

DOI: <https://doi.org/10.36910/6775-2524-0560-2020-41-19>

УДК: 519.86 + 519.21

Хомяк Марія Ярославівна, канд. фіз.-мат. наук, доцент

<https://orcid.org/0000-0002-9245-7993>

Волинський національний університет імені Лесі Українки, м. Луцьк, Україна

## ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ПОЛІНОМІАЛЬНОЇ МОДЕЛІ РЕГРЕСІЇ З ПОХИБКАМИ ВИМІРЮВАННЯ В ПРОГНОЗУВАННІ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ

**Хомяк М.Я. Особливості застосування поліноміальної моделі регресії з похибками вимірювання в прогнозуванні соціально-економічних процесів.** Розглянуто поліноміальну регресійну модель з похибками вимірювання. З допомогою методу виправлених найменших квадратів знайдено конзистентну оцінку для невідомого параметра. Досліджується взаємозв'язок між факторними змінними, які потенційно можуть входити до економетричної моделі. Обґрунтовується вплив мультиколінеарності на якість оцінок, отриманих за допомогою МНК. Розглянуто способи виявлення мультиколінеарності та методи тестування мультиколінеарності.

**Ключові слова:** модель регресії, похибки вимірювання, оцінка, метод найменших квадратів, економетрична модель, мультиколінеарність.

**Хомяк М.Я. Особенности применения полиномиальной модели регрессии с погрешностями измерения в прогнозировании социально-экономических процессов.** Рассмотрено полиномиальную модель регрессии с погрешностями измерения. С помощью метода исправленных наименьших квадратов найдено конзистентную оценку для неизвестного параметра. Исследуется взаимосвязь между факторными переменными, которые могут входить в эконометрической модели. Обосновывается влияние мультиколлинеарности на качество оценок, полученных с помощью МНК. Рассмотрены способы выявления мультиколлинеарности и методы тестирования мультиколлинеарности.

**Ключевые слова:** модель регрессии, погрешности измерения, оценка, метод наименьших квадратов, эконометрическая модель, мультиколлинеарность.

**Khomyak M.Ya. Features of application of polynomial regression model with measurement errors in forecasting of socio-economic processes.** A polynomial regression model with measurement errors is considered. Using the method of corrected least squares, a consistent estimate for an unknown parameter was found. The relationship between factor variables that could potentially be included in the econometric model is investigated. The influence of multicollinearity on the quality of estimators obtained with the help of MLS is substantiated. Methods of multicollinearity detection and multicollinearity testing methods are considered.

**Keywords:** regression model, errors of measurement, estimator, least squares method, econometric model, multicollinearity.

**Постановка проблеми.** Математична модель – це дієвий інструмент дослідження та прогнозування розвитку соціально-економічних процесів і явищ. Вона розвиває наші уявлення про закономірності та взаємозв'язки 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 [2]. Тому розробка адекватної та статистично значущої математичної моделі є одним із основних завдань при вивченні соціально-економічних процесів.

**Аналіз актуальних досліджень.** Оцінювання невідомих параметрів поліноміальної функціональної регресійної моделі з похибками у змінних вивчали Ченг та Шнеєвайс. В [7] для даної моделі методом виправлених найменших квадратів вони побудували конзистентну оцінку, яка має асимптотичний нормальний розподіл. Модифікація цієї оцінки для малої вибірки запропонована у [8], модифікована оцінка асимптотично еквівалентна немодифікованій, але більш стійка для малих та середніх вибірок. Огляд інших результатів щодо вищезгаданої поліноміальної моделі міститься у [9]. В [3] отримано оцінку для невідомого параметра векторної моделі регресії з похибками вимірювання.

Проте на даний час отримано досить мало результатів щодо вивчення взаємозв'язків між факторними змінними, які потенційно можуть входити до економетричної моделі. Умови, яким факторні змінні, що входять до досліджуваної моделі мають задовольняти, викладено у [1].

**Мета статті.** Розглянути поліноміальну регресійну модель з похибками вимірювання, побудувати оцінку невідомого параметра, а також розкрити роль мультиколінеарності на етапі розробки багатовимірної економетричної моделі, висвітлити способи тестування на мультиколінеарність.

**Виклад основного матеріалу.** Розглянемо поліноміальну функціональну модель з похибками вимірювання при фіксованому  $k \geq 1$ :

$$y_i = \sum_{j=0}^k \beta_j \xi_i^j + \varepsilon_i = \zeta_i' \beta + \varepsilon_i,$$

$$x_i = \xi_i + \delta_i, i = 1, 2, \dots, n,$$

де  $\xi_i$  - невідомі не випадкові величини,  $\delta_i$  та  $\varepsilon_i$  - похибки спостережень,  $\zeta_i := \begin{pmatrix} 1 \\ \xi_i \\ \xi_i^2 \\ \dots \\ \xi_i^k \end{pmatrix}$

$\zeta_i := (1, \xi_i, \xi_i^2, \dots, \xi_i^k)'$  і  $\beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{pmatrix}$ . Хоча  $i = 1, 2, \dots, n$  являються просто індексами нумерації, в

прогнозуванні вони зазвичай відповідають моментам часу, взятих з однаковим кроком, при цьому допускається, щоб деякі значення  $i$  були опущені.

Визначення оптимальної кількості  $x_i$ , пояснюючих змінних, що адекватно описують зміни залежної змінної  $y_i$  - одне з найбільш важливих завдань, що стоять перед прогнозістом. Завдання це, як відомо, зовсім не просте і не є одноразовим рішенням, а є радше процесом.

Припустимо, що виконуються наступні умови.

а)  $\{\delta_i, i \geq 1\}$  та  $\{\varepsilon_i, i \geq 1\}$  - дві послідовності незалежних однаково розподілених випадкових величин, причому  $\{\delta_i\}$  та  $\{\varepsilon_i\}$  незалежні між собою.

Надалі  $(x_i, \varepsilon_i, \delta_i) = (x, \varepsilon, \delta)$ ;  $\mathbf{E}$  - символ математичного сподівання.

б)  $\mathbf{E}\delta = 0$ ,  $\mathbf{E}\varepsilon = 0$ .

в) Дисперсія  $\sigma_\varepsilon^2 := \mathbf{E}\varepsilon^2 < \infty$  - невідома, і відомі всі моменти до порядку  $2k$  включно від похибки  $\delta$ , причому  $\sigma_\delta^2 := \mathbf{E}\delta^2 > 0$ .

Нехай  $t_r(x)$  - такий поліном, що при всіх  $\xi \in \square$ ,  $\mathbf{E}t_r(x) = \mathbf{E}t_r(\xi + \delta) = \xi^r$ ,  $r = 0, 1, \dots, 2k$ . Функція  $t_r(x)$  є многочленом  $r$ -го степеня, що виражається через відомі моменти  $\mathbf{E}\delta^l$ ,  $l = 1, 2, \dots, r$ .

Позначимо  $t(x) = \begin{pmatrix} t_0(x) \\ t_1(x) \\ \dots \\ t_k(x) \end{pmatrix}$ , і нехай  $H = H(x)$  - матриця розміру  $(k+1) \times (k+1)$ , з матричним

елементом  $h_{pq}(x) = t_{p+q}(x)$ ,  $p, q = 0, 1, \dots, k$ .

Оцінку  $\hat{\beta}$  параметра  $\beta$  будемо виправленим методом найменших квадратів як розв'язок лінійної системи  $\overline{H}\hat{\beta} = \overline{ty}$ , коли матриця  $\overline{H}$  є не виродженою. Якщо ж  $\overline{H}$  вироджена, то покладемо  $\hat{\beta} = 0$ .

Тут і надалі риска зверху означає осереднення за  $n$  спостереженнями, наприклад,  $\overline{ty} = \frac{1}{n} \sum_{i=1}^n t(x_i) y_i$ .

Для довільної функції  $f(\xi)$ ,  $\xi \in \square$ , позначатимемо  $M(f(\xi)) := \lim_{n \rightarrow \infty} \overline{f(\xi)}$  за умови, що границя існує та скінченна.

**Лема [7, 10].** Нехай виконуються умови:

(i)  $E\delta^{4k} < \infty$ ,

(ii) Існує границя  $M(\xi^r)$  для  $r=1, \dots, 4k$  і матриця  $S := M(\zeta(\xi)\zeta'(\xi))$  не вироджена.

(Штрих тут позначає транспонування)

Тоді  $\bar{N}$  не вироджена з імовірністю, що прямує до 1 при  $n \rightarrow \infty$ ,

$$\hat{\beta}^P \rightarrow \beta \text{ при } n \rightarrow \infty,$$

$$\hat{\sigma}_\varepsilon^2 := \overline{y^2} - (\overline{ty})' \hat{\beta}^P \rightarrow \sigma_\varepsilon^2 \text{ при } n \rightarrow \infty.$$

Однією з умов використання методу найменших квадратів (МНК) для знаходження параметрів моделі є те, що пояснювальні змінні  $\zeta_i$  мають бути незалежними між собою. Проте не практиці можливі випадки, коли пояснювальні змінні пов'язані між собою, що стає перешкодою до використання МНК. Отже, варто не забувати про вплив статистичної лінійної залежності між регресорами на стандартні відхилення коефіцієнтів регресії (мультиколінеарність). Оскільки  $\hat{\beta}$  є

оцінкою істинного, хоча і не спостережуваного коефіцієнта  $\beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{pmatrix}$ , то через фактори випадковості

компоненти  $\hat{\beta}$  можуть значно відрізнятись від  $\beta := \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{pmatrix}$ , що, в свою чергу, впливає на кількісні

характеристики моделі і може стати причиною неякісного прогнозу. Тобто, процес адекватно описується моделлю регресії навіть, можливо, з малою величиною  $\sigma_\varepsilon$ , а прогноз виявляється неякісним. Причиною є саме мультиколінеарність.

Найбільш комплексним методом дослідження мультиколінеарності є метод Феррара-Глобера [6], який 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 не виправдано великі значення [1; 5].

Найповніше вивчити мультиколінеарність, а також її усунути, надає змогу метод головних компонент [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). 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.

Проте відомо, що мультиколінеарність в тій чи іншій мірі завжди присутня в моделі множинної регресії. Щоб позбутися мультиколінеарності прогнозист може виключити з моделі регресори з

невиправдано великими стандартними помилками, але при цьому можуть бути виключені окремі регресори, вплив яких необхідно враховувати для розуміння процесу. Так що ми маємо тут дві тенденції з діаметрально протилежним ефектом. З економічної точки зору регресор може виявитися істотним для отримання більш адекватної моделі, але з суто статистичних причин його включення може створити проблеми, пов'язані з мультиколінеарністю. Отже, оптимальним варіантом могло б стати застосування *ex post* прогнозування, яке дозволить виявити модель, що забезпечує значно кращі результати.

### Висновки з дослідження і перспективи подальших розробок.

Розглянуто поліноміальну регресійну модель з похибками вимірювання. З допомогою методу виправленої оцінки найменших квадратів знайдено конзистентну оцінку для невідомого параметра, яка має асимптотичний нормальний розподіл. Однією з передумов того, що побудована модель відповідатиме всім вимогам і адекватно відображатиме досліджуваний процес, є відсутність сильної кореляції між факторними змінними, тобто, мультиколінеарності. Дана залежність може значно впливати на якість оцінки, отриманої за допомогою МНК, адже модель, в якій спостерігається мультиколінеарність, стає надзвичайно чутливою до конкретного набору даних, а отримана за МНК оцінка є зміщеною. Статистичними методами виправлення мультиколінеарності є факторний аналіз та гребінчаста регресія, на що будуть спрямовані подальші дослідження.

### Список бібліографічного опису.

1. Глушак О.М., Семеняка С.О. Передумови побудови багатфакторної економетричної моделі: дослідження на мультиколінеарність. Фізико-математична освіта. 2018. Вип. 1. С. 171–175.
2. Економіко-математичне моделювання: Навчальний посібник / За ред. О. Т. Іващука. Тернопіль: ТНЕУ «Економічна думка», 2008. 704 с.
3. Кукуш А.Г. та Полеха М.Я. Критерій згоди для багатовимірної моделі регресії з похибками у змінних. Теорія стохастичних процесів. 12 (28), 3-4, 2006. 67-79.
4. Метод головних компонент. Вікіпедія – вільна енциклопедія. URL: [https://uk.wikipedia.org/wiki/Метод\\_головних\\_компонент](https://uk.wikipedia.org/wiki/Метод_головних_компонент) (дата звернення: 03.11.2020).
5. Моисеев Н.А. Сравнительный анализ эффективности методов устранения мультиколлинеарности. Учет и статистика. 2017. № 2. С. 62–73.
6. Пул Р.К. Мультиколінеарність: причини, наслідки та засоби захисту. IASRI. 2006. No 35. С. 58–65.
7. Ченг К.-Л. та Шнеесвайс Г. Поліноміальна модель регресії з похибками у змінних. Дж. Р. Статист. Соц. Б, 60, 1998. С. 189-199.
8. Ченг К.-Л. та Шнеесвайс Г. Оцінка для малої вибірки поліноміальної моделі регресії з похибками у змінних. Дж. Р. Статист. Соц. Б, 62, 2000. С. 699-709.
9. Ченг К.-Л. та Шнеесвайс Г. Про модель поліноміальної помилки вимірювання, у: С. ван Хаффель та П. Леммерлінг (редактори), Моделювання загальних найменших квадратів та помилок у змінних. Клувер, Дордрехт, 2002. 131-143.
10. Хомяк М. Я. Застосування поліноміальної моделі з похибками у змінних в прогнозуванні соціально-економічних процесів // Сучасні виклики і актуальні проблеми науки, освіти та виробництва: міжгалузеві диспути [зб. наук. пр.]: матеріали ІХ міжнародної науково-практичної інтернет-конференції (м. Київ, 16 жовтня 2020 р.). Київ, 2020. – С. 723-726.
11. Khomyak M. A polynomial errors-in-variables model in forecasting of economic processes // Міжнародна наукова інтернет-конференція "Інформаційне суспільство: технологічні, економічні та технічні аспекти становлення (випуск 52) / Збірник тез доповідей: випуск 52 (м. Тернопіль, 14 жовтня 2020 р.). –Тернопіль. – С. 17-19.

### References.

1. Glushak OM, Semenyaka SO Prerequisites for constructing a multifactor econometric model: a study on multicollinearity. Physical and mathematical education. 2018. Vip. 1. S. 171–175.
2. Economic and mathematical modeling: Textbook / Ed. OT Ivashchuk. Ternopil: TNEU "Economic Thought", 2008. 704 p.
3. Kukush A.G. and Polekha M.Ya. A goodness-of-fit test for a multivariate errors-in-variables model. Theory of Stochastic Processes. 12(28), 3-4, 2006. 67-79.
4. The method of principal components. Wikipedia is a free encyclopedia. URL: [https://uk.wikipedia.org/wiki/Метод\\_головних\\_компонент](https://uk.wikipedia.org/wiki/Метод_головних_компонент) (access date: 03.11.2020).
5. Moiseev NA Comparative analysis of the effectiveness of methods to eliminate multicollinearity. Accounting and statistics. 2017. № 2. S. 62–73.
6. Paul R.K. Multicollinearity: causes, effects and remedies. IASRI. 2006. No. 35. P. 58–65.
7. Cheng C.-L. and Schneeweiss H. Polynomial regression with errors in the variables. J. R. Statist. Soc. B, 60, 1998. 189-199.
8. Cheng C.-L. and Schneeweiss H. A small sample estimator for a polynomial regression with errors in the variables. J. R. Statist. Soc. B, 62, 2000. 699-709.
9. Cheng C.-L. and Schneeweiss H. On the polynomial measurement error model, in: S. van Huffel and P. Lemmerling (editors), Total Least Squares and Errors-in-Variables Modeling. Kluwer, Dordrecht, 2002. 131-143.

10. Khomyak M. Ya. Application of the polynomial model with errors in variables in forecasting socio-economic processes // Modern challenges and current problems of science, education and production: intersectoral debates [collection. Science. pr.]: materials of the IX International Scientific and Practical Internet Conference (Kyiv, October 16, 2020). Kyiv, 2020. - P. 723-726.

11. Khomyak M. A polynomial errors-in-variables model in forecasting of economic processes // International Scientific Internet Conference "Information Society: Technological, Economic and Technical Aspects of Formation" (issue 52) / Collection of abstracts: issue 52 (m. Ternopil, October 14, 2020). - Ternopil. - P. 17-19.