МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Кафедра економічної кібернетики

Методичні вказівки та завдання до контрольної / домашньої роботи

з дисципліни

"Економетрика"

для студентів економічних спеціальностей всіх форм навчання

Київ 2019

ЗМІСТ

ПРАВИЛА ВИКОНАННЯ ТА ОФОРМЛЕННЯ КОНТРОЛЬНОЇ РОБОТИ

1. Студент повинен виконати контрольну роботу, яка містить два завдання.

Варіант для виконання контрольної роботи вибирається за двома останніми цифрами студентського шифру. Позначимо число, складене із цих цифр, через n. Тоді:

а) якщо $0 < n \le 30$, номер варіанту дорівнює n;

б) якщо $30 < n \le 60$, номер варіанту дорівнює n -30;

в) якщо $60 < n \le 90$, номер варіанту дорівнює n -60;

г) якщо $90 < n \le 99$, номер варіанту дорівнює n - 90;

д) якщо $n = 0$, номер варіанту дорівнює 30.

2. На титульному аркуші зошита мають бути написані прізвище та ініціали студента, номер його залікової книжки, назва групи, назва дисципліни, номер варіанта.

3. Перед розв'язанням задачі необхідно вказати її номер та повністю переписати її умову.

4. Після рецензування роботи викладачем студент повинен передивитися роботи та виконати повторно неправильно виконані завдання.

5. Контрольні роботи підлягають захисту.

ПРОГРАМА КУРСУ

Тема 1. *Основи економетричного моделювання*

Приклади економетричних моделей. Виробнича функція Кобба-Дугласа, її властивості й оцінка параметрів.

Тема 2. *Методи побудови загальної лінійної моделі*

Проста економетрична модель. Поняття моделі та етапи її побудови. Специфікація моделі.

Тема 3. *Дисперсійний аналіз економетричної моделі*

Передумови застосування методу найменших квадратів. Побудова економетричної моделі на основі покрокової регресії.

Тема 4. *Мультиколінеарність та її вплив на оцінки параметрів моделі*

Поняття мультиколінеарності. Основні наслідки мультиколінеарності. Ознаки мультиколінеарності. Алгоритм Фаррара-Глобера. Методи звільнення від мультиколінеарності.

Тема 5. *Гетероскедастичність*

Поняття гетероскедастичності. Наслідки гетероскедастичності. Методи визначення гетероскедастичності. Узагальнений метод найменших квадратів (метод Ейткена). Прогноз.

Тема 6. *Автокореляція*

Причини виникнення автокореляції в економетричних моделях. Перевірка наявності автокореляції. Оцінювання параметрів моделі з автокорельованими залишками. Прогноз

Тема 7. *Метод інструментальних змінних*

Властивості оцінок у разі стохастичних змінних. Метод інструментальних змінних. Визначення інструментальних змінних. Помилки вимірювання змінних.

Тема 8. *Моделі розподіленого лагу*

Поняття лагу і лагових змінних. Взаємна кореляційна функція. Лаги залежних і незалежних змінних. Модель адаптивних сподівань. Модель часткового коригування. Методи оцінювання.

Тема 9. *Економетричні моделі на основі системи структурних рівнянь*

Системи одночасових структурних рівнянь.

МЕТОДИЧНІ ВКАЗІВКИ ДО КОНТРОЛЬНОЇ РОБОТИ

1. Економетрична модель з двома змінними

Серед багаточисленних зв'язків між економічними показниками завжди можна виділити такий показник, вплив якого на результативну ознаку є основним, найбільш важливим. Щоб виміряти цей зв'язок кількісно, необхідно побудувати економетричну модель з двома змінними (просту модель). Загальний вигляд такої моделі:

$$Y = f(X, u),$$

де Y — залежна змінна (результативна ознака); X — незалежна змінна (фактор); u — стохастична складова.

Аналітична форма цієї моделі може бути різною залежно від економічної сутності зв'язків. Найбільш поширені форми залежностей:

$$Y = a_0 + a_1 X; \quad Y = a_0 e^{a_1 X}; \quad Y = a_0 X^{a_1};$$

$$Y = a_0 + \frac{a_1}{X},$$

де a_0, a_1 — невідомі параметри моделі.

Неважко переконатись, що наведені нелінійні форми залежностей за допомогою елементарних перетворень приводяться до лінійних. Якщо припустити, що економетрична модель з двома змінними є лінійною:

$$Y = a_0 + a_1 X + u,$$

в якій стохастична складова (залишки) має нульове математичне сподівання та постійну дисперсію, то параметри моделі можна оцінити на основі звичайного методу найменших квадратів (1МНК).

В основі методу 1МНК лежить принцип мінімізації суми квадратів залишків моделі. Реалізація цього принципу дає можливість отримати систему нормальних рівнянь:

$$\begin{cases} na_0 + \sum_i x_i a_1 = \sum_i y_i \\ \sum_i x_i a_0 + \sum_i x_i^2 a_1 = \sum_i x_i y_i \end{cases}$$

В даній системі n — кількість спостережень, $\sum_i x_i$, $\sum_i y_i$, $\sum_i x_i^2$, $\sum_i x_i y_i$ — величини, які можна розрахувати на основі вихідних спостережень над змінними Y і X.

Розв'язавши систему нормальних рівнянь, одержимо оцінки невідомих параметрів моделі $\hat{a}_0$ і $\hat{a}_1$:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X.$$

Достовірність побудованої економетричної моделі можна перевірити, користуючись елементами дисперсійного аналізу. Перш за все слід розрахувати залишки моделі

$$u_i = y_i - \hat{y}_i$$

та знайти їх дисперсію:

$$\sigma_u^2 = \frac{\sum_i u_i^2}{n-m},$$

де m — кількість змінних моделі ($m = 2$).

$$S_{\hat{a}_j} = \sigma_u \sqrt{c_{jj}}$$

необхідно визначити стандартну помилку кожного параметра моделі.

c_{jj} в цій формулі характеризує відповідний діагональний елемент матриці помилок (матриці, оберненої до матриці системи нормальних рівнянь).

На основі коефіцієнта детермінації

$$R^2 = \frac{\sigma_Y^2 - \sigma_U^2}{\sigma_Y^2}$$

можна зробити висновок про ступінь значущості вимірюваного зв'язку на основі економетричної моделі

$$R^2 \in \left]0,1\right[.$$

Оскільки коефіцієнт детермінації R^2 характеризує, якою мірою варіація залежної змінної визначається варіацією незалежної змінної, то чим ближче R^2 до одиниці, тим суттєвішим є зв'язок між цими змінними.

Коефіцієнт кореляції R $=\sqrt{R^2}$ характеризує тісноту зв'язку між змінними моделі. Він може знаходитись на множині $R \in \left]-1,1\right[$. Чим ближче R до одиниці по модулю, тим тіснішим є зв'язок. Від'ємний знак свідчить про обернений зв'язок, додатній — про прямий.

Якщо прийняти відповідну гіпотезу про закон розподілу залишків економетричної моделі, то параметри її можна оцінити на основі метода максимальної правдоподібності.

Нехай залишки моделі розподіляються за нормальним законом, тоді функція правдоподібності запишеться так:

$$F=\frac{1}{(\tilde{\sigma}_u^2 2\pi)^{n/2}} \exp\left[-\frac{1}{2\tilde{\sigma}_u^2}\sum_{i=1}^{n}(y_i-\tilde{a}_0-\tilde{a}_1 x_i)^2\right]$$

і

$$\ln F=-\frac{n}{2}\ln 2\pi-\frac{n}{2}\ln\tilde{\sigma}_u^2-\frac{1}{2\tilde{\sigma}_u^2}\sum_{i=1}^{n}(y_i-\tilde{a}_0-\tilde{a}_1 x_i)^2$$

Продиферецiюємо цю функцію за невідомими параметрами $\tilde{a}_0$, $\tilde{a}_1$ і $\tilde{\sigma}_u^2$ і, прирівнявши похідні до нуля, отримаємо систему рівнянь:

$$\begin{cases} n\tilde{a}_0+\tilde{a}_1\sum_i x_i = \sum_i y_i \\ \sum_i x_i\tilde{a}_0 + \tilde{a}_1\sum_i x_i^2 = \sum_i x_i y_i \\ \frac{1}{n}\sum_i (y_i - \tilde{a}_0 - \tilde{a}_1 x_i)^2 = \sigma_U^2 \end{cases}$$

Підставимо в цю систему величини $\sum_i x_i$, $\sum_i y_i$, $\sum_i x_i^2$, $\sum_i x_i y_i$, які розраховуються на основі вихідних даних, і розв'яжемо її відносно параметрів $\tilde{a}_0$, $\tilde{a}_1$ і $\tilde{\sigma}_u^2$. В результаті отримаємо оцінки параметрів моделі $\tilde{a}_0$ і $\tilde{a}_1$, а також оцінку дисперсії залишків.

2. Побудова загальної лінійної моделі

Для того щоб кількісно описати зв'язок між кількома або багатьма змінними, одна з яких є залежною, інші — незалежними змінними, необхідно розглянути лінійну економетричну модель, яка базується на регресійному аналізі.

У загальному вигляді цю модель можна записати так:

$$Y = f(X_1, X_2, ..., X_m, u),$$

де Y — залежна змінна; $X_j,\ (j=\overline{1,m})$ — незалежні змінні;

u — стохастична складова.

Залежна змінна Y називається також пояснюваною, ендогенною змінною, незалежні змінні X_j — пояснюючими, предетермінованими, екзогенними змінними.

Аналітична форма загальної лінійної економетричної моделі:

$$Y = a_0 + a_1X_1 + a_2X_2 + \dots + a_mX_m + u,$$

де $a_j\quad (j=\overline{0,m})$ — параметри моделі.

В матричній формі економетрична модель має такий вигляд:

$$Y = XA + u,$$

X — матриця незалежних змінних; A — вектор оцінок параметрів моделі; u — вектор залишків.

Щоб оцінити параметри моделі на основі методу 1МНК, необхідно дотримуватися таких передумов (гіпотез):

1) математичне сподівання залишків має дорівнювати нулю, тобто

$$M(u) = 0;$$

2) значення вектора залишків u незалежні між собою і мають постійну дисперсію:

$$M(uu') = \sigma^2 E;$$

незалежні змінні моделі не зв'язані із залишками, тобто

$$M(X'u) = 0;$$

4) незалежні змінні моделі створюють лінійно-незалежну систему векторів, тобто

$(X'_k X_j) = 0, \quad k \neq j; \quad k = \overline{1,m}$

$(X'_k X_j) = 1, \quad k = j; \quad j = \overline{1,m}$

Оператор оцінювання параметрів моделі на основі 1МНК:

$$A = (X' X)^{-1} X' Y.$$

Неважко довести, що оцінки $\hat{A}$, які можна отримати на основі оператора оцінювання 1МНК, мінімізують суму квадратів залишків u. При цьому значення вектора $\hat{A}$ є розв'язком нормальної системи рівнянь:

$$(X' X) \hat{A} = X' Y$$

Якщо незалежні змінні в матриці X взяті як відхилення кожного значення від своєї середньої, то матрицю $X' X$ називають матрицею моментів. Числа, що стоять на її головній діагоналі, характеризують величину дисперсій незалежних змінних, інші елементи відповідають взаємним коваріаціям.

Оцінки параметрів загальної економетричної моделі повинні мати такі властивості:

1) незміщеності;
2) обгрунтованості;
3) ефективності;
4) інваріантності.

Оцінка параметра моделі буде незміщеною, коли дотримується рівність:

$$M(\hat{A}) = A.$$

Якщо ця рівність не дотримується, то різниця $M(\hat{A}) - A = Q$ називається зміщенням оцінки.

Оцінка параметра моделі буде обгрунтованою, якщо при заданій малій величині $\varepsilon > 0$ справедливе відношення:

$$\lim_{n \to \infty} P\left\{ \left|\hat{A} - A\right| < \varepsilon \right\} = 1.$$

Оцінки $\hat{A}$ параметрів A називаються ефективними, коли вони мають найменшу дисперсію.

Якщо функція $g(\hat{A})$ відповідає функції $g(A)$, то оцінки $\hat{A}$ параметрів A є інваріантними.

3. Дисперсійний аналіз економетричної моделі

Між оцінками параметрів економетричної моделі та коефіцієнтом кореляції, що характеризує тісноту зв'язку, існує зв'язок. Для простої економетричної моделі його можна записати так:

$$\hat{a} = r \frac{\sigma_Y}{\sigma_X},$$

де — коефіцієнт парної кореляції; σ_Y, σ_X — середньоквадратичне відхилення відповідно залежної і незалежної змінної.

Це співвідношення було покладено в основу алгоритму визначення альтернативної оцінки параметрів моделі за методом 1МНК. Алгоритм має назву ***покрокової регресії*** і наступні кроки:

Крок 1. Стандартизація (нормалізація) всіх змінних моделі:

$$Y_i^* = \frac{Y_i - \overline{Y}}{\sigma_Y};$$

$$X_{ij}^* = \frac{X_{ij} - \overline{X}_j}{\sigma_{X_j}}.$$

Крок 2. Визначення кореляційної матриці , елементи якої розраховуються таким чином:

$$r_{YX_j} = \frac{1}{n} Y^* \; X_j^* \;;$$

$$r_{X_K X_j} = \frac{1}{n} X_k^* \; X_j^* \;; \quad k = \overline{1,m},\; j = \overline{1,m}.$$

Крок 3. Із усіх елементів матриці r_{YXj} вибирається той, якому відповідає $\max\left\{\left|r_{YX_j}\right|\right\}$. Це означає, що незалежна змінна X_j найтісніше зв'язана з залежною змінною Y. Будується економетрична модель:

$$\hat{Y}^* = \hat{\beta}_j X_j^*.$$

Крок 4. Серед інших елементів матриці YXj знову вибирається $\max\left\{\left|r_{YX_j}\right|\right\}$. Якщо даному коефіцієнту кореляції відповідає X^*_{j+1}, то ця змінна вводиться в побудовану раніше економетричну модель; в результаті дістанемо:

$$Y^* = \hat{\beta}_j X^*_j + \hat{\beta}_{j+1} X^*_{j+1}$$

і т.д.

Процес продовжується до тих пір, поки всі незалежні змінні поступово будуть включені в модель. Якщо є обмеження, яке вказує на недоцільність розширення економетричної моделі за рахунок змінних, що залишилися, то процес розрахунку закінчується раніше. Таким обмеженням може бути співвідношення між коефіцієнтом кореляції чи детермінації, виправленими й невиправленими на число ступеней свободи.

Система нормальних рівнянь у даному алгоритмі:

$$\begin{cases} r_{YX_1} = \hat{\beta}_1 + \hat{\beta}_2 r_{X_1X_2} + \hat{\beta}_3 r_{X_1X_3} + \ldots + \hat{\beta}_m r_{X_1X_m} \\ r_{YX_2} = \hat{\beta}_1 r_{X_1X_2} + \hat{\beta}_2 + \hat{\beta}_3 r_{X_2X_3} + \ldots + \hat{\beta}_m r_{X_2X_m} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ r_{YX_m} = \hat{\beta}_1 r_{X_1X_m} + \hat{\beta}_2 r_{X_2X_m} + \hat{\beta}_3 r_{X_3X_m} + \ldots + \hat{\beta}_m \end{cases}$$

Позначимо елементи $_{YXj}$ через вектор $_{YX}$, а інші елементи $_{XkXj}$ — через матрицю , тоді система рівнянь у матричному вигляді матиме такий вигляд:

$$r\hat{\beta} = r_{YX}.$$

Звідси $\hat{\beta} = r^{-1} r_{YX}$, тобто отримаємо альтернативний оператор оцінювання параметрів моделі за методом 1МНК.

Оскільки оцінки параметрів моделі $\hat{\beta}$ відносяться до стандарти-зованих змінних, то щоб перейти до оцінок параметрів моделі, в якій змінні мають свій початковий вимір, необхідно:

$$\hat{a}_j = \hat{\beta}_j \frac{\sigma_Y}{\sigma_{X_j}}, \quad j = \overline{1, m};$$

$$\hat{a}_0 = \bar{y} - \sum_j \hat{a}_j \bar{x}_j.$$

Множинний коефіцієнт детермінації, який визначає рівень варіації залежної змінної за рахунок незалежних, розраховується таким чином:

$$R^2 = \frac{\sigma_Y^2 - \sigma_U^2}{\sigma_Y^2}$$, чи

$$R^2 = \frac{\hat{A}'X'Y}{Y'Y} \cdot \frac{n-1}{m-1}.$$

Коефіцієнт детермінації без урахування числа ступеней свободи:

$$\overline{R}^2 = \frac{\hat{A}'X'Y}{Y'Y}.$$

Співвідношення між ними дорівнюватиме:

$$R^2 = 1 - \left(\frac{n-1}{n-m}\right)\left(1 - \overline{R}^2\right).$$

Множинний коефіцієнт кореляції $R = \sqrt{R^2}$ характеризує тісноту зв'язку між залежною і незалежними змінними. Множинний коефіцієнт детермінації і кореляції знаходяться на множині

$R^2 \in]0,1[$;

$R \in]0,1[$.

Якщо оцінка параметрів моделі отримана на основі покрокової регресії, то для визначення коефіцієнта детермінації можна використати такі співвідношення:

$$R^2 = 1 - \frac{|r|}{R_{11}};$$

$$R^2 = \hat{\beta}_1 r_{YX_1} + \hat{\beta}_2 r_{YX_2} + \hat{\beta}_3 r_{YX_3} + \ldots + \hat{\beta}_m r_{YX_m},$$

де $|r|$ – визначник матриці ;

R_{11} — алгебраїчне доповнення першого елемента матриці .

Гіпотеза про наявність чи відсутність зв'язку між залежною і незалежною змінними може бути перевірена на основі F-критерію:

$$F = \frac{\sigma_Y^2 - \sigma_U^2}{\sigma_U^2} = \frac{\sigma_P^2}{\sigma_U^2}.$$

Фактичне значення F — критерію порівнюється з табличним при сту-пенях свободи $n-m$ і $m-1$ і вибраному рівні довіри. Якщо

Fфакт > Fтабл , то гіпотеза про суттєвість зв'язку між залежною і незалежними змінними економетричної моделі підтверджується, в протилежному випадку — відкидається.

Альтернативна формула розрахунку F — критерію через коефіцієнт детермінації:

$$F = \frac{R^2/(m-1)}{(1-R^2)\ /(n-m)}.$$

Значущість оцінок параметрів моделі можна визначити на основі t — критерію:

$$t_j = \frac{a_j}{\sqrt{\sigma_U^2 c_{jj}}}.$$

Значення критерію t_j порівнюється з табличним при вибраному рівні значущості і $n-m$ ступенями свободи. Якщо t факт > t табл , то відповідний параметр економетричної моделі є достовірним.

На основі t — критерію і стандартної помилки будуються довірчі інтервали для параметрів $\hat{a}_j$:

$$a_j = \hat{a}_j \pm t_{(\alpha)} \sqrt{\sigma_U^2 c_{jj}}.$$

Прогноз залежної змінної на основі економетричної моделі при заданих залежних змінних можна виконати на основі такого співвідношення:

$$\hat{Y}_{\text{пр}} - t_{(\alpha)}\ S(\hat{y})\ \le Y \le \hat{Y}_{\text{пр}} + t_{(\alpha)}\ S(\hat{y}).$$

У цьому співвідношенні $S(\hat{y})$ є стандартною помилкою прогнозу:

$$S(\hat{y}) = \sqrt{\sigma_u^2\ X_p'\ (X'X)^{-1}\ X_p},$$

де X_p — прогнозні значення незалежних змінних.

4. Мультиколінеарність

На практиці при кількісній оцінці параметрів економетричної моделі досить часто зустрічаються з проблемою взаємозв'язку між предетермінованими (пояснюючими змінними). Якщо взаємозв'язок досить тісний, то оцінка параметрів моделі може мати велику похибку. Такий взаємозв'язок між пояснюючими змінними називається мультиколінеарністю. Мультиколінеарність незалежних змінних приводить до зміщення оцінок параметрів моделі, які розраховуються за методом 1МНК. На основі цих оцінок неможливо зробити конкретні висновки про результати взаємозв'язку між пояснюваною і пояснюючими змінними.

Ознаки мультиколінеарності

1. Якщо серед парних коефіцієнтів кореляції незалежних змінних є такі, рівень яких наближається або дорівнює множинному коефіцієнту кореляції, то це свідчить про можливість існування мультиколінеарності. Інформацію про парну залежність може дати симетрична матриця коефіцієнтів парної кореляції, або кореляції нульового порядку:

$$r=\begin{pmatrix} r_{X_1X_1} & r_{X_1X_2} & r_{X_1X_3} & \cdots & r_{X_1X_K} \\ r_{X_2X_1} & r_{X_2X_2} & r_{X_2X_3} & \cdots & r_{X_2X_K} \\ r_{X_3X_1} & r_{X_3X_2} & r_{X_3X_3} & \cdots & r_{X_3X_K} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{X_KX_1} & r_{X_KX_2} & r_{X_KX_3} & \cdots & r_{X_KX_K} \end{pmatrix}.$$

Але якщо в моделі фігурує більше двох незалежних змінних, вивчення питання про мультиколінеарність не може обмежуватись інформацією, що дає ця матриця. Явище мультиколінеарності ні в якому разі не зводиться тільки до існування парної кореляції між незалежними змінними.

Більш загальна перевірка передбачає визначення визначника (детермінанта) матриці r, який називається детермінантом кореляції і позначається $|r|$. Числові значення детермінанта кореляції знаходяться на множині: $|r| \in [0,1]$.

2. Якщо $|r| = 0$, то існує повна мультиколінеарність, якщо $|r| = 1$ — мультиколінеарність відсутня, чим ближче $|r|$ до нуля, тим певніше можна стверджувати, що між незалежними змінними існує мультиколінеарність. Незважаючи на те, що на числове значення $|r|$

впливає дисперсія незалежних змінних, цей показник можна вважати точковою мірою тісноти мультиколінеарності.

3. Якщо в економетричній моделі одержано мале значення параметра $\hat{a}_k$ при високому рівні коефіцієнта детермінації і при цьому F-критерій суттєво відрізняється від нуля, то це також свідчить про наявність мультиколінеарності.

4. Якщо коефіцієнт детермінації R^2, що розрахований для регресійних залежностей між однією незалежною змінною та іншими, має значення, яке близьке до одиниці, то можна говорити про наявність мультиколінеарності.

5. Якщо при побудові економетричної моделі на основі покрокової регресії включення нової незалежної змінної суттєво змінює оцінку параметрів моделі при незначному підвищенні (або зниженні) коефіцієнтів кореляції чи детермінації, то ця змінна, очевидно, знаходиться в лінійній залежності від інших, які введені в модель раніше.

Всі ці методи виявлення мультиколінеарності мають один загальний недолік: жоден із них не проводить чіткої межі між тим, що треба вважати «суттєвою» мультиколінеарністю, яку треба враховувати, і тим, коли мультиколінеарністю можна знехтувати.

Алгоритм Феррара—Глобера

Найбільш повне дослідження мультиколінеарності можна здійснити на основі алгоритму Феррара—Глобера. Цей алгоритм включає три види статистичних критеріїв, на основі яких перевіряється мультиколінеарність всього масиву незалежних змінних (χ^2, хі-квадрат); кожної незалежної змінної зі всіма незалежними змінними (F-критерій) і мультиколінеарність кожної пари незалежних змінних (t-критерій).

Всі ці критерії при порівнянні з їх критичними значеннями дають можливість зробити конкретні висновки відносно наявності чи відсутності мультиколінеарності незалежних змінних.

Опишемо алгоритм Феррара—Глобера.

Крок 1. Стандартизація (нормалізація) змінних.

Позначимо вектори незалежних змінних економетричної моделі через $X_1, X_2, X_3, \ldots, X_m$. Елементи стандартизованих векторів розрахуємо за формулою:

$$X_{ik}^* = \frac{X_{ik} - \overline{X}_k}{\sqrt{n\sigma^2_{X_k}}},$$

де n — число спостережень, $(i=\overline{1,n})$;

m — число незалежних змінних, $(k=\overline{1,m})$;

$\overline{X}_k$ — середня арифметична k— її незалежної змінної;

$\sigma^2_{X_k}$ — дисперсія k-ї незалежної змінної.

Крок 2. Знаходження кореляційної матриці (матриці моментів стандартизованої системи нормальних рівнянь):

$$R = X^{*'}\ X^*,$$

де X^* — матриця стандартизованих незалежних змінних;

$X^{*'}$ — матриця, транспонована до матриці X^*.

Крок 3. Визначення критерію χ^2 (хі-квадрат):

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

де $|R|$ — визначник кореляційної матриці R.

Значення цього критерію порівнюється з табличним при $\frac{1}{2}m\,(m-1)$ ступенях свободи і рівні значущості α. Якщо χ^2 факт $< \chi^2$ табл , в масиві незалежних змінних не існує мультиколінеарності.

Крок 4. Визначення оберненої матриці C (див.п.3):

$$C = R^{-1} = (X^{*'}\ X^*)^{-1}.$$

Крок 5. Розрахунок F- критеріїв:

$$F_k = (c_{kk} - 1)\ \frac{n-m}{m-1},$$

де c_{kk} — діагональні елементи матриці C. Фактичні значення критеріїв F_k порівнюються з табличними при $n-m$ і $m-1$ ступенях свободи і рівні значущості α. Якщо F_k факт > Fтабл , відповідна k-та незалежна змінна мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної розраховується таким чином:

$$R^2_{X_k} = 1 - \frac{1}{c_{kk}}.$$

Крок 6. Знаходження часткових коефіцієнтів кореляції:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} * c_{jj}}},$$

де c_{kj} — елемент матриці C, що заходиться в k-му рядку і j-му стовпці, $k = \overline{1,m}$, $j = \overline{1,m}$, c_{kk} і c_{jj} — діагональні елементи матриці C.

Крок 7. Розрахунок t критеріїв:

$$t_{kj} = \frac{r_{kj}\sqrt{n-m}}{\sqrt{1-r_{kj}^2}}.$$

Фактичні значення критеріїв t_{kj} порівнюються з табличними при $n-m$ ступенях свободи і рівні значущості α. Якщо $t_{kj\ факт} > t_{табл}$, між незалежними змінними X_k і X_j існує мультиколінеарність.

5. Гетероскедастичність

Передумови, які висуваються при оцінці параметрів моделі за методом 1МНК на практиці часто можуть порушуватись. Однією з таких передумов є незмінність дисперсії залишків для всіх спостережень вихідної сукупності. Це явище називається гомоскедастичністю. В практичних дослідженнях воно часто порушується. Наприклад, в економетричній моделі, що характеризує залежність витрат на споживання від доходу, дисперсія залишків може змінюватись для спостережень, які відносяться до різних груп населення за розміром доходів.

Якщо дисперсія залишків в економетричному моделюванні змінюється для кожного спостереження або для груп спостережень, то це явище називається гетероскедастичністю.

Наявність гетероскедастичності спричиняє порушення властивостей оцінок параметрів моделі при розрахунку їх за методом 1МНК. Тому завжди виникає необхідність вивчати це явище, і, якщо воно існує, для оцінки параметрів моделі використовувати узагальнений метод найменших квадратів (метод Ейткена).

Для визначення гетероскедастичності застосовуються чотири критерії:

1) $F_{критерій} \propto$;

2) параметричний тест Гольдфельда—Квандта;

3) непараметричний тест Гольдфельда—Квандта;

4) тест Глейсера.

1. Критерій $\propto$

Цей метод застосовується в тих випадках, коли вихідна сукупність спостережень досить велика. Розглянемо цей алгоритм.

Крок 1. Вихідні дані залежної змінної Y розбиваються на k груп $(r=\overline{1,k})$ згідно із зміною рівня величини Y.

Крок 2. По кожній групі даних розраховується сума квадратів відхилень:

$$S_r = \sum_{i=1}^{n_r} (y_{ir} - \overline{y_r})^2$$
.

Крок 3. Розраховується сума квадратів відхилень у цілому по всій сукупності спостережень:

$$\sum_{r=1}^{k} S_r = \sum_{i=1}^{n_r} \sum_{r=1}^{k} (y_{ir} - \overline{y}_r)^2$$
.

Крок 4. Обчислюється параметр λ:

$$\lambda = \prod_{r=1}^{k} \left(\frac{S_r}{n_r} \right)^{n_r/2} \Bigg/ \left(\frac{\sum_{r=1}^{k} S_r}{n} \right)^{n/2}$$
,

де n — загальна сукупність спостережень;

n_r — кількість спостережень r-ї групи.

Крок 5. Розраховується критерій $\propto$:

$$\propto = -2 \ln \lambda$$
,

який наближено буде відповідати розподілу X^2 при ступенях свободи k – 1, коли дисперсія всіх спостережень однорідна. Тобто, якщо значення $\propto$ менше табличного значення X^2

при вибраному рівні довіри і ступені свободи k – 1, то явище гетероскедастичності відсутнє.

2. Параметричний тест Гольдфельда—Квандта

Коли сукупність спостережень невелика, то розглянутий метод 1 застосовувати неможливо.

Тоді Гольдфельд і Квандт розглянули випадок, коли $M(uu') = \sigma^2 x_{ij}^2$, тобто дисперсія залишків зростає пропорційно квадрату однієї із незалежних змінних моделі:

$$Y = XA + u.$$

Вони запропонували для виявлення наявності гетероскедастичності параметричний тест, в якому треба виконати наступні кроки.

Крок 1. Упорядкувати спостереження згідно з величиною елементів вектора x_j.

Крок 2. Відкинути c спостережень, які будуть знаходитись у центрі вектора. На основі експериментальних розрахунків автори вирахували оптимальні співвідношення між параметрами c і n, де n — кількість елементів вектора x_j.

$$\frac{c}{n} = \frac{4}{15}.$$

Крок 3. Побудувати дві економетричні моделі на основі 1МНК по двох створених сукупностях спостережень $(n_1 - c)/2$ за умови, що $(n_2 - c)/2$ перевищує кількість змінних m.

Крок 4. Знайти суму квадратів залишків за першою (1) і другою (2) моделях S_1 і S_2.

$S_1 = \hat{u}_1' \hat{u}_1$, де $\hat{u}_1$ — залишки по моделі (1) ;

$S_2 = \hat{u}_2' \hat{u}_2$, де $\hat{u}_2$ — залишки по моделі (2).

Крок 5. Розрахувати критерій R:

$$R = \frac{S_2}{S_1},$$

який при виконанні гіпотези про гомоскедастичність буде відповідати F-розподілу з $(n_1 - c - 2m)/2$, $(n_2 - c - 2m)/2$ ступенями свободи. Це означає, що розраховане значення R*порівнюється з табличним значенням F-критерію при ступенях свободи $(n_1 - c - 2m)/2$ і $(n_2 - c - 2m)/2$ і вибраному рівні довіри. Якщо $R \leq F$ табл , то гетероскедастичність відсутня.

3. Непараметричний тест Гольдфельда—Квандта

Гольфельд і Квандт запропонували також для оцінки наявності гетероскедастичності непараметричний тест. Цей тест базується на числі піків у величині залишків після упорядкування спостережень по x_{ij}.

4. Тест Глейсера

Ще один тест для перевірки гетероскедастичності запропонував Глейсер: розглядати регресію абсолютних значень залишків $|u_i|$, які відповідають регресії найменших квадратів як деяку функцію від x_j, де x_j є тією незалежною змінною, яка відповідає зміні дисперсії σ^2_u. Для цього використовуються такі види функцій:

1) $|u| = a_0 + a_1 x_j$;

2) $|u| = a_0 + a_1 x_j^{-1}$;

3) $|u| = a_0 + a_1 x_j^{1/2}$ і т.п.

Рішення про відсутність гетероскедастичності залишків приймається на основі статистичної значущості коефіцієнтів a_0 й a_1. Переваги цього тесту визначаються можливістю розрізняти випадок чистої і змішаної гетероскедастичності. Чистій гетероскедастичності відповідають значення параметрів $a_0 = 0$,

$a_1 \neq 0$; а змішаній — $a_0 \neq 0$, $a_1 \neq 0$. Залежно від цього треба користуватись різними матрицями S. Нагадаємо, що:

$$M(uu') = \sigma_u^2 S$$.

Якщо при економетричному моделюванні для певних вихідних даних буде виявлено явище гетероскедастичності, то оцінку параметрів моделі треба виконувати на основі узагальненого методу найменших квадратів. Оператор оцінювання цим методом запишеться:

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

де

$$S^{-1} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$.

В даній матриці залежно від висунутої гіпотези:

або $\lambda_i = \dfrac{1}{x_{ij}}$;

або $\lambda_i = \dfrac{1}{x_{ij}^2}$;

або $\lambda_i = \left\{\left|\hat{u}_i\right|\right\}$.

Прогноз на основі економетричної моделі, в якій оцінка параметрів виконана узагальненим методом найменших квадратів, можна отримати на основі такого співвідношення:

$$\hat{Y}_{пр} = X_0 A + W' V^{-1} U$$,

де u — вектор залишків, який відповідає оцінці параметрів моделі на основі 1МНК;

W' — транспонований вектор коваріацій поточних і прогнозних значень залишків;

$V^{-1} = S^{-1}$, а $V = \sigma_U^2 S$.

6. Економетрична модель з автокорельованими залишками

В економетричних дослідженнях часто зустрічаються такі випадки, коли дисперсія залишків є постійною, але спостерігається їх коваріація. Це явище має назву автокореляції залишків.

Автокореляція залишків виникає частіше за все тоді, коли економетрична модель будується на основі часових рядів. Якщо існує кореляція між послідовними значеннями деякої незалежної змінної, то буде спостерігатись і кореляція послідовних значень залишків. Тобто в цьому випадку також порушується гіпотеза, згідно з якою $M(uu') = \sigma_U^2 E$, але при гетероскедастичності змінюється дисперсія залишків при відсутності їх коваріації, а при автокореляції — існує коваріація залишків при незмінній дисперсії.

При автокореляції залишків, як і при гетероскедастичності дисперсія залишків запишеться:

$$M(uu') = \sigma_U^2 S,$$

але матриця S матиме тут зовсім інший вигляд. Запишемо цю матрицю:

$$S = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{n-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 \end{pmatrix}.$$

В даній матриці параметр ρ характеризує коваріацію кожного наступного значення залишків із попереднім. Так, якщо для залишків записати авторегресійну модель першого порядку:

$$u_t = \rho \ u_{t-1} + \varepsilon_t,$$

то ρ характеризує силу зв'язку величини залишків у період t від величини залишків у період t – 1.

Якщо проігнорувати матрицю S при визначенні дисперсії залишків, і для оцінки параметрів моделі застосувати метод 1МНК, то можливі такі наслідки:

1) оцінки параметрів моделі можуть бути незміщеними, але неефективними, тобто вибіркові дисперсії вектора оцінок $\hat{A}$ можуть бути невиправдано великими;

2) статистичні критерії t і F- статистики, які отримані для класичної лінійної моделі, практично не можуть бути використані для дисперсійного аналізу, бо їх розрахунок не враховує наявності коваріації залишків;

3) неефективність оцінок параметрів економетричної моделі, як правило, призводить до неефективних прогнозів, тобто прогнозні значення матимуть велику вибіркову дисперсію.

Перевірка наявності автокореляції

Для перевірки наявності автокореляції залишків можна застосувати чотири методи:

1. Критерій Дарбіна—Уотсона.
2. Критерій фон Неймана.
3. Нециклічний коефіцієнт автокореляції.
4. Циклічний коефіцієнт автокореляції.

1. Критерій Дарбіна—Уотсона:

$$DW = \frac{\sum_{t=2}^{n}(u_t - u_{t-1})^2}{\sum_{t=1}^{n} u_t^2}.$$

Критерій Дарбіна—Уотсона може приймати значення на множині $DW \in [0,4]$. Якщо залишки ut є випадковими величинами, тобто не автокорельовані, то значення DW знаходиться поблизу 2. При додатній автокореляції $DW < 2$, при від'ємній $DW > 2$.

Значення критерія DW табульовані на інтервалі $DW_1 - DW_2$, де DW_1— нижня межа, DW_2— верхня межа. Фактичні значення критерію порівнюються з табличними (критичними) для числа спостережень n і числа незалежних змінних m при вибраному рівні довіри α. Якщо DW факт $< DW_1$, залишки мають автокореляцію. Якщо DW факт $> DW_2$, приймається гіпотеза про відсутність автокореляції. Якщо $DW_1 < DW < DW_2$ конкретних висновків зробити не можна.

2. Критерій фон Неймана:

$$Q=\frac{\sum_{t=2}^{n}(u_t-u_{t-1})^2}{\sum_{t=1}^{n}u_t^2}\cdot\frac{n}{n-1}\,.$$

Звідси $Q = DW\cdot\dfrac{n}{n-1}$, при $n\to\infty$, $Q = DW$. Фактичне значення критерію фон Неймана порівнюється з табличним при вибраному рівні довіри α і заданому числі спостережень. Якщо $Q_{факт} < Q_{табл}$, то існує додатня автокореляція.

3. Нециклічний коефіцієнт автокореляції:

$$r^* = \frac{\sum_{t=2}^{n}(u_t\ u_{t-1}) - \frac{1}{n-1}\left(\sum_{t=2}^{n}u_t\right)\left(\sum_{t=2}^{n}u_{t-1}\right)}{\sqrt{\left[\sum_{t=2}^{n}u_t^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n}u_t\right)^2\right]\left[\sum_{t=2}^{n}u_{t-1}^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n}u_{t-1}\right)^2\right]}}\,.$$

r^*може приймати значення в інтервалі $[-1;+1]$. Від'ємні значення свідчать про від'ємну автокореляцію, додатні — про додатню. Значення, що знаходяться в деякій критичній області біля нуля, свідчать про відсутність автокореляції.

4. Циклічний коефіцієнт автокореляції:

$$r^0 = \frac{\sum_{t=2}^{n}u_t\ u_{t-1} + u_n\ u_1 - \frac{1}{n}\left(\sum_{t=1}^{n}u_t\right)^2}{\sum_{t=1}^{n}u_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n}u_t\right)^2}\,.$$

Фактичне значення цього критерію порівнюється з табличним для вибраного рівня довіри і довжини ряду спостережень n. Якщо $r^0_{факт} \geq r^0_{табл}$, то існує автокореляція. Припускаючи, що

$$\sum_{t=2}^{n}u_t \approx \sum u_{t-1} \approx 0\,,$$

циклічний коефіцієнт автокореляції можна записати так:

$$r = \frac{\sum_{t=2}^{n} u_t \; u_{t-1}}{\sum_{t=1}^{n} u_t^2}.$$

Оцінку параметрів моделі з автокорельованими залишками можна виконувати на основі чотирьох методів:

1) Ейткена;

2) перетворення вихідної інформації;

3) Кочрена—Оркатта;

4) Дарбіна.

Перші два методи доцільно застосовувати тоді, коли залишки описуються авторегресійною моделлю першого ступеня:

$$u_t = \rho \; u_{t-1} + \varepsilon_t.$$

Ітеративні методи Кочрена—Оркатта і Дарбіна можна застосовувати для оцінки параметрів економетричної моделі і тоді, коли залишки описуються авторегресійною моделлю більш високого ступеня:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t;$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t.$$

1. Метод Ейткена

Оператор оцінювання цим методом запишеться так:

$$\hat{A} = (X' \; S^{-1} X) \;\; X' S^{-1} Y$$

або

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

де S^{-1} — матриця, обернена до матриці S (див. стор. 77);

V^{-1} — матриця, обернена до матриці $V = \sigma_U^2 \, S$.

Оскільки в матриці S коваріація залишків ρ^s при $s > 2$ наближається до нуля, то матриця, обернена до матриці S, буде мати наступного вигляду:

$$S^{-1}=\frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

На практиці для розрахунку ρ використовується співвідношення:

$$\rho \approx r = \frac{\sum_{t=2}^{n} u_t \; u_{t-1}}{\sum_{t=1}^{n} u_t^2}$$

або

$$\rho \approx r' = \frac{\sum_{t=2}^{n} u_t \; u_{t-1}}{\sum_{t=1}^{n} u_t^2} \cdot \frac{n}{n-1}$$

2. Метод перетворення вихідної інформації

Цей метод складається з двох кроків:

1) перетворення вихідної інформації при застосуванні для цього параметра ρ;

2) застосування методу 1МНК для оцінки параметрів моделі на основі перетворених даних.

Перетворення вихідної інформації виконується за допомогою матриць T_1 або T_2:

$$T_1=\begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix};$$

$$T_1=\begin{pmatrix} -\rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ 0 & 0 & -\rho & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

тобто замість матриці X використовуємо T_1X або T_2X, замість вектора Y — T_1Y або T_2X.

$\sqrt{1-\rho^2}\,Y_1$;

$Y_2 - \rho\ Y_1$;

$Y_3 - \rho\ Y_2$;

$Y_4 - \rho\ Y_3$;

....

$Y_n - \rho\ Y_{n-1}$.

3. Метод Кочрена—Оркатта

Цей метод є ітеративним методом наближеного пошуку параметрів a_j і ρ $(j=\overline{0,m})$ і, які мінімізують суму квадратів залишків.

Коли економетрична модель має вигляд:

$$Y_t = a_0 + a_1 X_t + u_t\ ;$$

$$u_t = \rho\ u_{t-1} + \varepsilon_t,\quad t = \overline{1,n},\ |\rho| < 1,$$

сума квадратів залишків запишеться так:

$$\sum_{t=2}^{n}\varepsilon_t = \sum\left[\left(Y_t - \rho Y_{t-1}\right) - a_0(1-\rho) - a_1\left(X_t - \rho X_{t-1}\right)\right]^2.$$

Алгоритм

Крок 1. Довільно вибирається значення $\rho < 1\ (\rho = r_1)$ і підставляється в формулу суми квадратів залишків.

Крок 2. На основі методу 1МНК знаходяться параметри $a_0^{(1)}$ і $a_1^{(1)}$.

Крок 3. Прийнявши $a_0 = a_0^{(1)}$ і $a_1 = a_1^{(1)}$, підставимо у формулу суми квадратів залишків і розрахуємо $\rho = {}_2$.

Крок 4. Підставивши $\rho = {}_2$, розрахуємо параметри $a_0^{(2)}$ і $a_0^{(2)}$ і т.д.

Процедура продовжується доти, доки послідовні значення параметрів не будуть відрізнятися менше ніж на задану величину.

4. Метод Дарбіна

Цей метод базується на простій двокроковій процедурі.

Крок 1. Підставимо значення залишків, яке підкоряється авторегресійній моделі першого порядку

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

в економетричну модель

$$Y_t = a_0 + a_1 X_t + u_t,$$

тоді одержимо:

$$Y_t = a_0 + a_1 X_t + \rho u_{t-1} + \varepsilon_t,$$

де $u_{t-1} = Y_{t-1} - a_0 - a_1 X_{t-1}$.

Звідси

$$Y_t = a_0(1-\rho) + \rho Y_{t-1} + a_1 X_t + a_1(\rho X_{t-1}) + \varepsilon_t.$$

ε_t в даному випадку має скалярну матрицю дисперсій:

$$M(\varepsilon_t \varepsilon_t') = \sigma_{\varepsilon_t}^2 E.$$

На основі методу 1МНК розраховуються оцінки параметрів $\hat{\rho}$, $\hat{a}_0$, $\hat{a}_1$.

Крок 2. Оцінка параметра $\hat{\rho}$ використовується для перетворення змінних $(Y_t - \hat{\rho} Y_{t-1})$ і $(X_t - \hat{\rho} X_{t-1})$, а метод 1МНК уже застосовується для перетворених даних.

Прогноз на основі економетричної моделі з автокорельованими залишками виконується так:

$$\hat{Y}_{n+1} = X_{n+1} A + \rho \varepsilon_n.$$

7. Метод інструментальних змінних

Автокореляція елементів часового ряду може бути описана на основі авторегресійної функції певного порядку. Але коли йдеться про моделювання взаємозв'язків на основі багатомірних часових рядів, де необхідно кількісно описати залежність однієї змінної від інших і врахувати автокореляцію залежної змінної, то економетрична модель буде мати лагову змінну Y, що входить до пояснюючих змінних, тобто:

$$Y_t = f(\ Y_{t-1}\ ,\ X_t\ ,\ u_t\ )$$
.

Наведемо економічні приклади :

Приклад 1. Нехай потрібно побудувати економетричну модель, яка характеризує залежність обсягу інвестицій від технічного рівня об'єкта інвестування та ефективності економічної діяльності. В цьому випадку вихідна інформація повинна мати три часових ряди, один з яких буде характеризувати обсяг зміни капітальних вкладень у часі, другий — технічний рівень об'єкта інвестування, а третій — ефективність економічної діяльності. При цьому кожне наступне значення часового ряду буде знаходитись у певній залежності від попереднього. Наприклад, якщо в попередньому році в даний об'єкт було вкладено велику частину інвестицій, то, очевидно, що в наступному році їх рівень може значно зменшитись через економічну та технічну недоцільність. А це означає, що економетрична модель наведеної залежності у загальному вигляді:

$$Y_t = f(\ Y_{t-1}\ ,\ X_{1t}\ ,\ X_{2t}\ ,\ u_t\ )$$
,

де Y_t — обсяг інвестицій у періоді t ; Y_{t-1} — обсяг інвестицій у періоді t – 1; X_{1t} — фондомісткість основних фондів у періоді t; $X_{2\,t}$ — рентабельність економічної діяльності в періоді t; ut — стохастична складова, залишки.

Приклад 2. Нехай треба побудувати економетричну модель, яка характеризує залежність між рівнем експорту певної продукції та обсягом її виробництва. Будуючи цю економетричну модель на основі двох часових рядів, потрібно також урахувати той факт, що рівень експорту в період t, як правило, корелює з його рівнем в період t – 1. Тому економетрична модель матиме загальний вигляд:

$$Y_t = f(\ Y_{t-1}\ ,\ X_t\ ,\ u_t\ )$$
,

де Y_t — рівень експорту продукції в період ;

Y_{t-1} — рівень експорту продукції в період $t-1$;

X_t — обсяг виробництва в період t ;

u_t — залишки.

Можна було б навести ще багато прикладів, коли залежна змінна (Y_t) економетричної моделі залежить не тільки від пояснюючих змінних X_t, але й від свого попереднього значення – Y_{t-1}. У таких випадках порушується третя передумова для застосування методу 1МНК при оцінці параметрів моделі, тобто пояснюючі змінні корелюють із залишками (u_t). В цьому випадку вони стають стохастичними змінними.

Зважаючи, що u_{t-1} впливає на Y_{t-1}, а Y_{t-1} – на Y_t, то й u_{t-1} впливає на Y_t навіть тоді, коли послідовні значення залишків є незалежними.

Якщо в цьому випадку оцінки параметрів моделі визначатимуться на основі методу 1МНК, то вони будуть необгрунтованими. Зауважимо, що навіть якщо лише один елемент вектора $p\lim\left\{\frac{1}{n}(X'u)\right\}\neq 0$, ми маємо змогу отримати всі елементи вектора $\hat{a}$ (всі оцінки моделі) необгрунтованими.

Другою причиною, яка може призвести до порушення третьої передумови, коли $M(X'u)=0 \Rightarrow p\lim\left\{\frac{1}{n}(X'u)\right\}\neq 0$, є наявність помилок у вихідній інформації.

До цих пір припускалось, що змінні економетричної моделі вимірюються без помилок, і тільки залишки u_t — це єдина форма помилок, яка допускається. Але дуже часто при вимірюванні змінних, які включаються в економетричну модель, допускаються помилки. Наявність цих помилок впливає на оцінку параметрів моделі. Так, якщо матриця пояснюючих змінних X має помилки, то її можна записати як суму двох матриць

$$X = \widetilde{X} + V,$$

де $\widetilde{X}$ — матриця дійсних значень пояснюючих змінних;

V — матриця помилок.

Звідси економетрична модель матиме вигляд:

$$Y = Xa + (u - Va).$$

Обгрунтованість оцінок параметрів цієї моделі залежить від того, чи дорівнює нулю матриця коваріацій змінних X із залишками $u-Va$, тобто:

$$p\lim\left[\frac{1}{n}X'(u-Va)\right]=p\lim\left(\frac{1}{n}X'u\right)-p\lim\left(\frac{1}{n}X'V\right)a$$.

$$p\lim\left(\frac{1}{n}X'V\right)a$$

, у свою чергу, можна записати так:

$$p\lim\left(\frac{1}{n}X'V\right)a=p\lim\left(\frac{1}{n}X'V\right)+p\lim\left(\frac{1}{n}V'V\right)$$.

Якщо в цих співвідношеннях допустити, що матрицю X (як її дійсні значення, так і помилки) гранично не корелюють із залишками, тобто:

$$p\lim\left(\frac{1}{n}X'u\right)=0$$ і $$p\lim X'V=0$$,

то матриця коваріації помилок частіше всього не дорівнює нулю,

$$p\lim\left(\frac{1}{n}V'V\right)\neq 0$$

, а це означає, що за наявності помилок виміру змінних оцінка параметрів моделі за методом 1МНК є необгрунтованою і асимптотично зміщеною.

Кореляція між пояснюючими змінними і залишками є серйозною перешкодою для застосування методу 1МНК.

8. Моделі розподіленого лага

Для взаємозв'язків багатьох економічних процесів типовим є той факт, що ефект від впливу одного показника на інший виявляється не одразу, а поступово, через деякий період часу. Це явище називається лагом (запізненням). Кількісний вираз взаємозв'язку між капітальними вкладеннями і введенням основних фондів, між затратами виробничих ресурсів і обсягом виробництва, між доходами й витратами і т.п. повинен базуватись на врахуванні запізнення впливу, або лага.

Вимірювання зв'язку між економічними показниками з урахуванням лагу виконується на основі побудови економетричної моделі розподіленого лага:

$$y_t=\sum_{j=0}^{\infty}a_j x_{t-\tau}+u_t$$.

Коефіцієнти a_j, $j = 0,1,2,3...$ називаються коефіцієнтами лага, а послідовність $a = \{a_j;\ j = 0,1,2,3...\}$ структурою лага.

Якщо економетрична модель включає не тільки лагові змінні, а й змінні, що характеризують поточні умови функціонування економічних систем, то така модель називається узагальненою моделлю розподіленого лага:

$$y_t = \sum_{\tau \ge 0} a_\tau x_{t-\tau} + \sum_{s=1}^{m} b_s x_{t,s} + u_t .$$

Оскільки параметри a_τ відносяться до однієї й тієї самої лагової змінної, що впливає на залежну змінну протягом певного часу, то $\sum_\tau a_\tau = W$ виражає сумарний вплив цієї лагової змінної на залежну, де W є кінцевим числом.

Щоб побудувати економетричну модель розподіленого лага, необхідно обґрунтувати величину лага. Для обґрунтування лага чи лагів доцільно використовувати взаємну кореляційну функцію, яка визначає ступінь зв'язку кожного елемента залежної змінної Y_t з елементом вектора незалежної змінної X_t, зрушеним відносно один одного на часовий лаг τ. При різних значеннях τ на основі взаємної кореляційної функції можна одержати $n+1$ значення (τ), які знаходяться на множині $r_{(\tau)} \in]-1\,1[$. Найбільше значення (τ) по модулю визначає зрушення або часовий лаг. Якщо таких зрушень декілька, то запізнення впливу змінної X_t відбувається протягом певного проміжку часу, що й відображає модель розподіленого лагу.

Наявність мультиколінеарності між лаговими змінними в економетричній моделі ускладнює її побудову. Щоб звільнитись від мультиколінеарності, необхідно ввести такі коефіцієнти при лагових змінних, які мали б однаковий знак і для них можна було б знайти суму, тоді економетрична модель:

$$y_t = a \sum_{j=0}^{1} w_j x_{t-\tau} + u_t .$$

Для зображення лагових коефіцієнтів Л. Койк запропонував форму спадаючої геометричної прогресії, тобто:

$$w_j = (1-\lambda)\,\lambda^{j}, \qquad 0 \le \lambda \le 1 .$$

Тоді економетрична модель набуде такого вигляду:

$$Y_t = w\,(1-\lambda)\,X_t + \lambda\,Y_{t-1} + (u_t - \lambda\,u_{t-1}),$$

тобто в правій частині моделі з'являється лагова змінна Y_{t-1}.

Лагову змінну в правій частині моделі мають також моделі часткового коригування:

$$Y_t = \alpha\,Y + \beta_\gamma\,X_t + (1-\gamma)\,Y_{t-1} + u_t, \qquad 0 \le \gamma \le 1$$

й адаптивних сподівань:

$$Y_t = \alpha\,(1-\lambda) + \beta\,(1-\lambda)\,X_t + \lambda\,Y_{t-1} + u_t\,(1-\lambda\,u_{t-1}) .$$

Наявність в економетричній моделі лагової змінної та прийняття гіпотези відносно залишків зумовлюють особливості оцінки параметрів моделі. Наведемо ці гіпотези.

Гіпотеза 1. Залишки є випадковими величинами і розподіляються нормально.

Гіпотеза 2. Залишки описуються авторегресійною схемою першого порядку:

$$v_t = u_t - \lambda\,u_{t-1}, \qquad 0 \le \lambda \le 1 .$$

Гіпотеза 3. Залишки описуються авторегресійною схемою першого порядку:

$$v_t = u_t - \lambda\,u_{t-1},$$

де $\quad v_t = \rho\,v_{t-1} + \varepsilon_t$.

Якщо відносно залишків приймається перша гіпотеза, то для оцінки параметрів можна застосувати 1МНК.

Якщо відносно залишків приймається друга гіпотеза (залишки автокорельовані), то застосовується метод Ейткена. В операторі $\hat{A} = (X'V^{-1}X)^{-1}X'V^{-1}Y$ матриця V матиме вигляд:

$$V = \sigma_u^2 \begin{pmatrix} 1+\lambda^2 & -\lambda & 0 & 0 & \cdots & 0 \\ -\lambda & 1+\lambda^2 & -\lambda & 0 & \cdots & 0 \\ 0 & -\lambda & 1+\lambda^2 & -\lambda & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1+\lambda^2 \end{pmatrix}$$

Коли модель $Y_t = a_0 + a_1 X_t + \lambda Y_{t-1} + u_t$ при другій гіпотезі треба записати:

$$Y_t = \lambda Y_{t-1} = a_0 + a_1 X_t + u_t,$$

можна застосувати 1МНК для перетворення даних залежної змінної Y_t на основі параметра λ. Параметр λ пропонується вибирати довільно на інтервалі $0 < \lambda < 1$ таким чином, щоб мінімізувати суму квадратів залишків $u V^{-1} u$.

Якщо відносно залишків моделі приймається третя гіпотеза, то для оцінки параметрів моделі можна використовувати:

1) 1МНК, коли вихідні дані перетворені на основі параметрів ρ і λ;

2) метод Ейткена;

3) ітеративний метод;

4) двокрокову процедуру:

а) 1МНК для вихідних даних, коли $\hat{\rho} = \dfrac{a_2 \rho}{a_2}$;

б) 1МНК для перетворених даних на основі $\rho = \hat{\rho}$.

5) метод інструментальних змінних;

6) алгоритм Уолліса.

Щоб застосувати для оцінки параметрів 1МНК, матриця вихідних даних матиме вигляд:

$$X(\rho) = \begin{pmatrix} \lambda & 1 & x_1(\rho) \\ \lambda^2 & 1+\lambda & x_2(\rho) + \lambda x_1(\rho) \\ \lambda^3 & 1+\lambda+\lambda^2 & x_3(\rho) + \lambda x_2(\rho) + \lambda^2 x_1(\rho) \\ \cdots & \cdots & \cdots \\ \lambda^n & 1+\lambda+\lambda^2+\ldots+\lambda^{n-1} & x_n(\rho) + \lambda x_{n-1}(\rho) + \ldots + \lambda^{n-1} x_1(\rho) \end{pmatrix}$$

Параметри ρ і λ вибираються довільно на множині $]0,1[$. Для кожної пари ρ і λ послідовно розраховуються й залишки. ρ і λ вибираються доти, доки не буде мінімізована сума відхилень.

Оператор оцінювання за методом Ейткена $\hat{A} = (X'S^{-1}X)^{-1}X'S^{-1}Y$ базується на матриці

$$S = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix},$$

а матриця X дорівнює:

$$X = \begin{pmatrix} 1 & y_0 & x_1 \\ 1 & y_1 & x_2 \\ 1 & y_2 & x_3 \\ \cdots & \cdots & \cdots \\ 1 & y_{n-1} & x_n \end{pmatrix}.$$

Цей метод аналогічний оцінкам 1МНК для моделі:

$$(Y_t - \rho\, Y_{t-1}) = a_0(1-\rho) + a_1(Y_{t-1} - \rho\, Y_{t-2}) + a_2(X_t - \rho\, X_{t-1}) + \varepsilon_t \qquad (1)$$

відносно перетворених даних.

Ітеративний метод є альтернативою методу Ейткена. Його алгоритм має чотири кроки:

Крок 1. Вибирається початкове значення $\rho = \hat{\rho}$ і підставляється у функцію (1).

Крок 2. Застосовується 1МНК для оцінки параметрів $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$.

Крок 3. У функцію (1) підставляють параметри $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$ і на основі 1МНК розраховують параметр $\hat{\hat{\rho}}$.

Крок 4. Задавши $\hat{\hat{\rho}}$ на основі 1МНК, розраховують параметри $\hat{\hat{a}}_0$, $\hat{\hat{a}}_1$, $\hat{\hat{a}}_2$ і т.д.

Метод інструментальних змінних для оцінки параметрів моделі застосовується тоді, коли залишки не автокорельовані, але існує залежність пояснюючих змінних від залишків. Якщо модель має вигляд:

$$Y_t = a_0 + a_1 X_t + a_2 X_{t-1} + a_3 Y_{t-1} + u_t ,$$

то можна замість змінної Y_{t-1} використати як інструментальну змінну $\hat{Y}_{t-1}$, яка розраховується як функція $\hat{Y}_{t-1} = f(X_{t-1})$.

Оцінка параметрів моделі виконується на основі алгоритму Уолліса, який складається з трьох етапів.

На першому етапі оцінка параметрів моделі виконується на основі методу інструментальних змінних, де X_{t-1} використовується як інструментальна змінна для Y_{t-1}.

На другому етапі розраховують коефіцієнт автокореляції першого порядку з урахуванням поправки на зміщення і формують матрицю S.

На третьому етапі виконують оцінку параметрів моделі на основі методу Ейткена.

9. Економетрична модель на основі системи одночасових структурних рівнянь

Наявність прямих та зворотніх зв'язків між економічними показниками в багатьох випадках вимагає використання системи одночасових рівнянь. Вони, як правило, містять лінійні рівняння. Нелінійність зв'язків апроксимується лінійними співвідношеннями. Динаміка економічних зв'язків ураховується за допомогою часових лагів або лагових змінних.

Система одночасових структурних рівнянь в матричному вигляді має такий вигляд:

$$Y = AY + BX + u .$$

Якщо кожне рівняння системи розв'язати відносно Y, то одержимо приведену форму моделі, яка має вигляд:

$$Y = R X + v ,$$

де залишки v є лінійною комбінацією залишків u.

Зв'язок між коефіцієнтами структурної і приведеної форми моделі визначиться:

$$R = (E - A)^{-1} B ,$$

або $R = -A^{-1}B$,

або $AR + B = 0$.

Оцінка параметрів моделі на основі одночасових рівнянь методом 1МНК буде давати зміщення, яке буде дорівнювати:

$$\frac{(1-a_1)\,\sigma^2 / \overline{m}_{zz}}{1+\sigma^2 / \overline{m}_{zz}}$$, де

$\overline{m}_{zz}$ — момент другого порядку залежної змінної, який прямує до деякої константи.

Чисельна оцінка параметрів моделі на основі одночасових структурних рівнянь пов'язана з проблемою ідентифікації.

Необхідна умова ідентифікації системи — справедливість нерівності для кожного рівняння:

$k_s - 1 \le m - m_s$,

де k_s — кількість ендогенних змінних, які входять в s-те рівняння структурної форми;

m — загальна кількість екзогенних змінних моделі;

m_s — кількість екзогенних змінних, які не входять в s-те рівняння структурної форми моделі.

Якщо записане вище співвідношення виконується як рівність, то відповідне рівняння є строгоідентифікованим, а коли — як нерівність, то відповідне рівняння є надідентифікованим.

Якщо в структурній формі моделі $Y = AY + BX + u$ матриця A є трикутною, а залишки характеризуються діагональною матрицею виду:

$$\Sigma = M(uu') = \begin{pmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{pmatrix},$$

то така система рівнянь називається рекурсивною і для оцінки параметрів такої моделі можна застосувати 1МНК.

Якщо кожне рівняння моделі є строгоідентифікованим, то для оцінки параметрів моделі можна застосувати непрямий метод найменших квадратів (НМНК). Алгоритм цього методу складається з чотирьох кроків:

Крок 1. Перевіряється умова ідентифікованості для кожного рівняння. Якщо кожне рівняння точноідентифіковане, то виконується перехід до кроку 2.

Крок 2. Перехід від структурної форми моделі до приведеної.

Крок 3. Оцінка параметрів кожного рівняння приведеної форми моделі 1МНК.

Крок 4. Розрахунок оцінок параметрів рівнянь структурної форми на основі співвідношення:

$$AR = -B,$$

де A і B — параметри структурних рівнянь, а R — матриця оцінок параметрів приведеної форми моделі.

Якщо рівняння структурної форми моделі надідентифіковані, то для оцінки параметрів моделі застосовується двокроковий метод найменших квадратів (2МНК). Система рівнянь для обчислення оцінок двокроковим методом найменших квадратів запишеться так:

$$\begin{pmatrix} Y_1{}'X(X'X)^{-1}X'Y_1 & Y'_1 X_1 \\ X'_1 Y_1 & X'_1 X_1 \end{pmatrix} \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} Y_1{}'X(X'X)^{-1}X'Y \\ X'_1 Y \end{pmatrix},$$

де Y — вектор залежної або ендогенної змінної;

Y_1 — матриця поточних ендогенних змінних, які входять у праву частину рівняння;

X — матриця всіх пояснюючих або екзогенних змінних;

X_1 — матриця пояснюючих або екзогенних змінних даного рівняння;

$\hat{a}$ — вектор структурних параметрів, які відносяться до змінних матриці Y_1;

$\hat{b}$ — вектор структурних параметрів, які відносяться до змінних матриці X_1.

Оператор оцінювання 2МНК:

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} Y_1{}'X(X'X)^{-1}X'Y_1 & Y'_1 X_1 \\ X'_1 Y_1 & X'_1 X_1 \end{pmatrix}^{-1} \begin{pmatrix} Y_1{}'X(X'X)^{-1}X'Y \\ X'_1 Y \end{pmatrix}$$

Дисперсія залишків для кожного рівняння:

$$\sigma_{u_S}^2 = \frac{1}{n-k-r+1} u_S{}' u_S$$
.

Матриця коваріацій параметрів кожного рівняння визначається на основі співвідношення:

$$asy\ \mathrm{var}\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} = \sigma_{u_S}^2\begin{pmatrix} Y'_1 X(X'X)^{-1} X'Y_1 & Y'_1 X_1 \\ X'_1 Y_1 & X'_1 X_1\end{pmatrix}^{-1}$$

Трикроковий метод найменших квадратів (3МНК), на відміну від попередніх, призначений для одночасної оцінки параметрів всіх рівнянь моделі. Оператор оцінювання 3МНК матиме вигляд:

$$\hat{\delta} = \begin{pmatrix} (S_{11}^2)^{-1} Z_1'(X'X)^{-1} X'Z_1 & \cdots & (S_{1r}^2)^{-1} Z_1'(X'X)^{-1} X'Z_r \\ \cdots & \cdots & \cdots \\ (S_{r1}^2)^{-1} Z_r'(X'X)^{-1} X'Z_1 & \cdots & (S_{rr}^2)^{-1} Z_r'(X'X)^{-1} X'Z_r \end{pmatrix}^{-1} \cdot$$

$$\cdot\begin{pmatrix} \sum_{j=1}^{r} (S_{1j}^2)^{-1} Z_1' X(X'X)^{-1} X'Y_j \\ \cdots \\ \sum_{j=1}^{r} (S_{rj}^2)^{-1} Z_r' X(X'X)^{-1} X'Y_j \end{pmatrix},$$

де $\hat{\delta} = \begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix}$ — оцінки параметрів моделі;

$Z_s = (\ Y_s\ ,\ X_s\ )\ ,(s=1,\bar{r})$, де Z_s — змінні моделі, які знаходяться у правій частині s-го рівняння;

S_{rj}^2 — дисперсії залишків для кожного рівняння, які є наближеною оцінкою σ_{rj}^2.

Щоб застосувати 3МНК на практиці, необхідно дотримання таких вимог:

1) приступаючи до оцінки параметрів моделі, необхідно виключити всі тотожності;

2) виключити із системи кожне неідентифіковане рівняння;

3) при наявності серед рівнянь системи точноідентифікованих та надідентифікованих, 3МНК доцільно застосувати до кожної з груп рівнянь окремо;

4) якщо група надідентифікованих рівнянь має тільки одне рівняння, то 3МНК перетворюється в 2МНК.

Якщо матриця коваріацій для структурних залишків є блочно-діагональною, то вся процедура оцінювання на основі 3МНК може бути застосована окремо для кожної групи рівнянь, які відповідають одному блоку.

Точковий прогноз залежних змінних визначається на основі приведеної форми економетричної моделі:

$$Y_f = \hat{R}\, X_f\ ,$$

де X_f — вектор прогнозних екзогенних змінних.

Визначення довірчих інтервалів для цього прогнозу залежить від методу, за допомогою якого була одержана матриця $\hat{R}$.

Довірчі інтервали для кожної ендогенної змінної задаються співвідношенням:

$$\hat{Y}_{jf} \pm t_{(\alpha/2)} \sqrt{S_{ss}^2}\ ,$$

де S_{ss}^2 — дисперсія залишків s- го рівняння моделі;

$$t_{(\alpha/2)} = F_{(\alpha)}\ .$$

Довірчі інтервали для всіх ендогенних змінних визначаються так:

$$Y_{sf} = \hat{Y}_{jf} \pm \sqrt{\frac{\left[1 + X'_f\ (X'X)^{-1} X_f\right](n-k)\ r}{n-k-r+1} \cdot S_{ss}^2}\ ,$$

де S_{ss}^2 — незміщена дисперсія залишків всіх рівнянь моделі.

Ці інтервали будуть ширшими, ніж у випадку, коли їх задавати для кожної ендогенної змінної окремо.

ЗАВДАННЯ ДО КОНТРОЛЬНОЇ РОБОТИ

Завдання 1. Моделювання динаміки зміни прибутку підприємства

На підставі даних табл. 1 побудувати економетричну модель динаміки зміни прибутку підприємства (y, млн. грн.) залежно від поточного періоду (x, рік) та проаналізувати її якість.

1. Побудувати поле кореляції.

2. Методом найменших квадратів (МНК) розрахувати параметри рівнянь парної регресії: а) лінійного $\hat{y}(x) = b_0 + b_1x$; б) квазілінійного $\hat{y}(x) = b_0 + b_1u(x)$ (функції $u(x)$ задані в табл. Намалювати графіки цих ліній.

3. Для визначення тісноти зв'язку між прибутком та поточним періодом розрахувати коефіцієнт і індекс кореляції, коефіцієнт детермінації. За t-критерієм оцінити значущість коефіцієнта кореляції для рівня надійності $P = 0{,}95$. Для кожного рівняння регресії зробити висновок про тісноту зв'язку.

4. За t-критерієм перевірити статистичну значущість коефіцієнтів лінійного рівняння з надійністю $P = 0{,}95$.

5. Знайти інтервали довіри для коефіцієнтів лінійного рівняння з надійністю $P = 0{,}95$, $P = 0{,}99$.

6. Оцінити якість рівнянь регресії через середню помилку апроксимації.

7. За F-критерієм Фішера оцінити статистичну надійність результатів регресійного моделювання для кожного типу рівнянь.

8. Розрахувати прогнозне значення прибутку для значення поточного періоду, збільшеного на 10 % від його максимального рівня. Визначити надійний інтервал прогнозу для рівня значущості $\alpha = 0{,}05$.

9. Для кожного типу рівнянь навести графіки залишків. За критерієм Дарбіна-Уотсона перевірити наявність або відсутність автокореляції залишків.

10. За результатами досліджень вибрати найкраще рівняння регресії. Обґрунтувати свій вибір

Таблиця 1

Варіант	i	1	2	3	4	5	6	7	8	9	10
1	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	-0,003	2,186	1,72	4,416	5,426	9,016	10,229	12,804	14,572	16,6337
2	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	2,9973	5,186	4,72	7,416	8,426	12,016	13,229	15,804	17,572	19,6337
3	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	1,498	2,2934	3,08882	2,56012	3,48854	3,77528	4,18112	5,2582	5,36303	7,3507
4	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	1,8268	6,696	9,88607	10,5493	12,197	12,7293	14,8434	14,9108	15,6051	17,2862
5	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	0,856	3,67059	5,56945	7,10318	7,42075	7,90404	8,84864	8,74577	10,0189	10,2803
6	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	1,1623	1,4482	2,7496	2,5494	3,4908	3,839	5,7362	6,2756	7,6014	10,1075
7	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	4,5232	13,748	12,097	10,273	9,04388	8,03315	7,49724	6,987	6,37454	4,6749
8	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	7,0192	3,98	2,17	1,554	1,41194	1,17677	1,15616	0,97	1,20231	1,1909
9	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	17,9973	19,786	18,12	18,816	17,026	17,016	13,829	11,204	6,972	2,2337
10	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	19,2492	10,8	8,03333	6,538	6,786	5,13067	5,22471	3,502	5,87678	5,7228

Варіант	i	1	2	3	4	5	6	7	8	9	10
11	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	5,498	5,8506	6,10518	4,91588	5,03746	4,33872	3,54088	3,1478	1,45697	1,2514
12	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	14,377	9,911	8,38915	6,5408	5,894	4,56973	3,74904	4,29699	3,04829	2
13	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	10,856	8,12541	6,78055	6,01282	4,54525	3,56996	3,28136	2,11023	2,4411	1,8596
14	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	8,3623	8,0658	8,5544	7,2726	6,8252	5,439	5,2178	3,2164	1,5406	0,5459
15	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	6,5232	0,748	2,097	3,273	3,98506	4,26392	4,63238	4,787	4,68223	3,3822
16	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	0,0192	1,98	3,17	3,554	3,82371	3,79215	3,88589	3,77	4,04846	4,069
17	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	2,9973	5,186	4,72	7,416	8,426	12,016	13,229	15,804	17,572	19,6337
18	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	0,8924	5,226	6,31133	6,969	7,42	7,81867	8,08014	7,965	7,70111	8,1846
19	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	4,498	5,73621	7,07247	7,20436	8,93962	10,2118	11,8214	14,3686	16,2691	20,45
20	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	13,5745	11,727	9,76336	8,33085	8,111	7,6058	6,88253	6,04975	5,16172	4,8755

Варіант	i	1	2	3	4	5	6	7	8	9	10
21	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	6,2112	4,53371	3,68278	3,04241	3,01512	2,20148	2,08418	1,24212	2,04555	1,7327
22	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	2,0826	1,793	2,409	2,624	3,831	4,611	5,961	8,616	10,078	11,6944
23	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	2,1341	6,997	5,784	5,123	4,39794	4,32608	3,82762	3,601	3,18477	2,2505
24	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	8,014	4,937	3,224	2,668	2,43694	2,27777	2,22516	2,167	2,07031	1,8111
25	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	3,3912	4,2	4,722	8,744	13,276	17,698	23,733	26,316	35,766	44,6538
26	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	0,1994	9,84	13,0933	14,631	16,468	16,2377	17,0939	16,301	18,6568	18,8996
27	x_i	0	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8
	y_i	1,7164	1,9174	3,30682	2,71112	3,42254	3,26728	4,91612	4,7372	5,41603	7,4463
28	x_i	0	1	2	3	4	5	6	7	8	9
	y_i	4,5745	8,727	9,24864	9,72315	11,111	12,0222	12,5795	12,9243	13,1323	13,8755
29	x_i	1	2	3	4	5	6	7	8	9	10
	y_i	0,2189	1,53529	1,72422	2,52459	2,98388	4,03052	4,15282	3,76688	4,22145	5,0503
30	x_i	1	1,2	1,4	1,6	1,8	2	2,2	2,4	2,6	2,8
	y_i	1,184	2,7916	2,9258	3,4882	5,1714	6,851	8,6386	10,6668	13,0212	16,4414

Таблиця 2

Варіант	**Функція $u(x)$**	**Варіант**	**Функція $u(x)$**
1, 9, 17, 25	x^2	5, 13, 21, 29	$\ln x$
2, 10, 18, 26	$1/x$	6, 14, 22, 30	x^3
3, 11, 19, 27	e^x	7, 15, 23	$x/(x^2+1)$
4, 12, 20, 28	$\sqrt{x}$	8, 16, 24	$1/(x^2+1)$

Основні формули та поняття до завдання 1

Теоретична парна лінійна регресійна модель:

$$y(x) = \mathbf{M}(Y|X = x) + \varepsilon = \beta_0 + \beta_1 x + \varepsilon, \qquad (1.1)$$

де y залежна змінна; x незалежна змінна; ε випадкова складова (похибка, залишок, відхилення); β_0, β_1 параметри моделі.

Теоретичне парне лінійне рівняння регресії:

$$\bar{y}(x) = \mathbf{M}(Y|X = x) = \beta_0 + \beta_1 x. \qquad (1.2)$$

Емпірична парна лінійна регресійна модель:

$$\tilde{y}(x) = b_0 + b_1 x + e. \qquad (1.3)$$

Емпіричне парне лінійне рівняння регресії:

$$\hat{y}(x) = b_0 + b_1 x. \qquad (1.4)$$

Оцінки параметрів за методом найменших квадратів (МНК)

Згідно з МНК параметри рівняння (1.4) визначаються за умов мінімуму суми квадратів відхилень теоретичних значень $\hat{y}_i$ від фактичних y_i:

$$\sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = f(b_0, b_1) \to \min. \qquad (1.5)$$

Необхідні умови екстремуму функції (1.5):

$$\begin{cases} \dfrac{\partial f(b_0, b_1)}{\partial b_0} = \dfrac{\partial}{\partial b_0}\sum_{i=1}^{n} e_i^{\,2} = -2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0 \\ \dfrac{\partial f(b_0, b_1)}{\partial b_1} = \dfrac{\partial}{\partial b_1}\sum_{i=1}^{n} e_i^{\,2} = -2\sum_{i=1}^{n} x_i (y_i - b_0 - b_1 x_i) = 0 \end{cases} \qquad (1.6)$$

З (1.6) одержимо нормальну систему лінійних рівнянь:

$$\begin{cases} n b_0 + b_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b_0 \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} y_i x_i \end{cases} \Leftrightarrow \begin{cases} b_0 + b_1 \bar{x} = \bar{y} \\ b_0 \bar{x} + b_1 \overline{x^2} = \overline{xy} \end{cases}, \qquad (1.7)$$

де $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$, $\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ середні значення; n кількість спостережень.

Із системи (1.7) знаходимо параметри рівняння (1.4):

$$b_1 = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\overline{x^2} - \bar{x}^2}; \qquad b_0 = \bar{y} - b_1 \bar{x}. \qquad (1.8)$$

Коефіцієнт кореляції є характеристикою лінійного взаємозв'язку між двома випадковими величинами (ВВ) X і Y та розраховується за формулою:

$$r_{xy} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sqrt{\overline{x^2} - \bar{x}^2}\sqrt{\overline{y^2} - \bar{y}^2}}. \qquad (1.9)$$

Коефіцієнт кореляції набуває значення в інтервалі $-1 \le r_{xy} \le 1$. Позитивне значення коефіцієнта кореляції свідчить про прямий зв'язок між X і Y (зі зростанням однієї ВВ зростає середнє значення іншої), від'ємне – про зворотний зв'язок (із зростанням однієї ВВ середнє значення іншої убуває). Якщо $r_{xy} \to \pm 1$ – зв'язок тісний, а якщо $r_{xy} \to 0$ – лінійного зв'язку немає.

t-тест Стьюдента для перевірки значущості коефіцієнта кореляції

1. Розрахувати t-відношення: $t_r = r_{xy}\sqrt{\dfrac{n-2}{1-r_{xy}^2}}$. (1.10)
2. З таблиць критичних точок розподілу Стьюдента знайти $t_{\alpha/2,\, n-2}$ (α – заданий рівень значущості, $n-2$ – кількість ступенів свободи).
3. Якщо $|t_r| > t_{\alpha/2,\, n-2}$ – коефіцієнт r_{xy} є статистично значущим.

***t*-тест Стьюдента для перевірки значущості параметрів b_0, b_1 лінійного рівняння регресії**

1. Розрахувати t-відношення: $t_{b_1} = b_1 / S_{b_1}$, $t_{b_0} = b_0 / S_{b_0}$, (1.11)

де $S_{b_1}^2 = S^2 / \sum(x_i - \bar{x})^2$, $S_{b_0}^2 = \overline{x^2} S_{b_1}^2$ – дисперсії параметрів b_1, b_0; $S^2 = \dfrac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ незміщена оцінка дисперсії залишків.

2. З таблиць критичних точок розподілу Стьюдента знайти $t_{\alpha/2,\, n-2}$ (α – рівень значущості, $n-2$ – кількість ступенів свободи).

3. Якщо $|t_{b_i}| > t_{\alpha/2,\, n-2}$ – коефіцієнт b_i є статистично значущим.

Інтервали довіри для параметрів β_0, β_1 теоретичної регресії

Інтервалом довіри називають інтервал, який із заданою надійністю P накриває невідомий параметр. Інтервали довіри для параметрів β_0, β_1 теоретичної регресії обчислюються за формулою:

$$b_i - \Delta b_i < \beta_i < b_i + \Delta b_i, \qquad (1.12)$$

де $\Delta b_i = t_{\alpha/2,\, n-2} S_{b_i}$ – граничне відхилення параметра (α – рівень значущості, зв’язаний з рівнем надійності P співвідношенням: $\alpha = 1 - P$).

Коефіцієнт детермінації є сумарною мірою відповідності рівняння регресії статистичним даним та розраховується за формулою:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \qquad (1.13)$$

де $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$ – загальна сума квадратів;

$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – сума квадратів помилок;

$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ – сума квадратів, що пояснюється регресією.

Коефіцієнт детермінації змінюється в межах: $0 \le R^2 \le 1$. Чим тісніше лінійний зв'язок між величинами X і Y, тим ближче до 1 коефіцієнт детермінації. Чим слабкіше такий зв'язок, тим ближче R^2 до нуля.

Коефіцієнт детермінації показує, на скільки процентів варіація залежної змінної визначається варіацією незалежних змінних.

Індекс кореляції є характеристикою взаємозв'язку між випадковими величинами (ВВ) та розраховується за формулою:

$$R = \sqrt{R^2} \qquad (0 \le R \le 1). \qquad (1.14)$$

Для парної лінійної регресії індекс кореляції співпадає з коефіцієнтом кореляції.

Середня помилка апроксимації характеризує середнє відхилення (у відсотках) розрахованих значень $\hat{y}_i$ від фактичних y_i:

$$A = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{y_i}\right| \cdot 100 = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \cdot 100. \qquad (1.15)$$

Якщо $0 \le A \le 10$ % – апроксимація відмінна;

$10 \% < A \le 20 \%$ – апроксимація добра;

$20 \% < A \le 30 \%$ – апроксимація задовільна;

$A > 30 \%$ – апроксимація незадовільна.

Перевірка моделі на адекватність за F-критерієм Фішера

Гіпотезу про рівень значущості зв'язку між залежною і незалежною змінними можна перевірити за допомогою F-критерію.

1. Розрахувати F-відношення:

$$F_{(1,\,n-2)} = \frac{SSR/1}{SSE/(n-2)} = \frac{R^2\ /1}{(1-R^2)/(n-2)}. \qquad (1.16)$$

2. За таблицями F-розподілу Фішера знайти критичне значення $F_{кр} = F_{\alpha;\ \ 1,\,n\text{-}2}$ (α – заданий рівень значущості; 1, n – 2 – ступені свободи).

3. Якщо розраховане значення $F > F_{кр}$, то гіпотеза про істотність зв'язку між залежною і незалежною змінними підтверджується, побудована регресійна модель адекватно апроксимує дані спостережень.

Точковий прогноз значення залежної змінної $\hat{y}_{n+1}$ **для прогнозного значення** x_{n+1} розраховується за допомогою емпіричного рівняння регресії:

$$\hat{y}_{n+1} = b_0 + b_1 x_{n+1}. \qquad (1.17)$$

Інтервал довіри для прогнозованого індивідуального значення залежної змінної розраховується за формулою:

$$\hat{y}_{n+1} - \Delta y_{n+1} < y_{n+1} < \hat{y}_{n+1} + \Delta y_{n+1}, \qquad (1.18)$$

де

$$\Delta y_{n+1} = t_{\alpha/2,\,n-2} S \sqrt{1 - \frac{1}{n} + \frac{(x_{n+1}-\bar{x})^2}{\sum (x_i - \bar{x})^2}}. \qquad (1.19)$$

Перевірка наявності автокореляції за методом Дарбіна-Уотсона

Важливою передумовою побудови якісної регресійної моделі за МНК є відсутність автокореляції залишків ($\mathbf{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$). Для перевірки наявності автокореляції використовується метод Дарбіна-Уотсона, який включає наступні етапи:

1. Визначити значення відхилень $e_i = y_i - \hat{y}_i$ для i = 1, 2, …, n.

2. Розрахувати статистику Дарбіна-Уотсона:

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}. \qquad (1.20)$$

3. З таблиць критичних точок критерію Дарбіна-Уотсона визначити нижню d_l та верхню d_u критичні межі для рівня значущості α, числа спостережень n та числа пояснювальних змінних $m = 1$ і зробити висновки за правилом:

$0 \le DW < d_l$ – існує позитивна автокореляція;

$d_l \le DW < d_u$ – висновок про наявність автокореляції не визначений;

$d_u \le DW < 4 - d_u$ – автокореляція відсутня;

$4 - d_u \le DW < 4 - d_l$ – висновок про наявність автокореляції не визначений;

$4 - d_l \le DW \le 4$ – існує від'ємна автокореляція.

Нелінійна парна регресія

За методикою оцінок параметрів розглядають 2 види нелінійних регресій:

1) нелінійні за факторами, але лінійні за параметрами (квазілінійні регресії):

$$\hat{y}(x) = b_0 + b_1 u(x); \qquad (1.21)$$

2) нелінійні за факторами і параметрами,

наприклад, $\hat{y} = b_0 b_1^x$, $\hat{y} = b_0 x^{b_1}$.

Квазілінійна регресія (1.21) заміною $z = u(x)$ приводиться до лінійної парної регресії $\hat{y}(x) = b_0 + b_1 z$.

Завдання 2. Побудова лінійної багатофакторної регресійної моделі

Нижче у таблицях для варіантів 1-11 наведені дані (у млрд. дол.) про діяльність компаній США в 2012 р. (y – чистий дохід, x_1 – обіг капіталу, x_2 – використаний капітал), а для варіантів 12-30 – дані про вартість житла y (тис. дол.) у м. Києві залежно від її загальної x_1 та житлової x_2 площі (м^2).

1. Розрахувати коефіцієнти лінійного рівняння множинної регресії:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 .$$

2. За допомогою t-критерію оцінити статистичну значущість коефіцієнтів.

3. Розрахувати індекс множинної кореляції та коефіцієнт детермінації.

4. Оцінити статистичну значущість коефіцієнта детермінації за F-критерієм Фішера. Зробити висновок про якість апроксимації.

5. Оцінити якість рівняння регресії через середню помилку апроксимації.

6. Розрахувати прогнозне значення результату, якщо прогнозні значення факторів складають 110 % від їхніх максимальних значень.

7. Розрахувати дисперсію прогнозу та інтервал довіри прогнозу для рівня значущості $\alpha = 0{,}05$.

8. Побудувати матриці парних і частинних коефіцієнтів кореляції. Зробити висновки про силу впливу факторів на результат і можливість мультиколінеарності факторів.

9. За допомогою частинних F-критеріїв Фішера оцінити доцільність включення в рівняння регресії фактора x_2 після фактора x_1 і фактора x_1 після фактора x_2.

10. Розрахувати середні частинні коефіцієнти еластичності і дати на їхній основі порівняльну оцінку сили впливу факторів на результат.

11. За критерієм Дарбіна-Уотсона перевірити автокореляцію залишків.

Варіант	Змінна	Значення									
1	y	6,6	3	6,5	3,3	0,1	3,6	1,5	5,5	2,4	3
	x_1	6,9	18	107,9	16,7	79,6	16,2	5,9	53,1	18,8	35,3
	x_2	83,6	6,5	50,4	15,4	29,6	13,3	5,9	27,1	11,2	16,4
2	y	6,5	3,3	0,1	3,6	1,5	5,5	2,4	3	4,2	2,7
	x_1	107,9	16,7	79,6	16,2	5,9	53,1	18,8	35,3	71,9	93,6
	x_2	50,4	15,4	29,6	13,3	5,9	27,1	11,2	16,4	32,5	25,4
3	y	0,1	3,6	1,5	5,5	2,4	3	4,2	2,7	1,6	2,4
	x_1	79,6	16,2	5,9	53,1	18,8	35,3	71,9	93,6	10	31,5
	x_2	29,6	13,3	5,9	27,1	11,2	16,4	32,5	25,4	6,4	12,5
4	y	1,5	5,5	2,4	3	4,2	2,7	1,6	2,4	3,3	1,8
	x_1	5,9	53,1	18,8	35,3	71,9	93,6	10	31,5	36,7	13,8
	x_2	5,9	27,1	11,2	16,4	32,5	25,4	6,4	12,5	14,3	6,5
5	y	2,4	3	4,2	2,7	1,6	2,4	3,3	1,8	2,4	1,6
	x_1	18,8	35,3	71,9	93,6	10	31,5	36,7	13,8	64,8	30,4
	x_2	11,2	16,4	32,5	25,4	6,4	12,5	14,3	6,5	22,7	15,8
6	y	4,2	2,7	1,6	2,4	3,3	1,8	2,4	1,6	1,4	0,9
	x_1	71,9	93,6	10	31,5	36,7	13,8	64,8	30,4	12,1	31,3
	x_2	32,5	25,4	6,4	12,5	14,3	6,5	22,7	15,8	9,3	18,9
7	y	3	6,5	3,3	0,1	3,6	1,5	5,5	2,4	3	4,2
	x_1	18	107,9	16,7	79,6	16,2	5,9	53,1	18,8	35,3	71,9

Варіант	Змінна	Значення									
	x_2	6,5	50,4	15,4	29,6	13,3	5,9	27,1	11,2	16,4	32,5
8	y	3,3	0,1	3,6	1,5	5,5	2,4	3	4,2	2,7	1,6
	x_1	16,7	79,6	16,2	5,9	53,1	18,8	35,3	71,9	93,6	10
	x_2	15,4	29,6	13,3	5,9	27,1	11,2	16,4	32,5	25,4	6,4
9	y	3,6	1,5	5,5	2,4	3	4,2	2,7	1,6	2,4	3,3
	x_1	16,2	5,9	53,1	18,8	35,3	71,9	93,6	10	31,5	36,7
	x_2	13,3	5,9	27,1	11,2	16,4	32,5	25,4	6,4	12,5	14,3
10	y	5,5	2,4	3	4,2	2,7	1,6	2,4	3,3	1,8	2,4
	x_1	53,1	18,8	35,3	71,9	93,6	10	31,5	36,7	13,8	64,8
	x_2	27,1	11,2	16,4	32,5	25,4	6,4	12,5	14,3	6,5	22,7
11	y	3	4,2	2,7	1,6	2,4	3,3	1,8	2,4	1,6	1,4
	x_1	35,3	71,9	93,6	10	31,5	36,7	13,8	64,8	30,4	12,1
	x_2	16,4	32,5	25,4	6,4	12,5	14,3	6,5	22,7	15,8	9,3
12	y	15,9	27	13,5	15,1	21,1	28,7	27,2	28,3	52,3	22
	x_1	39	68,4	34,8	39	54,7	74,7	71,7	74,5	137,7	40
	x_2	20	40,5	16	20	28	46,3	45,9	47,5	87,2	17,7
13	y	13,5	15,1	21,1	28,7	27,2	28,3	52,3	22	28	45
	x_1	34,8	39	54,7	74,7	71,7	74,5	137,7	40	53	86
	x_2	16	20	28	46,3	45,9	47,5	87,2	17,7	31,1	48,7
14	y	21,1	28,7	27,2	28,3	52,3	22	28	45	51	34,4
	x_1	54,7	74,7	71,7	74,5	137,7	40	53	86	98	62,6
	x_2	28	46,3	45,9	47,5	87,2	17,7	31,1	48,7	65,8	21,4
15	y	27,2	28,3	52,3	22	28	45	51	34,4	24,7	30,8
	x_1	71,7	74,5	137,7	40	53	86	98	62,6	45,3	56,4
	x_2	45,9	47,5	87,2	17,7	31,1	48,7	65,8	21,4	20,6	29,7
16	y	52,3	22	28	45	51	34,4	24,7	30,8	15,9	29
	x_1	137,7	40	53	86	98	62,6	45,3	56,4	37	67,5
	x_2	87,2	17,7	31,1	48,7	65,8	21,4	20,6	29,7	17,8	43,5
17	y	28	45	51	34,4	24,7	30,8	15,9	29	15,4	28,6

Варіант	Змінна	Значення									
	x_1	53	86	98	62,6	45,3	56,4	37	67,5	37	69
	x_2	31,1	48,7	65,8	21,4	20,6	29,7	17,8	43,5	17,8	42,4
18	y	51	34,4	24,7	30,8	15,9	29	15,4	28,6	15,6	27,7
	x_1	98	62,6	45,3	56,4	37	67,5	37	69	40	69,1
	x_2	65,8	21,4	20,6	29,7	17,8	43,5	17,8	42,4	20	41,3
19	y	24,7	30,8	15,9	29	15,4	28,6	15,6	27,7	34,1	37,7
	x_1	45,3	56,4	37	67,5	37	69	40	69,1	68,1	75,3
	x_2	20,6	29,7	17,8	43,5	17,8	42,4	20	41,3	35,4	41,4
20	y	15,9	29	15,4	28,6	15,6	27,7	34,1	37,7	41,9	24,4
	x_1	37	67,5	37	69	40	69,1	68,1	75,3	83,7	48,7
	x_2	17,8	43,5	17,8	42,4	20	41,3	35,4	41,4	48,5	22,3
21	y	15,4	28,6	15,6	27,7	34,1	37,7	41,9	24,4	21,3	36,7
	x_1	37	69	40	69,1	68,1	75,3	83,7	48,7	39,9	68,6
	x_2	17,8	42,4	20	41,3	35,4	41,4	48,5	22,3	18	35,5
22	y	15,6	27,7	34,1	37,7	41,9	24,4	21,3	36,7	21,5	26,4
	x_1	40	69,1	68,1	75,3	83,7	48,7	39,9	68,6	39	48,6
	x_2	20	41,3	35,4	41,4	48,5	22,3	18	35,5	20	31
23	y	34,1	37,7	41,9	24,4	21,3	36,7	21,5	26,4	53,9	34,2
	x_1	68,1	75,3	83,7	48,7	39,9	68,6	39	48,6	98	68,5
	x_2	35,4	41,4	48,5	22,3	18	35,5	20	31	56	30,7
24	y	41,9	24,4	21,3	36,7	21,5	26,4	53,9	34,2	35,6	34
	x_1	83,7	48,7	39,9	68,6	39	48,6	98	68,5	71,1	68
	x_2	48,5	22,3	18	35,5	20	31	56	30,7	36,2	41
25	y	21,3	36,7	21,5	26,4	53,9	34,2	35,6	34	19	46,6
	x_1	39,9	68,6	39	48,6	98	68,5	71,1	68	38	93,2
	x_2	18	35,5	20	31	56	30,7	36,2	41	19	49,5
26	y	21,5	26,4	53,9	34,2	35,6	34	19	46,6	58,5	24,2
	x_1	39	48,6	98	68,5	71,1	68	38	93,2	117	42
	x_2	20	31	56	30,7	36,2	41	19	49,5	55,2	21
27	y	53,9	34,2	35,6	34	19	46,6	58,5	24,2	35,7	51,2
	x_1	98	68,5	71,1	68	38	93,2	117	42	62	89

Варіант	Змінна	Значення									
	x_2	56	30,7	36,2	41	19	49,5	55,2	21	35	52,3
28	y	35,6	34	19	46,6	58,5	24,2	35,7	51,2	75,9	21,2
	x_1	71,1	68	38	93,2	117	42	62	89	132	40,8
	x_2	36,2	41	19	49,5	55,2	21	35	52,3	89,6	19,2
29	y	19	46,6	58,5	24,2	35,7	51,2	75,9	21,2	30,8	34
	x_1	38	93,2	117	42	62	89	132	40,8	59,2	65,4
	x_2	19	49,5	55,2	21	35	52,3	89,6	19,2	31,9	38,9
30	y	58,5	24,2	35,7	51,2	75,9	21,2	30,8	34	31,9	43,6
	x_1	117	42	62	89	132	40,8	59,2	65,4	60,2	82,2
	x_2	55,2	21	35	52,3	89,6	19,2	31,9	38,9	36,3	49,7

Основні формули та поняття до завдання 2

Теоретична лінійна двофакторна регресійна модель:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \qquad (2.1)$$

де β_1, β_2 – теоретичні коефіцієнти регресії; β_0 – вільний член.

Емпірична лінійна двофакторна регресійна модель:

$$\tilde{y} = b_0 + b_1 x_1 + b_2 x_2 + e, \qquad (2.2)$$

де b_0, b_1, b_2 – оцінки теоретичних коефіцієнтів регресії β_j ($j = 0,1,2$).

Оцінки параметрів лінійної двофакторної регресійної моделі

Для *і-го* спостереження маємо:

$$\tilde{y}_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i = \hat{y}_i + e_i. \qquad (2.3)$$

Параметри b_0, b_1, b_2 необхідно знайти так, щоб середнє квадратичне відхилення фактичних даних y_i від теоретичних $\hat{y}_i$ було мінімальним

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2}))^2 \to \min \qquad (n > 3). \ (2.4)$$

Для того, щоб знайти мінімум виразу (2.4), необхідно прирівняти до нуля часткові похідні за аргументами b_0, b_1, b_2:

$$\frac{\partial}{\partial b_0}\sum_{i=1}^{n} e_i^{\,2} = -2\sum_{i=1}^{n}\left(y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2})\right) = 0$$

$$\frac{\partial}{\partial b_1}\sum_{i=1}^{n} \varepsilon_i^{\,2} = -2\sum_{i=1}^{n}\left(y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2})\right)x_{i1} = 0.$$

$$\frac{\partial}{\partial b_2}\sum_{i=1}^{n} \varepsilon_i^{\,2} = -2\sum_{i=1}^{n}\left(y_i - (b_0 + b_1 x_{i1} + b_2 x_{i2})\right)x_{i2} = 0 \qquad (2.5)$$

З (2.5) отримаємо систему лінійних відносно параметрів b_0, b_1, b_2 рівнянь:

$$\begin{cases} n b_0 + b_1\sum\limits_{i=1}^{n} x_{i1} + b_2\sum\limits_{i=1}^{n} x_{i2} = \sum\limits_{i=1}^{n} y_i \\ b_0\sum\limits_{i=1}^{n} x_{i1} + b_1\sum\limits_{i=1}^{n} x_{i1}^2 + b_2\sum\limits_{i=1}^{n} x_{i2}x_{i1} = \sum\limits_{i=1}^{n} y_i x_{i1} \\ b_0\sum\limits_{i=1}^{n} x_{i2} + b_1\sum\limits_{i=1}^{n} x_{i1}x_{i2} + b_2\sum\limits_{i=1}^{n} x_{i2}^2 = \sum\limits_{i=1}^{n} y_i x_{i2} \end{cases} \qquad (2.6)$$

За допомогою матриць:

$$Z = \begin{pmatrix} n & \sum\limits_{i=1}^{n} x_{i1} & \sum\limits_{i=1}^{n} x_{i2} \\ \sum\limits_{i=1}^{n} x_{i1} & \sum\limits_{i=1}^{n} x_{i1}^2 & \sum\limits_{i=1}^{n} x_{i1}x_{i2} \\ \sum\limits_{i=1}^{n} x_{i2} & \sum\limits_{i=1}^{n} x_{i2}x_{i1} & \sum\limits_{i=1}^{n} x_{i2}^2 \end{pmatrix}, \quad B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}, \quad G = \begin{pmatrix} \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} y_i x_{i1} \\ \sum\limits_{i=1}^{n} y_i x_{i2} \end{pmatrix} \qquad (2.7)$$

систему (2.6) можна записати і розв'язати у матричній формі (за умови лінійної незалежності векторів X_1, X_2):

$$Z \cdot B = G \;\Rightarrow\; B = Z^{-1}G, \qquad (2.8)$$

де Z^{-1} – матриця, обернена до матриці Z.

Якщо представити матрицю Z та її визначник Δ у вигляді:

$$Z = \begin{pmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{pmatrix},$$

$$\Delta = \det Z = \begin{vmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{23} \\ z_{31} & z_{32} & z_{33} \end{vmatrix} =$$

$$= z_{11}z_{22}z_{33} + z_{12}z_{23}z_{31} + z_{21}z_{32}z_{13} -$$

$$-(z_{31}z_{22}z_{13} + z_{21}z_{12}z_{33} + z_{32}z_{23}z_{11}), \qquad (2.9)$$

то при $\Delta \neq 0$ обернену матрицю Z^{-1} можна обчислити за формулами

$$Z^{-1} = \frac{1}{\Delta}\begin{pmatrix} Z_{11} & Z_{21} & Z_{31} \\ Z_{12} & Z_{22} & Z_{32} \\ Z_{13} & Z_{23} & Z_{33} \end{pmatrix} = \begin{pmatrix} Z_{11}/\Delta & Z_{21}/\Delta & Z_{31}/\Delta \\ Z_{12}/\Delta & Z_{22}/\Delta & Z_{32}/\Delta \\ Z_{13}/\Delta & Z_{23}/\Delta & Z_{33}/\Delta \end{pmatrix} = \begin{pmatrix} z'_{11} & z'_{21} & z'_{31} \\ z'_{12} & z'_{22} & z'_{32} \\ z'_{13} & z'_{23} & z'_{33} \end{pmatrix}, \qquad (2.10)$$

де Z_{ij} – алгебраїчне доповнення елемента z_{ij}:

$$Z_{11} = \begin{vmatrix} z_{22} & z_{23} \\ z_{32} & z_{33} \end{vmatrix} = z_{22}z_{33} - z_{32}z_{23}, \quad Z_{21} = -\begin{vmatrix} z_{12} & z_{13} \\ z_{32} & z_{33} \end{vmatrix}, \quad Z_{31} = \begin{vmatrix} z_{12} & z_{13} \\ z_{22} & z_{23} \end{vmatrix},$$

$$Z_{12} = -\begin{vmatrix} z_{21} & z_{23} \\ z_{31} & z_{33} \end{vmatrix}, \quad Z_{22} = \begin{vmatrix} z_{11} & z_{13} \\ z_{31} & z_{33} \end{vmatrix}, \quad Z_{32} = -\begin{vmatrix} z_{11} & z_{13} \\ z_{21} & z_{23} \end{vmatrix},$$

$$Z_{13} = \begin{vmatrix} z_{21} & z_{22} \\ z_{31} & z_{32} \end{vmatrix}, \quad Z_{23} = -\begin{vmatrix} z_{11} & z_{12} \\ z_{31} & z_{32} \end{vmatrix}, \quad Z_{33} = \begin{vmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{vmatrix}. \qquad (2.11)$$

Дисперсії коефіцієнтів $\mathbf{b_j}$ обчислюються за формулами:

$$S_{b_0}^2 = S^2 z'_{11}, \quad S_{b_1}^2 = S^2 z'_{22}, \quad S_{b_2}^2 = S^2 z'_{33}, \qquad (2.12)$$

де z'_{jj} – діагональний елемент матриці Z^{-1};

$S^2 = \sum e_i^2 / (n - m - 1)$ – незміщена оцінка дисперсії відхилень;

$m = 2$ – кількість пояснюючих змінних (незалежних факторів).

t-тест Стьюдента для перевірки значущості параметрів b_0, b_1, b_2 лінійного рівняння багатофакторної регресії

1. Для кожного параметра b_j розрахувати t-відношення: $t_j = b_j / S_{b_j}$.

2. Для заданого рівня значущості α, числа пояснюючих змінних $m = 2$ із таблиць критичних точок розподілу Стьюдента знайти $t_{\alpha/2,\, n-m-1}$.

3. Якщо $|t_{b_j}| > t_{\alpha/2,\, n-m-1}$ – коефіцієнт b_j є статистично значущим.

Перевірка загальної якості рівняння регресії

Коефіцієнт детермінації $\mathbf{R^2}$ є характеристикою загального впливу всіх незалежних змінних на залежну і розраховується за формулами:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum e_i^2}{\sum (y_i - \bar{y})^2},\ R^2 \in [0, 1]. \quad (2.13)$$

Індекс кореляції (множинний коефіцієнт кореляції) характеризує тісноту зв'язку всіх незалежних змінних із залежною:

$$R = \sqrt{R^2},\ R \in [0, 1]. \quad (2.14)$$

Аналіз статистичної значущості коефіцієнта детермінації

1. Розрахувати F-статистику:

$$F = \frac{R^2 / m}{(1 - R^2)/(n - m - 1)}, \quad (2.15)$$

де $m = 2$ – кількість незалежних змінних.

2. Для заданого рівня значущості α, числа пояснюючих змінних $m = 2$ із таблиць критичних точок розподілу Фішера знайти $F_{кр}(\alpha;\, m,\, n - m - 1)$.

3. Якщо $F > F_{кр}$, то R^2 є статистично значущим, рівняння якісно описує зв'язок між залежною і незалежними змінними.

Прогнозування за допомогою двофакторної лінійної регресії

Точковий прогноз одержати підстановкою прогнозних значень факторів, заданих вектором $X_p = (1, x_{p1}, x_{p2})$, у рівняння регресії:

$$\hat{y}_p = b_0 + b_1 x_{p1} + b_2 x_{p2} \quad \Leftrightarrow \quad \hat{y}_p = X_p B. \quad (2.16)$$

Дисперсія точкового прогнозу обчислюється за формулою:

$$D(\hat{y}_p) \approx S^2_{\hat{y}_p} = S^2 X_p Z^{-1} X_p^T, \tag{2.17}$$

де $S^2 = \sum e_i^2 / (n - m - 1)$; X_p^T – вектор-стовпець, транспонований до вектора X_p.

Інтервали довіри регресії та прогнозу

Інтервал довіри *для теоретичної регресії*:

$$\hat{y}_p - t_{\alpha/2, n-m-1} S \sqrt{X_p Z^{-1} X_p^T} \le \hat{y}_p' \le \hat{y}_p + t_{\alpha/2, n-m-1} S \sqrt{X_p Z^{-1} X_p^T}. \tag{2.18}$$

Інтервал довіри *для індивідуального значення* y_p:

$$\hat{y}_p - t_{\alpha/2, n-m-1} S_p \le y_p \le \hat{y}_p + t_{\alpha/2, n-m-1} S_p,$$

$$S_p = S\sqrt{1 + X_p Z^{-1} X_p^T}. \tag{2.19}$$

Матриця парних коефіцієнтів кореляції (кореляційна матриця) у випадку двох незалежних факторів має вигляд:

$$r = \begin{pmatrix} 1 & r_{y1} & r_{y2} \\ r_{y1} & 1 & r_{12} \\ r_{y2} & r_{12} & 1 \end{pmatrix}, \tag{2.20}$$

де r_{y1}, r_{y2} – парні коефіцієнти кореляції між залежною змінною y і незалежними змінними x_1, x_2 відповідно;

r_{12} – коефіцієнт кореляції між незалежними змінними x_1, x_2:

$$r_{y1} = \frac{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_{i1} - \bar{x}_1)^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}},$$

$$r_{y2}=\frac{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2\sum_{i=1}^{n}(y_i-\bar{y})^2}},$$

$$r_{12}=\frac{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2}},\quad \bar{x}_1=\frac{1}{n}\sum_{i=1}^{n}x_{i1},\quad \bar{x}_2=\frac{1}{n}\sum_{i=1}^{n}x_{i2}.$$

Матриця частинних коефіцієнтів кореляції

$$r^*=\begin{pmatrix}1 & r_{y1.2} & r_{y2.1}\\ r_{y1.2} & 1 & r_{12.y}\\ r_{y2.1} & r_{12.y} & 1\end{pmatrix}, \tag{2.21}$$

де $r_{y1.2}=\dfrac{r_{y1}-r_{y2}r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}}$,

$$r_{y2.1}=\frac{r_{y2}-r_{y1}r_{12}}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}},$$

$$r_{12.y}=\frac{r_{12}-r_{y1}r_{y2}}{\sqrt{(1-r_{y1}^2)(1-r_{y2}^2)}}.$$

Частинні коефіцієнти кореляції характеризують тісноту зв'язку між двома змінними за умови, що інші змінні сталі.

Якщо значення $|r_{12.y}|$ близьке до 1, це вказує на мультиколінеарність факторів x_1, x_2. Для перевірки наявності мультиколінеарності необхідно:

1) розрахувати t-відношення: $t_{12}=\dfrac{r_{12.y}\sqrt{n-k}}{\sqrt{1-r_{12.y}^2}}$,

де $k = 3$ – кількість параметрів (b_0, b_1, b_2);

2) із таблиць критичних точок розподілу Стьюдента знайти $t_{\alpha/2,\, n-k}$ (α – заданий рівень значущості, $n-k$ – кількість ступенів свободи);

3) якщо $|t_{12}| > t_{\alpha/2,\, n-k}$ – коефіцієнт $r_{12,y}$ є статистично значущим, між пояснюючими змінними x_1, x_2 існує мультиколінеарність.

Оцінка доцільності включення в рівняння множинної регресії фактора x1 після фактора x2 за *F*-критерієм Фішера:

1. Розрахувати *F*-відношення: $F_{x_1} = (R^2 - r_{y2}^2)(n-k)/(1-R^2)$,

де $k = 3$ – кількість параметрів (b_0, b_1, b_2).

2. З таблиць критичних точок розподілу Фішера знайти $F_{кр}(\alpha; 1, n-k)$.

3. Якщо $F_{x_1} > F_{кр}$ – гіпотеза про доцільність включення фактора x_1 після фактора x_2 приймається. Аналогічно перевіряється гіпотеза про доцільність включення фактора x_2 після фактора x_1.

Середні частинні коефіцієнти еластичності показують, на скільки відсотків зміниться середнє значення показника y, якщо середнє значення одного з факторів зміниться на один відсоток при незмінних значеннях інших факторів і обчислюються за формулами:

$$Э_{yx_1} = b_1 \overline{x_1} / \overline{y}, \quad Э_{yx_2} = b_2 \overline{x_2} / \overline{y}. \qquad (2.22)$$

Якщо $|Э_{yx_2}| < |Э_{yx_1}|$ – фактор x_1 сильніше впливає на y, ніж x_2.

Перевірка наявності автокореляції за методом Дарбіна-Уотсона

Проводиться за тим же алгоритмом, що і для парної регресії (див. основні формули та поняття до завдання 1). Для множинної регресії нижня d_l та верхня d_u критичні межі визначаються за таблицею критичних точок Дарбіна-Уотсона для заданого рівня значущості α, числа спостережень n та числа пояснювальних змінних $m = 2$.

ЛІТЕРАТУРА

1. Наконечний С.І., Терещенко Т.О., Романюк Т.П Економетрія: Підручник / С.І. Наконечний, Т.О. Терещенко, Т.П.Романюк– К.: КНЕУ, 2006. -528 с.
2. Жлуктенко В.І., Водзянова Н.К., Савіна С.С., Колодінська Економетрія: Навч. посіб. / В.І. Жлуктенко, Н.К. Водзянова, С.С. Савіна, О.В. Колодінська — К.: Вид-во Європ. Ун-ту, 2005. — 552 с.
3. Назаренко О.М. Основи економетрики: Підручник / О.М.Назаренко– К.: «Центр навчальної літератури», 2005. – 392 с.
4. Винн Р., Холден К. Введение в прикладной эконометрический аналіз / Р. Винн, К.Холден - М.: Финансы и статистика, 2001. - 294с.
5. Грубер Й. Эконометрия: Том 1. Введение в эконометрию Том 2. Эконометрические прогнозные и оптимизационные модели / Й. Грубер - Киев: Астарта, 2006. – 397 с.
6. Гурова К.Д., Сивый В.Б. Эконометрия / Учебно-методическое пособие для самостоятельной работы студентов и преподавателей / К.Д. Гурова, В.Б. Сивый– Харьков: Константа, 2007. – 92 с.
7. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учебник / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий– М.: Дело, 2006. – 400с.
8. Долгіх В.М., Я.В. Долгіх. Методичні вказівки та завдання для контрольної роботи / В.М. Долгіх, Я.В. Долгіх. – Суми: УАБС НБУ, 2005. – 28 с.