$, ($i = \overline{1, n}$). Величини λ_i визначається для цієї гіпотези:

$$\lambda_i = \frac{1}{x_{ji}^2}.$$

6.4 Верифікація економетричної моделі і прогнозування у випадку гетероскедастичності

Оцінки (52), отримані методом Ейткена, є BLUE – оцінками і мають дисперсійно - коваріаційну матрицю

$$\text{var} - \text{cov}(b) = \hat{\sigma}_\varepsilon^2 (X'S^{-1}X)^{-1} \quad (54)$$

Незміщена оцінка дисперсії залишків $\hat{\sigma}_\varepsilon^2$ визначається для цього випадку наступним чином:

$$\hat{\sigma}_\varepsilon^2 = \frac{e'S^{-1}e}{n - k}, \quad (55)$$

де e – вектор залишків моделі, параметри якої обчислені за МНК.

Таким чином у випадку гетероскедастичності, якщо відома матриця S , оцінки параметрів узагальненої моделі можна визначити методом Ейткена (УМНК) за формулою (52), оцінку дисперсії залишків $\hat{\sigma}_\varepsilon^2$ за формулою (55), а оцінки дисперсій параметрів моделі – з дисперсійно-коваріаційної матриці (54). Це дає можливість у подальшому застосувати t – статистику для перевірки статистичної значимості параметрів моделі і побудови інтервалів довіри для них.

Що стосується перевірки загальної статистичної значимості моделі, то це можна зробити на основі відомих сум квадратів, розглянутих раніше у дисперсійному аналізі загальної лінійної економетричної моделі. Коли параметри економетричної моделі оцінюються за УМНК дисперсійний аналіз дає наступні суми квадратів:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2 = Y'S^{-1}Y,$$

$$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = B'X'S^{-1}Y,$$

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = e'S^{-1}e,$$

де

B – вектор оцінок параметрів моделі, отриманих узагальненим методом найменших квадратів (УМНК),

e – вектор залишків моделі, параметри якої обчислені за МНК ,

Y – вектор спостережень за залежною змінною моделі,

X - матриця спостережень за пояснюючими змінними моделі,

S – матриця (48).

Використовуючи зазначені суми квадратів можна визначити множений (або парний) коефіцієнт детермінації, множинний (або парний) коефіцієнт кореляції і перевірити побудовану модель за F – критерієм на статистичну значимість у цілому.

Найкращий незміщений лінійний точковий прогноз у випадку гетероскедастичності обчислюється за наступною залежністю :

$$\hat{y}_{pr} = X'_{pr} B + \lambda_n e_n ,$$

де

B – вектор оцінок параметрів моделі, отриманих узагальненим методом найменших квадратів (УМНК);

λ_n – останній параметр з матриці S (для останнього спостереження у вибірці);

e_n - залишок в останньому спостереженні, обчислений для моделі, параметри якої оцінені на основі МНК;

X_{pr} - вектор прогнозних значень пояснюючих змінних моделі.

Лабораторна робота № 8 «Гетероскедастичність»

Мета роботи: Набуття практичних навичок тестування наявності гетероскедастичності і оцінювання параметрів економетричної моделі узагальненим методом найменших квадратів.

Завдання роботи і вихідні дані

Для деякого регіону виконати дослідження залежності заощаджень населення (y) від доходу на душу населення (x). Вибіркові статистичні дані за 18 років наведені нижче у таблиці 13.

Таблиця 13 – Вихідні дані

Рік	1	2	3	4	5	6
Заощадження (млн. грошових одиниць)	2,3	2,5	2,08	2,2	2,1	2,7
Доход на душу населення (млн. грошових одиниць)	15	68	16	17	17	85
Рік	7	8	9	10	11	12
Заощадження (млн. грошових одиниць)	3,99	2,5	3,94	2,5	3,1	2,2
Доход на душу населення (млн. грошових одиниць)	100	20	90	22	64	15
Рік	13	14	15	16	17	18
Заощадження (млн. грошових одиниць)	2,82	3,04	2,32	2,2	3,1	2,45
Доход на душу населення (млн. грошових одиниць)	72	80	18	20	95	19

Грунтуючись на наведених статистичних даних:

1. Виходячи з ймовірності існування гетероскедастичності виконати параметричний тест Голдфелда–Квондта (для рівня значимості $\alpha=0,05$).
2. Знайти оцінки параметрів моделі узагальненим методом найменших квадратів.

Порядок виконання роботи

1. Виконати ранжування (впорядкування) даних статистичних спостережень у порядку зростання значень величини доходу (незалежної змінної x).
2. З середини впорядкованої вибірки відкидається c спостережень, значення c визначається зі співвідношення (43).
3. На основі МНК побудувати дві лінійні парні регресії для двох утворених сукупностей спостережень, кожна з яких має обсяг

$$n_1 = n_2 = \frac{n-c}{2}.$$

4. На основі отриманих рівнянь регресії для кожної з двох моделей обчислюються розрахункові значення залежної змінної $\hat{y}_i, (i = \overline{1, n_1})$ (заощадження) і залишки $e_i, (i = \overline{1, n_1})$.

5. Для кожної побудованої моделі визначити суми квадратів залишків за формулами (44) та (45).

6. Обчислити статистику F^* за формулою (46).

7. За статистичними таблицями F -розподілу Фішера для ступенів вільності $\nu_1 = \nu_2 = [(n-c)/2] - k$ (де k – кількість оцінених у кожній регресії параметрів) і рівня значимості $\alpha = 0,05$ знайти критичне значення критерію Фішера $F_{кр}$.

8. Порівняти значення F^* і $F_{кр}$ та зробити висновок про наявність або відсутність гетероскедастичності.

9. Застосувати для оцінки параметрів узагальнений метод найменших квадратів (методом Ейткена). Прийняти гіпотезу про те, що дисперсія залишків пропорційна до зміни пояснюючої змінної (фактора) x , тому елементи матриці S отримати за формулою (53).

10. Сформуувати матрицю спостережень за незалежними змінними моделі для цього задати матрицю $H (18 * 2)$:

$$i := 0..17$$

$$H_{i,0} := 1 \quad H_{i,1} := X_i$$

11. Знайти вектор оцінок параметрів узагальненої моделі B за формулою (52).

12. Здійснити верифікацію отриманого рівняння регресії.

Завдання та запитання для самоконтролю

1. Що таке гетероскедастичність і її природа?
2. Як впливає гетероскедастичність на оцінки параметрів моделі, отриманих за МНК?
3. Основна ідея і алгоритм параметричного тесту Голдфелда–Квондта.
4. Основна ідея і алгоритм узагальненого методу найменших квадратів (методу Ейткена) у випадку гетероскедастичності.
5. Як визначається матриця перетворень S у випадку гетероскедастичності?

7 АВТОКОРЕЛЯЦІЯ ЗАЛИШКІВ

7.1 Визначення автокореляції залишків, її природа, причини виникнення і наслідки

Одним з основних припущень класичного лінійного регресійного аналізу є припущення про відсутність взаємозв'язку між значеннями стохастичної складової моделі ε у різних спостереженнях, тобто припущення

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j.$$

Якщо це припущення порушується виникає явище, яке носить назву **автокореляції залишків**.

Автокореляцією залишків називається залежність між послідовними значеннями стохастичної складової моделі .

У випадку автокореляції залишків маємо :

$$\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0, \quad i \neq j,$$

і , як у випадку гетероскедастичності, формально можна записати :

$$M(\varepsilon\varepsilon') = \sigma^2 S,$$

де σ^2 - деяка невідома константа,

S – відома квадратна, додатньо визначена матриця розмірністю $n \times n$.

Але на відміну від випадку гетероскедастичності матриця S не є діагональною, а є повною, діагональ якої містить одиниці, оскільки дисперсія випадкової величини ε у цьому випадку є сталою, а інші елементи, як було показано у попередній темі представляють собою ненульові коваріації значень випадкової величини ε у різних спостереженнях. Слід також зазначити , що вигляд і „наповнення” матриці S залежить від виду залежності між залишками.

У загальному випадку залежність між значеннями стохастичної складової ε у різних спостереженнях для випадку автокореляції можна подати наступним чином :

$$\varepsilon_i = \rho_1 \varepsilon_{i-1} + \rho_2 \varepsilon_{i-2} + \dots + \rho_s \varepsilon_{i-s} + u_i,$$

де : $\rho_1, \rho_2, \dots, \rho_s$ – коефіцієнти автокореляції 1,2 і s-го порядку відповідно ;

u_i – випадкова величина, яка відповідає усім припущенням класичного лінійного регресійного аналізу – тобто вона розподілена за нормальним законом із сталою дисперсією і має нульове математичне сподівання.

Найпростішим і найбільш поширеним випадком автокореляції залишків є випадок, коли залежність між послідовними значеннями стохастичної складової описується так званою **авторегресійною схемою першого порядку – AR(1)** яка має наступний вигляд :

$$\varepsilon_i = \rho_1 \varepsilon_{i-1} + u_i, \quad (-1 \leq \rho \leq 1). \quad (56)$$

Якщо ρ додатне ($\rho > 0$), то автокореляція залишків є **позитивною**, якщо ρ від'ємне ($\rho < 0$), то автокореляція залишків є **негативною**. При $\rho = 0$ автокореляція залишків відсутня.

Графічно випадки позитивної і негативної автокореляції залишків, а також її відсутності можна представити наступним чином (рис. 9).

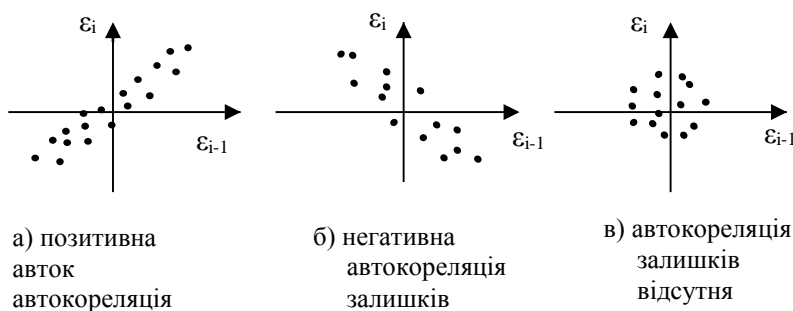


Рисунок 9- Графічна ілюстрація автокореляції залишків

Коефіцієнт автокореляції ρ у виразі (56) не може бути визначеним безпосередньо, оскільки неможливо визначити дійсні (у генеральній сукупності спостережень) значення випадкової величини ε_i . Але його можна оцінити звичайним методом найменших квадратів (МНК) на основі відомих залишків для статистичної вибірки. Тоді отримаємо :

$$\hat{\rho} = \frac{\text{cov}(e_{i-1}, e_i)}{\text{var}(e_{i-1})} \approx \frac{\sum_{i=2}^n e_{i-1} \cdot e_i}{\sum_{i=1}^n e_i^2}. \quad (57)$$

На практиці ж замість (57) частіше обчислюють наступну оцінку коефіцієнта автокореляції ρ :

$$\hat{\rho} = \frac{\frac{1}{n-1} \sum_{i=2}^n e_{i-1} \cdot e_i}{\frac{1}{n} \sum_{i=1}^n e_i^2} = \frac{n}{n-1} \cdot \frac{\sum_{i=2}^n e_{i-1} \cdot e_i}{\sum_{i=1}^n e_i^2}. \quad (58)$$

Оцінку (58) називають ще **циклічним коефіцієнтом автокореляції**.

Автокореляція залишків найчастіше спостерігається у наступних двох випадках :

1) коли економетрична модель будується на основі часових рядів (у цьому випадку, якщо існує кореляція між послідовними значеннями деякої незалежної змінної, то буде спостерігатися і кореляція між послідовними значеннями стохастичними складової ϵ , особливо ,якщо використовуються лагові змінні) ;

2) коли допущена помилка специфікації економетричної моделі – до моделі не включена істотна пояснююча змінна.

При наявності автокореляції залишків в принципі можливо оцінити параметри узагальненої економетричної моделі звичайним однокроковим методом найменших квадратів (МНК). Але отримані при цьому оцінки параметрів моделі не будуть BLUE – оцінками, оскільки хоча вони і будуть незміщеними, обґрунтованими, але вони ,як і у випадку гетероскедастичності будуть **неефективними**. Негативними наслідками цього, як і у випадку гетероскедастичності, буде :

- 1) завищені значення дисперсії параметрів моделі ;
- 2) помилки при використанні t – тестів і F – тестів ;
- 3) неефективність прогнозів, тобто отримання прогнозів з дуже великою дисперсією.

7.2 Тестування наявності автокореляції залишків

Оскільки автокореляція є негативним явищем потрібно вміти його тестувати. На даний час найбільш розповсюдженими тестами ,які використовуються для тестування автокореляції залишків є наступні статистичні тести :

- 1) тест Дарбіна - Уотсона ;
- 2) тест фон Неймана ;
- 3) тест на основі нециклічного коефіцієнта автокореляції ;
- 4) тест на основі циклічного коефіцієнта автокореляції .

Найбільш відомим і поширеним тестом перевірки моделі на наявність автокореляції залишків є тест Дарбіна-Уотсона. Цей тест використовується для авторегресійних схем 1-го порядку і має наступний алгоритм .

Алгоритм тесту Дарбіна - Уотсона

Крок 1. Виходячи з відсутності автокореляції залишків на основі МНК будується економетрична модель і обчислюються її залишки e_i ($i = \overline{1, n}$).

Крок 2. Розраховується статистика (критерій) Дарбіна-Уотсона за наступною залежністю :

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} . \quad (59)$$

Крок 3. Задаючись рівнем значимості α , для числа факторів моделі m і числа спостережень n за статистичними таблицями DW - розподілу Дарбіна-Уотсона визначаються два значення d_L , і d_U .

Крок 4. Будуються зони автокореляційного зв'язку, які схематично можна представити у наступному вигляді :



Рисунок 10 - Зони автокореляційного зв'язку

Крок 5. На основі розрахункового значення критерію DW робиться висновок про наявність або відсутність автокореляції залишків :

- якщо $0 < DW < d_L$ - це свідчить про наявність позитивної автокореляції залишків ;
- якщо $4 - d_L < DW < 4$ - це свідчить про наявність негативної автокореляції залишків;
- якщо $d_L \leq DW \leq d_U$ $\hat{=}$ $4 - d_U \leq DW \leq 4 - d_L$ - неможливо зробити висновок ні про наявність ні про відсутність автокореляції залишків ;
- якщо $d_U < DW < 4 - d_U$ - автокореляція залишків відсутня .

7.3 Оцінювання параметрів економетричних моделей у разі наявності автокореляції залишків

Для оцінювання параметрів економетричних моделей з автокорельованими залишками в основному використовуються наступні методи:

- 1) метод Ейткена (УМНК) ;
- 2) метод перетворення вихідної інформації ;
- 3) метод Кочрена – Оркатта ;
- 4) метод Дарбіна .

Перші два методи використовуються у випадку, коли залишки задовольняють авторегресійній схемі першого порядку, третій і четвертий можна застосовувати тоді, коли залишки описуються авторегресійною схемою вищого порядку.

Розглянемо докладніше метод Ейткена (узагальнений метод найменших квадратів). Цей метод як і у випадку гетероскедастичності базується на перетворенні вихідної моделі з урахуванням коваріації залишків (дисперсійно-коваріаційної матриці залишків) у модель без корельованих залишків, до якої потім застосовується оператор оцінювання МНК.

Нехай в економетричній моделі $Y = XB + \varepsilon$ випадкова величина задовольняє авторегресійній схемі першого порядку $\varepsilon_i = \rho_1 \varepsilon_{i-1} + u_i, (-1 \leq \rho \leq 1)$, де u_i – нормально розподілені залишки. Тоді оператор оцінювання параметрів моделі, як і випадку гетероскедастичності, буде мати вигляд :

$$B = (X' S^{-1} X)^{-1} X' S^{-1} Y, \quad (60)$$

де матриця S має вигляд :

$$S = \begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \rho^2 & \rho & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix}, \quad \dim S = n \times n. \quad (61)$$

Параметр ρ наближено можна визначити на основі залишків вибіркової моделі, оціненої за звичайним МНК. З цією метою розраховується так званий скоригований циклічний коефіцієнт кореляції :

$$r_{\text{скоп}} = \frac{n}{n-1} \cdot \frac{\sum_{i=2}^n e_i \cdot e_{i-1}}{\sum_{i=1}^n e_i^2} + \frac{m+1}{n}. \quad (62)$$

Тоді параметри ρ дорівнюють $\rho \approx r_{\text{скор}}$.

Оскільки коваріація залишків ρ^α ($\alpha = 0, 1, \dots, n-1$) у матриці S при $\alpha > 2$ часто наближається до нуля, у практичних розрахунках зразу ж можна визначити матрицю S^{-1} , обернену до матриці S :

$$S^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \quad \dim S = n \times n. \quad (63)$$

Використання виразу (63) значно полегшує і прискорює розрахунки оцінок параметрів моделі.

Оцінки параметрів узагальненої економетричної моделі з автокорельованими залишками, отримані методи Ейткена є BLUE - оцінками.

Незміщена оцінка дисперсії залишків $\hat{\sigma}_\varepsilon^2$, дисперсійно-коваріаційна матриця оцінок параметрів моделі, а також відповідні суми квадратів SST, SSR, SSE у цьому випадку можна обчислити за такими ж формулами, як і у випадку гетероскедастичності.

Найкращий незміщений лінійний точковий прогноз у випадку автокореляції залишків обчислюється за наступною залежністю:

$$\hat{y}_{\text{pr}} = X'_{\text{pr}} B + \rho \hat{e}_n,$$

де B – вектор оцінок параметрів моделі, отриманих узагальненим методом найменших квадратів (УМНК);

ρ – параметр з матриці S ;

\hat{e}_n - залишок в останньому спостереженні, обчислений для моделі, параметри якої оцінені на основі УМНК;

X_{pr} - вектор прогнозних значень пояснюючих змінних моделі.

Лабораторна робота № 9 «Автокореляція залишків»

Мета роботи: Набуття практичних навичок тестування наявності автокореляції залишків і оцінювання параметрів економетричної моделі з автокорельованими залишками узагальненим методом найменших квадратів.

Завдання роботи і вихідні дані

Для деякого регіону виконати дослідження залежності Роздрібний товарообіг (y) від доходу населення (x). Вибіркові статистичні дані за 10 років наведені нижче у таблиці 14.

Таблиця 14 – Вихідні дані

Рік	1	2	3	4	5
Роздрібний товарообіг (млрд. гр. од.)	24	25	25,7	27	28,8
Доходи населення (млрд. гр. од.)	27,1	28,2	29,3	31,3	34
Рік	6	7	8	9	10
Роздрібний товарообіг (млрд. гр. од.)	30,8	33,8	38,1	49,4	51,5
Доходи населення (млрд. гр. од.)	36	38,7	43,2	50	52,1

Грунтуючись на наведених статистичних даних:

1. Перевірити наявність автокореляції залишків за допомогою тесту Дарбіна–Уотсона.
2. Визначити оцінки параметрів моделі узагальненим методом найменших квадратів.
3. Перевірити статистичну значимість оціненої узагальненої моделі з автокорельованими залишками.

Порядок виконання роботи

1. На основі вибірових статистичних спостережень за 10 років побудувати модель парної регресії. Для цього за методом найменших квадратів визначаються оцінки параметрів моделі.
2. На основі оціненого рівняння регресії обчислити розрахункові значення залежної змінної $\hat{y}_i, (i = \overline{1,10})$ і залишки $e_i, (i = \overline{1,10})$.
3. Використовуючи обчислені залишки, розраховується критерій Дарбіна–Уотсона формула (59).
4. За статистичними таблицями DW-розподілу Дарбіна-Уотсона для рівня значимості $\alpha = 0,05$, числа спостережень $n = 10$ і числа факторів моделі $m=1$ знайти критичні точки d_L і d_U .
5. На основі знайдених значень DW, d_L і d_U зробити висновок про відсутність або наявність автокореляції залишків.

6. Прийняти гіпотезу про те, що залишки моделі відповідають авторегресійній схемі першого порядку за формулою (56).

7. Обчислити оцінку коефіцієнта автокореляції ρ за формулою (62).

8. Сформувати матрицю спостережень за незалежними змінними моделі для цього задати матрицю H ($18 * 2$):

$$i := 0..9$$

$$H_{i,0} := 1 \quad H_{i,1} := X_i$$

9. Сформувати матрицю S^{-1} за формулою (63).

10. Знайти вектор оцінок параметрів узагальненої моделі B за формулою (60).

9. Записати оцінене рівняння регресії та здійснити оцінку статистичної значущості узагальненої моделі з автокорельованими залишками.

Завдання та запитання для самоконтролю

1. Що таке автокореляція залишків економетричної моделі, природа і причини цього явища?

2. Як впливає автокореляція залишків на оцінки параметрів економетричної моделі, які оцінені за 1МНК?

3. Алгоритм і розрахункові залежності тесту Дарбіна–Уотсона на автокореляцію залишків.

4. Основна ідея і алгоритм узагальненого методу найменших квадратів (методу Ейткена) у випадку автокореляції залишків.

5. Як визначається матриця перетворень S у випадку автокореляції залишків?

8 АНАЛІЗ ЧАСОВОГО РЯДУ ІЗ ВИКОРИСТАННЯМ РЯДІВ ФУР'Є

Якщо в часовому ряді проявляється тривала тенденція зміни економічного показника, то говорять, що має місце тренд. Таким чином, під трендом розуміється зміна, що визначає загальний напрямок розвитку, основну тенденцію часових рядів. У зв'язку із цим економіко-математична динамічна модель, у якій розвиток модельованої економічної системи відображається через тренд її основних показників, називається трендовою моделлю. Для виявлення тренда в часових рядах (зокрема в ряді рівнів попиту), а також для побудови й аналізу трендових моделей використовується апарат теорії ймовірностей і математичної статистики, розроблений для простих статистичних сукупностей. Відмінність часових економічних рядів від простих статистичних сукупностей полягає, насамперед, у тому, що послідовні значення рівнів часового ряду залежать один від одного. Тому застосування висновків і формул теорії ймовірностей і математичної статистики вимагає відомої обережності при аналізі часових рядів, особливо при економічній інтерпретації результатів аналізу.

У часових рядах економічних процесів можуть мати місце більш-менш регулярні коливання. Якщо вони носять строго періодичний або близький до нього характер і завершуються протягом одного року, то їх називають сезонними коливаннями. У тих випадках, коли період коливань становить кілька років, то говорять, що в часовому ряді присутній циклічний компонент. Для моделювання й прогнозування сезонних і циклічних економічних процесів використовуються спеціальні методи (індексний і спектральний аналізи, вирівнювання по ряду Фур'є й ін.).

Тренд, сезонний і циклічний компонент називаються регулярними, або систематичними компонентами часового ряду. Складова частина часового ряду, що залишається після виділення з нього регулярних компонентів, являє собою випадковий, нерегулярний компонент. Вона є обов'язковою складовою частиною будь-якого часового ряду в економіці, тому що випадкові відхилення неминуче супроводжують будь-якому економічному явищу. Якщо систематичні компоненти часового ряду визначені правильно, що саме й становить одну з головних цілей при розробці трендових моделей, то залишкова послідовність (ряд залишків), що залишається після виділення з часового ряду цих компонентів, буде випадковим компонентом ряду, тобто буде мати наступні властивості:

- випадковість коливань рівнів залишкової послідовності;
- відповідність випадкового компонента нормальному закону розподілу;
- рівність математичного очікування випадкового компонента нулю;

- незалежність значень рівнів випадкової послідовності, тобто відсутність істотної автокореляції.

Перевірка зазначених властивостей здійснюється з використанням ряду статистичних критеріїв.

Для здійснення прогнозування необхідно одержати моделі часових рядів.

Часовий ряд із сезонною складовою представляють у вигляді:

$$F_i(t) = f(t) + \varphi(t) + \varepsilon(t), \quad i=1, \dots, m$$

де $F_i(t)$ - спостережуване значення i -го фактору (ряду) у момент часу t ;

$f(t)$ – трендова складова, що визначає загальну тенденцію розвитку спостережуваного процесу;

$\varphi(t)$ - сезонна складова, що визначає періодичні коливання значень фактору;

$\varepsilon(t)$ - випадкові коливання.

У загальному виді процедура побудови моделей здійснюється в кілька етапів:

- 1) виявлення тренда;
- 2) виділення трендової складової;
- 3) виділення сезонної складової;
- 4) перевірка адекватності отриманої моделі.

Розглянемо ці етапи послідовно.

8.1 Перевірка наявності тренда

На першому етапі аналізу тренда розвитку необхідно перевірити його існування.

Перевірку наявності тренда можна проводити за допомогою будь-якого відомого критерію, наприклад, метод Фостера-Стюарта, метод Мостеллера, знаковий критерій Діксона-Муда, критерій Зігеля-Т'юкі й інших. Розглянемо ці методи.

Метод Фостера-Стюарта дозволяє визначити як наявність тенденції явища, так і тренд дисперсії рівнів ряду динаміки.

На підставі порівняння кожного рівня ряду, починаючи із другого, з попередніми рівнями визначаються величини:

$$S = \sum_{i=1}^n S_i = \sum_{i=1}^n (U_i + e_i),$$

$$d = \sum_{i=1}^n d_i = \sum_{i=1}^n (U_i - e_i),$$

де $U = 1$, якщо i - тий рівень більше всіх попередніх;

$U = 0$ у противному випадку;
 $e = 1$, якщо i - тий рівень менше всіх попередніх;
 $e = 0$ у противному випадку.

З використанням t-критерію Ст'юдента перевіряються гіпотези про випадковості відхилення величин:

- S від μ – математичного очікування величини S для ряду, у якому рівні розташовані випадковим образом;
- d від нуля.

Ця перевірка проводиться з використанням розрахункових значень t-критерію Ст'юдента для середньої й для дисперсії:

$$t_S = \frac{S - \mu}{\sigma_1};$$

$$\sigma_1 = \sqrt{2 \ln(n) - 3,4253};$$

$$t_d = \frac{d - 0}{\sigma_2};$$

$$\sigma_2 = \sqrt{2 \ln(n) - 0,846},$$

де μ – середнє значення величини S , визначене для ряду, у якому рівні розташовані випадковим образом;

σ_1, σ_2 – стандартні помилки (середньоквадратичне відхилення) відповідно для величин S, d .

Отримані значення t-критерію порівнюються з табличними.

Якщо $t_S < t_{табл}$, то гіпотеза про відсутність тренда підтверджується.

Якщо $t_d < t_{табл}$, то гіпотеза про відсутність тенденції в дисперсії підтверджується.

Критерій Зігеля-Т'юкі припускає, що об'єднана вибірка, що містить $n_1 + n_2 = n$ значень, упорядковується по рангах, причому найменше значення одержує ранг 1, два найбільших значення одержують ранги 2 і 3, наступні два найменших значення – ранги 4 і 5, ранги 6 і 7 одержують наступні два найбільших значення й т.д.. Потім для кожної вибірки визначається сума рангових чисел (R_1, R_2) . При $n_1 = n_2$ нульовій гіпотезі (обидві підвибірки належать одній і тій же генеральній сукупності) відповідає співвідношення $R_1 \approx R_2$. Для безпосереднього кількісного порівняння використовуються відповідні таблиці.

Однак ці розглянуті методи виявлення тренда неточні, більш точними методами є знаковий критерій Діксона-Муда, метод Мостеллера й метод, заснований на порівнянні вибірових середніх.

Метод, заснований на порівнянні вибірових середніх, полягає в наступному. Вихідний часовий ряд розбивається на 2 підвибірки однакового розміру:

$$Y_1 = \left\{ y_1, y_2, \dots, y_{\frac{n}{2}} \right\};$$

$$Y_2 = \left\{ y_{\frac{n+1}{2}}, y_{\frac{n+2}{2}}, \dots, y_n \right\}.$$

Для кожної з підвбірок обчислюються середні значення \bar{y}_1 , \bar{y}_2 і дисперсії σ_1^2 , σ_2^2 по формулах:

$$\bar{y}_1 = \frac{\sum_{j=1}^{n_1} y_{1j}}{n_1}, \quad \bar{y}_2 = \frac{\sum_{j=1}^{n_2} y_{2j}}{n_2}$$

$$\sigma_1^2 = \frac{\sum_{j=1}^{n_1} (y_{1j} - \bar{y}_1)^2}{n_1}, \quad \sigma_2^2 = \frac{\sum_{j=1}^{n_2} (y_{2j} - \bar{y}_2)^2}{n_2}$$

де n_1 й n_2 – число елементів, що входять у першу й другу підвибірки відповідно.

Потім уводиться нульова гіпотеза H_0 про рівність середніх генеральних сукупностей, що лежать в основі порівнюваних підвбірок $Y_1 = \left\{ y_1, y_2, \dots, y_{\frac{n}{2}} \right\}$,

$Y_2 = \left\{ y_{\frac{n+1}{2}}, y_{\frac{n+2}{2}}, \dots, y_n \right\}$. Рівність цих середніх означає, що обидві підвбірки належать однієї й тій же генеральній сукупності, тобто тренда немає. Альтернативна гіпотеза H_1 про розходження середніх підвбірок Y_1 і Y_2 еквівалентна припущенню про наявність тренда.

Для перевірки гіпотез використовується t-статистика Ст'юдента:

$$\hat{t}_{\text{набл}} = \frac{|\bar{y}_1 - \bar{y}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Обчислене значення $\hat{t}_{\text{набл}}$ порівнюється зі значенням $t_{\text{кр}}$, обраним з таблиці розподілу Ст'юдента для заданого рівня значимості α й числа ступенів волі $\nu = n_1 + n_2 - 2$. Якщо $\hat{t}_{\text{набл}} < t_{\text{кр}}$, то гіпотеза H_0 приймається, у противному випадку – відкидається.

Метод Мостеллера, як і попередній метод, припускає, що вихідний часовий ряд розбивається на 2 підвибірki однакового розміру:

$$Y_1 = \left\{ y_1, y_2, \dots, y_{\frac{n}{2}} \right\}, Y_2 = \left\{ y_{\frac{n}{2}+1}, y_{\frac{n}{2}+2}, \dots, y_n \right\}.$$

Потім вводиться нульова гіпотеза (обидві вибірки належать одній й тій же генеральній сукупності), що відхиляється з імовірністю помилки $\alpha=0,05$, якщо 6 найбільших або найменших значень містить якась одна підвибірka.

Знаковий критерій Діксона-Муда використовується у випадку, коли $n_1 = n_2$ (як і попередні методи). Утворюються пари послідовних спостережень $(y_1, y_{\frac{n}{2}+1})$, $(y_2, y_{\frac{n}{2}+2})$, ..., $(y_{\frac{n}{2}}, y_n)$... Для кожної пари обчислюється різниця $d_i = y_i - y_{\frac{n}{2}+i}$, потім розраховується число пар, для яких $d_i \geq 0$, і число пар, для яких $d_i < 0$. Нехай S_+ - число пар, для яких $d_i \geq 0$, а S_- - число пар для яких $d_i < 0$. Статистикою \hat{T} служить абсолютне значення різниці числа позитивних і числа негативних пар, тобто $\hat{T} = |S_+ - S_-|$.

Якщо $\hat{T} > a\sqrt{n}$, то нульова гіпотеза про відсутність тренда повинна бути відкинута.

Всі перераховані методи виявлення тренда не дають точних результатів. Самим точним методом, що дозволяє не тільки виявити тренд, але й виділити його є побудова рівняння регресії.

Трендову складову узагальнено можна представити в наступному виді:

$$f(t) = \sum_{i=0}^d a_i t^i, \quad (64)$$

де $f(t)$ - ендогенна змінна (у нашому випадку рівень попиту);

t - екзогенна змінна (час).

При цьому виникає необхідність оцінки невідомих параметрів a моделі (64). Із цією метою використовується стандартний метод найменших квадратів (МНК). Показник ступеня d полінома (64) вибирають із міркувань одержання адекватного опису за критерієм Фішера.

8.2 Вирівнювання часових рядів по ряду Фур'є

Дуже часто рівні економічних рядів динаміки коливаються, при цьому тенденція розвитку економічного явища в часі схована випадковими відхиленнями рівнів у ту або іншу сторону. З метою більш чітко виявити тенденцію розвитку досліджуваного процесу, у тому числі для подальшого прогнозування на основі трендових моделей, роблять згладжування (вирівнювання) часових рядів.

Основна мета створення трендових моделей економічної динаміки - на їхній основі зробити прогноз про розвиток досліджуваного процесу на майбутній проміжок часу. Прогнозування на основі часового ряду економічних показників відноситься до одномірних методів прогнозування, що базуються на екстраполяції, тобто на продовженні на майбутнє тенденції, що спостерігалася в минулому. При такому підході передбачається, що прогнозований показник формується під впливом великої кількості факторів, які або неможливо виділити, або по яких відсутня інформація. У цьому випадку хід зміни даного показника зв'язують не з факторами, а із часом, що проявляється в утворенні одномірних часових рядів. Використання методу екстраполяції для прогнозування базується на двох припущеннях:

- часовий ряд економічного показника дійсно має тренд, тобто переважну тенденцію;
- загальні умови, що визначали розвиток показника в минулому, залишаються без істотних змін протягом періоду попередження.

Окреме місце в аналітичному вирівнюванні рядів динаміки займає згладжування за допомогою ряду Фур'є, що описується рівнянням:

$$Y_t = a_0 + \sum_{k=1}^m \left(a_k \cdot \cos\left(\frac{2\pi}{T_k} \cdot t\right) + a_{k+1} \cdot \sin\left(\frac{2\pi}{T_k} \cdot t\right) \right),$$

де T_k - період k -тої гармоніки;

t - час, виражений у радіанах або градусах.

Згладжування по наведеній формулі доречно, коли в емпіричному ряді є явна періодичність змін його рівнів, що має вигляд синусоїдних коливань, які є

гармонійними коливаннями. Синусоїди, отримані внаслідок згладжування рядом Фур'є, називають гармоніками випадкових порядків.

У випадку згладжування рядом Фур'є періодичні коливання рівнів динамічного ряду мають вигляд суми декількох гармонік, нашарованих одна на одну. Так, при $m=1$ ряд Фур'є має вигляд:

$$Y_t = a_0 + a_1 \cos\left(\frac{2\pi}{T_1}t\right) + a_2 \sin\left(\frac{2\pi}{T_1}t\right)$$

де T_1 - період спостережень.

При $m=2$ відповідно:

$$Y_t = a_0 + a_1 \cos\left(\frac{2\pi}{T_1}t\right) + a_2 \sin\left(\frac{2\pi}{T_1}t\right) + a_3 \cos\left(\frac{2\pi}{T_2}t\right) + a_4 \sin\left(\frac{2\pi}{T_2}t\right),$$

де $T_2 = \frac{T_1}{2}$.

Згладжування по ряду Фур'є найчастіше використовують для спостереження сезонності різних соціально-економічних явищ і процесів.

Параметри рівняння теоретичних рівнів визначають способом найменших квадратів.

У результаті перетворень, одержуємо систему нормальних рівнянь, по яких можна обчислити параметри:

$$a_0 = \frac{\sum y}{n};$$

$$a_1 = \frac{2\sum y \cos t}{n};$$

$$a_2 = \frac{2\sum y \sin t}{n}$$

Аналогічно розраховують рівняння ряду Фур'є із застосуванням другої, третьої і четвертої гармоніки з перевіркою збігів їх теоретичних і фактичних значень.

Щоб підвищити точність згладжування часових рядів за допомогою ряду Фур'є, у рівняння (2.17) вводять зрушення по фазі φ_n . Тоді ряд Фур'є має вигляд:

$$Y_t = a_0 + \sum_{k=1}^m \left(a_k \cdot \cos\left(\frac{2\pi}{T_k} t + \varphi_k\right) + a_{k+1} \cdot \sin\left(\frac{2\pi}{T_k} t + \varphi_{k+1}\right) \right)$$

Після здійснення згладжування необхідно перевірити отриману модель досліджуваного процесу на адекватність із використанням описаних критеріїв Фішера або детермінації. Якщо ця перевірка покаже, що моделі неадекватні, необхідно заново проводити перевірку на наявність автокореляції залишків і будувати нову модель, що також перевіряється на адекватність.

Тренд і сезонна складові є регулярними, або систематичними компонентами часового ряду. Складова частина часового ряду, що залишається після виділення з нього регулярних компонентів, являє собою випадковий, нерегулярний компонент ($\varepsilon(t)$). Вона є обов'язковою складовою частиною будь-якого часового ряду в економіці, тому що випадкові відхилення неминуче супроводжують будь-якому економічному явищу. Прогноз випадкової компоненти складніше, однак, якщо є припущення про те, як вона розподілена, у неї можна виділити складову й описати її законом за допомогою щільності розподілу. Частина, що залишилася, не цікава і являє собою так званий «білий шум».

Лабораторна робота № 9 «Аналіз сезонності»

Мета роботи: Набуття практичних навичок побудови та використання моделі часового ряду.

Задачі роботи:

1. Специфікація моделі.
2. Побудова трендової складової моделі.
3. Побудова циклічної складової моделі
4. Верифікація моделі.
5. Використання моделі часового ряду у прогнозуванні.

Вихідні дані:

Маються дані за обсягами продажів за три роки у помісячному розрізі
Таблиця 15 – Вихідні дані

Період	1	2	3	4	5	6	7	8	9
Обсяг	421,2	405	412,4	375,6	390,1	370,1	391,1	387,5	388,9
Період	10	11	12	13	14	15	16	17	18
Обсяг	380,5	406,3	417,6	405,9	382,1	397,7	363,6	375,5	360
Період	19	20	21	22	23	24	25	26	27
Обсяг	380,1	376	375,2	367,5	395,4	397	385	359,2	377,4
Період	28	29	30	31	32	33	34	35	36
Обсяг	349,9	353,1	330,2	354,2	352,5	353,9	345,5	376,2	373

Завдання роботи:

1. Побудувати модель часового ряду, що описує динаміку обсягів продажів підприємства з часом.
2. Оцінити якість, адекватність і статистичну значимість побудованої моделі для рівня значимості $\alpha = 0,05$.
3. Розрахувати прогноз обсягів продажів на три наступні роки.

Методичні вказівки до порядку виконання роботи

1. З урахуванням номеру варіанту (кожне значення обсягу помножити на N) ввести вихідні дані.
2. Побудувати трендову складову моделі. Тобто за методом найменших квадратів визначити оцінки параметрів моделі a_0, a_1 , (деталі див. у вказівках до лабораторної роботи «Парна регресія»).
3. Перевірити наявність автокореляції залишків за допомогою тесту Дарбіна–Уотсона (деталі див. у вказівках до лабораторної роботи «Автокореляція

залишків»).

4. Наявність залишків може бути обумовлена наявністю сезонних коливань у вихідних даних. Для того, щоб в цьому переконатися, необхідно зобразити графічно залишки часового ряду та здійснити їх візуальний аналіз.

5. За результатами аналізу поведінки залишків висунути гіпотезу про вигляд моделі залишків, а саме кількості гармонік, періоди їх коливань.

6. На підставі висунутої гіпотези про вигляд сезонної складової моделі (кількості гармонік, їх періоди), задати матрицю $H1$:

$$H1 = \begin{pmatrix} \cos\left(\frac{2\pi}{T_1} t_1\right) & \sin\left(\frac{2\pi}{T_1} t_1\right) & \cos\left(\frac{2\pi}{T_2} t_1\right) & \sin\left(\frac{2\pi}{T_2} t_1\right) & \cdots & \cos\left(\frac{2\pi}{T_m} t_1\right) & \sin\left(\frac{2\pi}{T_m} t_1\right) \\ \cos\left(\frac{2\pi}{T_1} t_2\right) & \sin\left(\frac{2\pi}{T_1} t_2\right) & \cos\left(\frac{2\pi}{T_2} t_2\right) & \sin\left(\frac{2\pi}{T_2} t_2\right) & \cdots & \cos\left(\frac{2\pi}{T_m} t_2\right) & \sin\left(\frac{2\pi}{T_m} t_2\right) \\ \cos\left(\frac{2\pi}{T_1} t_3\right) & \sin\left(\frac{2\pi}{T_1} t_3\right) & \cos\left(\frac{2\pi}{T_2} t_3\right) & \sin\left(\frac{2\pi}{T_2} t_3\right) & \cdots & \cos\left(\frac{2\pi}{T_m} t_3\right) & \sin\left(\frac{2\pi}{T_m} t_3\right) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cos\left(\frac{2\pi}{T_1} t_n\right) & \sin\left(\frac{2\pi}{T_1} t_n\right) & \cos\left(\frac{2\pi}{T_2} t_n\right) & \sin\left(\frac{2\pi}{T_2} t_n\right) & \cdots & \cos\left(\frac{2\pi}{T_m} t_n\right) & \sin\left(\frac{2\pi}{T_m} t_n\right) \end{pmatrix}.$$

де T_1, T_2, \dots, T_m – періоди гармонік

m – кількість пар гармонік

7. Оцінити невідомі параметри сезонної складової моделі з використанням МНК (приклад схеми розрахунку див. у вказівках до лабораторної роботи «Автокореляція залишків»)

$$B = (H1^T H1)^{-1} H1^T \xi$$

де ξ - вектор значень залишків часового ряду;

8. Побудувати графік залишків часового ряду та значення залишків, що передбачаються моделлю. Візуально оцінити якість моделі.

9. Оцінити адекватність сезонної складової моделі

$$R^2 = \frac{\sum_{j=1}^n (\hat{\xi}_j - \bar{\xi})^2}{\sum_{j=1}^n (\xi_j - \bar{\xi})^2},$$

де $\bar{\xi}$ - середнє значення залишків часового ряду;

$\hat{\xi}_j$ - значення залишків часового ряду, отримані за допомогою циклічної

складової моделі, які розраховуються наступним чином: $\hat{\xi} = H1B$

10. Якщо якість моделі не задовільна, то корегувати періоди коливань T_1, T_2, \dots, T_m . Причому в парі гармонік $\sin()$ та $\cos()$ періоди можуть бути різними. Та повернутися до п. 7.

11. Об'єднати трендову і сезонну складові моделі. Для цього додати до значень, що передбачаються трендовою складовою моделі значення, що передбачаються сезонною складовою моделі: $\hat{Y} = HA + H1B$, тобто використовувати результати отримані у п.2 та у п.9.

12. Побудувати графік вихідного часового ряду та значення обсягів продажів, що передбачаються моделлю. Візуально оцінити якість отриманої моделі часового ряду.

15. Оцінити адекватність загальної моделі

16.

$$R^2 = \frac{\sum_{j=1}^n (\hat{y}_j - \bar{y})^2}{\sum_{j=1}^n (y_j - \bar{y})^2},$$

де \bar{y} - середнє значення вихідного часового ряду;

\hat{y}_j - значення часового ряду, що передбачаються загальною моделлю, які отримані у п.10.

16. Здійснити прогноз на наступні три періоди

$$Y(t_{np}) = a_0 + a_1 \cdot t_{np} + b_1 \cos\left(\frac{2\pi}{T_1} t_{np}\right) + b_2 \sin\left(\frac{2\pi}{T_1} t_{np}\right) + b_3 \cos\left(\frac{2\pi}{T_2} t_{np}\right) + b_4 \sin\left(\frac{2\pi}{T_2} t_{np}\right)$$

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Багриновский К.А., Сумин Г.А. Математические методы в экономике и планировании народного хозяйства. М.: Изд-во РУДН, 1993.
2. Бородич С.А. Эконометрика: Учебное пособие.- Мн.: Новое знание, 2001. -408 с.
3. Грубер Й. Економетрія. Том 1. Вступ до множинної регресії та економетрії. – К.: Ніч лава, 1998. – 381 с.
4. Грубер Й. Економетрія. Том 2. Економічні прогнозні та оптимізаційні моделі. – К.: Ніч лава, 1999. – 295 с.
5. Доугерти К. Введение в эконометрику. - М.: ИНФРА. - М, 2001. - 402 с.
6. Дубина А.Г., Орлова С.С., Шубина И.Ю., Хромов А.В. Excel для экономистов и менеджеров.- СПб.: Питер, 2004.-295 с.
7. Замков О.О., Толстопятенко А.В., Черемных Ю.Н. Математические методы в экономике. М.: АО ДИС, 1997.
8. Карасев А.И., Кремер Н.Ш., Савельев Т.И. Математические методы и модели в планировании. М.: Экономика, 1987.
9. Карлберг К. Бизнес-анализ с помощью Excel / Пер. с англ. К.: Диалектика, 1997.
10. Кубонива М. и др. Математическая экономика на персональном компьютере / Пер. с яп. М.: Финансы и статистика, 1991.
11. Лебедев В.В. Математическое моделирование социально-экономических процессов. М.: Изограф, 1997.
12. Лопатников Л. Экономико-математический словарь. М.: АБФ, 1996.
13. Лук'яненко І.Г., Краснікова Л.І., Економетрика: підручник. - К: товариство "Знання", КОО, 1998. - 494 с.
14. Лук'яненко І.Г., Краснікова Л.І. Економетрика: практикум з використанням комп'ютера. – К.: Товариство «Знання», КОО, 1998. – 220 с.
15. Магнус Я.Р., Катышев П.Л., Пересецкий А.А. Эконометрика. Начальный курс. М.: Дело, 1997. – 248 с.
16. Макаренко Т.І. Моделювання та прогнозування у маркетингу: Навчальний посібник. – Київ, „Центр навчальної літератури”, 2005. – 160с.
17. Матюшок В.М. Excel 7.0. Решение общих и экономических задач. М.: Изд-во РУДН, 1997.
18. Матюшок В.М. и др. Персональный компьютер: диалог и программные средства. М.: Изд-во УДН, 1989.
19. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: підручник. - К.: КНЕУ, 2000.-296 с.

20. Наконечний С.І., Терещенко Т.О. Економетрія: Навч.-метод. посібник для самостійного вивчення дисципліни-К.:КНЕУ,2001.-192 с.
21. Раскин Л.Г., Харченко А.А. Эконометрия. Учебное пособие для студентов заочной формы обучения. – Харьков, 1999. – 79 с.
22. Статистическое моделирование и прогнозирование. Учебное пособие (Под ред. А.Г. Гранберга) – М.: Финансы и статистика, 1990. – 383 с.
23. Толбатов Ю.А. Эконометрика: Підручник для студентів екон. спеціальн. вищ. навч. зал. - К.: Четверта хвиля, 1997. - 320 с.
24. Федосеев В.М. Экономико-математические методы и модели в маркетинге - М.: Финстатинформ, 1996.

ДОДАТОК А Статистичні таблиці

Таблиця А.1 - Таблиця F-розподілу

v ₂	v ₁								
	1	2	3	4	5	6	7	8	9
1	161	200	216	225	230	234	237	239	241
2	18,5	19,0	19,2	19,2	19,3	19,3	1,94	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,82	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39

$v_1 = m, v_2 = n-k, m$ - кількість факторів (пояснюючих змінних),
 n - кількість спостережень, k - кількість параметрів моделі.

Таблиця А.2 - Критичні точки розподілу χ^2

Ступінь вільності df	Довірча імовірність p					
	0,99	0,975	0,95	0,05	0,025	0,01
1	6,6	5,0	3,8	0,0039	0,001	0,0002
2	9,2	7,4	6,0	0,103	0,051	0,621
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872

$df = \frac{1}{2} m(m-1)$, де m - кількість факторів (пояснюючих змінних).

Таблиця А.3 - Таблиця t-розподілу Ст'юдента

v	Рівень значимості α (для двостороннього тесту)						
	0.50	0.20	0.10	0.05	0.02	0.01	0.002
1	1,000	3,078	6,314	12,706	31,821	63,657	318,3
2	0,861	1,886	2,920	4,303	6,965	9,925	22,33
3	0,765	1,638	2,353	3,182	4,541	5,841	10,21
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025
12	0,695	1,356	1,782	2,179	2,681	3,055	3,93
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552

$v = n - k$, n – кількість спостережень, k – кількість параметрів моделі.

Таблиця А.4 - DW-статистика Дарбіна-Уотсона.

Критичні точки d_L і d_U при рівні значимості $\alpha = 0,05$

Число Спостережень n	Число факторів							
	m=1		m=2		m=3		m=4	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40	---	---	---	---	---	---
7	0,70	1,36	0,47	1,90	---	---	---	---
8	0,76	1,33	0,56	1,78	0,37	2,29	---	---
9	0,82	1,32	0,63	1,70	0,46	2,13	0,30	2,59
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41

**Таблиця А.5 - Таблиця стандартизованого нормального розподілу
(функція Лапласа)**

u	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0,0	,0000	,0040	,0080	,0120	,0160	,0199	,0239	,0279	,0319	,0359
0,1	,0398	,0438	,0478	,0517	,0557	,0596	,0636	,0675	,0714	,0753
0,2	,0793	,0832	,0871	,0910	,0948	,0987	,1026	,1064	,1103	,1141
0,3	,1179	,1217	,1255	,1293	,1331	,1368	,1406	,1443	,1480	,1517
0,4	,1554	,1591	,1628	,1664	,1700	,1736	,1772	,1808	,1844	,1879
0,5	,1915	,1950	,1985	,2019	,2054	,2088	,2123	,2157	,2190	,2224
0,6	,2257	,2291	,2324	,2357	,2389	,2422	,2454	,2486	,2517	,2549
0,7	,2580	,2611	,2642	,2673	,2704	,2734	,2764	,2794	,2823	,2852
0,8	,2881	,2910	,2939	,2967	,2995	,3023	,3051	,3078	,3106	,3133
0,9	,3150	,3186	,3212	,3238	,3264	,3289	,3315	,3340	,3365	,3389
1,0	,3413	,3438	,3461	,3485	,3508	,3531	,3554	,3577	,3599	,3621
1,1	,3643	,3665	,3686	,3708	,3729	,3749	,3770	,3790	,3810	,3830
1,2	,3849	,3869	,3888	,3907	,3925	,3944	,3962	,3980	,3997	,4015
1,3	,4032	,4049	,4066	,4082	,4099	,4115	,4131	,4147	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4251	,4265	,4279	,4292	,4306	,4319
1,5	,4332	,4345	,4357	,4370	,4382	,4394	,4406	,4418	,4429	,4441
1,6	,4452	,4463	,4474	,4484	,4495	,4505	,4515	,4525	,4535	,4545
1,7	,4554	,4564	,4573	,4582	,4591	,4599	,4608	,4616	,4625	,4633
1,8	,4611	,4649	,4656	,4664	,4671	,4678	,4686	,4693	,4699	,4706
1,9	,4713	,4719	,4726	,4732	,4738	,4744	,4750	,4756	,4761	,4767
2,0	,4772	,4778	,4783	,4788	,4793	,4798	,4803	,4808	,4812	,4817

Навчальне видання

ХАРЧЕНКО Алла Олександрівна;
КОНОХОВА Зоя Петрівна;

Методичні вказівки
до виконання лабораторних робіт
з дисципліни «**Економетрика**»
для студентів освітньо-кваліфікаційного рівня бакалавр
спеціальностей **051 «Економіка» та 075 «Маркетинг»**

Роботу до видання рекомендував проф. Заруба В.Я.

В авторській редакції

Підписано до друку 26.06.2019 Формат 60×90 1/16.
Папір офсетний. Умов.-друк. арк.6,4.
Тираж 100 прим.

.....

Надруковано у копії-центрі «МОДЕЛІСТ»
ФОП Миронов М.В.
Свідоцтво про реєстрацію №022953
М. Харків, вул. Мистецтв, 3 літер Б-1
Тел. (057) 717-03-54
www.modelist.in.ua