

МУЛЬТИКОЛІНЕАРНІСТЬ ТА ЇЇ ВПЛИВ НА ОЦІНКУ ПАРАМЕТРІВ МОДЕЛІ

Анотація. Стаття присвячена дослідженню мультиколінеарності та її впливу на оцінку параметрів багатofакторних моделей. У процесі дослідження визначено суть поняття мультиколінеарності як наявності сильного кореляційного зв'язку між обраними для аналізу ознаками, які спільно впливають на цільову функцію, що ускладнює оцінку регресійних параметрів моделі та виявлення залежності між ознаками й цільовою функцією загалом. Надано розширений опис ознак явища мультиколінеарності, на основі чого з'ясовано її наслідки. Доведено, що присутність явища мультиколінеарності насамперед впливає на достовірність оцінки параметрів економетричної моделі, що ідентифікуються зміщенням оцінок, розрахованих за методом найменших квадратів, прямуюванням значення t-статистики Стьюдента до нуля, збільшенням довірчих інтервалів оцінюваних параметрів. Вказано й охарактеризовано основні методи виявлення мультиколінеарності та способи її усунення. Встановлено, що найповніше вивчити мультиколінеарність, а також її усунути, надає змогу метод головних компонент (МПК), а тому більш детально описано механізм цього методу й запропоновано до використання у багатofакторних моделях.

Ключові слова: мультиколінеарність, параметри, багатofакторна економетрична модель, статистичні критерії, методи, регресійний аналіз.

Begun Svetlana, Sakharuk Mariana

Lesya Ukrainka Eastern European National University

MULTICOLLINEARITY AND ITS INFLUENCE
ON THE ESTIMATION OF MODEL PARAMETERS

Summary. The article is devoted to the study of multicollinearity and its influence on the estimation of parameters of multifactor models. In the course of the study it is stated that when studying possible scenarios of socio-economic development, when the statistical apparatus establishes relationships between characteristics of production, consumption, social and financial policies, etc., we use multifactorial regression models to track how planned (possible) changes certain managerial parameters of production or consumption will influence the value of the "original" characteristics. For the purpose of interpreting this model, average partial elasticity coefficients, beta and delta coefficients are used, which allow to order the factors according to their level of influence on the endogenous variable. It is substantiated that these coefficients can be interpreted only on condition that the other variables in the regression model are unchanged, when changing one variable does not lead to transformational changes of the other variables. In the case of multicollinearity of the variables by regression coefficients it is impossible to draw conclusions about the influence of these variables on the studied function. It is determined that multicollinearity is a phenomenon in which there is a dense linear relationship between two or more factor variables of the model, or, in other words, factor variables have a high degree of correlation. An extended description of the features of the multicollinearity phenomenon is provided and its consequences are explained. It is proved that the presence of the phenomenon of multicollinearity first of all influences the reliability of estimation of the parameters of the model, which are identified by the displacement of the estimates calculated by the method of least squares, by directing the values of t-statistics to zero, by increasing the confidence intervals of the estimated parameters. To determine multicollinearity and its elimination in the process of regression analysis, a number of methods are used, in particular: analysis of matrices of coefficients of pair and partial correlation; the Ferrar-Gloubert method, incremental regression method, method of inflation factors, principal component analysis, etc. It is established that the principal component analysis is the most comprehensive way to study and eliminate multicollinearity, and therefore the mechanism of this method is described detail and proposed for use in multifactorial models.

Keywords: multicollinearity, parameters, multifactor econometric model, statistical criteria, methods, regression analysis.

Постановка проблеми. Застосування математичних методів у різних сферах знань дає змогу виокремити та формально описати найважливіші, найсуттєвіші зв'язки між предметами й явищами, а також індуктивним шляхом отримати нові знання про об'єкт. Особливо затребуваними є математичні методи в економіці при дослідженні тенденцій економічного розвитку, їх використовують також і в інших суспільних дисциплінах, зокрема при прогнозуванні соціально-економічних феноменів і процесів. Політики чи менеджери, приймаючи управлінські рішення, обираючи одну із можливих стратегій, намагаються зрозуміти наслідки управлінських

дій, з'ясувати кінцевий результат. Поганий він чи гарний та чи можна було досягти кращого результату, перевірити дуже важко, оскільки соціально-економічні ситуації ідентично майже ніколи не повторюються. Тому на порядку денному особливо гостро постає питання моделювання фактів економічного життя та вивчення тенденцій розвитку соціальних процесів, для чого застосовується інструментарій економетрії – порівняно нового напрямку економічної науки, що утворився від поєднання теоретичної економіки, математики та статистики. Основний метод, яким оперує дана дисципліна, – це регресійний аналіз шляхом побудови економетричних моде-

лей. При цьому дуже важливим є те, наскільки побудована модель відповідає усім вимогам й адекватно відображає досліджуваний процес чи явище, основним показником чого є відсутність сильної кореляції між факторними змінними, тобто мультиколінеарності.

Аналіз останніх досліджень і публікацій. Проблемі мультиколінеарності, можливим способам її виявлення й уникнення присвячена значна кількість праць як вітчизняних, так і зарубіжних учених. Серед останніх найбільш значимих та актуальних наукових здобутків можна виокремити дослідження Р. Аскіна та Ч. Стендріджа [7], М. Мойсеева [5], І. Орлової [6], Р. Пауля [8], О. Глушак і С. Семеняка [2] й ін.

Виділення невирішених раніше частин загальної проблеми. Незважаючи на значну кількість досліджень явища мультиколінеарності, механізм її виникнення та вплив на оцінку параметрів багатфакторної моделі потребують спеціального розширеного аналізу з метою оптимізації загального результату моделювання.

Мета статті. Головною метою даного дослідження є з'ясування природи явища мультиколінеарності, його основних ознак та характеру впливу на оцінку параметрів моделі.

Виклад основного матеріалу. Економетричне моделювання використовується у різних сферах та для вирішення найрізноманітніших прикладних завдань. Об'єктом економетрики може бути як окрема компанія чи фірма, так і економіка різних галузей економічної діяльності, окремих держав та світова економіка у цілому [1, с. 910]. Якщо розглядати кінцеві цілі застосування економетричного апарату, то виділяють два основних завдання:

1) прогнозування економічних і соціально-економічних показників (змінних);

2) імітація можливих сценаріїв соціально-економічного розвитку досліджуваної системи, коли статистично виявлені взаємозв'язки між характеристиками виробництва, споживання, соціальної й фінансової політики тощо застосовуються для відстеження того, як планові (можливі) зміни тих чи інших управлінських параметрів виробництва чи споживання матимуть вплив на значення «вихідних» характеристик [6, с. 58].

При вирішенні завдань другого типу найчастіше використовуються багатфакторні регресивні моделі у цілях виявлення й ранжування факторів за рівнем їх впливу на результативну змінну. Найперше, при оцінці впливу факторів на ендогенну змінну враховують значення коефіцієнтів (параметрів) регресивної моделі. Однак, безпосередньо орієнтуючись на їх значення, не можна зіставити фактори за рівнем їх впливу на залежну змінну через відмінності одиниць виміру і різну ступінь коливання.

Для усунення зазначених вище відмінностей при інтерпретації моделі застосовуються середні часткові коефіцієнти еластичності, бета-коефіцієнти та дельта-коефіцієнти [6, с. 59].

Коефіцієнти еластичності ($E_j = \hat{a}_j \times \frac{\bar{x}_j}{\bar{y}}$) дозволяють порівнювати фактори у відсотках й можуть бути розраховані не тільки для лінійної функції. Вони вказують на середню відсоткову зміну залежної змінної при зміні пояснюючої

змінної на один відсоток при фіксованих значеннях інших пояснюючих змінних. Але даний коефіцієнт не враховує ступінь коливання факторів. Для вирішення питання порівняння сили впливу факторів, що мають різну ступінь коливання, використовують бета-коефіцієнти, або коефіцієнти регресії у стандартизованому вигляді ($\beta_j = \hat{a}_j \times \frac{S_{x_j}}{S_y}$). Ці коефіцієнти порівнювані між собою, тому за їх допомогою можна ранжувати фактори за силою впливу на результуючу змінну. Бета-коефіцієнт показує, на яку частину величини середнього квадратичного відхилення S_y зміниться ендогенна змінна зі зміною відповідної екзогенної змінної x_j на величину свого середнього відхилення при фіксованому значенні інших незалежних змінних.

Перераховані коефіцієнти дозволяють впорядкувати фактори за рівнем їх впливу на ендогенну змінну. При цьому слід звернути увагу на важливий момент – коефіцієнти еластичності, бета-коефіцієнти, як і коефіцієнти регресії, можуть інтерпретуватися тільки за умови, що інші змінні в моделі регресії незмінні, тобто коли зміна однієї змінної не призводить до трансформаційних змін інших. У разі наявності мультиколінеарності змінних за коефіцієнтами регресії не можна робити висновки про вплив цих змінних на цільову функцію.

Найчастіше у сучасній науковій літературі мультиколінеарність трактується як сильний кореляційний зв'язок між обраними для аналізу ознаками, які спільно впливають на цільову функцію, що ускладнює оцінку регресійних параметрів моделі та виявлення залежності між ознаками й цільовою функцією загалом. Також мультиколінеарність призводить до зменшення стійкості оцінок векторів параметрів, оскільки оцінка вектора параметрів вважається стійкою, якщо невелика кількість змін деякої компоненти цього вектора призводить до невеликих змін відповідної компоненти оцінки цільової функції [3, с. 188]. Отже, мультиколінеарність – це явище, при якому між двома або більше факторними змінними моделі існує щільна лінійна залежність, або, іншими словами, факторні змінні мають високий ступінь кореляції:

$$|r_{ij}| \rightarrow 1, \quad (1)$$

де r_{ij} – коефіцієнт кореляції між змінними X_i та X_j ($i \neq j$) [2, с. 172].

Присутність явища мультиколінеарності насамперед впливає на достовірність оцінки параметрів моделі, що проявляється у наступному:

– відбувається зміщення оцінок параметрів моделі, які розраховуються за допомогою МНК (методу найменших квадратів), що унеможливає достовірність висновків про результати взаємозв'язку між результативним показником і факторними змінними;

– значення t-статистики Стьюдента (пов'язана з перевіркою рівності середніх значень у двох вибірках) прямує до нуля, тому неможливо з'ясувати значущість окремих параметрів моделі;

– відбувається збільшення довірчих інтервалів параметрів (збільшується середній квадрат відхилення параметрів) [2; 7–8].

Все зазначене у сукупності створює значні проблеми для достовірної оцінки досліджуваної моделі, оскільки присутні значні стандартні похибки в оцінці модельованих параметрів. При цьому доцільно зауважити, що мультиколінеарність не є проблемою, якщо єдиною ціллю регресійного аналізу виступає прогнозування, оскільки чим більше значення R^2 (коефіцієнта детермінації – показника, що використовується у моделях як міра залежності варіації залежної змінної від варіації незалежних змінних), тим точнішим є прогноз [2, с. 172].

Явище мультиколінеарності проілюструємо на конкретному прикладі: у разі дослідження залежності ціни акції, величини дивідендів на неї та розміру отриманого прибутку, такі параметри як дивіденди й прибуток на акцію, можуть мати високий ступінь кореляції. Тобто виникає ситуація, коли два колінеарні фактори (у даному випадку дивіденди й прибуток) змінюються в одному напрямку, що унеможливило достовірну оцінку впливу кожного з них на результуючий показник (ціну акції).

Класичною ознакою присутності явища мультиколінеарності є наступне протиріччя в аналізі параметрів досліджуваної моделі: значне значення R^2 та незначущість t -статистики. Перший фактор засвідчує, що слід із великою ймовірністю брати до уваги F -тест Фішера, який відхиляє нульову гіпотезу, тобто залежність між оцінюваними параметрами певним чином існує. У свою чергу, другий чинник свідчить, що один або більше оцінених параметрів статистично майже не відрізняються від нуля [8, с. 60]. Також мультиколінеарність ідентифікується у разі великих значень парних коефіцієнтів кореляції (якщо значення хоча б одного з них становить більше 0,8). Одночасно зауважимо, що велике значення парних коефіцієнтів кореляції є достатньою, але не обов'язковою умовою наявності мультиколінеарності. Це явище може мати місце навіть при невеликих значеннях парних коефіцієнтів кореляції у більш, ніж двофакторній регресивній моделі.

У цілому для визначення мультиколінеарності та її усунення у процесі регресійного аналізу застосовують ряд методів, зокрема: аналіз матриць коефіцієнтів парної і часткової кореляції, метод Феррара-Глобера, метод додаткових регресій, метод інфляційних факторів, аналіз головних компонент тощо [5–6].

Одним із найбільш розповсюджених способів визначення щільності регресійного зв'язку є побудова регресійної залежності кожного фактора x_i з усіма іншими факторами (метод додаткових регресій). Обчислення відповідного коефіцієнта детермінації для допоміжного регресійного рівняння та його подальша перевірка за допомогою F -тесту дають змогу виявити лінійні зв'язки між незалежними змінними.

Також часто у процесі аналітичного дослідження явища мультиколінеарності використовується матриця парних коефіцієнтів кореляції, у якій необхідно обчислити визначник. При цьому можуть виникнути наступні ситуації:

– за відсутності кореляції між факторами матриця парних коефіцієнтів кореляції є одиничною, а в ній всі елементи, що знаходяться поза головної діагоналі, дорівнюють нулю;

– якщо між факторами з'ясована абсолютно лінійна залежність і всі коефіцієнти кореляції дорівнюють одиниці, то визначник такої матриці дорівнює нулю;

– чим меншим є значення визначника (наближається до нуля), тим сильнішою вбачається мультиколінеарність факторів і ненадійнішими результати множинної регресії;

– чим більше визначник наближається до одиниці, тим меншим є прояв явища мультиколінеарності факторів [3, с. 188].

Зазначимо, що дві змінні вважаються колінеарними, коли вони знаходяться між собою у лінійній залежності при значенні коефіцієнта кореляції більше ніж 0,7–0,8.

У свою чергу, метод Феррара-Глобера є найбільш комплексним методом дослідження мультиколінеарності та передбачає такі умови й дії: 1) за наявності міжфакторної кореляції один з пари взаємопов'язаних факторів виключається, або ж в якості пояснюючого фактора береться якась їх функція; 2) якщо незначним виявився тільки один з факторів, то його можна виключити або замінити іншим (при цьому залишається ймовірність важливості й значимості цього фактора на більш короткому проміжку часу). Найбільш характерними ознаками мультиколінеарності за методом Феррара-Глобера вважаються наступні:

– невелика зміна вихідних даних (наприклад, додавання нових спостережень) призводить до істотної зміни оцінок параметрів моделі;

– оцінки параметрів мають великі стандартні помилки, малу значимість, у той час як модель у цілому є значущою (велике значення коефіцієнта детермінації і відповідної F -статистики);

– оцінки параметрів характеризуються помилковими знаками або мають невиправдано великі значення [2; 5].

Найповніше вивчити мультиколінеарність, а також її усунути, надає змогу метод головних компонент, винайдений К. Пірсоном у 1901 р. та доповнений і розширений Г. Готелінгом у 1933 р. [4]. Зауважимо, що даний метод (алгоритм) використовується, як правило, для моделей великого розміру. Він передбачає за m -числом початкових ознак виділення r -головних компонент, або узагальнених ознак. При цьому простір головних компонент ортогональний (тобто такий, що не перетинається, не перекривається змістом окремих елементів-параметрів). Математична модель методу головних компонент базується на логічному припущенні, що значення множини взаємозалежних ознак породжують деякий загальний результат (рис. 1).

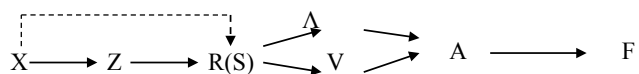


Рис. 1. Схема математичних перетворень за МГК [4]

Вказані на рис. 1 позначення інтерпретуються у наступний спосіб:

X – матриця вихідних даних розмірністю $n \times m$ (n – число об'єктів спостереження, m – число елементарних аналітичних ознак);

Z – матриця центрованих і нормованих значень ознак;

$R(S)$ – матриця парних кореляцій;

Λ – діагональна матриця власних (характеристичних) чисел;

V – нормована матриця власних векторів;

A – матриця факторного відображення, її елементи a_{ij} – вагові коефіцієнти;

F – матриця значень головних компонент розмірністю $r \times n$.

Отже, спочатку матриця A має розмірність $m \times m$ – за кількістю елементарних ознак X_i , потім в аналізі залишається r найвагоміших компонент, $r \leq m$. Обчислюється матриця A за відомими даними матриці власних чисел Λ і нормованими власними векторами V за формулою: $A = V\Lambda^{1/2}$ [4].

Як можна бачити з рис. 1, розв'язування задачі методом головних компонент зводиться до поетапного перетворення матриці початкових даних X . А точніше механізм методу полягає у перетворенні множини змінних X на нову множину попарно некорельованих змінних, серед яких перша відповідає максимально можливій дисперсії, а друга – максимально можливій дисперсії в підпросторі, який є ортогональним до першого і т.д. У результаті визначення усіх головних компонентів і відкидання тих з них, які відповідають невеликим значенням кореляції, встановлюється зв'язок залежної змінної Y з основними головними компонентами, а далі за допомогою оберненого перетворення від параметрів моделі з головними компонентами повертаються до визначення оцінок параметрів змінних X .

Підсумовуючи опис застосування алгоритму головних компонент у процесі дослідження та усунення мультиколінеарності, необхідно акцентувати, що основна перевага даного методу полягає у тому, що це один з основних способів зменшення розмірності даних із найменшою втратою кількості вихідної інформації, тому він є найбільш оптимальним алгоритмом уникнення мультиколінеарності у процесі оцінки параметрів регресійної моделі.

Вважаємо за доцільне також зазначити, що використовують й інші, більш простіші, способи усунення мультиколінеарності, зокрема:

– використання додаткових або первинних вихідних даних;

– об'єднання вихідних даних;

– відкидання змінної з великою кореляцією;

– перетворення даних (застосування перших різниць);

– збільшення кількості вихідних даних (спостережень) [8].

Використання того чи іншого способу уникнення мультиколінеарності залежить від властивостей конкретної моделі, її цільової спрямованості.

Висновки і пропозиції. Отже, мультиколінеарність – це кореляція незалежних факторних змінних, наявність якої значно ускладнює оцінку параметрів регресійної множинної моделі й аналіз загального результату моделювання явищ і процесів. Тобто у разі існування залежності двох незалежних змінних йдеться мова про виникнення явища мультиколінеарності. Дане явище оцінюється негативно, оскільки регресійна модель містить надлишкові факторні ознаки, що може свідчити про наступне: ускладнюється інтерпретація параметрів множинної регресії як величин впливу факторів – параметри регресії загалом втрачають сенс і слід розглядати інші факторні змінні; оцінки параметрів ненадійні – присутні великі стандартні похибки, які змінюються зі зміною обсягу спостережень, що робить модель регресії непридатною для адекватного використання. З метою визначення мультиколінеарності та її усунення у процесі проведення регресійного аналізу використовують ряд методів, зокрема: аналіз матриць коефіцієнтів парної і часткової кореляції, метод Феррара-Глоубера; метод додаткових регресій, метод інфляційних факторів, метод головних компонент тощо. Пропонується з метою усунення впливу мультиколінеарності на оцінку параметрів регресійної моделі виключати з її складу один із факторів. Як правило, це фактор, який за наявності сильного зв'язку з результуючою ознакою (цільовою функцією / цільовим вектором) має найбільш тісний зв'язок з іншими факторними змінними. При цьому обґрунтовано доцільність застосування саме алгоритму головних компонент, як такого, що дозволяє зменшити розмірність даних із найменшою втратою кількості вихідної інформації у багатофакторних регресійних моделях.

Список літератури:

1. Begun S., Levchuk A. Concept of econometrics and main stages of econometric analysis. *Глобальні та національні проблеми економіки*. 2017. № 17. С. 909–913. URL: <http://global-national.in.ua/archive/17-2017/182.pdf> (дата звернення: 17.04.2020).
2. Глушак О.М., Семеняка С.О. Передумови побудови багатофакторної економетричної моделі: дослідження на мультиколінеарність. *Фізико-математична освіта*. 2018. Вип. 1. С. 171–175.
3. Ковальова І.Л. та ін. Економетрія : навч. посіб. Одеса : ОДАБА, 2019. 423 с.
4. Метод головних компонент. *Вікіпедія – вільна енциклопедія*. URL: https://uk.wikipedia.org/wiki/Метод_головних_компонент (дата звернення: 18.04.2020).
5. Мойсеев Н.А. Сравнительный анализ эффективности методов устранения мультиколлинеарности. *Учет и статистика*. 2017. № 2. С. 62–73.
6. Орлова И.В. Подход к решению проблемы мультиколлинеарности при анализе влияния факторов на результирующую переменную в моделях регрессии. *Фундаментальные исследования*. 2018. № 3. С. 58–63.
7. Askin R, Standridge CR. Modeling and Analysis of Manufacturing Systems. *European Journal of Engineering Education*. 1994. Vol. 19. No. 1. P. 122–125.
8. Paul R.K. Multicollinearity: causes, effects and remedies. *IASRI*. 2006. No. 35. P. 58–65.

References:

1. Begun, S., & Levchuk, A. (2017). Concept of econometrics and main stages of econometric analysis. *Hlobalni ta natsionalni problemy ekonomiky* [Global and national problems of economy] (electronic journal), no. 17, pp. 909–913. Available at: <http://global-national.in.ua/archive/17-2017/182.pdf> (accessed 17 April 2020).

2. Ghlushak, O.M., & Semenyaka, S.O. (2018). Peredumovy pobudovy baghatofaktornoji ekonometrychnoji modeli: doslidzhennja na mul'tykolinearnistj [Prerequisites for constructing a multifactor econometric model: multicollinearity study]. *Physical and mathematical education: scientific journal*, vol. 1, pp. 171–175.
3. Kovaleva, I.L. (ed.) (2019). *Ekonometriia: navch. posib.* [Econometrics: teach. tool]. Odessa: ODABA. (in Ukrainian)
4. Metod holovnykh komponent [Principal Component Method]. Wikipedia: free Encyclopedia. Available at: https://uk.wikipedia.org/wiki/Метод_головних_компонент (accessed 18 April 2020).
5. Moiseev, N.A. (2017). Sravnitel'nyj analiz ehffektivnosti metodov ustraneniya mul'tikollinearnosti [Comparative analysis of the effectiveness of multicollinearity elimination methods]. *Accounting and statistics: scientific and practical journal*, no. 2, pp. 62–73.
6. Orlova, I.V. (2018). Podkhod k resheniyu problemy mul'tikolinearnosti pri analize vliyaniya faktorov na rezul'tiruyushchuyu peremennuyu v modelyakh regressii [An approach to solving the problem of multicollinearity in the analysis of the influence of factors on the resultant variable in regression models]. *Basic research: scientific journal*, no. 3, pp. 58–63.
7. Askin, R., & Standridge, C.R. (1994). *Modeling and Analysis of Manufacturing Systems*. *European Journal of Engineering Education*, vol. 19, no. 1, pp. 122–125.
8. Paul, R.K. (2006). Multicollinearity: causes, effects and remedies. *IASRI*, no. 35, pp. 58–65.