УДК 330.4:658

М. О. МЕЛЬНИКОВА

ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ АСПЕКТИ ЗАСТОСУВАННЯ ПРОГНОСТИЧНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ ДЛЯ ОЦІНКИ РІВНЯ ФІНАНСОВОГО СТАНУ ПІДПРИЄМСТВА

***Р**озглянуто економіко-математичні умови коректного застосування прогностичної моделі оцінки рівня фінансового стану підприємства в задачах прогнозування рівня фінансового стану підприємства методом парної лінійної регресії в умовах нестабільного розвитку економіки.*

Ключові слова: лінійна регресія, стохастичне моделювання, фінансовий стан.

***C**onsidered economic and mathematical terms correctly applying predictive model assessment of financial situation in problems forecasting the level of financial status of the pair by linear regression in an unstable economy.*

Key words: linear regression, stochastic modeling, financial condition.

Пошук найкращої моделі прогнозування рівня фінансового стану підприємства, яка б дала найменшу помилку прогнозування, має принципове значення при прогнозуванні коротких часових рядів. Серед багатьох видів регресії для прогнозування оцінок рівня фінансового стану на базі спостереження п'яти-шести періодів фінансової діяльності підприємства підходить тільки лінійна регресія при нестабільному розвитку економіки країни в силу значних змін фінансових показників від періоду до періоду. При цьому регресором є час, який вважається дискретною невипадковою змінною.

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), 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) [1]. При цьому помилки не повинні бути незалежними та однаково розподіленими (independent and identically distributed – i.i.d.). Помилки повинні бути тільки некорельованими та гомоскедастичними [3], оскільки четверта 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 в загальному випадку і дозволяє коректно застосовувати лінійну регресійну модель для прогнозування оцінки рівня фінансового стану підприємства.

Мето роботи є аналіз економіко-математичних умов коректного застосування парної лінійної регресії для прогнозування оцінки рівня фінансового стану підприємства з урахуванням того, що регресор є невипадковим, а фінансові показники можуть бути розподілені не за нормальним законом розподілу.

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.

Поперед всього приймається гіпотеза про те, що в генеральної сукупності (ГС) має бути лінійний статистичний зв’язок між відкликом (Y), тобто досліджуваної змінної, та невипадковим регресором, тобто часом (T):

$$Y = \beta_0 + \beta_1 T + u \quad . \tag{1}$$

Таким чином задається лінійна модель у ГС. У цій моделі β_0 і β_1 – значення параметрів моделі в ГС, які нам відомі і не можуть бути визначені у принципі, але можуть бути оцінені статистично з певною точністю (задача математичної статистики саме 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 . У моделі (1) величини *Y, T* та *u* вважаються неперервними, що дозволяє вважати *Y* і *u* розподіленими за певним законом. Таке приближення допустимо 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. До показників, які допускають неперервне представлення в ГС, відноситься більшість абсолютних і відносних показників мікроекономіки, таких як: власний і позиковий капітал, 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-мірна дискретна реалізація неперервної випадкової величини (НВВ), яку досліджує економіст.

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

$$M(u_i) = 0, i = 1 : n \quad . \tag{2}$$

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 у вигляді такої формули:

$$\overline{u}_i = 0 \Leftrightarrow \frac{1}{N}\sum_{m=1}^{N} u_m^{(i)} = 0 \quad \cdot \tag{3}$$

Суттєво (3) являє собою усереднення від реалізації до реалізації або усереднення по ансамблю (cross-sectional averaging). Для прив'язування такого усереднення до реалії економічних досліджень припустимо, що залежна змінна (Y) являє собою 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). Регрессор 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, не можна повторити експеримент для отримання другої реалізації обсягу продажів за той ж період часу. Реально, однак, досліджуваний вид продукції виробляється одночасно кількома виробниками. Тому існує можливість отримати декілька реалізацій обсягу продажів за один і той ж період. Неважко також уявити 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 дискретної реалізацією НВВ за i-й період. Усі допустимі значення цієї ВВ складають ансамбль. Усереднення по ансамблю в цьому випадку має вигляд:

$$\overline{y}_i = \frac{1}{N}\sum_{m=1}^{N} y_i^{(m)} \cdot$$

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

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

можна вважати, що середнє по ансамблю значення i-й компоненти випадкової помилки є незміщена та обґрунтована оцінка її МС:

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

Саме в цьому сенсі слід розуміти першу умову теореми Г-М. 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

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

означає, що дисперсії всіх компонентів випадкової помилки, які розглядаються як НВВ, однакові та скінченні [1, с. 80]. 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 умові (2), друга умова (4) може й не виконуватися на практиці. Ступінь 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 йдеться 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; 4]. Дослідження довели, що розрахунок дисперсії параметрів по теоретичним формулам [1, с. 83], які використовують припущення про нормальність розподілу параметрів моделі по ансамблю, дають перевищене значення дисперсії. У роботах деяких науковців розглянуто вплив гетероскедастичності на точність розрахунку оцінок параметрів лінійної моделі в реальних економічних дослідженнях [4; 5]. Про важливість і складність проблеми врахування гетероскедастичності в задачах прогнозування свідчить той факт, що один із авторів роботи [4], саме Robert Engle, отримав Нобелевську премію за дослідження застосування регресійного аналізу при наявності гетероскедастичності (ARCH modeling technique) в задачах прогнозування. Тим не менш, неврахування гетероскедастичності не завжди спричиняє значні помилки в оцінках параметрів лінійної моделі. Це трапляється тільки тоді, коли дисперсії випадкової помилки змінюються досить значно [5]. Наявність зміни дисперсії помилки з часом можливо оцінити по графіку залишкової помилки після рішення задачі лінійної регресії. Якщо зміни не суттєві, можна не застосовувати спеціальні методи врахування гетероскедастичності [Там само].

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 в наступному періоді значною мірою визначається його значеннями в минулі періоди, що й дозволяє здійснити обґрунтований прогноз його значення в 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 в дійсності точність буде вищою. Якщо ж точність, яка отримана звичайним МНК, недостатня, то необхідно звертатися до таких спеціальних методів оцінки дисперсії параметрів моделі, як бутстрап.

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

Як видно з детального розгляду 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 відомі 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 вибірку, за якою 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 в ГС. При цьому адекватна модель теж невідома. Тому в реальній ситуації в будь-якому випадку слід 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 приведена К. Доугерті [1, с. 74]. У цьому експерименті генеруються n-мірні 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-мірних реалізацій відгуку учиняється $M >> 1$ разів по різним реалізаціям випадкової помилки u із 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 $(\widehat{b}_0, \widehat{b}_1)$. У результаті отримуємо M пар таких оцінок: $\{\widehat{b}_0^{(i)}\}_{i=1:M}$, $\{\widehat{b}_1^{(i)}\}_{i=1:M}$, 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 [Там само, с. 83]:

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

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 ця схема описана лише в загальних рисах, а чисельні розрахунки економічного прикладу приведені не повністю [1, табл. 3.1, с. 76]. Метою приклада 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 нібито підтверджується те, що моделювання за методом Монте-Карло дає той ж самий результат, що і теоретичні формули (6) [Там само]. 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 приклад використовує тільки $M = 10$ експериментів, чого край недостатньо [5]. Прості розрахунки доводять, 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 приведеним в [1] σ_1. Але при цьому не приведено 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 члену b_0. Якщо, наприклад, лінійна модель використовується для визначення постійних витрат [2, с. 221], то моделювання з $M = 10$ приведе до помилки в 10 %, що не допустимо. Тому звичайно використовується число експериментів не менше ніж 50000 [5]. Для розглянутого в [1] приклада, при $M = 50000$ отримаємо 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%-е зміщення оцінки постійних витрат. Якщо така точність не задовольняє аналітика, кількість експериментів слід збільшити. Варто відмітити, що такі зусилля не варті того, щоб тільки показати збіжність результатів моделювання з теоретичними формулами (6) [1]. У дійсності, розглянута 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), крос-валідація (cross-validation) та бутстрап (bootstrap) [Там само].

Проведений аналіз теоретичних положень математичної статистики щодо коректного застосування МНК для оцінки параметрів лінійної регресії довів таке: закон розподілу досліджуваної змінної, також 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.

Для прогнозування оцінок рівня фінансового стану, таким чином, використовується парна лінійна регресійна модель, яка в генеральній сукупності має такий вигляд:

$$R_i = \beta_0 + \beta_1 t_i + \varepsilon_i \ , \tag{7}$$

де R_i є оцінками рівня фінансового стану в періоди t_i $(t_i = t_1, t_2, ..., t_T)$, β_0 та β_1 є коефіцієнтами регресії, які визначають тренд залежності оцінки від часу, а ε_i є випадкова помилка, яка відображає вплив ендогенних та екзогенних факторів на функціонування підприємства. Коефіцієнти регресії β_0 та β_1 знайти не можна в принципі. Замість їх використовують оцінки цих коефіцієнтів за даними вибірки, або, що одно й теж, за розрахунками оцінок рівня фінансового стану, які визначені за фінансовими коефіцієнтами досліджуваного підприємства. Оцінки коефіцієнтів регресії розраховуються на базі вибіркової моделі парної лінійної регресії:

$$R_i = b_0 + b_1 t_i + e_i\,, \qquad t_i = t_1, t_2, \ldots, t_T\,, \tag{8}$$

де b_0 та b_1 є вибірковими коефіцієнтами регресії, а e_i – квазівипадковою помилкою. За конкретними значеннями оцінок рівня фінансового стану в періоди спостереження, значення оцінок коефіцієнтів регресії $\widehat{b}_0$ та $\widehat{b}_1$ знаходяться методом найменших квадратів за такими формулами:

$$\widehat{b}_1 = \frac{\sum_{i=1}^{T}(R_i - \overline{R})(t_i - \overline{t}\,)}{\sum_{i=1}^{T}(t_i - \overline{t}\,)^2} = \frac{S_{Rt}}{S_{tt}}\,, \tag{9}$$

$$\widehat{b}_0 = \overline{R} - \widehat{b}\overline{t_1}\,. \tag{10}$$

У цих формулах, як звичайно, позначено $\overline{R}$ та $\overline{t}$ як середні значення оцінок рівня фінансового стану та періодів спостереження. При цьому тренд часової поведінки оцінки рівня фінансового стану визначається розрахунковою функцією:

$$R^*(t) = \widehat{b}_0 + \widehat{b}_1 t\,. \tag{11}$$

Наявність розрахункової функції усередненої поведінки оцінки рівня фінансового стану від періоду до періоду, яка отримана за даними спостереженої вибірки, дозволяє визначити значення квазівипадкової помилки у кожному періоді:

$$e_i = R_i - \widehat{b}_0 - \widehat{b}_1 t_i\,\cdot \tag{12}$$

Саме існування цієї помилки свідчить про те, що вибіркові оцінки параметрів лінійної регресійної моделі $\widehat{b}_0$ та $\widehat{b}_1$ змінюються від вибірки до вибірки, тобто містять випадкову складову, яка й визначає помилку прогнозування. Незважаючи на те, що досліджується тільки одна вибірка, тобто оцінки рівня фінансового стану одного підприємства за декілька часових періодів, теоретично припускається можливість того, що це могло б бути не саме те підприємство, яке досліджується, а будь-яке інше підприємство з генеральної сукупності підприємств, якому належить досліджуване підприємство. Для будь-якого іншого підприємства з генеральної сукупності вибіркові оцінки параметрів лінійної регресійної моделі $\widehat{b}_0$ та $\widehat{b}_1$ будуть відрізняться від тих значень, (10 та 11), які отримані за даними спостереженої вибірки.

Оскільки прогнозування здійснюється на основі розрахункової функції (11), при змінах параметрів лінійної регресійної моделі $\widehat{b}_0$ та $\widehat{b}_1$ від вибірки до вибірки буде змінюватися орієнтація прямої лінії тренду, від якої залежить швидкість зміни розрахункової оцінки рівня фінансового стану підприємства періоду до періоду. Як наслідок, буде змінюватися величина прогнозного значення оцінки рівня фінансового стану на наступний період, що й призводить до помилки прогнозування. Це означає, що помилка прогнозування цілком визначається дисперсією параметрів лінійної регресійної моделі $\widehat{b}_0$ та $\widehat{b}_1$ від вибірки до вибірки. Розрахунок дисперсій або середніх квадратичних відхилень, проводиться теоретично, оскільки економічні дані не можуть бути багаторазово повторені для вибіркового розрахунку цих дисперсії.

Звичайно при цьому припускається нормальний закон розподілу випадкової помилки в генеральній сукупності. Також припускається нормальний закон розподілу параметрів регресії від вибірки до вибірки. Крім того, припускається відсутність гетероскедастичності та автокореляції. Якщо ці умови виконуються, то вибіркові оцінки параметрів регресії $\widehat{b}_0$ та $\widehat{b}_1$, (10; 11), є незміщеними, тобто їх математичні сподівання збігаються з параметрами лінійної моделі β_0 та β_1 (8) в генеральної сукупності:

$$E(\widehat{b}_0) = \beta_0; \quad E(\widehat{b}_1) = \beta_1. \tag{13}$$

Ці умови є дуже важливими для оцінки помилки прогнозування рівня фінансового стану підприємства з теоретичних засад. Якщо вони виконуються, то можливо розрахувати дисперсії параметрів лінійної моделі в генеральній сукупності за такими формулами:

$$D(b_0) = \sigma_e^2 \left(\frac{1}{T} + \frac{\overline{t}^2}{S_{tt}} \right); \tag{14}$$

$$D(\widehat{b}_1) = \frac{\sigma_e^2}{S_{tt}} \quad , \tag{15}$$

де $S_{tt} = \sum_{i=1}^{T} (t_i - \overline{t})^2$, а σ_e являє собою середнє математичне відхилення випадкової помилки. Як відомо, теоретичні формули такого типу мають суттєвий недолік, який полягає в тому, що дисперсії в генеральній сукупності взагалі невідомі. Тому дисперсію випадкової помилки σ_e^2 в формулах (14) та (15) можливо тільки оцінити за допомогою визначення квазівипадкової помилки (2.22) таким чином:

$$\sigma^2 \approx \sigma_e^2 = \frac{\sum_{i=1}^{T} (R_i - \widehat{b}_0 - \widehat{b}_1 t_i)}{T - 2} \,. \tag{16}$$

Оцінка дисперсії випадкової помилки (16) дозволяє оцінити помилку прогнозування за допомогою побудови довірчого інтервалу на певному ймовірнісному рівні та тестувати відповідну статистичну гіпотезу відносно параметрів лінійної моделі.

Оскільки випадкова помилка вважається розподіленою за нормальним законом з математичним сподіванням рівним нулю та стандартним відхиленням рівним σ_ε, тобто скорочено: $\mathbb{N}\,(0, \sigma_e)$, то такі статистики

$$\frac{\widehat{b}_0 - b_0}{\sigma_\varepsilon \sqrt{\left(\frac{1}{T} + \frac{\overline{t}^2}{S_{tt}}\right)}} \quad \text{та} \quad \frac{\widehat{b}_1 - b_1}{\sigma_\varepsilon}\sqrt{S_{tt}} \tag{17}$$

також розподілені за нормальним законом з математичним сподіванням рівним нулю та стандартним відхиленням рівним одиниці, тобто $\mathbb{N}(0,1)$.

Ураховуючи те, що стандартне відхилення випадкової помилки σ_ε невідомо, його заміняють на вибіркове значення квазівипадкової помилки σ_e:

$$\frac{\widehat{b}_0 - b_0}{\sigma_e \sqrt{\left(\frac{1}{T} + \frac{\overline{t}^2}{S_{tt}}\right)}} \quad \text{та} \quad \frac{\widehat{b}_1 - b_1}{\sigma_e}\sqrt{S_{tt}}\,. \tag{18}$$

У цьому випадку закон розподілу статистик (18) змінюється на закон розподілу Стьюдента з N-2 ступенями свободи.

Перша статистика в (18) використовується для побудови довірчого інтервалу для коефіцієнта регресії b_0. Згідно зі стандартним методом побудови довірчих інтервалів для статистик з відомим законом розподілу, $100(1-\alpha)$ відсотковий довірчий інтервал для коефіцієнта регресії b_0 розраховується за такою формулою:

$$\widehat{b}_0 - t_{\alpha/2,N-2}\sigma_e\sqrt{\frac{1}{T} + \frac{\overline{t}^2}{S_{tt}}} \le b_0 \le \widehat{b}_0 + t_{\alpha/2,N-2}\sigma_e\sqrt{\frac{1}{T} + \frac{\overline{t}^2}{S_{tt}}}\,. \tag{19}$$

Для коефіцієнта регресії b_1, $100(1-\alpha)$ відсотковий довірчий інтервал розраховується за формулою:

$$\widehat{b}_1 - t_{\alpha/2,N-2}\sigma_e / \sqrt{S_{tt}} \le b_1 \le \widehat{b}_1 + t_{\alpha/2,N-2}\sigma_e / \sqrt{S_{tt}}\,. \tag{20}$$

Згідно зі статистичними стандартами, звичайно будуються або 95 %, або 99 % довірчі інтервали. Для 95 % довірчого інтервалу $\alpha = 0{,}05$, а для 99 % довірчого інтервалу $\alpha = 0{,}01$. Критичне значення статистики Стьюдента для 95 % довірчого інтервалу є $t_{0{,}025,T}$, а для 99 % довірчого інтервалу $t_{0{,}005,T}$, де T є число періодів спостереження фінансових коефіцієнтів досліджуваного підприємства.

Довірчі інтервали для параметрів лінійної регресійної моделі показують, в яких межах змінюються прямі лінії тренду $R^*(t) = b_0 + b_1 t$ оцінки рівня фінансового стану підприємства зі ймовірністю $100(1-\alpha)$ відсотків від вибірки до вибірки, що й визначає помилку прогнозування наступним чином.

Позначимо абсолютні значення відхилень b_0 та, b_1 у формулах (19), (20)

$$\delta b_0 = t_{\alpha/2,N-2}\sigma_e\sqrt{\frac{1}{T} + \frac{\overline{t}^2}{S_{tt}}}\;; \qquad \delta b_1 = t_{\alpha/2,N-2}\sigma_e / \sqrt{S_{tt}}\,. \tag{21}$$

Тоді розрахункова функція оцінки рівня фінансового стану підприємства від вибірки до вибірки змінюється зі ймовірністю $100(1-\alpha)$ відсотків, згідно з формулами (19), (20) для довірчих інтервалів, у таких межах.

Нижня $100(1-\alpha)$ відсоткова границя у часі t:

$$R_-^*(t) = \widehat{b}_0 - \delta b_0 + (\widehat{b}_1 - \delta b_1)t \ . \tag{22}$$

Верхня $100(1-\alpha)$ відсоткова границя у часі t:

$$R_-^*(t) = \widehat{b}_0 + \delta b_0 + (\widehat{b}_1 + \delta b_1)t \cdot \tag{23}$$

Прогнозні значення оцінки рівня фінансового стану підприємства в період $t = T+1$ згідно з формулами (22) та (23) є для нижньої границі:

$$R_-^*(T+1) = \widehat{b}_0 - \delta b_0 + (\widehat{b}_1 - \delta b_1)(T+1) \tag{24}$$

і, відповідно, для верхньої границі:

$$R_+^*(T+1) = \widehat{b}_0 + \delta b_0 + (\widehat{b}_1 + \delta b_1)(T+1) \cdot \tag{25}$$

З формул (24) та (25) слідує, що помилка прогнозування в наступний період дорівнює:

$$\delta R(T+1) = (R_+^*(T+1) - R_-^*(T+1)) / 2 = \delta b_0 + \delta b_1(T+1) \tag{26}$$

і визначає прогнозне значення оцінки рівня фінансового стану підприємства в період $t = T+1$ з помилкою:

$$R^*(T+1) = b_0 + b_1(T+1) \pm \delta R(T+1) \ . \tag{27}$$

Крім визначення помилки прогнозування необхідно також з'ясувати значимість параметрів лінійної регресійної моделі. Якщо параметри лінійної регресійної моделі не значимі, то це означає, що в іншій вибірці вони можуть бути настільки іншими, що ні про яке прогнозування не може бути й мови.

Для перевірки параметрів лінійної регресійної моделі на значимість слід задати нульову та конкуруючу гіпотезу, тобто H_0 та H_1. Як правило, ці гіпотези задають таким чином:

$$H_0 : b_1 = 0; \quad H_1 : b_1 \neq 0 \cdot \tag{28}$$

У даному випадку, якщо не можна відхилити нульову гіпотезу H_0, це означає, що між регресором і відкликом, тобто між часом та оцінкою рівня фінансового стану підприємства, ніякого лінійного зв'язку не існує. У цьому випадку не можна взагалі використати лінійну регресійну модель для прогнозування рівня фінансового стану підприємства.

Тест на перевірку нульової гіпотези може бути здійсненим за критеріями Фішера або Стьюдента. Критерій Фішера базується на вихідній вибірковій моделі випадкового явища (10), з якої випливає основна тотожність для дисперсій:

$$S_R^2 = S_{R^*}^2 + S_e^2 \,, \tag{29}$$

за допомогою якої визначається коефіцієнт детермінації

$$R^2 = \frac{S_{R^*}^2}{S_t^2} \tag{30}$$

та статистика Фішера

$$F = (T-2)\frac{R^2}{1-R^2} \,. \tag{31}$$

Справедливість нульової гіпотези перевіряється порівнянням статистики F зі критичним значенням статистики Фішера $F_{\alpha,1,T-2}$. Якщо виповнюється нерівність

$$F > F_{\alpha,1,T-2} \,, \tag{32}$$

то нульова гіпотеза $H_0 : b_1 = 0$ відкидається. У цьому випадку лінійна регресія зі знайденими вибірковими коефіцієнтами регресії (10, 11) вважається значимою на рівні 100α %, що дає підставу обґрунтовано використати розрахункову функцію оцінки рівня фінансового стану підприємства (11) в якості тренду та визначити за її допомогою прогнозного значення оцінки на наступний період за формулою (27).

Результати досліджень довели таке. Для невеликих вибірок ($n = 5 \div 100$) дисперсії параметрів лінійної моделі по імовірності не залежать від закону розподілу випадкової помилки і можуть бути розраховані за теоретичними формулами для нормального закону.

Систематична залежність дисперсії параметрів лінійної моделі від закону розподілу випадкової помилки починається з розміру вибірки ~ 100.

Для невеликих вибірок ($n = 5 \div 100$) із ГС з нормальним розподілом спостерігається значний розкид емпіричних асиметрій, що приводить до значного 95 % довірчого інтервалу для асиметрії, який досить повільно зменшується зі зростанням розміру вибірки.

Розрахунок емпіричної асиметрії для “спостережених” вибірок з ГС, яка розподілена за симетричним, або за асиметричним законом Фішера, доводить, що для невеликих вибірок ($n = 5 \div 100$) значення асиметрії таких вибірок входить в 95 % довірчий інтервал вибірок того ж розміру з нормального закону. Те ж саме спостерігається для вибірок з бета-розподілу. Це означає, що ми не можемо ідентифікувати зі стандартною імовірністю закон розподілу спостереженої вибірки, якщо її розмір менше 100.

Наведені результати дослідження залежності моделювання лінійної регресії від закону розподілу випадкової помилки дозволяють зробити такий висновок: якщо розмір спостереженої вибірки менше 100, то згідно з принципом відсутності достатньої підстави, ми не можемо стверджувати, що вибірка вилучена не з ГС, яка розподілена за нормальним законом, бо нормальний закон розподілу є найбільш поширеним у випадкових економічних явищах.

Такий висновок дає підставу для проведення статистичної інференції на основі нормального закону розподілу, якщо розмір спостереженої вибірки менше 100. Це означає, що для невеликих вибірок дисперсії коефіцієнтів лінійної регресії можна розраховувати за теоретичними формулами нормального розподілу, а довірчі інтервали для параметрів регресії будувати на базі розподілу Стьюдента або нормального закону.

Література:

1. *Доугерти К.* Введение в економетрику / К. Доугерти ; пер. с англ. – М. : ИНФРА-М, 1999. – XIV, 402 с.

2. *Савицкая Г. В.* Экономический аналіз / Г. В. Савицкая. – М. : Новое знание, 2004. – 640 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. – P. 42–48

3. *Engle R. F.* Measuring and testing the impact of news on volatility / R. F. Engle // Journal of Finance. – 1993. – Vol. 48 (5). – P. 1749–1778. – Режим доступу : 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.

Надійшла до редколегії 30.09.2011р.