

ОПТИМІЗАЦІЯ ВІДБОРУ ФАКТОРНИХ ЗМІННИХ У ПРОЦЕСІ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ: ПРАКТИЧНІ РЕКОМЕНДАЦІЇ

Оксана ГЛУШАК ✉

Київський столичний університет
імені Бориса Грінченка, Україна
o.hlushak@kubg.edu.ua
<https://orcid.org/0000-0001-9849-1140>

Світлана СЕМЕНІАКА

Київський столичний університет
імені Бориса Грінченка, Україна
s.semeniaka@kubg.edu.ua
<https://orcid.org/0000-0001-5083-1433>

Надія ЗІНЧЕНКО

Київський столичний університет
імені Бориса Грінченка, Україна
nm.zinchenko@kubg.edu.ua
<https://orcid.org/0000-0003-1124-4636>

OPTIMIZATION OF FACTOR VARIABLE SELECTION IN ECONOMETRIC MODELING: PRACTICAL RECOMMENDATIONS

Oksana HLUSHAK ✉

Borys Grinchenko Kyiv Metropolitan University, Ukraine
o.hlushak@kubg.edu.ua
<https://orcid.org/0000-0001-9849-1140>

Svitlana SEMENIACA

Borys Grinchenko Kyiv Metropolitan University, Ukraine
s.semeniaka@kubg.edu.ua
<https://orcid.org/0000-0001-5083-1433>

Nadiia ZINCHENKO

Borys Grinchenko Kyiv Metropolitan University, Ukraine
nm.zinchenko@kubg.edu.ua
<https://orcid.org/0000-0003-1124-4636>

АНОТАЦІЯ

Формулювання проблеми. Правильний відбір факторних змінних у сучасних економетричних дослідженнях є ключовою умовою забезпечення точності, стабільності та прогностичної здатності моделей, проте у студентських роботах часто трапляються типові помилки, пов'язані з надмірною кількістю факторів, слабким теоретичним обґрунтуванням їх вибору або наявністю мультиколінеарності. У статті обґрунтовано необхідність створення практичних рекомендацій, який поєднує наукову обґрунтованість із простотою застосування та забезпечує коректний вибір факторних змінних у процесі моделювання економічних явищ.

Матеріали і методи. У дослідженні використано комплекс теоретичних, статистичних та економетричних методів, серед яких ключову роль відіграли кореляційний і регресійний аналіз, що дали змогу виявити та кількісно оцінити взаємозв'язки між змінними. Для підвищення достовірності результатів і стабільності моделі проведено перевірку на мультиколінеарність за алгоритмом Фаррара–Глобера, який передбачає застосування χ^2 -, F- та t-критеріїв для оцінки взаємозалежностей між незалежними змінними.

Результати. У результаті дослідження запропоновано поетапний алгоритм відбору факторних змінних, що передбачає: (1) розрахунок коефіцієнтів кореляції для формування вектора та матриці кореляцій; (2) відбір змінних, які мають сильний зв'язок із результативною змінною та слабку взаємну кореляцію; (3) перевірку інформативності змінних за допомогою методу показників інформаційної місткості та коефіцієнта множинної кореляції; (4) додатковий аналіз на мультиколінеарність для виявлення надлишкових факторів. Алгоритм забезпечує обґрунтованість вибору предикторів, підвищує точність і стабільність моделей, а також є зручним для навчального використання.

Висновки. Запропонований підхід дозволяє системно організувати процес відбору факторних змінних у студентських економетричних дослідженнях, поєднуючи наочність, доступність та наукову коректність. Його застосування сприяє формуванню у студентів навичок аналітичного мислення, статистичної інтерпретації результатів та усвідомленого підходу до побудови моделей.

ABSTRACT

Formulation of the problem. The correct selection of factor variables in modern econometric studies is a key condition for ensuring the accuracy, stability, and predictive power of models. However, student papers often contain typical errors related to an excessive number of factors, weak theoretical justification for their selection, or the presence of multicollinearity. The article substantiates the need to create practical recommendations that combines scientific validity with ease of use and ensures the correct selection of factor variables in the process of modeling economic phenomena.

Materials and methods. The study uses a set of theoretical, statistical, and econometric methods, among which correlation and regression analysis played a key role, allowing us to identify and quantitatively assess the relationships between variables. To increase the reliability of the results and the stability of the model, a multicollinearity check was performed using the Farrar–Glauber algorithm, which involves the use of χ^2 , F, and t criteria to assess the interdependencies between independent variables.

Results. As a result of the study, a step-by-step algorithm for selecting factor variables is proposed, which involves: (1) calculating correlation coefficients to form a vector and correlation matrix; (2) selecting variables that have a strong relationship with the outcome variable and weak mutual correlation; (3) verification of the informativeness of variables using the method of information capacity indicators and the multiple correlation coefficient; (4) additional analysis for multicollinearity to identify redundant factors. The algorithm ensures the validity of the predictor choice, enhances the accuracy and stability of models, and is convenient for educational purposes.

Conclusion. The proposed approach enables the systematic organization of the process of selecting factor variables in student econometric research, combining clarity, accessibility, and scientific rigor. Its application contributes to the formation of students' analytical thinking skills, statistical interpretation of results, and a conscious approach to model building.

КЛЮЧОВІ СЛОВА: економетричне моделювання; факторні змінні; відбір факторів; мультиколінеарність; кореляційний аналіз.

KEYWORDS: econometric modeling; factor variables; factor selection; multicollinearity; correlation analysis.

ДЛЯ ЦИТУВАННЯ: Глушак О., Семеняка С., Зінченко Н. Оптимізація відбору факторних змінних у процесі економетричного моделювання: практичні рекомендації. *Фізико-математична освіта*, 2025. Том 40. № 5. С. 29-35. <https://doi.org/10.31110/fmo2025.v40i5-04>.

FOR CITATION: Hlushak, O., Semenaka, S., & Zinchenko, N. (2025). Optimization of factor variable selection in econometric modeling: practical recommendations. *Physical and Mathematical Education*, 40(5), 29-35. <https://doi.org/10.31110/fmo2025.v40i5-04>.

ВСТУП

Постановка проблеми. Сучасна підготовка економістів, фінансистів і менеджерів передбачає формування у студентів здатності здійснювати ґрунтовний кількісний аналіз соціально-економічних процесів на основі економетричних методів (Глушак & Семеняка, 2017; Глушак та ін., 2018). Ефективність таких досліджень значною мірою залежить від коректного відбору факторних змінних, які визначають структуру та пояснювальну силу побудованої моделі. Неправильний вибір факторів, що часто трапляється у студентських роботах, зумовлює низьку якість економетричних моделей, викривлення результатів і помилкові управлінські висновки. Типовими проблемами є надмірне включення змінних без економічного обґрунтування, ігнорування взаємозалежності між ними, а також наявність мультиколінеарності, яка порушує статистичну надійність оцінених параметрів.

Аналіз актуальних досліджень. Відбір факторних змінних є одним із ключових етапів побудови економетричних моделей, оскільки саме він визначає точність, інтерпретованість і прогностичну здатність отриманих результатів. У сучасних дослідженнях запропоновано широкий спектр методів відбору факторів, які можна класифікувати як класичні (Hall & Miller, 2009), регуляризаційні (Turlach et al., 2005), байєсівські (García-Donato & Paulo, 2021) та спеціалізовані, орієнтовані на особливості факторного аналізу (Abramov et al., 2021; Obushnyi et al., 2022; Virovets et al., 2023).

У праці Kuo & Mallick (1998) представлено узагальнений підхід до відбору факторних змінних у регресійних моделях, що поєднує елементи класичних і байєсівських методів. Автори запропонували просту процедуру підмножинного відбору незалежних змінних, яка базується на розширенні стандартного рівняння регресії шляхом введення індикаторних змінних. Ці змінні визначають, які фактори мають бути включені до моделі. У роботі розглянуто традиційні методи – покрокову регресію, послідовне виключення та додавання змінних, засновані на мінімізації середньоквадратичної помилки. Також проведено порівняння з байєсівськими підходами, зокрема з використанням байєсівського та псевдо-байєсівського факторів для вибору найкращої моделі.

Подальший розвиток цієї проблематики представлено у дослідженні Desboulets (2018), де здійснено ґрунтовний огляд сучасних процедур відбору факторних змінних з позицій практичного застосування. Автор наголошує, що зростання обсягів даних сприяло поширенню автоматизованих методів моделювання, які дають змогу формувати моделі без попередніх теоретичних обмежень. Незважаючи на активний розвиток теоретичних аспектів, у більшості прикладних робіт, як зазначає Desboulets, досі широко використовуються базові процедури – передусім покрокова регресія. Дослідник запропонував класифікацію методів залежно від типу моделей (лінійні, групові, адитивні, частково лінійні, непараметричні) та розглянув доступні програмні засоби, що забезпечують їх реалізацію. Крім того, автор навів рекомендації щодо вибору методів для різних цілей моделювання та запропонував підходи до підвищення точності й стабільності процесу відбору факторів.

Інший напрям розвитку цієї тематики представлено в роботі Fan et al. (2018), де розглянуто проблему відбору змінних у високовимірних регресійних задачах за умов наявності просторової та часової залежності між даними. Автори підкреслюють, що більшість поширених методів втрачають ефективність у разі сильної кореляції між факторами. Для розв'язання цієї проблеми вони запропонували метод Factor-Adjusted Regularized Model Selection (FarmSelect), який використовує факторну модель для виділення прихованих (латентних) факторів і індивідуальних компонент змінних. Це дає змогу зменшити взаємну залежність між ознаками та підвищити якість відбору змінних. Доведено, що метод FarmSelect забезпечує узгодженість відбору моделі та оптимальні швидкості збіжності навіть за слабких статистичних умов. Результати чисельних експериментів показують, що запропонований підхід є ефективним як у межах вибірки, так і при прогнозуванні на нових даних, забезпечуючи стабільність навіть за слабкої або відсутньої кореляції між змінними.

Окрему увагу питанню відбору факторних змінних приділено у дослідженні Kvalheim (2020). У роботі порівнюються два показники – вектор та матриця коефіцієнтів кореляції – для ранжування змінних за їхнім внеском у латентно-змінні регресійні моделі, побудовані на основі методу часткових найменших квадратів. Обидва показники базуються на проекції матриці пояснювальних змінних на нормалізований вектор регресії.

Подальший розвиток досліджень у напрямі відбору факторних змінних демонструє зростаючий інтерес до методів, що ґрунтуються на аналізі кореляційних зв'язків між змінними. Зокрема, у роботі Dudáš et al. (2025) було представлено новий підхід до «очищення» кореляційної матриці шляхом застосування нечітких масок (fuzzy masks). Такий підхід дозволяє не лише виділяти пари змінних із високим рівнем взаємозв'язку, але й зберігати поступовість переходу між «значущими» та «незначущими» коефіцієнтами кореляції. Це відкриває нові можливості для точнішої побудови множини факторів, що мають найвищий прогностичний потенціал.

У контексті навчального процесу результати подібних досліджень доцільно використовувати для формування у студентів практичних навичок застосування методів відбору факторних змінних, які спираються на аналіз вектора та матриці коефіцієнтів кореляції. Зокрема, варто приділити увагу методу показників інформаційної місткості, який дозволяє оцінити внесок кожної змінної у загальну варіацію досліджуваного процесу, а також методу коефіцієнта множинної кореляції, що дає змогу визначити ступінь лінійного зв'язку між кожною потенційною факторною змінною та результативною ознакою в умовах багатовимірної залежності.

Опанування цих методів сприятиме розвитку у студентів здатності:

- інтерпретувати структуру взаємозв'язків між змінними;
- здійснювати вибір найбільш інформативних факторів на основі кількісних показників;
- критично оцінювати кореляційну надлишковість у даних і запобігати мультиколінеарності.

Поєднання сучасних підходів, описаних у дослідженнях Fan et al. (2018), Kvalheim (2020) та Dudáš et al. (2025), із базовими методами кореляційного аналізу, створює комплексну методологічну основу для навчання майбутніх економістів, фінансистів і менеджерів методам раціонального відбору факторних змінних у економетричних моделях.

Мета статті полягає у формулюванні практичних рекомендацій для студентів щодо раціонального відбору факторних змінних у процесі побудови економетричних моделей, спрямованих на підвищення достовірності результатів аналізу та якості економічних інтерпретацій.

МЕТОДИ ДОСЛІДЖЕННЯ

У дослідженні використано комплекс теоретичних, статистичних та економетричних методів, спрямованих на обґрунтування, порівняння та оцінку ефективності підходів до відбору факторних змінних у регресійних моделях. Теоретичну основу становили аналіз, синтез і порівняльний аналіз, які дали змогу систематизувати сучасні підходи до відбору змінних та узагальнити результати, подані у працях: Abramov et al. (2021), Desboulets (2018), Dudáš (2025), Fan et al. (2018), García-Donato & Paulo (2021), Hall & Miller (2009), Kuo & Mallick (1998), Kvalheim (2020), Obushnyi et al. (2021), Turlach et al. (2005), Virovets et al. (2023). Для кількісного аналізу застосовано методи кореляційного та регресійного аналізу, зокрема побудову вектора та матриці коефіцієнтів кореляції, а також метод показників інформаційної місткості, що дає змогу визначити внесок кожного предиктора у варіацію результативної ознаки. Застосовані статистичні χ^2 -, F- та t-критерії для оцінки взаємозалежностей між незалежними змінними в алгоритмі дослідження на мультиколінеарність (Глушак & Семеняка, 2018).

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Результати проведеного дослідження дали змогу розробити практичні рекомендації, які представлені у вигляді алгоритму для студентів щодо відбору факторних змінних у процесі побудови економіко-математичних моделей. Відібрані факторні змінні, які входять до досліджуваної моделі, мають задовольняти ряду ознак, а саме: мати високу варіабельність, сильно корелювати з результативною змінною, слабо корелювати між собою, а також сильно корелювати зі змінними, які не використовуються в моделі як факторні, але пов'язані з результативною змінною. Алгоритм відбору факторів передбачає кілька етапів. На початку здійснюється розрахунок коефіцієнтів кореляції, що дозволяє побудувати вектор або матрицю кореляційних зв'язків між змінними. Потім відбувається вибір тих змінних, які мають високу кореляцію з результативною змінною та низьку взаємну кореляцію між собою. Далі проводиться перевірка адекватності вибору факторів із використанням методу показників інформаційної місткості та методу коефіцієнта множинної кореляції. Для підвищення достовірності результатів здійснюється також додатковий аналіз на мультиколінеарність, що дозволяє виявити надлишкові або тісно пов'язані змінні. Перевагами запропонованого алгоритму є його простота, наочність та швидкість застосування для невеликих наборів даних, що робить його зручним для навчальних цілей і практичних завдань студентів. Водночас обмеженням є те, що цей підхід не враховує нелінійні зв'язки між змінними й може бути недостатньо ефективним при наявності високої мультиколінеарності чи великої кількості потенційних факторів.

Розглянемо докладніше алгоритм відбору факторних змінних (Рис.1), що входять до складу економетричної моделі. Дана процедура складається з наступних кроків:

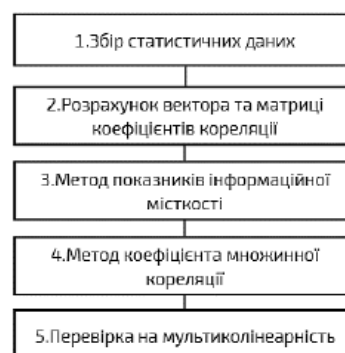


Рис. 1. Алгоритм відбору факторних змінних

Джерело: розроблено авторами

1. Збір статистичних даних про реалізацію результативної змінної Y та факторних змінних X_j ($j = \overline{1, n}$). Отримані статистичні дані y_j та x_{ij} ($i = \overline{1, m}$, $j = \overline{1, n}$). На цьому кроці систематизують та записують дані у вигляді таблиці або вектора спостереження залежної змінної Y і матриці спостереження незалежних змінних $X_1, X_2, \dots, X_n$:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix}.$$

2. Наступним етапом дослідження виступає аналіз щільності взаємозв'язків між результативною ознакою та відібраними факторними змінними, а також між самими факторами, що включені до моделі. Така перевірка дозволяє

оцінити ступінь відповідності факторів основним вимогам, зокрема другій та третій ознакам, яким мають задовольняти обрані змінні, – наявності тісного зв'язку з результативною змінною та слабкої кореляції між собою. Для кількісного вимірювання сили цих зв'язків застосовується розрахунок коефіцієнтів кореляції, що дозволяє виявити ступінь взаємозалежності показників і забезпечити коректність подальшого моделювання.

Для визначення сили зв'язку обчислюємо коефіцієнти кореляції за формулами:

$$r_j = \frac{\overline{x_j y} - \overline{x_j} \cdot \overline{y}}{\sqrt{\overline{y^2} - (\overline{y})^2} \cdot \sqrt{\overline{x_j^2} - (\overline{x_j})^2}} \text{ та } r_{kj} = \frac{\overline{x_k x_j} - \overline{x_k} \cdot \overline{x_j}}{\sqrt{\overline{x_k^2} - (\overline{x_k})^2} \cdot \sqrt{\overline{x_j^2} - (\overline{x_j})^2}}, \text{ де } k = \overline{1, n}, j = \overline{1, n}$$

Перша формула дозволяє визначити силу зв'язку між відповідними факторними та результативною змінними, а друга – між потенційними факторними змінними. Результати, отримані після використання першої та другої формул, можна записати відповідно у вигляді вектора кореляції R_0 та матриці кореляції R :

$$R_0 = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}, X = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & 1 \end{pmatrix}.$$

Метою цього етапу дослідження є відбір таких факторних змінних, які характеризуються високим рівнем кореляції із залежною змінною та водночас низьким ступенем взаємної кореляції факторів між собою. Для оцінювання статистичної значущості отриманих коефіцієнтів кореляції застосовується критерій Стюдента: розраховані значення порівнюються з критичним значенням коефіцієнта кореляції, визначеним на основі t-розподілу Стюдента для заданого рівня значущості α і $n - 2$ ступенів вільності (в даному випадку n – це обсяг відповідної вибірки). Такий підхід дає змогу обґрунтовано виокремити найбільш інформативні фактори для подальшого включення до моделі.

$$r_{\text{крит}} = \sqrt{\frac{t_{\text{табл}}^2}{t_{\text{табл}}^2 + n - 2}}, \text{ де } t_{\text{табл}} = t\left(\frac{\alpha}{2}; n - 2\right).$$

У разі, якщо потенційні факторні змінні не проходять перевірку за критерієм $r_{kj} < r_{\text{крит}}$, вони вилучаються з подальшого аналізу. Серед змінних, що залишилися, обирається та, яка має найбільший коефіцієнт кореляції з результативною змінною, оскільки саме вона містить найбільший обсяг інформації про поведінку залежної змінної. Далі здійснюється аналіз кореляційних зв'язків між вибраною факторною змінною та іншими змінними, що залишилися у моделі. Із розгляду виключаються ті змінні, для яких виконується умова $r_{kj} > r_{\text{крит}}$, оскільки вони дублюють інформацію, вже відображену вибраною змінною. Такий підхід дозволяє мінімізувати мультиколінеарність і забезпечити включення до моделі лише найбільш інформативних факторів.

3. *Метод показників інформаційної місткості* є одним із ефективних підходів до відбору факторних змінних у процесі побудови економетричних моделей. Його суть полягає у виборі тих змінних, які демонструють високий рівень кореляції з результативною змінною та водночас характеризуються низьким ступенем взаємної кореляції між собою. Такий підхід дає змогу уникнути дублювання інформації та підвищити точність моделі. Початковими даними для реалізації методу виступають вектор кореляцій результативної змінної з факторними R_0 та матриця парних кореляцій між самими факторними змінними R . Саме на основі цих статистичних показників здійснюється оцінювання інформаційної місткості кожної змінної та формування оптимального набору факторів для побудови моделі.

Суть даного методу полягає в наступному: розглядають усі комбінації потенційних факторних змінних $X_1, X_2, \dots, X_n$. Загальна кількість таких комбінацій $C = 2^n - 1$. Для кожної з таких комбінацій розраховують *індивідуальні* та *інтегральні* показники інформаційної місткості.

Індивідуальні показники інформаційної місткості для кожної конкретної ситуації визначаються за формулою

$$h_{cj} = \frac{r_j^2}{1 + \sum_{i=1, i \neq j}^{n_c} |r_{ji}|},$$

де c – номер комбінації ($c = \overline{1, C}$), j – номер факторної змінної, що входить в дану комбінацію та індивідуальний показник якої знаходимо ($j = \overline{1, n}$), n_c – кількість змінних в певній комбінації, r_{ji} – коефіцієнт кореляції між факторними змінними X_j та X_i , r_j – коефіцієнт кореляції між факторною змінною X_j та результативною змінною. Інтегральні

показники інформаційної місткості обчислюються за формулою $H_c = \sum_{j=1}^{n_c} h_{cj}$.

Індивідуальні та інтегральні показники інформаційної місткості нормуються в межах інтервалу $[0; 1]$, що забезпечує можливість їх порівняння між собою. Чим ближче значення цих показників до одиниці, тим вищою є інформативність відповідних факторних змінних. У випадку, коли факторні змінні демонструють сильний кореляційний зв'язок із результативною змінною та водночас характеризуються слабкими взаємними кореляціями, індивідуальні та інтегральні показники інформаційної місткості наближаються до своїх максимальних значень. Відповідно, для включення до моделі обираються ті змінні, які мають найбільші значення інтегрального показника інформаційної місткості, оскільки

саме вони забезпечують найповніше відображення інформації про результативну ознаку при мінімальній надлишковості даних.

4. Наступний метод відбору факторних змінних ґрунтується на використанні *коефіцієнта множинної кореляції*. Коефіцієнт множинної кореляції може бути критерієм відбору найкращої комбінації факторних змінних серед комбінацій, що містять *однакову* кількість факторів.

Даний коефіцієнт є мірою лінійного зв'язку результативної змінної Y з факторними змінними $X_1, X_2, \dots, X_n$. Очевидно, максимальне значення коефіцієнта відповідає найбільшій щільності зв'язку. Позначається коефіцієнт множинної кореляції R_c і задається на інтервалі $[0; 1]$.

Якщо до комбінації факторних змінних входять *всі* факторні змінні, то коефіцієнт множинної кореляції можна визначити за допомогою вектора кореляції R_0 та матриці кореляції R за відповідною загальною формулою:

$$R_c = \sqrt{1 - \frac{\det W}{\det R}}, \text{ де } W = \begin{pmatrix} 1 & R_0^T \\ R_0 & R \end{pmatrix}$$

5. Наступним кроком для підвищення достовірності отриманих результатів після первинного відбору факторних змінних доцільно провести додатковий аналіз на мультиколінеарність. Такий аналіз спрямований на виявлення надлишкових або тісно взаємопов'язаних змінних, які можуть спотворювати результати оцінювання параметрів моделі та знижувати її пояснювальну здатність. Мультиколінеарність виникає тоді, коли дві або більше незалежних змінних мають високий рівень взаємної кореляції, унаслідок чого їхній внесок у пояснення варіації результативної змінної стає неоднозначним. Для виявлення цього явища застосовують алгоритм Фаррара — Глобера, який включає три види статистичних критеріїв, за якими перевіряється мультиколінеарність усього масиву незалежних змінних (χ^2 -критерій); кожної незалежної змінної з усіма іншими (F-критерій); кожної пари незалежних змінних (t-критерій). Та складається з семи кроків: нормалізація змінних; знаходження кореляційної матриці; визначення критерію χ^2 ; визначення оберненої матриці; обчислення F-критеріїв; знаходження частинних коефіцієнтів кореляції; обчислення t-критеріїв (Глушак & Семеняка, 2018). Проведення такого аналізу дозволяє забезпечити стабільність параметрів регресійної моделі, підвищити точність прогнозування та уникнути дублювання інформації між факторами.

ОБГОВОРЕННЯ

Під час виконання студентських економетричних досліджень без застосування запропонованого алгоритму найчастіше спостерігаються типові помилки, пов'язані з некоректним урахуванням взаємозв'язків між факторними змінними. Зокрема, студенти часто не аналізують щільність кореляційних зв'язків між факторами, що призводить до появи мультиколінеарності в моделі. Унаслідок цього оцінки параметрів регресійного рівняння стають статистично незначущими, а точність прогнозування суттєво знижується. Використання запропонованого алгоритму дає змогу своєчасно виявити такі взаємозалежності, оптимізувати набір факторів і підвищити достовірність отриманих результатів.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШОГО ДОСЛІДЖЕННЯ

Проведене дослідження дозволило сформулювати системний підхід до відбору факторних змінних у студентських економетричних дослідженнях, що поєднує сучасні теоретичні напрацювання та практичні рекомендації для навчального застосування. Обґрунтовано важливість правильного вибору факторів як ключової умови побудови достовірних, інтерпретованих і прогностично ефективних економетричних моделей. На основі аналізу сучасних наукових праць встановлено, що актуальні підходи до відбору факторних змінних еволюціонують від класичних методів до інтегрованих моделей, які враховують високу вимірність даних, латентні фактори та складні кореляційні залежності.

Запропонований алгоритм забезпечує послідовність дій при формуванні системи факторів, зокрема обчислення коефіцієнтів кореляції, визначення найбільш інформативних змінних за методами показників інформаційної місткості та коефіцієнта множинної кореляції, а також перевірку моделі на наявність мультиколінеарності. Такий підхід дозволяє студентам усвідомлено та обґрунтовано здійснювати вибір змінних, підвищуючи якість побудованих моделей.

Практичне значення одержаних результатів полягає у можливості використання розробленого алгоритму як навчально-методичного інструменту під час виконання кваліфікаційних робіт з економетричного моделювання. Застосування запропонованих методів сприяє розвитку в студентів навичок аналітичного мислення, статистичного моделювання та інтерпретації даних.

Перспективи подальших досліджень полягають у розширенні методологічної бази за рахунок врахування нелінійних залежностей між змінними, а також у розробленні алгоритмів для роботи з великими даними. Такий напрям забезпечить поглиблення теоретичних засад і підвищить практичну ефективність економетричних досліджень у підготовці майбутніх економістів, фінансистів і менеджерів.

КОНФЛІКТ ІНТЕРЕСІВ

Автори підтверджують відсутність фінансових, особистих чи інших інтересів, що можуть розглядатися як потенційний конфлікт інтересів щодо публікації цієї статті.

ФІНАНСУВАННЯ

Робота виконана за відсутності фінансової підтримки з боку будь-яких організацій.

ДОСТУПНІСТЬ ДАНИХ

Це теоретичне дослідження не передбачає використання додаткових наборів даних.

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ

Інструменти штучного інтелекту не використовувались при написанні цієї роботи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Abramov, V., Astafieva, M., Boiko, M., Bodnenko, D., Bushma, A., Vember, V., Hlushak, O., Zhylytsov, O., Ilich, L., Kobets, N., Kovaliuk, T., Kuchakovska, H., Lytvyn, O., Lytvyn, P., Mashkina, I., Morze, N., Nosenko, T., Proshkin, V., Radchenko, S., & Yaskevych, V. (2021). *Theoretical and practical aspects of the use of mathematical methods and information technology in education and science*. <https://doi.org/10.28925/9720213284km>
2. Dudáš, A., Michalíková, A., & Jašek, R. (2025). Fuzzy masks for correlation matrix pruning. *IEEE Access*, 13, 35387–35400. <https://doi.org/10.1109/access.2025.3544027>
3. Desboulets, L. (2018). A review of variable selection in regression analysis. *Econometrics*, 6(4). <https://doi.org/10.3390/econometrics6040045>
4. Fan, J., Ke, Y., & Wang, K. (2018). Factor-adjusted regularized model selection. *Econometrics: Econometric Model Construction*. <https://doi.org/10.2139/ssrn.3248047>
5. García-Donato, G., & Paulo, R. (2021). Variable selection in the presence of factors: A model selection perspective. *Journal of the American Statistical Association*, 117(540), 1847–1857. <https://doi.org/10.1080/01621459.2021.1889565>
6. Hall, P., & Miller, H. (2009). Using generalized correlation to effect variable selection in very high dimensional problems. *Journal of Computational and Graphical Statistics*, 18(3), 533–550. <https://doi.org/10.1198/jcgs.2009.08041>
7. Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, 60(1), 65–81.
8. Kvalheim, O. (2020). Variable importance: Comparison of selectivity ratio and significance multivariate correlation for interpretation of latent-variable regression models. *Journal of Chemometrics*, 34(9). <https://doi.org/10.1002/cem.3211>
9. Obushnyi, S., et al. (2021). Ensuring data security in the peer-to-peer economic system of the DAO. In *Cybersecurity Providing in Information and Telecommunication Systems II* (Vol. 3187, pp. 284–292).
10. Obushnyi, S., et al. (2022). Autonomy of economic agents in peer-to-peer systems. In *Cybersecurity Providing in Information and Telecommunication Systems* (Vol. 3288, pp. 125–133).
11. Turlach, B., Venables, W., & Wright, S. (2005). Simultaneous variable selection. *Technometrics*, 47(3), 349–363. <https://doi.org/10.1198/004017005000000139>
12. Virovets, D., et al. (2023). Ways of interaction of autonomous economic agents in decentralized autonomous organizations. In *Cybersecurity Providing in Information and Telecommunication Systems* (Vol. 3421, pp. 182–190).
13. Глушак, О. М., & Семеняка, С. О. (2017). Економіко-математичне моделювання – перспективний напрямок прикладної математики. *Фізико-математична освіта*, 1(11), 28–31. http://fmo-journal.fizmatsspu.sumy.ua/journals/2017-v1-11/2017_1-11-GlushakSemenyaka_Scientific_journal_FMO.pdf
14. Глушак, О. М., & Семеняка, С. О. (2018). Передумови побудови багатофакторної економетричної моделі: дослідження на мультиколінеарність. *Фізико-математична освіта*, 1(15), 171–175. <https://doi.org/10.31110/2413-1571-2018-015-1-031>
15. Глушак, О. М., Бодненко, Д., & Семеняка, С. О. (2018). Формування інформатичної компетентності майбутніх фінансистів під час вивчення дисципліни «Економетрика». *Освітологічний дискурс*, 1–2(20–21), 325–340. <http://od.kubg.edu.ua/index.php/journal/article/view/492>

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Abramov, V., Astafieva, M., Boiko, M., Bodnenko, D., Bushma, A., Vember, V., Hlushak, O., Zhylytsov, O., Ilich, L., Kobets, N., Kovaliuk, T., Kuchakovska, H., Lytvyn, O., Lytvyn, P., Mashkina, I., Morze, N., Nosenko, T., Proshkin, V., Radchenko, S., & Yaskevych, V. (2021). *Theoretical and practical aspects of the use of mathematical methods and information technology in education and science*. <https://doi.org/10.28925/9720213284km>
2. Dudáš, A., Michalíková, A., & Jašek, R. (2025). Fuzzy masks for correlation matrix pruning. *IEEE Access*, 13, 35387–35400. <https://doi.org/10.1109/access.2025.3544027>
3. Desboulets, L. (2018). A review on variable selection in regression analysis. *Econometrics*, 6(4). <https://doi.org/10.3390/econometrics6040045>
4. Fan, J., Ke, Y., & Wang, K. (2018). Factor-adjusted regularized model selection. *Econometrics: Econometric Model Construction*. <https://doi.org/10.2139/ssrn.3248047>
5. García-Donato, G., & Paulo, R. (2021). Variable selection in the presence of factors: A model selection perspective. *Journal of the American Statistical Association*, 117(540), 1847–1857. <https://doi.org/10.1080/01621459.2021.1889565>
6. Hall, P., & Miller, H. (2009). Using generalized correlation to effect variable selection in very high dimensional problems. *Journal of Computational and Graphical Statistics*, 18(3), 533–550. <https://doi.org/10.1198/jcgs.2009.08041>
7. Kuo, L., & Mallick, B. (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, 60(1), 65–81.
8. Kvalheim, O. (2020). Variable importance: Comparison of selectivity ratio and significance multivariate correlation for interpretation of latent-variable regression models. *Journal of Chemometrics*, 34(9). <https://doi.org/10.1002/cem.3211>
9. Obushnyi, S., et al. (2021). Ensuring data security in the peer-to-peer economic system of the DAO. In *Cybersecurity Providing in Information and Telecommunication Systems II* (Vol. 3187, pp. 284–292).
10. Obushnyi, S., et al. (2022). Autonomy of economic agents in peer-to-peer systems. In *Cybersecurity Providing in Information and Telecommunication Systems* (Vol. 3288, pp. 125–133).
11. Turlach, B., Venables, W., & Wright, S. (2005). Simultaneous variable selection. *Technometrics*, 47(3), 349–363. <https://doi.org/10.1198/004017005000000139>
12. Virovets, D., et al. (2023). Ways of interaction of autonomous economic agents in decentralized autonomous organizations. In *Cybersecurity Providing in Information and Telecommunication Systems* (Vol. 3421, pp. 182–190).

13. Hlushak, O. M., & Semenyaka, S. O. (2017). Economic and mathematical modeling is a promising area of applied mathematics. *Physical and Mathematical Education*, 1(11), 28–31. http://fmo-journal.fizmatsspu.sumy.ua/journals/2017-v1-11/2017_1-11-GlushakSemenyaka_Scientific_journal_FMO.pdf (in Ukrainian).
14. Hlushak, O. M., & Semenyaka, S. O. (2018). Prerequisites for building a multifactor econometric model: a study on multicollinearity. *Physical and Mathematical Education*, 1(15), 171–175. <https://doi.org/10.31110/2413-1571-2018-015-1-031> (in Ukrainian).
15. Hlushak, O. M., Бодненко, Д., & Semenyaka, S. O. (2018). Developing the IT skills of future financiers while studying the discipline "Econometrics." *Educational discourse*, 1–2(20–21), 325–340. Borys Grinchenko Kyiv University. <http://od.kubg.edu.ua/index.php/journal/article/view/492> (in Ukrainian).

| Матеріал надійшов до редакції: 02.10.2025 р. | Прийнято до друку: 10.11.2025 р. | Опубліковано: 28.11.2025 р. |

