Назва дисципліни: ЕКОНОМЕТРИКА

Тема 5:

МУЛЬТИКОЛІНЕАРНІСТЬ

Лектор: **к.е.н., доцент кафедри економетрії та статистики ДЕМЧИШИН М.Я.**

Навчальна мета:

Після вивчення теми студент повинен знати :

- поняття **мультиколінеарності**;
- **методи оцінки** ступеня мультиколінеарності;
- **вплив мультиколінеарності** на характеристики економетричної моделі;
- **методи усунення** мультиколінеарності;
- **алгоритм Фаррара-Глобера**.

План лекції

1. Поняття, ознаки і наслідки існування мультиколінеарності

2. Виявлення мультиколінеарності в економетричній моделі. Алгоритм Фаррара-Глобера

3. Способи усунення мультиколінеарності. Метод головних компонент

Основна література

- Єлейко В. Основи економетрії. – Львів: Марка Лтд, 1995. -191 с.
- Єлейко В.І., Копич І.М., Боднар Р.Д., Демчишин М.Я. Економетрія: Навч.посібн. – Львів: вид-во Львівської комерційної академії, 2007. – 420 с.
- Корольов О.А. Економетрія: Лекції, питання, тести, задачі, ситуації, проблеми: Навч. посібник. -К.: КДТЕУ, 2000. - 724 с.
- Лещинський О.Л. Економетрія: Навч. посібн. для студ. вищ. навч. закл. / О.Л. Лещинський, В.В. Рязанцева, О.О. Юнькова. – К.: МАУП, 2003. – 208 с.
- Лук'яненко І.Г., Краснікова Л.І. Економетрика: Підручник. - К.: Т-во "Знання", КОО, 1998.-494 с.
- Наконечний С.І., Терещенко Т.О. Економетрія: Навч. посіб. для сам ост. вивч. диск. - К.: КНЕУ, 2001.- 192 с.
- Толбатов Ю.А. Економетрика: Підручник. – К.: Четверта хвиля, 1997. – 320 с.

Питання 1. Поняття, ознаки і наслідки існування мультиколінеарності

- ***Мультиколінеарність*** означає існування тісної лінійної залежності, або кореляції, між двома чи більше пояснювальними змінними***.***

Ознаки мультиколінеарності

1. Якщо серед **парних коефіцієнтів кореляції пояснювальних змінних** є такі, рівень яких наближається до **множинного коефіцієнта кореляції** або дорівнює йому, це свідчить про можливість існування мультиколінеарності.

Інформацію про парну залежність може дати **симетрична матриця коефіцієнтів парної кореляції**, або, як її ще називають, **матриця кореляції нульового порядку (*p* - кількість пояснювальних змінних)**:

$$r = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} & \cdots & r_{x_1x_p} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} & \cdots & r_{x_2x_p} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} & \cdots & r_{x_3x_p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_px_1} & r_{x_px_2} & r_{xpx_3} & \cdots & r_{x_px_p} \end{pmatrix}$$

!!! Явище мультиколінеарності в жодному разі не зводиться лише до існування парної кореляції між пояснюючими змінними.

2. Загальніша перевірка передбачає застосування визначника (детермінанта) матриці *r*, який називається детермінантом кореляції і позначається $|r|$

Числові значення детермінанта кореляції містяться на інтервалі $|r| \in [0;1]$

$|r| = 0$ - існує повна мультиколінеарність.

$|r| = 1$ - мультиколінеарність відсутня.

Чим ближче $|r|$ до нуля, тим певніше можна стверджувати, що між пояснюючими змінними існує мультиколінеарність.

Незважаючи на те, що числове значення $|r|$ зазнає впливу дисперсії пояснюючих змінних, цей показник можна вважати **точковою мірою рівня мультиколінеарності**.

3. Коли коефіцієнт детермінації R^2, який обчислено для регресійних залежностей між однією пояснюючою змінною та іншими такими змінними, близький до **одиниці,** то можна говорити про наявність мультиколінеарності.

4. Якщо $F_{k\,факт} > F_{табл}$, то x_k мультиколінеарна з іншими, тобто залежить від інших незалежних змінних і треба вирішити питання про вилучення з переліку змінних однієї з них.

5. Якщо $t_{kj\,факт} > t_{табл}$, то x_k і x_j тісно пов'язані між собою.

Аналізуючи ***F***- і ***t***-критерії, можна зробити висновок, яку із змінних треба вилучити з розгляду в побудованій моделі для усунення мультиколінеарності (треба при цьому виходити і з економіко-логіко-теоретичних міркувань).

Основні наслідки мультиколінеарності в економетричній моделі:

- *Зміщення оцінок параметрів моделі*, обчислених методом найменших квадратів.
- *Збільшення дисперсії* та коваріації оцінок параметрів моделі, обчислених методом найменших квадратів, і, відповідно, *збільшення довірчих інтервалів для параметрів і зниження точності їх оцінювання*.
- *Незначущість параметрів моделі*, обчислених методом найменших квадратів.
- *Чутливість оцінок параметрів моделі до сукупності спостережень,* зокрема до обсягу вибірки.

Питання 2.

Виявлення мультиколінеарності в економетричній моделі.

Алгоритм Фаррара-Глобера

ПОКРОКОВИЙ АЛГОРИТМ ФАРРАРА-ГЛОБЕРА

Алгоритм містить **три види статистичних критеріїв**, згідно з якими перевіряється відповідно мультиколінеарність:

- а) усього масиву пояснюючих змінних (**$\chi 2$-критерій** “хі-квадрат”);
- б) кожної пояснюючої змінної з рештою пояснюючих змінних (**F-критерій**);
- в) кожної пари пояснюючих змінних (**t-критерій**).

Усі ці критерії при порівнянні з їх критичними значеннями дають змогу зробити конкретні висновки щодо наявності чи відсутності мультиколінеарності пояснюючих змінних.

Крок 1. НОРМАЛІЗАЦІЯ ЗМІННИХ

Нехай $x_1, x_2, x_3, ..., x_m$ – вектори пояснювальних змінних економетричної моделі.

Елементи стандартизованих векторів обчислимо за формулою:

$$x_{ik}^{\ *} = \frac{x_{ik} - \overline{x}_k}{\sigma_{x_k}}$$

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

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

$\overline{x}_k$ – середня арифметична *k*-ї пояснювальної змінної;

$\sigma_{x_k}^{\ 2}$ - дисперсія *k*-ї пояснювальної змінної.

Крок 2. Знаходження КОРЕЛЯЦІЙНОЇ МАТРИЦІ 1-го порядку

$$r = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_k} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} & \dots & r_{x_2x_k} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} & \dots & r_{x_3x_k} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} & \dots & r_{x_kx_k} \end{pmatrix}$$

Матриця складається з **парних коефіцієнтів кореляції**, які вказують на **щільність кореляційного зв'язку** між **факторними ознаками.**

Знаходиться матриця покроково:

- знайти матрицю $r^* = X^{*T} X^*$, де X^* – матриця нормалізованих пояснювальних змінних, а X^{*T} – матриця, транспонована до матриці X^*;
- кореляційна матриця r – результат ділення r^* на n.

Крок 3.

1. Визначити $|r|$ – детермінант кореляції - **ВИЗНАЧНИК КОРЕЛЯЦІЙНОЇ МАТРИЦІ *r***

2 . Обчислити **χ2-КРИТЕРІЙ** (“хі-квадрат”):

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

3. Значення цього критерію порівнюється з табличним за таких умов: $\frac{1}{2}m(m-1)$ ступенів свободи і рівні значущості α.

Якщо $\chi_{факт}{}^2 < \chi_{табл}{}^2$, то в масиві незалежних змінних мультиколінеарність відсутня.

Крок 4. ВИЗНАЧЕННЯ МАТРИЦІ ПОМИЛОК С

Матриця помилок С – це матриця, обернена до кореляційної матриці *r*:

$$C = r^{-1}$$

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}$$

Крок 5.

1. **РОЗРАХУНОК F-КРИТЕРІЇВ**:

$$F_{k\,факт} = (c_{kk} - 1)\frac{n-m}{m-1}$$

де c_{kk} – *діагональні елементи* матриці **С,** тобто $\mathbf{c_{11}}$, $\mathbf{c_{22}}$, $\mathbf{c_{33}}$ і т.д.

2. Фактичні значення критеріїв порівнюються з табличним при (***m***-**1**) і (***n-m***) ступенях свободи і рівні значущості α

   Якщо $F_{k\,факт} > F_{табл}$, відповідна ***k***-та пояснююча змінна мультиколінеарна з іншими.

3. Обчислити ***коефіцієнти детермінації*** для кожної змінної:

$$R_k^2 = 1 - \frac{1}{c_{kk}}$$

Крок 6.

Знаходження ЧАСТИННИХ КОЕФІЦІЄНТІВ КОРЕЛЯЦІЇ

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

де c_{kj} – елементи матриці *C*, що містяться 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 порівнюються з **табличним** за ***n-m*** **ступенів свободи** і рівня значущості α.

Якщо $t_{kj} > t_{табл}$, між змінними x_k і x_j існує мультиколінеарність.

Питання 3.

Способи усунення мультиколінеарності.

Метод головних компонентів

Методи усунення мультиколінеарності

- Використання додаткової або первинної інформації
- Об’єднання інформації.
- Відкидання змінної з високою кореляцією.
- Перетворення даних (використання перших різниць).
- Збільшення кількості спостережень

- *Метод головних компонентів* застосовується для оцінювання параметрів регресійних моделей з великою кількістю факторних змінних у випадку, коли ці фактори мають однакові одиниці вимірювання.

- *Суть методу головних компонентів* полягає в заміні сукупності факторних змінних на нові змінні, які між собою були б попарно некорельованими і впорядкованими в порядку спадання їх дисперсій.

Алгоритм методу головних компонентів

Крок 1.
Нормалізація факторних змінних

Нехай $x_1, x_2, x_3, \ldots, x_p$ – вектори пояснювальних змінних економетричної моделі.

Елементи стандартизованих векторів обчислимо за формулою:

$$x^*_{ji} = \frac{x_{ji} - \bar{x}_j}{\sigma_{x_j}}$$

i=1, 2, …, *n*; *j*=1, 2, …, *p*,

де ***n*** – кількість спостережень;

p – кількість факторів у моделі,

$\bar{x}_j$ – середнє арифметичне *j*-го фактора;

σ_{x_j} - середньоквадратичне відхилення *j*-ої факторної змінної. .

Крок 2.
Знаходження матриці нормованих факторних змінних

Матриця нормованих факторних змінних X^* має вигляд:

$$X^* = \begin{pmatrix} x_{11}^* & x_{21}^* & \dots & x_{p1}^* \\ x_{12}^* & x_{22}^* & \dots & x_{p2}^* \\ \dots & \dots & \dots & \dots \\ x_{1n}^* & x_{2n}^* & \dots & x_{pn}^* \end{pmatrix}$$

Крок 3.

Знаходження кореляційної матриці

Кореляційна матриця або матриця моментів нормалізованої системи нормальних рівнянь ***r*** знаходиться за формулою:

$$r = \frac{1}{n}\left(X^{*T} X^{*}\right)$$

де X^{*T}– матриця, транспонована до матриці .

Кореляційна матриця матиме вигляд:

,

$$r = \begin{pmatrix} 1 & r_{21} & r_{31} & \dots & r_{p1} \\ r_{12} & 1 & r_{32} & \dots & r_{p2} \\ r_{13} & r_{23} & 1 & \dots & r_{p3} \\ \dots & \dots & \dots & \dots & \dots \\ r_{1p} & r_{2p} & r_{3p} & \dots & 1 \end{pmatrix}$$

де $r_{ij} = r_{x_i x_j}$ – парні коефіцієнти кореляції, *i*≠*j*, *i*=1, 2, …, *n*; *j*=1, 2, …, *p*.

Крок 4.

Знаходження власних чисел кореляційної матриці

Власні числа $\lambda_1, \lambda_2, \ldots, \lambda_p$ є розв'язками рівняння

$$\left| r - \lambda I \right| = 0$$

де *I* – одинична матриця розмірності $p \times p$.

Крок 5.

Ранжування власних чисел кореляційної матриці

Впорядкуємо власні числа λ_1 , λ_2 , … , λ_p

в порядку спадання їх абсолютних значень.

.

Крок 6.

Знаходження власних векторів кореляційної матриці

З системи рівнянь

$$|r - \lambda I| a = 0$$

знаходимо власні вектори a_i, *i* = 1, 2, …, *p*,
при умові, що виконується співвідношення

$$a_i^T a_j = \begin{cases} 1, & \text{якщо} \quad i = j; \\ 0, & \text{якщо} \quad i \neq j, \end{cases}$$

де a_i^T – вектор, транспонований до a_i.

Крок 7.

Знаходження головних компонентів власних векторів кореляційної матриці

Головні компоненти знаходимо зі співвідношення

$$Z_i = X^* a_i \qquad i = 1, 2, \ldots, p$$

Вони повинні задовольняти наступні умови

$$\sum_{j=1}^{n} z_{ij} = 0 \qquad i = 1, 2, \ldots, p;$$

$$\frac{1}{n} Z_i^T Z_i = \lambda_i \qquad i = 1, 2, \ldots, p;$$

$$Z_k^T Z_i = 0 \qquad i \neq k, \quad i, k = 1, 2, \ldots, p.$$

Крок 8.

Знаходження параметрів економетричної моделі з головними компонентами

Якщо ввести позначення

$$x_0 = \begin{pmatrix} x_{01} \\ x_{02} \\ \dots \\ x_{0n} \end{pmatrix} \quad Z_1 = \begin{pmatrix} z_{11} \\ z_{12} \\ \dots \\ z_{1n} \end{pmatrix} \quad Z_2 = \begin{pmatrix} z_{21} \\ z_{22} \\ \dots \\ z_{2n} \end{pmatrix} \quad \cdots \quad Z_p = \begin{pmatrix} z_{p1} \\ z_{p2} \\ \dots \\ z_{pn} \end{pmatrix}$$

то регресійне рівняння з головними компонентами

$$x_0 = \widetilde{b}_1 Z_1 + \widetilde{b}_2 Z_2 + \dots + \widetilde{b}_p Z_p$$

у матричному вигляді можна записати як

$$x_0 = Z\widetilde{b}$$

Крок 8.

Знаходження параметрів економетричної моделі з головними компонентами

Невідомі параметри $\tilde{b}_1$, $\tilde{b}_2$, …, $\tilde{b}_p$ знаходяться із співвідношення

$$\tilde{b} = Z^{-1} x_0$$

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

Крок 9.

Визначення параметрів вихідної моделі

Для моделі

$$x_0 = Xb$$

де

$$X = \begin{pmatrix} x_{11} & x_{21} & \dots & x_{p1} \\ x_{12} & x_{22} & \dots & x_{p2} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{pn} \end{pmatrix} \quad \text{і} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_p \end{pmatrix},$$

невідомі параметри b_1, b_2, …, b_p знаходяться із співвідношення

$$b = A\tilde{b}$$

де матриця *A* складається з власних векторів a_i, *i* = 1, 2, …, *p*:

$$A = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{p1} \\ a_{12} & a_{22} & \dots & a_{p2} \\ \dots & \dots & \dots & \dots \\ a_{1p} & a_{2p} & \dots & a_{pp} \end{pmatrix}$$