Scientific journal
PHYSICAL AND MATHEMATICAL EDUCATION
Has been issued since 2013.

Науковий журнал
ФІЗИКО-МАТЕМАТИЧНА ОСВІТА
Видається з 2013.

ISSN 2413-158X (online)
ISSN 2413-1571 (print)

http://fmo-journal.fizmatsspu.sumy.ua/

Глушак О.М., Семеняка С.О. Передумови побудови багатофакторної економетричної моделі: дослідження на мультиколінеарність. Фізико-математична освіта. 2018. Випуск 1(15). С. 171-175.

Glushak O., Semenyaka S. Preconditions Of Construction Of Multifactory Econometric Model: Research Of Multicolinearity. Physical and Mathematical Education. 2018. Issue 1(15). P. 171-175.

УДК 519.233.5

О.М. Глушак[1], С.О. Семеняка[2]
Київський університет імені Бориса Грінченка, Україна
[1]*o.hlushak@kubg.edu.ua*, [2]*s.semeniaka@kubg.edu.ua*
DOI 10.31110/2413-1571-2018-015-1-031

ПЕРЕДУМОВИ ПОБУДОВИ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ: ДОСЛІДЖЕННЯ НА МУЛЬТИКОЛІНЕАРНІСТЬ

Анотація. *Статтю присвячено детальному аналізу одного із етапів, який передує побудові економетричної моделі множинної регресії, а саме, аналізу та відбору факторних змінних, що входитимуть до досліджуваної моделі. Встановлено характеристики, яким повинні задовольняти потенційні факторні змінні: бути кількісно виміряними, мати високу варіабельність, сильно корелювати з результативною змінною та слабко корелювати між собою, а також сильно корелювати зі змінними, які не використовуються в моделі як факторні змінні, але пов’язані з результативною змінною.*

Розкрито роль мультиколінеарності на етапі розробки економетричної моделі. Визначено, що відсутність мультиколінеарності є ключовою передумовою для побудови багатофакторної економетричної моделі, яка адекватно відображатиме досліджуваний процес. Акцентовано увагу, що відсутність мультиколінеарності дає можливість використовувати метод найменших квадратів (МНК) для знаходження оцінок параметрів моделі, оскільки не відбувається зміщення оцінок. Розглянуто способи виявлення мультиколінеарності, зважаючи на зовнішні ознаки, та методи тестування мультиколінеарності. Сформульовано та обґрунтовано конструктивну схему дослідження мультиколінеарності за допомогою покрокового алгоритму Фаррара-Глобера, який застосовує три види статистичних критеріїв (χ^2, F*-критерій та t-критерій). Дані критерії дозволяють виявити мультиколінеарність, як усього масиву незалежних змінних (критерій* χ^2, *), так і кожної незалежної змінної з усіма іншими (*F*-критерій) та кожної пари незалежних змінних (t-критерій).*

Відображено задачі, які розв’язуються за допомогою методу математичного моделювання. На прикладі конкретної задачі (дослідження залежності між витратами обігу підприємства та обсягом вантажообігу, запасами по вантажообігу, трудомісткістю його одиниці) проілюстровано конструктивність та ефективність реалізації алгоритму Фаррара-Глобера при виявленні наявності між факторними змінними мультиколінеарності. В результаті проведеного дослідження було встановлено, що між факторними змінними, які потенційно можуть входити до економетричної моделі, існує сильний взаємозв’язок, як по всьому масиву змінних, так і попарно.

Ключові слова: *економіко-математичне моделювання, економетрична модель, множинна регресія, коефіцієнт кореляції, мультиколінеарність, критерій Стьюдента, критерій Фішера, алгоритм Фаррара-Глобера.*

Постановка проблеми. Математична модель – це дієвий інструмент дослідження та прогнозування розвитку економічних процесів і явищ. Вона розвиває наші уявлення про закономірності та взаємозв’язки економічних процесів і допомагає формуванню наукового мислення та навичок порівняльного аналізу на новому, більш високому рівні.

Використання математичних методів в економічних дослідженнях дає можливість виділити та формально описати найбільш важливі й суттєві закономірності функціонування економічних систем і об’єктів у вигляді моделей; на основі сформульованих за певними правилами логіки вхідних даних і співвідношень, методами дедукції зробити висновки, які адекватні до об’єкта дослідження стосовно зроблених припущень. Математичні методи дають можливість отримати дедуктивним шляхом нові дані про об’єкт дослідження, окрім того, використання мови математики дозволяє компактно описати основні положення економічної теорії,сформулювати їх змістовний апарат і робити відповідні висновки [1].

Тому розробка адекватної та статистично значущої математичної моделі є одним із основних завдань при дослідженні задач економіки.

Аналіз актуальних досліджень. Будуючи економетричну модель, в якій випадкова величина *У* залежить від кількох факторних змінних потрібно звернути увагу на декілька моментів [2].

По-перше, факторні змінні мають бути кількісно виміряними. Якщо фактори не мають кількісної міри, то їм потрібно надати кількісної визначеності. Так, наприклад, задати якість у вигляді балів або встановити наявність чи відсутність ознаки у вигляді балів: відповідно 1 та 0 та ін..

По-друге, факторні змінні, які входять до досліджуваної моделі мають задовольняти ряд ознак, а саме:

– мати високу варіабельність (здатність до змінюваності);
– сильно корелювати з результативною змінною;
– слабко корелювати між собою;
– сильно корелювати зі змінними, які не використовуються в моделі як факторні змінні, але пов'язані з результативною змінною.

Отже, однією з передумов того, що побудована модель відповідатиме всім вимогам і адекватно відображатиме досліджуваний процес, є відсутність сильної кореляції між факторними змінними, тобто, мультиколінеарності.

Мета статті. Розкрити роль мультиколінеарності на етапі розробки багатофакторної економетричної моделі, висвітлити способи тестування на мультиколінеарність та продемонструвати застосування алгоритму Фаррара-Глобера на прикладній задачі.

Виклад основного матеріалу. 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: $|r_{ij}| \to 1$, де r_{ij} – коефіцієнт кореляції між змінними X_i та X_j ($i \neq j$) [3].

Наявність мультиколінеарності створює певні проблеми для розробки моделей. Насамперед, мультиколінеарність призводить до зміщення оцінок параметрів моделі, які розраховуються за допомогою МНК (методу найменших квадратів). На основі таких оцінок (зміщених) неможливо зробити достовірні висновки про результати взаємозв'язку між результативним показником і факторними змінними. Окрім того, наявність мультиколінеарності не дає змоги з'ясувати значущість окремих параметрів моделі, оскільки значення *t*-статистики Стьюдента стає близьким до нуля. А також не варто забувати, що при наявності мультиколінеарності, відбувається збільшення довірчих інтервалів параметрів.

Звичайно, мультиколінеарність не є проблемою, якщо єдиною метою регресійного аналізу є прогнозування, оскільки чим більше значення R^2 (коефіцієнта детермінації), тим імовірніший прогноз. Якщо ж метою дослідження є не тільки прогнозування, а й аналіз моделі, де використовуються значення параметрів (оцінки параметрів), то мультиколінеарність перетворюється на проблему, оскільки її наявність призводить до великих стандартних похибок в оцінці параметрів.

Єдиного способу визначення мультиколінеарності, на жаль, не існує. Зовнішніми ознаками існування мультиколінеарності є наявність суперечності, яка виникає при поєднанні двох факторів, а саме, високого значення коефіцієнту детермінації R^2 (що свідчить, згідно критерію Фішера, про відсутність нульових значень параметрів) та незначущості *t*-статистики (згідно критерію Стьюдента це означає, що один або більше оцінених параметрів статистично мало відрізняються від нуля).

Високе значення коефіцієнтів кореляції r_{ij} також може бути ознакою мультиколінеарності. Якщо значення хоча б одного коефіцієнта кореляції $|r_{ij}| > 0{,}8 \; (i \neq j)$, то мультиколінеарність є серйозною проблемою. Зауважимо, що дана умова є достатньою, але не необхідною для появи мультиколінеарності, яка може існувати і при невеликих значеннях коефіцієнтів кореляції у більш, ніж двофакторній регресійній моделі.

Існуючі методи тестування мультиколінеарності не є універсальними. Всі вони мають спільний недолік: жоден з них не проводить чіткої межі між тим, що треба вважати «суттєвою» мультиколінеарністю, яку необхідно враховувати, і тим, коли мультиколінеарністю можна знехтувати.

Найповніше дослідження мультиколінеарності проводять за алгоритмом Фаррара-Глобера [4], який застосовує три види статистичних критеріїв і дозволяє виявити мультиколінеарність:

- усього масиву незалежних змінних (критерій χ^2);
- кожної незалежної змінної з усіма іншими (F-критерій);
- кожної пари незалежних змінних (*t*-критерій).

Проілюструємо застосування даного алгоритму на приладі конкретної прикладної задачі

Задача. Нехай на витрати обігу впливають: обсяг вантажообігу, запаси по вантажообігу та трудомісткість його одиниці. Статистичну інформацію, яка необхідна для дослідження подано у таблиці:

№	Вантажообіг	Запаси	Трудомісткість
1	16,9	33,7	2,13
2	14,5	13,8	2,56
3	19,5	31,5	1,91
4	11,5	52,5	3,00
5	17,1	37,0	2,23
6	19,6	43,6	1,96
7	12,5	48,3	2,82
8	16,5	44,7	2,29
9	16,0	35,7	2,33
10	16,1	49,3	2,31

Щоб побудувати економетричну модель, необхідно бути впевненим, що між факторами вантажообігу, запасів та трудомісткості не існує мультиколінеарності. Завдання полягає в тому, щоб дослідити наявність мультиколінеарності між цими факторами і, в разі підтвердження її існування зробити висновки щодо варіантів її усунення і можливості використання МНК (методу найменших квадратів) для оцінки параметрів моделі.

Ідентифікуємо змінні: Y – витрати обігу (результативна змінна), X_1 – обсяг вантажообігу, X_2 – запаси по вантажообігу, X_3 – трудомісткість (факторні змінні). Запишемо проміжні розрахунки, які використовуватимуться в подальшому дослідженні, у вигляді таблиці.

№	X_1	X_2	X_3	X_1^2	X_2^2	X_3^2	X_1^*	X_2^*	X_3^*
1	16,9	33,7	2,13	285,61	1135,69	4,5369	0,111548	-0,15572	-0,21344
2	14,5	13,8	2,56	210,25	190,44	6,5536	-0,19267	-0,73932	0,196285
3	19,5	31,5	1,91	380,25	992,25	3,6481	0,441122	-0,22024	-0,42306
4	11,5	52,5	3	132,25	2756,25	9	-0,57295	0,395612	0,615534
5	17,1	37	2,23	292,41	1369	4,9729	0,1369	-0,05895	-0,11815
6	19,6	43,6	1,96	384,16	1900,96	3,8416	0,453798	0,134608	-0,37542
7	12,5	48,3	2,82	156,25	2332,89	7,9524	-0,44619	0,272441	0,444023
8	16,5	44,7	2,29	272,25	1998,09	5,2441	0,060844	0,166867	-0,06098
9	16	35,7	2,33	256	1274,49	5,4289	-0,00254	-0,09707	-0,02287
10	16,1	49,3	2,31	259,21	2430,49	5,3361	0,010141	0,301767	-0,04192
Середнє	16,02	39,01	2,354	262,864	1638,055	5,65146			

Використовуючи алгоритм Фаррара-Глобера, з’ясуємо, чи існує між факторними змінними сильний кореляційний зв'язок.

На першому кроці даного алгоритму нормалізуємо факторні змінні X_1, X_2, X_3, даної економетричної моделі за формулою:

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{n \cdot \left(\overline{x_j^2} - (\bar{x}_j)^2\right)}}, \quad i = \overline{1, n}; \quad j = \overline{1, m},$$

де n – кількість спостережень відповідних факторних змінних; m – кількість факторних змінних.

Будуємо матрицю X^*, елементами якої є значення x_{ij}^*

X^* =

0,111548	-0,15572	-0,21344
-0,19267	-0,73932	0,196285
0,441122	-0,22024	-0,42306
-0,57295	0,395612	0,615534
0,1369	-0,05895	-0,11815
0,453798	0,134608	-0,37542
-0,44619	0,272441	0,444023
0,060844	0,166867	-0,06098
-0,00254	-0,09707	-0,02287
0,010141	0,301767	-0,04192

та відповідну їй транспоновану матрицю $\left(X^*\right)^T$

0,111548	-0,19267	0,441122	-0,57295	0,1369	0,453798	-0,44619	0,060844	-0,00254	0,010141
-0,15572	-0,73932	-0,22024	0,395612	-0,05895	0,134608	0,272441	0,166867	-0,09707	0,301767
-0,21344	0,196285	-0,42306	0,615534	-0,11815	-0,37542	0,444023	-0,06098	-0,02287	-0,04192

Обчислюємо кореляційну матрицю за формулою $R = \left(X^*\right)^T \cdot X^*$:

R =

1	-0,25383	-0,98966
-0,25383	1	0,281601
-0,98966	0,281601	1

Звідси, визначник кореляційної матриці $\det R$ = 0,018332 і значення χ^2-критерію буде дорівнювати: $\chi^2 = -\left[n - 1 - \frac{1}{6} \cdot (2m+5)\right] \cdot \ln|R|$ = 28,66039, де n – кількість спостережень ($n = 10$), m – кількість факторних змінних ($m = 3$).

Порівнюючи отримане значення χ^2 з табличним χ^2_{tabl} = 7,814728 при $\frac{1}{2}m(m-1)$ степенях вільності і заданому рівні значущості α (α = 0,05), робимо наступний висновок: оскільки $\chi^2 > \chi^2_{tabl}$, то у даному масиві факторних змінних наявне явище мультиколінеарності.

Подальшим кроком дослідження буде розрахунок F-критерію за формулою: $F_k = \frac{(c_{kk}-1)(n-m)}{m-1}$, де c_{kk} – діагональні елементи матриці $C = R^{-1} = \left(\left(X^*\right)^T \cdot X^*\right)^{-1}$

C =

50,22454	-1,35601	50,08685
-1,35601	1,12274	-1,65815
50,08685	-1,65815	51,03567

Знаходимо F_1 = 172,2859; F_2 = 0,42959; F_3 = 175,1248 і порівнюємо відповідні значення з F_{tabl} = 19,35322 при $(n-m)$ та $(m-1)$ степенях вільності і заданому рівні значущості α = 0,05.

Очевидно, що $F_1 > F_{\text{tabl}}$ і $F_3 > F_{\text{tabl}}$, а, отже, існує залежність факторних змінних X_1 і X_3 від сукупності інших змінних.

Знаходимо часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що інші змінні не впливають на цей зв'язок: $r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} \cdot c_{jj}}}$, де c_{kj} – елемент матриці C, розміщений на перетині k-го рядка та j-го стовпця, c_{kk} та c_{jj} – діагональні елементи матриці C.

r_{kj} =

-1	0,180579	-0,9893
0,180579	-1	0,219052
-0,9893	0,219052	-1

Розраховуємо t-критерій за формулою $t_{kj} = \left|r_{kj}\right| \cdot \sqrt{\frac{n-m}{1-r_{kj}^2}}$

t_{kj} =

	0,485752	17,94213
0,485752		0,593983
17,94213	0,593983	

Порівнюючи отримані значення t_{kj} з табличними t_{tabl} = 2,364624 при $(n-m)$ степенях вільності та заданому рівні значущості α = 0,05, робимо наступний висновок: оскільки $t_{13} > t_{\text{tabl}}$, то у парі змінних X_1 і X_3 існує мультиколінеарність.

Висновок: в результаті проведеного дослідження було встановлено, що між факторними змінними, які потенційно можуть входити до економетричної моделі, існує сильний взаємозв'язок, як по всьому масиву змінних (результат перевірки χ^2-критерію), так і попарно. Дана залежність, яку ми називаємо мультиколінеарністю, може значно впливати на якість оцінок, отриманих за допомогою МНК, адже моделі, в яких спостерігається мультиколінеарність, стають надзвичайно чутливими до конкретного набору даних, а отримані за МНК оцінки є зміщеними.

Список використаних джерел

1. Економіко-математичне моделювання: Навчальний посібник / За ред. О. Т. Іващука. Тернопіль: ТНЕУ «Економічна думка», 2008. 704 с.
2. Глушак О. М., Семеняка С. О. Економіко-математичне моделювання – перспективний напрямок прикладної математики. Физико-математична освіта. 2017. №1. С.28-31
3. Лещинський О. Л., Рязанцева В. В., Юнькова О. О. Економетрія: Навч. посіб. для студ. вищ. навч. закл. Київ: МАУП, 2003. 208с.
4. Наконечний С. І., Терещенко Т. О., Романюк Т. П. Економетрія: Підручник. Київ: КНЕУ, 2000. 296 с.

References

1. Economic-mathematical modeling: Textbook / Ed. O. T. Ivashchuk. Ternopil: TNEU "Economic Thought", 2008. 704 pp.
2. Glushak O. M., Semenyaka S. O. Economic-mathematical modeling - perspective direction of applied mathematics. Physical-mathematical education. 2017. №1. P.28-31
3. Leshchyns'kyy O. L., Ryazantseva V. V., Yun'kova O. O. Econometrics. K.: MAUP, 2003. 208 pp.
4. Nakonechnyy S. I., Tereshchenko T. O., Romanyuk T. P. Econometrics: Textbook. K.: KNEU, 2000. 296 pp.

PRECONDITIONS OF CONSTRUCTION OF MULTIFACTORY ECONOMETRIC MODEL: RESEARCH OF MULTICOLINEARITY

Glushak O. M, Semenyaka S. O.
Boris Grinchenko Kyiv University, Ukraine

***Abstract**. The article is devoted to a detailed analysis of one of the stages preceding the construction of a econometric model of multiple regression, namely, the analysis and selection of factor variables included in the model under study. The characteristics that potential factor variables must satisfy are: to be quantitatively measured, to have high variability, to strongly correlate with the resultant variable and to weakly correlate with each other, and also to strongly correlate with variables that are not used in the model as factor variables but related to productive variable.*

The role of multicollinearity at the stage of development of econometric model is revealed. It is determined that the lack of multicollinearity is a key prerequisite for constructing a multivariate econometric model that adequately reflects the investigated process. Attention is drawn to the fact that the lack of multicollinearity makes it possible to use the method of least squares) to find estimates of the model parameters, since there is no bias in estimates. Methods of detecting multicollinearity, taking into account external signs, and methods of testing multicollinearity are considered. The constructive scheme of multicollinearity research is formulated and substantiated by means of a step-by-step Farrar-Globe algorithm, which uses three types of statistical criteria (χ^2*,* F*-criterion and t-criterion). These criteria allow us to detect multicollinearity, as an entire array of independent variables (* χ^2 *criterion), and each independent variable with all other (F-criteria) and each pair of independent variables (t-criteria).*

The tasks, which are solved by means of mathematical modeling, are shown. On the example of a concrete task (the study of the relationship between the expenses of the company's turnover and the volume of cargo turnover, cargo turnover, the complexity of its unit), the constructivity and efficiency of the Farrar-Globard algorithm implementation were demonstrated when the presence of the multiplicity variables was found. As a result of the study, it was found that there are strong interconnections between the factor variables potentially entering the econometric model, both in the whole array of variables, and in pairs.

***Key words:** economic-mathematical modeling, econometric model, multiple regression, correlation coefficient, multicollinearity, Student's criterion, Fisher's criterion, Farrar-Glober's algorithm.*