

6 МУЛЬТИКОЛІНЕАРНІСТЬ

При побудові структури регресії, з одного боку, потрібно включити в регресію всі фактори, які мають суттєвий статистичний вплив на показник, а з іншого боку, потрібно, щоб була виконана умова лінійної незалежності між факторами. Якщо існує лінійна залежність хоча б між двома факторами, то говорять, що між цими факторами існує *мультиколінеарність*. Якщо між факторами x_i і x_j існує лінійна залежність $x_i = kx_j$, то говорять, що між цими факторами існує *строга мультиколінеарність*. Отже, *мультиколінеарність* – явище, при якому між факторними ознаками існує зв'язок близький до функціонального.

Враховуючи, що x_i і x_j – випадкові величини то, як правило, між ними існує приблизна лінійна залежність $x_i = kx_j + \varepsilon$, де ε – відхилення. В таких випадках *мультиколінеарність* між факторами *нестрога*. Оскільки одним із основних припущень МНК є відсутність лінійної залежності між факторами, то постає питання: як позбутися мультиколінеарності. Якщо в регресії присутнє явище мультиколінеарності, то не виконується умова $\det(X^T X) \neq 0$, і при строгій мультиколінеарності неможливо отримати оцінки параметрів МНК. Якщо мультиколінеарність не строга, то отримані оцінки параметрів мало надійні. В цьому випадку незначні зміни вибірових даних призводять до значних змін оцінок параметрів. Очевидно, що у випадку, коли $\det(X^T X)$ близький до нуля, то в регресії присутнє явище мультиколінеарності.

Взагалі при визначенні лінійної структури корисно будувати кореляційну матрицю, в яку включені фактори і показник, і в першу чергу у регресію потрібно включати фактори, які корегують з показником і не корегують між собою. Якщо зв'язок між незалежними змінними не функціональний, проте статистично істотний, то незважаючи на те, що теоретично допустима оцінка параметрів методом найменших квадратів, вона може привести до таких помилок параметрів, що сама модель стає беззмістовною. Тому, будуючи економетричну модель, яка повинна кількісно описати залежність економічних процесів та явищ, потрібно мати інформацію про те, що між незалежними змінними не існує колінеарності.

Якщо серед парних коефіцієнтів кореляції незалежних змінних є такі, рівень яких наближається або дорівнює множинному коефіцієнту кореляції, то це свідчить про можливість існування мультиколінеарності.

При двофакторній моделі для оцінки наявності мультиколінеарності достатньо розрахувати парний лінійний коефіцієнт кореляції. При багатфакторній моделі високі коефіцієнти кореляції ($r_{x_i x_j} > 0,8$) – достатня, але не необхідна умова наявності мультиколінеарності. ($r_{x_i x_j} \rightarrow 1$).

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:

$$r = \begin{pmatrix} r_{X_1 X_1} & r_{X_1 X_2} & \dots & r_{X_1 X_k} \\ r_{X_2 X_1} & r_{X_2 X_2} & \dots & r_{X_2 X_k} \\ \dots & \dots & \dots & \dots \\ r_{X_k X_1} & r_{X_k X_2} & \dots & r_{X_k X_k} \end{pmatrix}.$$

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.

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

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.

Таким чином природа мультиколінеарності полягає у неможливості статистично оцінити і обґрунтувати вплив кожної пояснюючої змінної на

залежну зміну моделі, що, в свою чергу, робить ненадійною економічну інтерпретацію оціненого рівняння регресії.

Це в свою чергу призводить до наступних негативних практичних наслідків:

1. Збільшення інтервалів довіри параметрів моделі.

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

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

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

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 $X_1, X_2, \dots, X_m$. Елементи стандартизованих векторів розрахуємо за формулою:

$$X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{\sqrt{n\sigma_{X_k}^2}}, \quad i = \overline{1, n}, \quad k = \overline{1, m},$$

де n – число спостережень;

m – число незалежних змінних;

$\bar{X}_k$ – середня арифметична k -ї незалежної змінної;

$\sigma_{X_k}^2$ – дисперсія k -ї незалежної змінної.

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

$$R = X^{*T} X^*,$$

де X^* – матриця стандартизованих незалежних змінних;

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

Крок 3. Дослідження загальної мультиколінеарності.

Використовується тест χ^2 . Для цього треба обчислити визначник кореляційної матриці $|r|$ і знайти розрахункове значення

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

Для заданої довірчої ймовірності γ та $k = \frac{m(m-1)}{2}$ – що є числом ступенів вільності, треба знайти табличне значення $\chi^2_{k \text{ табл}}$. Якщо $\chi^2 \leq \chi^2_{k \text{ табл}}$, то із заданою надійністю γ можна вважати, що загальна мультиколінеарність відсутня і на цьому закінчується дослідження мультиколінеарності. Якщо $\chi^2 > \chi^2_{k \text{ табл}}$, то з прийнятою ймовірністю γ можна вважати, що між факторами існує мультиколінеарність

Крок 4. 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 $F_{k \text{ факт}} > F_{k \text{ табл}}$, то відповідна k -та незалежна змінна мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної розраховується таким чином:

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

Крок 5. Знаходження часткових коефіцієнтів кореляції:

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}, \quad k = \overline{1, m}, \quad j = \overline{1, m},$$

де c_{kj} – елемент матриці C , що заходиться в k -му рядку і j -му стовпці, c_{kk} і c_{jj} – діагональні елементи матриці C .

Крок 6. 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 $t_{kj \text{ факт}} > t_{\text{табл}}$, між незалежними змінними X_k і X_j існує мультиколінеарність.

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

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

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

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

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

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

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