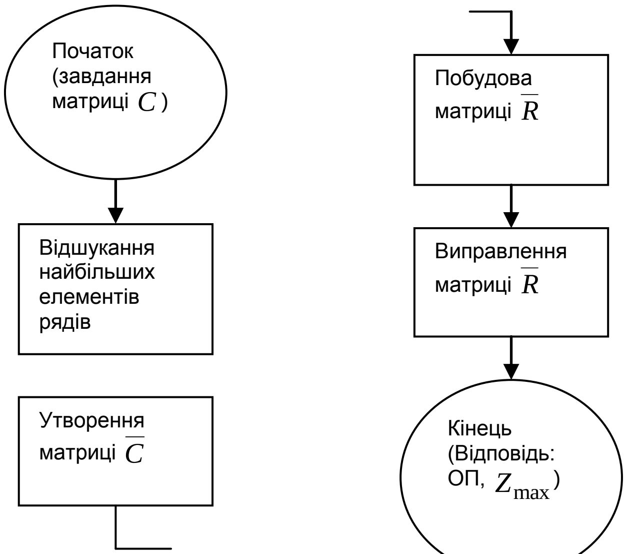

Рис. **Структурна схема розв'язання задачі**

Якщо розв'язується задача на мінімум ($Z \to \min$), то утворюють одну з матриць $\underline{C} = (\underline{c}_{ij})_{n\times n}$ за мінімальними елементами рядків (стовпців):

$$\underline{C} = (\underline{c}_{ij})_{n\times n},$$

де $\forall i = \overline{1,n}: \underline{c}_{ij} = \min_{\forall j \in I}\{c_{ij}\}$ (7)

та $\forall j = \overline{1,n}: \underline{c}_{ij} = \min_{\forall i \in I}\{c_{ij}\}$.

Потім будують матрицю $\underline{R} = C - \underline{C}$ і виправляють її для отримання правильної конфігурації нулів.

Отже, маємо такі висновки. Перевагою розв'язання задачі про призначення способом "зсунення нулів" – перевагою над методом Фогеля – є суттєво менше число операцій, необхідних для установлення оптимального плану. Ця обставина дає можливість відповідно зменшити кількість комірок, що змінюються, при використанні пакетів прикладних програм (наприклад, у надбудові "Поиск решения" в MS Excel).

Запропонований спосіб припускає його використання в інших задачах дискретної оптимізації (транспортна задача, задача комівояжера).

Розглянутий підхід до розв'язання задачі про призначення (чи задачі про закріплення транспортних засобів за маршрутами, якщо відома матриця прибутків), можна впровадити у навчальний процес при вивченні дисциплін "Економіко-математичні методи і моделі" та "Вища і прикладна математика" у разі відносно невеликих розмірів матриці корисності (прибутків) і розв'язання відповідних задач уручну.

____

Література: 1. Пономаренко В. С. Багатовимірний аналіз соціально-економічних систем : навчальний посібник / В. С. Пономаренко, Л. М. Малярець. – Харків : Вид. ХНЕУ, 2009. – 384 с. 2. Сенчуков В. Ф. К вопросу о составлении оптимальных маршрутов доставки грузов / В. Ф. Сенчуков // Економіка розвитку. – 2009. – № 1(49). – С. 88–91. 3. Кузнецов А. В. Высшая математика: Мат. программир. : учебник / А. В. Кузнецов, В. А. Сакович, Н. И. Холод. – Мн. : Высш. шк., 1994. –286 с. 4. Карпелевич Ф. И. Элементы линейной алгебры и линейного программирования / Ф. И. Карпелевич, Л. Е. Садовский. – М. : Физматгиз, 1963. – 276 с. 5. Акулич И. Л. Математическое программирование в примерах и задачах : учебн. пособ. для студ. эконом. спец. вузов. – М. : Высш. шк., 1986. – 320 с.

Стаття надійшла до редакції 05.11.2010 р.

УДК 338.27

Тижненко О. Г.

Тижненко Л. О.

ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ АСПЕКТИ ЗАСТОСУВАННЯ ПАРНОЇ ЛІНІЙНОЇ РЕГРЕСІЇ В ЗАДАЧАХ ЕКОНОМІЧНОГО ПРОГНОЗУВАННЯ

Анотація. Розглянуто економіко-математичні умови коректного застосування звичайного методу найменших квадратів у задачах прогнозування методом парної лінійної регресії.

Аннотация. Рассмотрены экономико-математические условия корректного применения обычного метода наименьших квадратов в задачах прогнозирования методом парной линейной регрессии.

Annotation. Correct implementation of ordinary least square method in forecast problems is considered from economic-mathematics' point of view.

Ключові слова: лінійна регресія, стохастичне моделювання.

Лінійна регресія з невипадковим регресором історично є першою моделлю оцінки статистичної залежності економічної випадкової величини (ВВ) від часу. У тому випадку, коли математичні оцінки параметрів моделі знаходяться звичайним методом найменших квадратів (ЗМНК), проблема інференції вирішується на основі теореми Гаусса – Маркова (ГМ) [1], яка стверджує, що якщо помилки в лінійної моделі мають нульове математичне сподівання, некорельовані та мають однакові дисперсії, то ЗМНК дає ефективну та незміщену оцінку параметрів моделі (Best Linear Unbiased Estimator – BLUE). При цьому помилки не повинні бути незалежними та однаково розподіленими (independent and identically distributed – i.i.d.). Помилки повинні бути тільки некорельованими та гомоскедастичними [2], оскільки 4-та умова теореми ГМ [1, с. 81], яка стосується некорельованості помилки та регресорів, виконується автоматично для невипадкових регресорів, що знімає проблему обґрунтованості при наявності такої кореляції у загальному випадку.

Для з'ясування особливостей застосування лінійної регресії з невипадковим регресором в економічних дослідженнях необхідно зв'язати математичні дефініції теореми ГМ з реальними процесами обробки економічної інформації. Розглянемо для цього математичну постанову задачі парної лінійної регресії.

Перш за все, приймається гіпотеза про те, що в генеральної сукупності (ГС) має бути лінійний статистичний зв'язок між відкликом (Y), тобто досліджуваною змінною, та невипадковим регресором, тобто часом (T):

$$Y = \beta_0 + \beta_1 T + u\,. \qquad (1)$$

Таким чином задається лінійна модель у ГС. У цієї моделі β_0 і β_1 – значення параметрів моделі в ГС, які нам не звісні і не можуть бути визначені у принципі, але можуть бути оцінені статистично з певною точністю (задача математичної статистики як раз і полягає в тому, щоб знайти найкращі оцінки цих параметрів). Оскільки зв'язок статистичний, то в моделі (1) до невипадкової величини $\beta_0 + \beta_1 T$ додається випадкова помилка u, яка враховує випадкові впливи неврахованих у моделі факторів на Y. В моделі (1) величини Y, T та u вважаються неперервними, що дозволяє вважати Y і u розподіленими за певним законом. Таке приближення допустиме в економічних дослідженнях у тому випадку, коли економічні показники можуть, у принципі, приймати випадковим чином будь-яке значення в певному інтервалі, який є носієм закону розподілу цього показника. К показникам, які допускають неперервне представлення в ГС, відноситься більшість абсолютних та відносних показників мікроекономіки, таких, як: власний і позиковий капітали, рентабельність капіталу, коефіцієнти покриття тощо.

Неперервні змінні в (1) необхідно зв'язати з дискретною вибіркою обсягу n, яка досліджується на практиці. Для цього використовуються дискретні реалізації неперервної ВВ. Так, якщо Y є неперервним аналогом ВВ у ГС, то $y_{(n)} = \{y_1, y_2, ..., y_n\}$ є *n*-вимірною дискретною реалізацією неперервної випадкової величини (НВВ), яку досліджує економіст.

Усі твердження теореми ГМ стосуються властивостей випадкової помилки u в (1). Дискретна її реалізація в ГС, $u_{(n)} = \{u_1, u_2, ..., u_n\}$, є вибіркою з ГС. Першою умовою теореми ГМ є вимога рівності нулю математичного сподівання (МС) кожної вибіркової компоненти помилки:

$$M(u_i) = 0,\ i = 1:n\,. \qquad (2)$$

Суттєво це означає, що кожна компонента u_i розглядається як реалізація НВВ з певним законом розподілу і рівним нулю МС. Щоб зв'язати таке представлення з реальною методикою проведення економічних досліджень, слід апроксимувати НВВ її дискретною реалізацією: $u_i = \{u_1^{(i)}, u_2^{(i)}, ..., u_N^{(i)}\}$. У цьому разі умова (2) теореми ГМ може бути наближено записана:

$$\bar{u}_i = 0 \Leftrightarrow \frac{1}{N}\sum_{m=1}^{N} u_m^{(i)} = 0\,. \qquad (3)$$

Суттєво (3) є усередненням від реалізації до реалізації або усередненням по ансамблю (cross-sectional averaging). Для прив'язування такого усереднення до реалії економічних досліджень припустимо, що залежна змінна (Y) становить обсяг продажів за тиждень товару певного виду, зробленого на досліджуваному підприємстві. При цьому, обсяг продажів вивчається за певний відрізок часу, розбитий на рівні періоди (тижні). Регресор у такому випадку має вигляд: $T = \{t_1, t_2, ..., t_n\}$. Середнє значення попиту за n спостережених періодів створює вибірку: $y_{(n)} = \{y_1, y_2, ..., y_n\}$. Ці дані використовуються в рівнянні вибіркової регресії:

$$y_i = b_0 + b_1 t_i + e_i\,,$$

яке вирішується за допомогою ЗМНК. У результаті рішення отримуємо оцінки параметрів регресії $\widehat{b_0}$ та $\widehat{b_1}$, які визначають тренд статистичної залежності тижневого обсягу продажів від часу для досліджуваного підприємства.

При цьому використовується тільки одна реалізація ($y_{(n)}$) НВВ – обсяг продажів, яка складає ГС. Якщо врахувати те, що економічні дані мають невідтворний характер, ми не можемо повторити експеримент для отримання другої реалізації обсягу продажів за той же період часу. Реально, однак, досліджуваний вид продукції виробляється одночасно кількома виробниками. Тому, в принципі, існує можливість отримати декілька реалізацій обсягу продажів за один і той же період. Неважко також уявити собі абстрактну ситуацію, коли виробників багато і вони складають ГС об'єктів. У цьому випадку за *i*-й період (тиждень) можна отримати досить велике число значень залежної змінної (обсягу продажів), $y_i^{(m)},\ m = 1:N$, яких можна вважати дискретною реалізацією НВВ за *i*-й період. Усі допустимі значення цієї ВВ складають ансамбль. Усереднення по ансамблю в цьому випадку має вигляд:

$$\bar{y}_i = \frac{1}{N}\sum_{m=1}^{N} y_i^{(m)}\,.$$

Тоді у рівнянні лінійної регресії в ГС (1) для *i*-ї дискретної реалізації НВВ

$$y_i = \beta_0 + \beta_1 t_i + u_i\,,$$

ми можемо вважати, що середнім по ансамблю значенням *i*-ї компоненти випадкової помилки є незміщена та обґрунтована оцінка її МС:

$$\bar{u}_i = \frac{1}{N}\sum_{i=1}^{N} u_i{}^{m} \xrightarrow{P} M(u_i)\cdot$$

Саме в цьому сенсі слід розуміти першу умову теореми ГМ. З точки зору розглянутого прикладу, це означає, що середні значення по ансамблю відхилень обсягів продажів від тренду для кожного періоду дорівнюють нулю.

Відмітимо, що ця умова виконується автоматично, якщо модель містить постійний член (β_0).

Друга умова теореми ГМ [1, с. 80]

$$D(u_i) = \sigma^2 < \infty \tag{4}$$

означає, що дисперсії всіх компонентів випадкової помилки, які розглядаються як НВВ, однакові та скінченні. Дискретний варіант цієї умови свідчить про те, що всі компоненти випадкової помилки мають однакові та обмежені дисперсії по ансамблю. Щодо розглянутого економічного прикладу це означає, що дисперсії відхилень обсягу продажів від тренду для кожного періоду однакові та скінченні. На відміну від першої умови (2), друга умова (4) може й не виконуватися на практиці. Степінь коливань обсягу продажів для ансамблю об'єктів-виробників може суттєво змінюватися від одного до іншого періоду, оскільки він залежить від багатьох факторів, таких, як сезон, нестабільність ринку або політичних колізій. У цьому випадку кажуть про наявність гетероскедастичності (heteroscedasticity). Доведено, що її присутність не змінює незміщеності та обґрунтованості оцінок параметрів моделі за ЗМНК, але оцінки можуть бути неефективними [1 − 3]. Дослідження показали, що розрахунок дисперсії параметрів за теоретичними формулами [1, с. 83], які використовують припущення про нормальність розподілу параметрів моделі по ансамблю, дають перевищене значення дисперсії. У роботах [3; 4] розглянуто вплив гетероскедастичності на точність розрахунку оцінок параметрів лінійної моделі в реальних економічних дослідженнях. Про важливість та складеність проблеми врахування гетероскедастичності в задачах прогнозування свідчить той факт, що один із авторів роботи [3], саме Robert Engle, отримав Нобелевську премію за дослідження застосування регресійного аналізу при наявності гетероскедастичності (ARCH modeling technique) в задачах прогнозування. Тим не менше, неврахування гетероскедастичності не завжди приводить до значних помилок в оцінках параметрів лінійної моделі. Це трапляється тільки тоді, коли дисперсії випадкової помилки змінюються досить значно [4]. Наявність зміни дисперсії помилки з часом можливо оцінити за графіком залишкової помилки після рішення задачі лінійної регресії. Якщо зміни не суттєві, можна не застосовувати спеціальні методи врахування гетероскедастичності [4].

Третя умова теореми ГМ [1, с. 80]

$$cov(u_i, u_j) = 0, i \neq j \tag{5}$$

означає відсутність систематичного зв'язку між значеннями помилки в ГС, тобто відсутність автокореляції. Ця умова рідко коли виконується в часових рядах. Як правило, значення показника у наступному періоді значною мірою визначається його значеннями в минулі періоди, що й дозволяє здійснити обґрунтований прогноз його значення у наступному періоді з певною точністю. Точність прогнозу визначається дисперсією параметрів моделі. Присутність автокореляції, тобто невиконання умови (5), призводить до зміщення оцінок параметрів. Дисперсія параметрів, яка розрахована за допомогою теоретичних формул, як правило, завищена. У дійсності точність оцінки виявляється вище за теоретичних формул. Якщо точність оцінки, яка отримана за допомогою теоретичних формул, влаштовує дослідника, то він може бути впевненим, що у дійсності точність буде вищою. Якщо ж точність, яка отримана звичайним МНК, недостатня, то необхідно звертатися до спеціальних методів оцінки дисперсії параметрів моделі, таким, як бутстрап, наприклад.

Четверта умова теореми ГМ [1, с. 80] стосується стохастичних регресорів і у випадку регресії з невипадковим регресором не розглядається.

Як видно з детального розглядання умов Гаусса − Маркова, вони не накладають ніяких обмежень на закон розподілу випадкової помилки і, як наслідок, на закон розподілу економічного показника, який досліджується. Якщо, однак, ми впевнені, що закон розподілу помилки і показника є нормальним, то для проведення обґрунтованої інференції можливо скористатися теоретичними формулами оцінки дисперсії параметрів моделі, які звісні тільки для нормального розподілу. Якщо ми не знаємо закону розподілу помилки, то для проведення інференції необхідно використати методи стохастичного моделювання. Саме цей аспект обмежує застосування регресійних методів в економічних дослідженнях просто завдяки тому, що стохастичні методи моделювання регресії не отримали необхідного поширення в економічних дослідженнях на пострадянському просторі.

У реальності закон розподілу помилки в ГС невідомий і не існує ніякої можливості його визначити. Єдине, що має дослідник, − це вибірка, за котрою визначаються оцінки параметрів застосованої моделі, а також залишкова помилка, яка залежить від виду моделі. Залишкову помилку можливо, при адекватній моделі, вважати дискретною реалізацією випадкової помилки у ГС. При цьому, адекватна модель чи ні апріорі теж не відомо. Тому в реальній ситуації в будь-якому випадку слідує використати не теоретичні формули нормального розподілу, а методи стохастичного моделювання для проведення обґрунтованої інференції. У такій ситуації необхідно, перш за все, дослідити особливості моделювання регресії на ідеальній схемі за методом Монте-Карло.

Ідеальна схема стохастичного експерименту за методом Монте-Карло в задачах лінійної регресії (1) з невипадковим регресором при відсутності неадекватності та гетероскедастичності схематично наведена в роботі [1, с. 74]. У цьому експерименті генеруються *n*-вимірні реалізації u_n випадкової помилки u, котра вважається нормально розподіленою: $u \in \mathcal{N}(0, \sigma_u)$. Відповідна реалізація відгуку, y_n, генерується виходячи з рівняння (1) при заданих апріорі значеннях параметрів моделі (β_0, β_1) в ГС: $y_n = \beta_0 + \beta_1 x_n + u_n$. Генерація *n*-вимірних реалізацій відгуку здійснюється $M >> 1$ разів по різним реалізаціям випадкової помилки u з заданим стандартним відхиленням σ_u. Кожен раз при цьому вирішується задача лінійної регресії:

$$y_n = b_0 + b_1 x_n + e_n$$

за допомогою ЗМНК, який дозволяє отримати оцінки параметрів моделі $(\hat{b}_0, \hat{b}_1)$. У результаті отримуємо M пар таких оцінок: $\{\hat{b}_0^{(i)}\}_{i=1:M}$, $\{\hat{b}_1^{(i)}\}_{i=1:M}$, які дозволяють розрахувати з заданою точністю середнє по ансамблю значення параметрів моделі та їх дисперсію, тобто характеристики параметрів регресії у ГС (β_0, β_1 та σ_0^2, σ_1^2) при заданому стандартному відхиленні помилки σ_u. У свою чергу, ці характеристики дозволяють побудувати довірчі інтервали для параметрів регресії, тобто провести обґрунтовану інференцію, якщо ми знаємо закон розподілу цих параметрів. Відмітимо, що для нормально розподіленої помилки закон розподілу параметрів повинен бути теж нормальним. У цьому випадку для побудови інференції можливо використати теоретичні формули для дисперсій параметрів моделі [1, с. 83]:

$$\sigma_0 = \frac{\sigma_u}{\sqrt{n}}\left(1 + \frac{\bar{x}^2}{\sigma_x^2}\right), \quad \sigma_1 = \frac{\sigma_u}{\sqrt{n}\sigma_x}. \qquad (6)$$

Взагалі, ідеальна схема стохастичного моделювання має важливе значення при теоретичному дослідженні регресії як з невипадковим, так і з стохастичним регресором, як у лінійному, так і в нелінійному випадках, оскільки вона реалізує ту гіпотетичну ситуацію, яка найбільш пристосована для статистичних досліджень, коли аналітик має можливість отримувати необмежене число вибірок відгуку з однієї і тієї ж ГС. У літературі, наприклад, у роботі [1], ця схема описана лише у загальних рисах, а чисельні розрахунки економічного прикладу ([1, табл. 3.1, с. 76]) наведені не повністю. Метою прикладу в роботі [1] є демонстрація того, що подвоєння стандартного відхилення помилки приводить до подвоєння стандартного відхилення параметрів регресії. Цей результат є тривіальним, оскільки він слідує з формул (6). Тобто в роботі [1] нібито підтверджується те, що моделювання за методом Монте-Карло дає той же самий результат, що і теоретичні формули (6). Це дійсно так, але з одним принциповим зауваженням: результати моделювання є тільки обґрунтованими оцінками значень (6), тобто вони наближаються до них зі збільшенням числа експериментів (M). Розглянутий у роботі [1] приклад використовує тільки $M=10$ експериментів, чого край недостатньо [4]. Прості розрахунки показують, що для $n=20$, $\beta_0=2$, $\beta_1=0{,}5$ та $\sigma_u=1$, теоретичні дисперсії параметрів є $\sigma_0=0{,}4555$, $\sigma_1=0{,}0378$, що співпадає з наведеним у роботі [1] σ_1. Але при цьому не наведено результатів моделювання. Для $M=10$ усереднені оцінки параметрів такі: $\tilde{b}_0^{(M)}=2{,}1483$, $\tilde{b}_1^{(M)}=0{,}4828$. Оцінки їх стандартних відхилень: $\tilde{s}_0^{(M)}=0{,}2695$, $\tilde{s}_1^{(M)}=0{,}0283$. Друга реалізація з 10 експериментів дасть інший результат, наприклад: $\tilde{b}_0^{(M)}=2{,}1195$, $\tilde{b}_1^{(M)}=0{,}4946$; $\tilde{s}_0^{(M)}=0{,}3942$, $\tilde{s}_1^{(M)}=0{,}0354$. Ми бачимо, що результати значно змінюються, особливо для постійного члена b_0. Якщо, наприклад, лінійна модель використовується для визначення постійних витрат [5, с. 221], то моделювання з $M=10$ приведе до помилки у 10 %, що не допустимо. Тому звичайно використовується число експериментів не менше ніж 50 000 [4]. Для розглянутого в роботі [1] прикладу, при $M=50\,000$ получимо такі результати: $\tilde{b}_0^{(M)}=1{,}9913$, $\tilde{b}_1^{(M)}=0{,}5005$; $\tilde{s}_0^{(M)}=0{,}4677$, $\tilde{s}_1^{(M)}=0{,}0390$. У цьому випадку помилка моделювання дає 0,4-відсоткове зміщення оцінки постійних витрат. Якщо така точність не задовольняє аналітика, число експериментів слід збільшити. Варто відмітити, що такі зусилля не варті того, щоб тільки показати збіжність результатів моделювання з теоретичними формулами (6), як це зроблено в роботі [1]. У дійсності, розглянута в роботі [1] ідеальна схема моделювання регресії дозволяє дослідити такі важливі аспекти, як вплив неадекватності моделі на інференцію, залежність оцінки параметрів та їх дисперсій від закону розподілу випадкової помилки та від наявності гетероскедастичності, якщо використати цю схему сумісно з непараметричними методами моделювання, такими, як джекнайф (jackknife), кросвалідація (cross-validation) та бутстрап (bootstrap).

Проведений аналіз теоретичних положень математичної статистики щодо коректного застосування МНК для оцінки параметрів лінійної регресії показав: закон розподілу досліджуваної змінної, як і випадкової помилки, не повинен бути обов'язково нормальним; якщо закон розподілу випадкової помилки нормальний, дисперсія параметрів моделі може бути розрахована за теоретичними формулами; якщо закон розподілу випадкової помилки не відомий, необхідно користуватися стохастичними методами оцінки дисперсій параметрів моделі; коректне застосування ЗМНК припустимо не тільки при наявності гомоскедастичності, але й при помірній гетероскедастичності; при наявності автокореляції оцінки параметрів можуть бути зміщені; при наявності помірної гетероскедастичності або автокореляції застосування ЗМНК приводить до завищення дисперсій параметрів моделі.

У роботі наведені також методичні рекомендації щодо застосування ідеальної схеми моделювання регресії за методом Монте-Карло, яка може використовуватися для дослідження неадекватності лінійної моделі та залежності оцінок параметрів моделі від закону розподілу випадкової помилки.

Література: 1. Доугерти К. Введение в эконометрику / К. Доугерти ; пер. с англ. – М. : ИНФРА-М, 1999. – XIV. – 402 с. 2. Aitken A. C. "On Least Squares and Linear Combinations of Observations" / A. C. Aitken // Proceedings of the Royal Society of Edinburgh. – 1935. – Vol. 55. – Pp. 42–48. 3. Engle R. F. Measuring and testing the impact of news on volatility [Electronic resource] / R. F. Engle // Journal of Finance. – 1993. – Vol. 48 (5). – Pp. 1749–1778. – Access mode : http://papers.ssrn.com-/sol3/papers.cfm?abstract_id=262096. 4. Fox J. Applied Regression Analysis, Linear Models, and Related Methods. Thousand Oaks / J. Fox. – CA : Sage Publications, 1997. – 597 p. 5. Савицкая Г. В. Экономический анализ / Г. В. Савицкая. – М. : НОВОЕ ЗНАНИЕ, 2004. – 640 с.

Стаття надійшла до редакції 18.11.2010 р.