УДК 004.942
DOI: 10.20535/SRIT.2308-8893.2020.2.05

ЗАСТОСУВАННЯ РЕГРЕСІЙНИХ МОДЕЛЕЙ ДЛЯ АНАЛІЗУ І ПРОГНОЗУВАННЯ ПОКАЗНИКІВ ЯКОСТІ ФІНАНСОВОЇ ДІЯЛЬНОСТІ ПІДПРИЄМСТВА

Н.В. КУЗНЄЦОВА, З.С. ЧЕРНИШ

Анотація. Досліджено задачу прогнозування успішності діяльності компанії на основі її фінансових показників на базі регресійних моделей. Побудовано множину моделей на основі лінійної множинної регресії, авторегресії з ковзним середнім, авторегресії з інтегрованим ковзним середнім та сезонної моделі авторегресії з інтегрованим ковзним середнім для прогнозування абсолютної величини фінансових показників. Проведено експериментальне дослідження на реальних даних і виконано прогнозування на основі регресійних моделей, методу групового урахування аргументів та авторегресійної нейронної мережі. Для прогнозування волатильності фінансового ряду застосовано гетероскедастичні моделі зі змінною волатильністю типу ARCH та GARCH. Застосовано попереднє оброблення даних з використанням методу Хольта–Вінтерса та фільтра Калмана, що дозволило істотно покращити якість моделей і точності прогнозування. Запропоновано і розроблено комбінацію моделей сезонної авторегресії з інтегрованим ковзним середнім та гетероскедастичної, що дало змогу врахувати наявні сезонні ефектів і тренди, притаманні фінансовим рядам, і отримати високі прогнозні оцінки для фінансових показників.

Ключові слова: регресійні моделі, сезонна модель авторегресії з інтегрованим ковзним середнім, лінійна множинна регресія, попереднє оброблення даних, гетероскедастичні моделі.

ВСТУП

Діяльність будь-якого підприємства в умовах конкурентного середовища та впливів ззовні має постійно переглядатися та коригуватися через появу нових факторів, продуктів конкурентів, зміни фінансового обрахунку всередині компанії, зміни нормативної бази тощо. Об'єктивною можливістю постійного віддзеркалення успішності діяльності компанії в ринкових умовах є перевірка та глибинний аналіз фінансової звітності [1–3]. Виконання такого аналізу для працюючого підприємства потребує, по-перше, оброблення великого масиву різноманітних даних, не завжди ідеальних за якістю, і, по-друге, наявності адекватних математичних моделей, які враховують особливості фінансових процесів і впливів на них [4, 5]. Застосування відомого математичного апарату регресійних моделей [1, 4, 5, 6], доповненого засобами попереднього оброблення вхідних даних [3], може надати аналітикам підприємства зручний інструмент для швидкого поточного оцінювання стану і прогнозування фінансових показників. Це дозволить не лише спрогнозувати розвиток діяльності компанії на основі основних економічних факторів, а й оцінити фінансову спроможність компанії до розвитку і модернізації, закупівлі нових засобів та сировини, коригування послідовності та ефективності діяльності її менеджменту.

ПОСТАНОВКА ЗАВДАННЯ

Визначити основні методи і моделі для аналізу та прогнозування фінансових показників. Запропонувати алгоритм побудови прогнозу фінансових показників стійкості компанії із застосуванням регресійних моделей для його подальшої імплементації у систему підтримання прийняття рішень. Виконати моделювання та прогнозування основних показників із застосуванням регресійних моделей, методу групового урахування аргументів (МГУА), нейронних мереж тощо. Оцінити волатильність залишків фінансового процесу на основі гетероскедастичних моделей. Забезпечити можливість попереднього оброблення даних, виконання фільтрації та згладжування даних у разі наявності аномальних, пропущених даних або викидів, застосувати її до реальних даних та порівняти результати прогнозування на побудованих моделях до і після додаткового оброблення вхідних даних.

МЕТОДИ АНАЛІЗУ ФАКТОРІВ ФІНАНСОВОЇ ДІЯЛЬНОСТІ

Одним з найважливіших факторів успішності підприємства є його фінансовий стан, тобто певна сукупність показників, що характеризують його конкурентоспроможність (платоспроможність або кредитоспроможність), використання капіталу та фінансових ресурсів, здатність виконувати зобов'язання перед державою або іншими підприємствами. Існують цілі підрозділи, аудиторські відділи, які здійснюють фінансовий аналіз підприємства як ззовні, так і всередині. Основним етапом фінансового аналізу є саме прогнозування фінансово-економічного стану компанії. Фінансове прогнозування — це процес оцінювання і формування прогнозів, метою яких є ефективна організація діяльності підприємства, передбачення наслідків прийняття певних рішень. Завданнями фінансового прогнозування є оцінка очікуваного обсягу фінансових ресурсів, пошук джерел формування та напрямів ефективного використання фінансових ресурсів, напрацювання рекомендацій для раціональної фінансової стратегії і тактики менеджменту компанії. Для отримання адекватних прогнозів потрібно виконати велику кількість експериментальних досліджень із застосуванням різноманітних методів прогнозування, щоб визначити найбільш імовірне значення або довірчий інтервал для фінансового показника.

Серед методів прогнозування зазвичай виділяють дві категорії: якісні та кількісні. Якісні методи — методи експертних оцінок, метод Делфі, прогнозування на основі очікування — базуються на суб'єктивних оцінках та інтуїтивно-логічному мисленні і використовуються тоді, коли на прогнозований процес впливає велика кількість факторів, усі з яких неможливо врахувати, або за наявності високого ступеня невизначеності та повної відсутності інформації про об'єкт прогнозування. Кількісні методи прогнозування ґрунтуються на математичних методах і підходах, передбачають розроблення математичних моделей, можуть застосовувати каузальні методи (однобічне згладжування, ковзне середнє, подвійне експоненціальне згладжування), методи часових рядів (із застосуванням адитивного або мультиплікативного методу декомпозиції), методи штучного інтелекту (нейронні мережі, мережі Байєса, групові методи оброблення даних, методи опорних векторів тощо).

Регресійні моделі

Моделі авторегресії широко застосовують для опису стаціонарних процесів в економічних дослідженнях. Побудова авторегресійних моделей в економіці заснована на такій важливій властивості рядів економічних явищ і процесів, як взаємозалежність рівнів одного і того ж ряду один від одного. Умова нормальності розподілу ряду для побудови його економічної моделі не є обов'язковою. Регресія — функціональна залежність математичного сподівання залежної змінної від однієї або декількох інших пояснювальних (незалежних) змінних. Серед регресійних моделей можна виділити:

– *однопараметричні моделі* $y = f(x)$ (що залежать від однієї змінної):

- лінійні вигляду $y = \beta_0 + \beta_1 x$;
- нелінійні (експоненціальна, степенева, квадратична, логістична тощо);

– *багатопараметричні моделі* $y = f(x_1, x_2, ..., x_n) + \varepsilon$ (залежність від декількох змінних):

- лінійні моделі вигляду $y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \varepsilon$, де $x_1, x_2, ..., x_n$ — незалежні змінні;
- нелінійні моделі за змінними та параметрами.

Моделі лінійної регресії

Якщо економічний процес можна подати за допомогою кількісних характеристик, між якими існує певна математична залежність, то цей процес може бути описаний в аналітичній формі через рівняння регресії. У такому випадку зв'язок між певними факторами та параметром, що досліджується (прибуток компанії), може бути показаний не лише у вигляді графіка, а і записаний з використанням емпіричної формули. Із використанням такого рівняння на основі набору вхідних ознак і може бути побудований прогноз цільової змінної, тобто виконане прогнозування методами регресійного аналізу.

Лінійна множинна регресія може бути подана таким чином [1]:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \varepsilon,$$

де y — залежна змінна; $x_1, x_2, ..., x_n$ — фактори (незалежні змінні); ε — випадкова похибка. Для кожного i-го спостереження залежність між змінними визначається так:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_n x_{in} + \varepsilon_i .$$

Модель авторегресії з ковзним середнім

На практиці іноді буває доцільним уведення у модель як елементів авторегресії, так і складової ковзного середнього для того, щоб оцінити характеристики часового ряду, використовуючи меншу кількість параметрів. Такий процес отримав назву авторегресії з ковзним середнім (АРКС (p,q) або ARMA (p,q)) [1]:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - ... - \beta_t \varepsilon_{t-p} \ ;$$

$$y_t (1 - \alpha_1 L - \alpha_2 L^2 - ... - \alpha_p L^p) = \varepsilon_t (1 - \beta_1 L - \beta_2 L^2 - ... - \beta_q L^q) \ ,$$

$$\beta(L)\varepsilon_t = \alpha(L) y_t \ ,$$

де $\alpha(L)$ — оператор авторегресії; $\beta(L)$ — оператор ковзного середнього.

Описаний клас моделей є характерним для стаціонарних процесів. Проте на практиці більшість економічних і фінансових рядів є нестаціонарними. Тому необхідно застосовувати інший клас регресійних моделей, що враховують наявність трендової складової або змінної волатильності.

Авторегресійна модель з інтегрованим ковзним середнім (АРІКС)

У 1970 р. Джордж Бокс та Гвілім Дженкінс [2] запропонували методику, яка дозволяє виділити клас нестаціонарних рядів, які за допомогою процедури взяття послідовних різниць можна звести до вигляду АРКС. Якщо після взяття d послідовних різниць ряд зводиться до стаціонарного вигляду, для прогнозування його рівнів можна застосувати комбіновану авторегресійну модель з інтегрованим ковзним середнім — АРІКС (p,d,q) або ARIMA (p,d,q) [3].

Методологія підбору моделі АРІКС для конкретного ряду спостережень складається із чотирьох етапів: вибір моделі, яка найбільше відповідає реальному процесу; оцінювання моделі — використання статистичних методів для отримання оцінок структури і параметрів моделі; тестування моделі — перевірка адекватності з використанням тестів на нормальний розподіл, автокореляцію залишків (тест Дарбіна-Уотсона), на якість специфікації моделі (F-тест); використання моделі для оцінювання прогнозу [1, 3]. Для того щоб оцінити невідомі параметри моделі АРКС (p,q), застосовують звичайний або нелінійний метод найменших квадратів, або метод максимальної правдоподібності. Параметри моделі оцінюються таким чином, щоб сума квадратів залишків була мінімальною.

Сезонна авторегресійна модель з інтегрованим ковзним середнім (САРІКС)

Сезонна модель авторегресії з ковзним середнім, Seasonal ARIMA або SARIMA — це розширення авторегресійної моделі з інтегрованим ковзним середнім, яка дозволяє виконати оброблення одновимірних часових рядів даних із сезонною складовою. Модель включає також три нові гіперпараметри для уточнення порядку авторегресії (AR), диференціювання (I) та ковзного середнього (MA) для сезонної складової ряду, а також додатковий параметр, який показує порядок (періодичність) сезонності. Існують такі елементи тренду, які потребують додаткового налаштування для застосування моделі на практиці: p — порядок авторегресії тренду, d — порядок різниць для вилучення трендів, q — порядок ковзного середнього тренду. Сезонна частина моделі складається з елементів, які також потребують налаштування: P — порядок сезонної авторегресії; D — порядок сезонних різниць; Q — порядок сезонного ковзного середнього; s — кількість часових кроків за один сезонний період. Формально модель SARIMA записується як $SARIMA(p,d,q)(P,D,Q)[s]$ і визначається таким чином:

$$\Phi_P(B^s)\varphi_p(B)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\vartheta_Q(B^s)e_t,$$

де B — оператор зсуву, а

$$\Phi_P(B^s) = 1-\Phi_1 B^s - \Phi_2 B^{2s} - ... - \Phi_P B^{Ps},$$

$$\varphi_p(B) = 1-\varphi_1 B - \varphi_2 B^2 - ... - \varphi_p B^p,$$

$$\theta_q(B) = 1-\theta_1 B - \theta_2 B^2 - ... - \theta_q B^q,$$

$$\vartheta_Q(B^s) = 1-\vartheta_1 B^s - \vartheta_2 B^{2s} - ... - \vartheta_Q B^{Qs}.$$

Ця модель враховує сезонні ефекти, тому авторами запропоновано застосувати такий клас моделей для задачі прогнозування показників фінансової діяльності підприємства.

Моделі зі змінною волатильністю

Процеси, які не є стаціонарними, оскільки мають змінну дисперсію, називають гетероскедастичними. Це означає, що дисперсія процесу змінюється в часі, або є більш складною функцією часу і для створення моделі процесу потрібно знайти закон, за яким дисперсія змінюється. Іноді використовують припущення, що гетероскедастичність має таку форму:

$$\sigma_{\varepsilon(k)}{}^2 = k^2 x^2,$$

де k — константа, яку необхідно оцінити за допомогою експериментальних даних та вибраного методу оцінювання параметрів.

Для моделювання волатильності дохідності фінансових активів на практиці часто використовують моделі умовної гетероскедастичності, її модифікації та узагальнення. *Авторегресійна умовно гетероскедастична модель* (АРУГ) або AutoRegressive Conditional Heteroscedasticity (ARCH) уперше згадується у праці Роберта Енгла [4] і полягає в тому, що дисперсія вільного члена ε_t у деякий момент часу t залежить від квадратів вільних членів попередніх моментів.

Тоді модель ARCH(p) є функцією, яка описує умовну дисперсію за допомогою квадратів попередніх значень спостережуваної величини і записується у вигляді

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-1}^2,$$

де p — максимальний порядок АРУГ; σ_t — умовна дисперсія; y_{t-1} — попередні значення; α_0, α_i — параметри моделі і $\alpha_0 > 0$, $\alpha_i \geq 0, i = 1..p$, оскільки дисперсія завжди має бути додатною, проте за великої кількості лагів ця умова може бути порушена.

Уведемо позначення $v_t = y_t^2 - \sigma^2$ і запишемо модель $\mathrm{ARCH}(p)$ [5]:

$$y_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-1}^2 + v_t \,.$$

Оскільки y_{t-i} і v_t теоретично є незалежними величинами, можна бачити, що модель відповідає моделі AR(p) для квадратів залишків.

Узагальнена авторегресійна умовно гетероскедастична модель (УАРУГ) (Generalized ARCH — GARCH модель) запропонована Тімом Боллерслевом [5] для моделювання волатильності, що залежить від часу. Вона містить значення попередніх умовних дисперсій, завдяки чому замість великих значень p моделі АРУГ(p) використовують невеликі значення p і q. Модель УАРУГ визначає умовну дисперсію як лінійну комбінацію p попередніх квадратів залишків і q лагів попередніх значень умовної дисперсії:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-1}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \,,$$

де α_0, α_i, β_j — параметри моделі, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1..p$, $\beta_j \geq 0$, $j = 1...q$.

Модель УАРУГ (p, q) можна записати у вигляді

$$\sigma_t^2 = \alpha_0 + \alpha(L) y_{t-1}^2 + \beta(L) \sigma_{t-1}^2 \,,$$

де $\alpha(L)$, $\beta(L)$ — поліноми оператора зсуву.

ПІДГОТОВКА ДАНИХ ДО МОДЕЛЮВАННЯ

Реальні фінансово-економічні дані через певні збурення та випадкові впливи на фінансові процеси можуть мати пропуски даних, а самі ряди даних не відповідати очікуваним стійким закономірностям і розподілам. Тоді необхідно виконувати попередній аналіз і оброблення вхідних даних, які дозволяють здійснювати пошук стійких закономірностей, застосовувати методи згладжування або фільтрації даних з метою заміни фактичних значень прогнозними, що мають менший рівень коливань, зашумленості та кількості зайвої інформації. Одним з найпоширеніших оптимальних фільтрів є фільтр Калмана, який застосовується для обчислення оптимальних оцінок вектора стану і побудови короткострокових прогнозів, заснованих на обраних моделях [7]. На практиці досить поширеним є застосування цифрових фільтрів, які дозволяють перетворити дискретний набір вхідних даних x_t у дискретний набір вихідних даних y_t за допомогою лінійного співвідношення вигляду $y_t = \sum_{i=-\infty}^{+\infty} \alpha_i x_{t-i}$. Цифрові фільтри використовують для розв’язання задач стиснення та згладжування даних, усунення трендів, поділу часових рядів на компоненти, пробного оцінювання спектрів. Виділяють декілька видів цифрових фільтрів: ковзного середнього, експоненціального згладжування, поліноміальний фільтр, медіанний фільтр.

Фільтр ковзного середнього — фільтр зі скінченною кількісною характеристикою (на відфільтроване значення впливають лише N останніх значень), що базується на розрахунку середнього на певному ковзному інтервалі:

$$y_t = \sum_{i=0}^{N-1} \alpha_i x_{t+i} ,$$

де x_t — вхідний сигнал; y_t — вихідний сигнал; N — кількість точок, що використовується для обчислення ковзного середнього.

Експоненціальне згладжування (метод Хольта–Вінтерса) — фільтр, для якого на відфільтроване значення впливають усі попередні значення (з нескінченною імпульсною характеристикою). Перевагами фільтра є простота та здатність усувати великі скачки, а також немає потреби зберігати у пам'яті декілька попередніх значень.

Записують експоненціальний фільтр за допомогою рівняння:

$$y_t = \theta x_t + (1-\theta) y_{t-1} ,$$

де x_t — обчислене значення у момент t; y_t — відфільтроване значення у момент t; θ — коефіцієнт фільтрації, що набуває значень від 0 до 1. Серед експоненціальних фільтрів виділяють подвійний експоненціальний фільтр, застосування якого еквівалентне двом фільтрам першого порядку, та нелінійний фільтр, для якого $\theta = \min\left[1, \frac{x_t - y_{t-1}}{R}\right]$, R — коефіцієнт налаштування фільтра.

У цьому експериментальному дослідженні було використано метод найближчих сусідів для відновлення пропусків даних, а після цього застосовано фільтр Калмана та метод Хольта–Вінтерса для фільтрації та згладжування даних [8]. Розглянемо детальніше основні етапи експериментального дослідження, яке проводилось авторами.

ЕКСПЕРИМЕНТАЛЬНЕ МОДЕЛЮВАННЯ ДОХІДНОСТІ КОМПАНІЇ

Для проведення досліджень обрано дані компанії Intel Corporation: звіти її фінансових результатів за 1975–2019 рр. та фінансові коефіцієнти [9], обчислені на їх основі. Для експериментального моделювання використано мову програмування для статистичного оброблення даних R та програмне середовище R Studio, зокрема бібліотеки: tseries (для роботи з часовими рядами), forecast (для побудови моделей типу АР, АРКС, АРІКС), rugarch (для побудови моделей гетероскедастичних процесів), DescTools, Metrics, lmtest (для проведення статистичних тестів).

Аналізувались такі показники:

- Y — Net revenue — виручка, обсяг товарів та послуг, помножений на їх ціну;
- X_1 — Earning per Share (EPS) — чистий прибуток компанії, розділений на середньозважену кількість випущених акцій протягом облікового періоду;
- X_2 — Gross Margin — коефіцієнт валового прибутку або маржинальності — частина загального обсягу виручки компанії, що залишається після відрахування прямих витрат, пов'язаних із виробництвом та реалізацією товару або послуги;

- X_3 — EBIT Margin — показник EBIT-маржинальності, використовується для визначення рентабельності;
- X_4 — Interest coverage ratio — коефіцієнт покриття відсотків — здатність позичальника покривати відсотки за позиками та облігаціями;
- X_5 — Return on Assets (ROA) — показник рентабельності активів, характеризує ефективність використання доступних активів для отримання виручки;
- X_6 — Return on Equiety (ROE) — показник рентабельності власного капіталу, характеризує ефективність використання не всього капіталу підприємства, а тієї частини, що належить його власникам;
- X_7 — Assets Turnover Ratio — показник оборотності активів, визначає ефективність використання виробничих ресурсів.

Виконано прогнозування показника дохідності з використанням лінійної множинної регресії, авторегресії з ковзним середнім, авторегресії з інтегрованим ковзним середнім із застосуванням методів попереднього оброблення даних (зокрема, методів заповнення пропусків, фільтрації, згладжування) і на основі сукупності критеріїв для обрання кращої моделі. Для перевірки адекватності розроблених математичних моделей використовувались коефіцієнт множинної детермінації R^2, *інформаційний критерій Акайке*: $AIC = N\ln\left(\sum_{k=1}^{N} e^2(k)\right) + 2n$ та *критерій Байєса–Шварца*: $BSC = N\ln\left(\sum_{k=1}^{N} e^2(k)\right) + n\ln(N)$, де $n = p + q + 1$ — кількість параметрів моделі (p — кількість параметрів авторегресійної частини моделі); q — кількість параметрів ковзного середнього; N — довжина (потужність) вибірки. Критерії Акайке та Байєса–Шварца пов'язані з квадратом похибок, а тому мають бути мінімальними для кращої моделі. За критерії якості оцінок прогнозів для аналізу точності прогнозування фінансових показників обрано такі характеристики [1]:

середню абсолютну похибку $MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i |$;

середню відсоткову абсолютну похибку $MAPE = \frac{1}{N}\sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{|y_i|} \cdot 100$;

середній квадрат похибок $MSE = E((y - \hat{y})^2) = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}$;

коефіцієнт Тейла: $U = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i)^2} + \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i)^2}}$.

Модель множинної регресії

Першим кроком побудови моделі є визначення основних регресорів. В експериментальному моделюванні було використано метод виключення, який передбачає, що на початковому етапі включаються всі фактори до моделі, оскільки достеменно невідомо, які з них є значущими, а далі послідовно по одному вилучаються фактори і перевіряється, чи покращується якість моделі. У разі, якщо якість моделі погіршується, виведений з моделі фактор повертається назад, а виводиться наступний фактор і т.д. Побудовано кореляційну матрицю і визначено суттєві показники для подальшої побудови моделі: Gross Margin, ICR, ROA, ROE, ANR, Price, які відповідають змінним $X_1, X_2, X_3, X_4, X_5, X_6$. Отримано таку модель множинної регресії:

$$Y = 8771{,}749 X_1 + 10{,}165 X_2 - 166{,}813 X_3 + 136{,}720 X_4 - 32{,}718 X_5 +$$

$$+ 206{,}680 X_6 + 2755{,}078 + \varepsilon$$

Значення критеріїв адекватності побудованої моделі: $R^2 = 0{,}824$; $DW = 0{,}676$; $AIC = 3324{,}81$; $BIC = 3350{,}3532$ показують, що лінійна модель є незадовільною, а оцінки прогнозів моделі виявились низькими, а тому вирішено не використовувати цю модель для подальшого дослідження.

Побудова авторегресійних моделей

Побудова авторегресійних моделей для прогнозування фінансових показників на прикладі компанії Intel Corporation проводилась згідно з етапами, показаними на рис. 1.

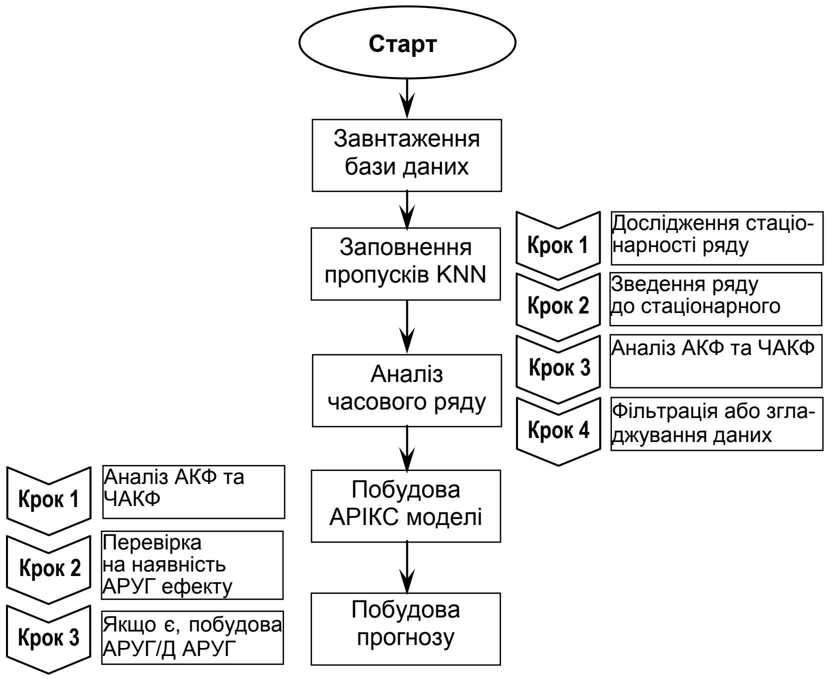

Рис. 1. Алгоритм побудови прогнозу фінансових показників на основі авторегресійних моделей

Крок 1. Дослідження стаціонарності ряду. Використовують такі статистичні тести, як розширений тест Дікі–Фуллера (ADF тест), Філліпса–Перрона (PP тест), Квятковського–Филліпса–Шмідта–Шина (KPSS тест) [3]. На основі отриманих експериментальних результатів для ряду цільової змінної Net Revenue отримано значення статистичних тестів, що свідчать про нестаціонарність ряду і необхідність використання більш складних моделей.

Крок 2. Візуальний аналіз ряду показав, що ряд є нестаціонарним, у моделі ймовірно наявний тренд першого або навіть другого порядку, тому ряд потребує певних додаткових маніпуляцій та оброблення. Існують декілька методів перетворення часового ряду (логарифмування, диференціювання — перехід до попарних різниць сусідніх значень ряду, та послідовне логарифмування із диференціюванням), метою яких є зведення ряду до стаціонарного. У випадках аналізу економічних та фінансових процесів найчастіше застосовують перетворення типу логарифмування, а потім беруть першу різницю. Це зумовлено тим, що для таких часових рядів характерне експоненціальне зростання, яке можна згладити за допомогою операції логарифмування, у той час як диференціювання стабілізує середнє значення ряду і дозволяє позбутися тренду та сезонності. Для часового ряду виконано відповідні перетворення, щоб звести його до стаціонарного (рис. 2).

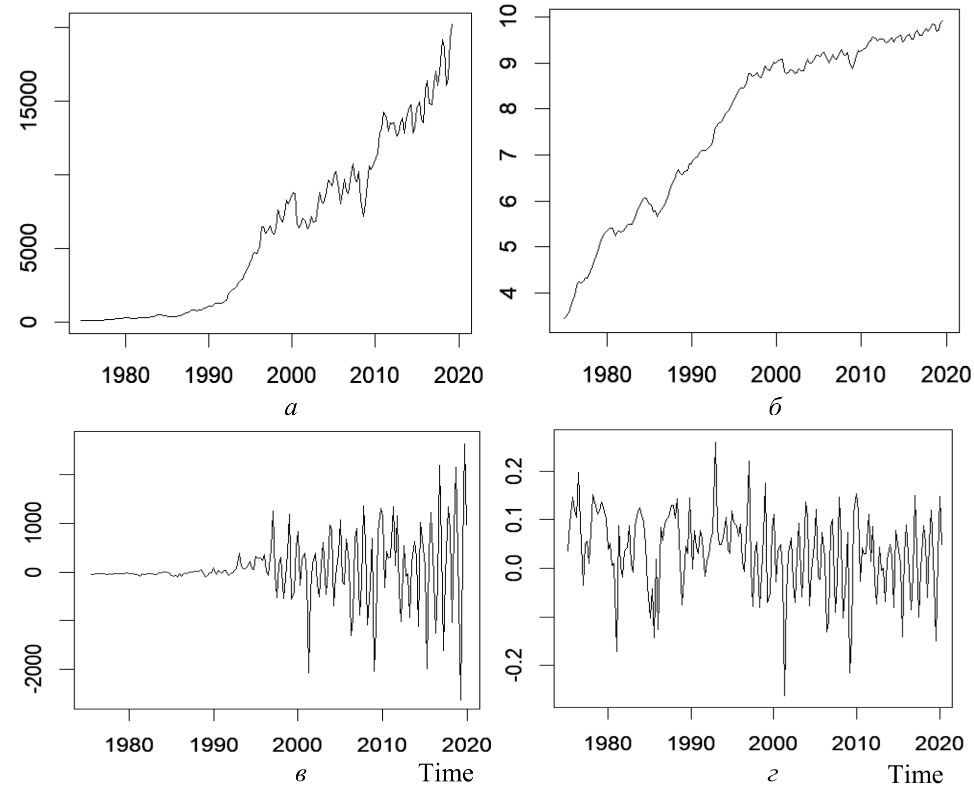

Рис. 2. Часовий ряд зміни фінансового показника Net Revenue: *а* — вхідні дані; *б* — прологарифмований Net Revenue; *в* — продиференційований Net Revenue; *г* — прологарифмований та продиференційований Net Revenue

Крок 3. Автокореляційна функція (АКФ) та часткова автокореляційна функція (ЧАКФ) прологарифмованого і прологарифмованого та продиференційованого ряду мають вигляд, показаний на рис. 3.

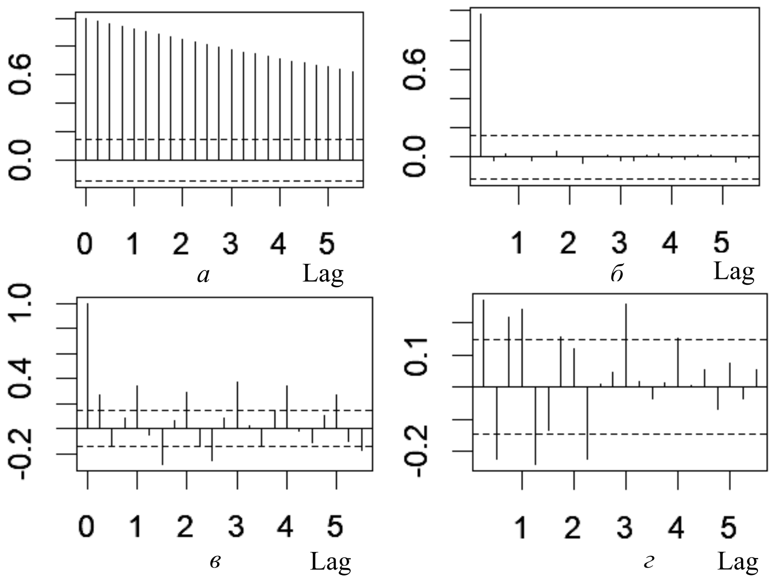

Рис. 3. а — АКФ; *б* — ЧАКФ прологарифмованого ряду; *в* — АКФ; *г* — ЧАКФ прологарифмованого та продиференційованого ряду

Аналіз ЧАКФ прологарифмованого та продиференційованого ряду показав, що значущими для моделі є лаги 0,1, 2 і 3, тому відповідні лаги і мають бути включені в модель, а відповідні порядки моделей бути проаналізовані. Оскільки візуальний аналіз ряду свідчив про наявність тренду, то було виконано побудову низки моделей авторегресії з інтегрованим ковзним середнім. Специфіка обраних вхідних даних свідчить про наявність певної сезонної складової. Тому авторами було вирішено врахувати сезонну складову і побудувати сукупність сезонних авторегресійних моделей. Результати аналізу якості побудованих авторегресійних моделей наведено у табл. 1.

Таблиця 1. Порівняльний аналіз побудованих авторегресійних моделей

Модель	**AIC**	**BIC**
ARIMA(1,1,1)	-373,157	-363,6626
ARIMA(2,1,1)	-371,1632	-358,5041
ARIMA(0,1,2)	-373,2225	-363,7282
ARIMA(0,2,1)	-373,0245	-366,7064
ARIMA(1,2,0)	-315,3137	-308,9956
SARIMA(0,2,1)(2,0,1)[4]	-406,4964	-390,7011

Отже, найкращою моделлю виявилася сезонна авторегресійна модель з інтегрованим ковзним середнім SARIMA(0,2,1)(2,0,1)[4]. Саме рівняння сезонної авторегресійної моделі після перетворення має вигляд: $(1-\Phi_1 B^4 - \Phi_2 B^8)(1-B)^2 Z_t = (1-\theta_1 B)(1-\vartheta_1 B^4)e_t$ з такими значеннями її параметрів за сезонної та несезонної складових, відповідно: $\mathrm{sar1} = 0{,}9005$, $\mathrm{sar2} = 0{,}0506$, $\mathrm{ma1} = -0{,}9849$, $\mathrm{sma1} = -0{,}7649$.

Крок 4 —перевірка моделі на сталість дисперсії її залишків. Наявність авторегресійної умовної гетероскедастичності у залишків моделі можна перевірити за допомогою тестів Бокса та Льюнга–Бокса, нульовою гіпотезою яких є випадковий розподіл даних. Значення критерію виявилося меншим за 0,05, а отже, на рівні впевненості 95% можна стверджувати, що для залишків характерна взаємна корельованість. Побудовано гетероскедастичні моделі (ARCH/GARCH) для залишків моделі SARIMA (табл. 2).

Таблиця 2. Порівняння гетероскедастичних моделей ARCH/GARCH

Модель	AIC	BIC
ARCH(1)	-2,4089	-2,3729
ARCH(2)	-2,4106	-2,3566
GARCH(1,1)	-2,3968	-2,3427
GARCH(1,2)	-2,3855	-2,3134
GARCH(2,1)	-2,3854	-2,3134
GARCH(2,2)	-2,3741	-2,2841

Найкращою моделлю за сукупністю критеріїв виявилася модель ARCH(2) із характеристиками: omega=0,004598, alpha1=0,018598, alpha2=0,101433 [1].

Виконано прогнозування показника Net Revenue на наступні 4 квартали з використанням регресійних моделей: обраної кращої моделі SARIMA(0,2,1)(2,0,1)[4], комбінації моделей SARIMA(0,2,1)(2,0,1)[4] + + ARCH(2) для прогнозування абсолютного значення дохідності і опису залишків моделі SARIMA, і порівняння з іншими методами інтелектуального аналізу даних, зокрема моделі на основі МГУА та нейронних мереж [10], для прогнозування показника дохідності компанії Intel Corporation за 2019 р. Результати прогнозування, отримані за всіма методами, та їх порівняння з реальними значеннями наведено у табл. 3.

Таблиця 3. Порівняння якості оцінок прогнозу різними методами інтелектуального аналізу даних для прогнозування показника дохідності

Модель	MSE	MAE	MAPE	U
Без попереднього оброблення і згладжування даних				
SARIMA(0,2,1)(2,0,1)[4]	0,00386	0,05109	0,521%	0,00632
SARIMA(0,2,1)(2,0,1)[4] + ARCH(2)	0,00413	0,05152	0,526%	0,00653
МГУА	0,00948	0,08847	0,9%	0,00991
Авторегресійна нейронна мережа	0,00874	0,09063	0,926%	0,00955
Із застосуванням методів попереднього оброблення і згладжування даних				
Метод Хольта–Вінтерса SARIMA(1,2,2)(0,0,1)[4]	0,01882	0,13532	1,364%	0,01383
Метод Хольта–Вінтерса і SARIMA(1,2,2)(0,0,1)[4] + ARCH(4)	0,02647	0,16069	1,615%	0,01636
Метод Хольта–Вінтерса і МГУА	0,00402	0,05246	0,533%	0,00645
Метод Хольта–Вінтерса і авторегресійна нейронна мережа	0,00367	0,05085	0,517%	0,00617
Метод Калмана і SARIMA(2,2,3)(1,0,1)[4]	4,2e-05	0,00485	0,049%	0,00066
Метод Калмана і SARIMA(2,2,3)(1,0,1)[4] + ARCH(4)	6,04e-05	0,00568	0,058%	0,00079
Метод Калмана і МГУА	0,00095	0,02568	0,2632%	0,00316
Метод Калмана і авторегресійна нейронна мережа	0,00045	0,01633	0,167%	0,00217

Отже, найкращі результати прогнозування дохідності компанії без попереднього оброблення і згладжування отримано за допомогою сезонної моделі SARIMA(0,2,1)(2,0,1)[4], а найгірші — для моделі, побудованої на основі методу групового урахування аргумента. Найкращою моделлю після попереднього оброблення виявилась модель на основі застосування фільтрації методом Калмана і SARIMA(2,2,3)(1,0,1)[4]. Результати моделювання підтвердили доцільність застосування попереднього оброблення даних, якість прогнозування фінансових показників підвищилась на декілька порядків і досягла точності до п'ятого знака після коми.

ВИСНОВКИ

Для розв'язання задачі вибору найкращих (адекватніших) моделей аналізу фінансової діяльності виконано аналіз квартальних звітів про фінансові результати компанії Intel Corporation за 1975–2019 рр. і розроблено такі математичні моделі: множинної регресії, авторегресії з інтегрованим ковзним середнім; сезонної авторегресії з інтегрованим ковзним середнім, моделі з умовною гетероскедастичністю, моделі на основі методу групового урахування аргументів та нейронних мереж. Було проведено дослідження, у яких попередньо оброблено вхідні дані із застосуванням згладжування та фільтрації даних методом Хольта–Вінтерса і фільтра Калмана, а пропуски заповнено методом найближчих сусідів. Під час проведення обчислювальних експериментів та розрахунків основних показників показано, що таке попереднє оброблення дало значне покращення адекватності всіх описаних моделей і дозволило підвищити прогнозні оцінки. Комбінація гетероскедастичних та сезонних авторегресійних моделей з інтегрованим ковзним середнім дозволяє здійснювати прогнозування для нестаціонарних рядів, характерних для фінансово-економічної діяльності, враховувати сезонні ефекти та отримувати прогнозні оцінки високої якості. Перспективним автори вважають застосування комбінованих моделей і інтегрованого підходу, запропонованого у праці [11]. Обраний у роботі математичний апарат регресійних моделей та проведене експериментальне дослідження показали ефективність і доцільність їх застосування для прогнозування макроекономічних та мікроекономічних індексів.

ЛІТЕРАТУРА

1. П.І. Бідюк, В.Д. Романенко, та О.Л. Тимощук, *Аналіз часових рядів: навч. посіб.* Київ: НТУУ «КПІ», 2013.
2. Дж. Бокс и Г. Дженкинс, *Анализ временных рядов, прогноз и управление*, пер. с англ. Москва: Мир, 1974.
3. Н.В. Кузнєцова та П.І. Бідюк, *Теорія і практика аналізу фінансових ризиків: системний підхід: моногр.* Київ: Вид-во «Ліра-К», 2020.
4. R.F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, vol. 50, pp. 987–1007, 1982.
5. T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.
6. Л.Г. Саетова, "Основные модификации ARCH моделей", *Проблемы экономики и менеджмента*, № 6(34), с. 61–63, 2014.

7. М.З. Згуровский и В.Н. Подладчиков, *Аналитические методы калмановской фильтрации для систем с априорной неопределенностью*. Киев: Наукова думка, 1995.
8. N.V. Kuznietsova and P.I. Bidyuk, "Business Intelligence Techniques For Missing Data Imputation", *Research bulletin of NTUU "KPI"*, no. 5, pp. 47–56, 2015.
9. Th.R. Robinson, H. Greuning, E. Henry, and M.A. Broihahn, *International Financial Statement Analysis*. New Jersey, USA: John Wiley & Sons, Inc, 2009, pp. 864.
10. М.З. Згуровский и Ю.П. Зайченко, *Основы вычислительного интеллекта*. Киев: Наукова Думка, 2013.
11. Н.В. Кузнєцова, "Інтегрований підхід до оцінювання кредитних ризиків", *Тр. Одес. политехн. ун-та,* вып. 1(33)–2(34), с. 187–192, 2010.

Надійшла 16.07.2020

INFORMATION ON THE ARTICLE

N.V. Kuznietsova, ORCID: 0000-0002-1662-1974, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: natalia-kpi@ukr.net.

Z.S. Chernysh, ORCID: 0000-0002-5589-0018, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: zlata.chernysh@gmail.com.

REGRESSION MODELS APPLICATION FOR ANALYSIS AND FORECASTING OF THE FINANCIAL ACTIVITY QUALITY INDICATORS OF THE COMPANY / N.V. Kuznietsova, Z.S. Chernysh

Abstract. The company's success forecasting problem based on its financial indicators by regression models was studied in this research. Models based on linear multiple regression, autoregression with moving average, autoregression with integrated moving average, and seasonal model of autoregression with integrated moving average were built to predict the absolute value of financial indicators. An experimental study was performed on real data, and forecasting was made based on regression models. The models based on the method of group method of data handling and autoregressive neural network were developed. Heteroskedastic models with variable volatility such as ARCH and GARCH type were used to predict the volatility of the financial series. Preliminary data processing using the Holt-Winters method and the Kalman filter were applied to improve the model's quality and forecasting accuracy significantly. Authors suggested and developed a combination of seasonal autoregression with integrated moving average and heteroskedastic models that allowed them to consider the seasonal effects and trends inherent in the financial series and obtain high forecasts for financial indicators.

Keywords: regression models, seasonal autoregression model with integrated moving average, linear multiple regression, data processing, heteroskedastic models.

ПРИМЕНЕНИЕ РЕГРЕССИОННЫХ МОДЕЛЕЙ ДЛЯ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ФИНАНСОВОЙ ДЕЯТЕЛЬНОСТИ КОМПАНИИ / Н.В. Кузнецова, З.С. Черныш

Аннотация. Решена задача прогнозирования успешности деятельности компании на основе ее финансовых показателей на базе регрессионных моделей. Построено множество моделей: линейная множественная регрессия, авторегрессия со скользящим средним, авторегрессия с интегрированным скользящим средним и сезонная модель авторегрессии с интегрированным скользящим средним для прогнозирования абсолютных величин финансовых показателей. Проведено экспериментальное исследование на реальных данных и выполнено прогнозирование на базе регрессионных моделей, метода группового учета аргументов и авторегрессионной нейронной сети. Для прогнозирования волатильности финансового ряда применены гетероскедастических модели с переменной волатильностью типа ARCH и GARCH. Выполнена предварительная обработка данных с использованием метода Хольта–Винтерса и фильтра Калмана, что позволило существенно улучшить качество моделей и точность прогнозирования. Предложена и разработана комбинация моделей сезонной авторегрессии с интегрированным скользящим средним и гетероскедастических, что позволило учесть имеющиеся сезонные эффекты и тренды, присущие финансовым рядам, и получить высокие прогнозные оценки для финансовых показателей.

Ключевые слова: регрессионные модели, сезонная модель авторегрессии с интегрированным скользящим средним, линейная множественная регрессия, предварительная обработка данных, гетероскедастические модели.

REFERENCES

1. P.I. Bidyuk, V.D. Romanenko, and O. L. Timoshchuk, *Time Series Analysis*. Kyiv: Polytechnika, NTUU "KPI", 2013.
2. G.E.P. Box and G.M. Jenkins, *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day, 1970.
3. N.V. Kuznietsova and P.I. Bidyuk, *Theory and practice of financial risk analysis: systemic approach.* Kyiv: Lira-K, 2020.
4. R.F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", in *Econometrica*, vol. 50, pp. 987–1007, 1982.
5. T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity", in *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.
6. L.G. Saetova, "The Main Modifications ARCH Models", in *Problems of Economics and Management*, no. 6 (34), pp. 61–63, 2014.
7. M.Z. Zgurovsky and V.N. Podladchikov, *Analytical methods of Kalman filtering for systems with a priori uncertainty*. Kiev: Naukova Dumka, 1995.
8. N.V. Kuznietsova and P.I. Bidyuk, "Business Intelligence Techniques For Missing Data Imputation", in *Research bulletin of NTUU "KPI"*, no. 5, pp. 47–56, 2015.
9. Th.R. Robinson, H. Greuning, E. Henry, and M.A. Broihahn, *International Financial Statement Analysis*. New Jersey: John Wiley & Sons, Inc. 2009.
10. M.Z. Zgurovsky and Y.P. Zaychenko, *The Fundamentals of Computational Intelligence*. Kyiv: Naukova Dumka, 2013.
11. N.V. Kuznyetsova, "Integrated approach to credit risks estimation", in *Proceedings of Odessa Polytechnic University*, no. 1(33)–2(34), pp. 187–192.