

Т. В. ШАБЕЛЬНИК, д-р екон. наук, проф.

С. О. ЄВСЄЄВА, студентка

# КВАНТИТАТИВНЕ МОДЕЛЮВАННЯ ПОРТФЕЛЬНИХ СТРАТЕГІЙ ІЗ ВИКОРИСТАННЯМ ALPHA-СИГНАЛІВ НА ОСНОВІ МАШИННОГО НАВЧАННЯ ТА БЕКТЕСТИНГУ

**Резюме.** У статті запропоновано квантитативну модель побудови ML-based alpha-сигналів та їхньої інтеграції в процедуру портфельного бектестингу в межах середньострокового інвестиційного горизонту. Дослідження базується на щоденних фінансових часових рядах ліквідних інструментів фондового ринку США та передбачає формування структурованого простору ознак, що охоплює цінові (price-based), волатильні (volatility-based) та агреговані внутрішньоденні (intraday-aggregated) характеристики. Навчання моделей організовано в часово узгодженому режимі з використанням rolling або expanding схем і строгого розділення train/validation/test вибірок, що мінімізує витік інформації та забезпечує коректність оцінювання прогностичної здатності.

Alpha-сигнали генеруються в дискретні моменти ребалансування як умовні оцінки очікуваної дохідності на заданому горизонті та інтегруються в процедуру формування портфеля. Ефективність підходу оцінюється через портфельний бектестинг із використанням стандартних ризик-орієнтованих метрик. Запропонований алгоритмічний pipeline поєднує машинне навчання та портфельний аналіз в єдину експериментальну систему та забезпечує методологічне підґрунтя для оцінювання інвестиційної доцільності ML-based стратегій.

**Ключові слова:** кількісне моделювання, машинне навчання, alpha-сигнали, портфельний бектестинг, фондовий ринок, середньострокові стратегії, ризик-орієнтовані метрики.

## ВСТУП

Сучасні фінансові ринки характеризуються високим рівнем волатильності, частими змінами ринкових режимів і зростаючою складністю взаємозв'язків між фінансовими активами. За таких умов традиційні підходи до аналізу та управління інвестиційними портфелями, засновані на статичних припущеннях щодо розподілу дохідностей і стабільності кореляцій, дедалі частіше виявляються недостатніми для прийняття ефективних інвестиційних рішень. Це зумовлює зростання інтересу до квантитативного аналізу, що розглядає фінансовий ринок як динамічну стохастичну систему та ґрунтується на аналізі великих масивів даних і статистично значущих закономірностей.

Одним із ключових напрямів розвитку сучасних квантитативних підходів є використання методів машинного навчання (ML) для побудови прогнозних alpha-сигналів. На відміну від класичних моделей, зорієнтованих переважно на оцінювання середніх характеристик портфеля, ML-підходи дають змогу формувати сигнали, що враховують нелінійні залежності, зміну ринкових режимів і складну динаміку фінансових часових рядів. Особливого значення набувають alpha-

сигнали, які оцінюють не лише напрямок руху ціни, а й очікувану величину майбутньої дохідності активів, що створює передумови для їх безпосереднього використання в задачах портфельної оптимізації та динамічного ребалансування.

Водночас застосування ML у фінансових дослідженнях часто супроводжується надмірним акцентом на таких аналітичних метриках якості прогнозу, як середньоквадратична похибка або коефіцієнт детермінації. Хоча ці показники є важливими для оцінювання статистичної коректності моделей, вони не відображають реальної інвестиційної цінності побудованих alpha-сигналів. У практиці управління активами критичним є не лише сам факт точного прогнозування, а й здатність сигналів забезпечувати стійку портфельну ефективність з урахуванням динаміки ринку, ребалансування та транзакційних витрат.

У цьому контексті портфельний бектестинг набуває ролі необхідного методологічного продовження аналітичних метрик якості прогнозу. Саме бектестинг дає змогу оцінити економічну доцільність використання ML-alpha-сигналів, перевірити їхню стійкість у різних ринкових умовах і виявити потенційні ризики, пов'язані зі

зміною ринкових режимів. Навіть статистично стабільні  $\alpha$ -сигнали можуть втрачати ефективність за умов структурних зламів на ринку, що зумовлює необхідність регулярного перегляду та повторного калібрування моделей у процесі бектестингу.

Таким чином, поєднання квантитативного аналізу, ML і портфельного бектестингу формує цілісний підхід до дослідження інвестиційних стратегій, зорієнтований не лише на якість прогнозування, а й на практичну інвестиційну ефективність. Саме такий підхід відповідає сучасним вимогам до кількісних досліджень у фінансах і створює передумови для побудови адаптивних і економічно обґрунтованих інвестиційних рішень.

### ПОСТАНОВКА ПРОБЛЕМИ

Попри активний розвиток квантитативних методів аналізу фінансових ринків і широке впровадження алгоритмів ML у сфері управління активами, у сучасних дослідженнях залишається низка нерозв'язаних методологічних проблем. Значна частина праць зосереджена на підвищенні точності прогнозування цінних рухів або напрямку дохідності фінансових активів, причому інвестиційна інтерпретація отриманих результатів часто залишається поза межами аналізу. У результаті висока статистична якість прогнозів не завжди трансформується в практично значущу портфельну ефективність.

Класичні підходи до портфельного аналізу, зокрема моделі середньодисперсійної оптимізації, базуються на статичних оцінках очікуваної дохідності та ризику й передбачають відносну стабільність ринкових параметрів у часі. Проте в умовах змінних ринкових режимів, структурних зламів і дедалі більшої ролі нелінійних залежностей такі моделі демонструють обмежену адаптивність. Це зумовлює потребу в переході від статичних портфельних рішень до динамічних підходів, заснованих на прогнозних  $\alpha$ -сигналах, які оновлюються відповідно до поточної ринкової інформації.

Водночас застосування методів ML у фінансових дослідженнях часто супроводжується фокусом на таких аналітичних метриках якості прогнозу, як похибка регресії чи коефіцієнти поясненої дисперсії. Хоча ці показники є важливими для оцінювання статистичних властивостей моделей, проте вони не враховують портфельний контекст використання  $\alpha$ -сигналів, зокрема вплив ребалансування, часової стабільності сигналів та взаємодії між активами. Унаслідок цього виникає розрив між результатами модельної оцінки та реальною інвестиційною ефективністю.

Окрему проблему становить вибір часової шкали аналізу. Більшість досліджень або зосереджуються на короткострокових внутрішньоденних стратегіях, що потребують спеціалізованих даних і інфраструктури, або розглядають довгострокові інвестиційні горизонти, де вплив прогнозних сигналів суттєво згладжується. Натомість середньостроковий інвестиційний горизонт, характерний для практики кількісного управління портфелями, часто залишається недостатньо формалізованим з точки зору побудови та оцінювання ML- $\alpha$ -сигналів.

Таким чином, наукова проблема полягає у відсутності узгодженого підходу до формування ML-based  $\alpha$ -сигналів, зорієнтованих на оцінювання очікуваної величини дохідності фінансових активів, а також у недостатній інтеграції аналітичних метрик якості прогнозу з портфельним бектестингом як інструментом оцінювання їхньої інвестиційної доцільності. Прикладна проблема полягає в необхідності розроблення методології, що дає змогу пов'язати результати ML із практичними портфельними рішеннями в межах середньострокового інвестиційного горизонту.

### АНАЛІЗ АКТУАЛЬНИХ ДОСЛІДЖЕНЬ ТА ПУБЛІКАЦІЙ

У сучасній літературі з квантитативних фінансів та управління активами простежується стійкий перехід від класичних факторних і середньодисперсійних моделей до підходів, у яких очікувана дохідність (expected return) розглядається як прогнозована величина, що може бути оцінена на основі широкого набору предикторів із використанням методів ML. Одним із найвпливовіших напрямів є застосування ML для задачі емпіричного ціноутворення активів (empirical asset pricing), де ціллю є прогнозування премій за ризик і побудова сигналів, що мають економічну цінність у портфельному контексті. У праці S. Gu, B. Kelly та D. Xiu показано, що низка ML-алгоритмів здатна формувати прогнози очікуваної дохідності на основі великої кількості характеристик (firm characteristics) та забезпечувати істотні економічні вигоди інвестору, що підкреслює перспективність ML-based  $\alpha$ -сигналів як інструменту для прийняття інвестиційних рішень [1].

Паралельно зі зростанням популярності ML-методів виник критичний дискурс щодо їх коректного використання в asset pricing та factor zoo. У сучасних оглядових дослідженнях наголошується, що гнучкість ML може підвищувати прогнозу здатність, але водночас загострює ризику перенавчання, нестабільності результатів і помилкових висновків за умов множинного

тестування гіпотез і слабкої відтворюваності. Такі дослідження підкреслюють необхідність методологічно строгих процедур відбору ознак, валідації та інтерпретації результатів, оскільки статистична “успішність” моделі не гарантує її стійкої економічної ефективності [2].

Окремий вагомий пласт досліджень формує прикладна методологія побудови інвестиційних стратегій на основі ML-сигналів, де акцент переноситься з вибору “найточнішої” моделі на правильну організацію дослідження: структурування даних, контроль витоків інформації (look-ahead bias), уникнення overfitting, коректну постановку експерименту та повноцінний бектестинг. Праці M. López de Prado систематизують ці принципи та розглядають backtesting як центральний елемент процесу перетворення статистичного сигналу на практично релевантну інвестиційну стратегію [3].

Критично важливо, що в професійній та академічній літературі з backtesting підкреслюється проблема data mining / multiple testing: у середовищі, де генерується багато ідей і тестуються численні варіації стратегій, навіть оптимістичні показники з історичного тесту можуть бути випадковим результатом. Водночас C. R. Harvey та Y. Liu пропонують статистичний підхід до корекції (haircut) оцінених показників ефективності, пояснюючи, чому “історичний” коефіцієнт Шарпа (Sharpe ratio) часто потребує зниження для отримання реалістичної оцінки інвестиційної якості стратегії [4].

Ще один напрям досліджень стосується вдосконалення метрик оцінювання результатів бектестингу. Так, D. H. Bailey та M. López de Prado вводять імовірнісне трактування Sharpe ratio (Probabilistic Sharpe Ratio), що враховує невизначеність оцінки та відхилення дохідностей від нормальності. Це дає змогу більш коректно порівнювати стратегії та інтерпретувати значущість отриманого показника з урахуванням тривалості трек-рекорду та властивостей розподілу дохідностей [5].

Окрім того, сучасні публікації демонструють, що ML-підходи до прогнозування премій за ризик можуть узагальнюватися на ширший набір ринків та активів (наприклад, міжнародні вибірки), однак питання стійкості сигналів у часі, перенесення результатів між ринками та залежності від ринкових режимів залишається ключовим. Дослідники S. Gu, B. Kelly та D. Xiu підкреслюють економічну значущість ML-прогнозів, але водночас актуалізується потреба в обережній валідації та порівнянності результатів [1].

Отже, аналіз актуальних досліджень дає змогу зробити два принципові висновки. По-перше, сучасна література підтверджує пер-

спективність ML-based alpha-сигналів, зорієнтованих на оцінювання очікуваної величини дохідності як безперервної прогнозовної змінної, а не лише напрямку руху ціни [1; 6; 10]. По-друге, надійне оцінювання практичної цінності таких сигналів вимагає методологічно коректного процесу дослідження, у якому портфельний бектестинг є необхідним етапом поряд із аналітичними метриками якості прогнозу, з урахуванням ризиків множинного тестування та статистичної невизначеності показників ефективності [3–5]. Саме ці положення формують науково-методичне підґрунтя для подальшого обґрунтування запропонованого підходу в межах цього дослідження.

**Метою** статті є розроблення та обґрунтування квантитативного підходу до формування ML-based alpha-сигналів, зорієнтованих на оцінювання очікуваної величини дохідності фінансових активів, а також визначення ролі портфельного бектестингу як необхідного методологічного продовження аналітичних метрик якості прогнозу.

Досягнення поставленої мети передбачає аналіз сучасних підходів до використання методів ML у задачах прогнозування дохідності активів, формування безперервних alpha-сигналів на основі щоденних фінансових даних із використанням агрегованих внутрішньоденних ознак, а також дослідження їхньої практичної інвестиційної цінності в межах середньострокового інвестиційного горизонту. Особлива увага приділяється узгодженню результатів статистичної оцінки прогнозних моделей з результатами портфельного бектестингу, що дає змогу оцінити стійкість та економічну доцільність застосування запропонованого підходу.

## ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

Запропонований у дослідженні підхід ґрунтується на ідеї інтеграції методів квантитативного аналізу та ML у єдиний модельний комплекс, зорієнтований на формування прогнозних alpha-сигналів і подальшу оцінку їх інвестиційної цінності в портфельному контексті. На відміну від класичних портфельних моделей, у яких очікувана дохідність активів задається статично або оцінюється на засадах історичних середніх значень, у цьому дослідженні очікувана дохідність розглядається як динамічна величина, що може бути оцінена за допомогою ML-моделей на основі поточної ринкової інформації.

Ключовим елементом запропонованого підходу є чітке розмежування ролей між окремими компонентами модельного процесу. Методи ML використовуються лише для побудови alpha-сигналів, тобто прогнозних оцінок очікуваної

величини дохідності фінансових активів на заданому середньостроковому часовому горизонті. Формування торговельних або інвестиційних рішень безпосередньо ML-моделями не передбачається. Натомість alpha-сигнали виступають інформаційним входом для подальшого портфельного аналізу.

У межах підходу реалізується послідовний ланцюг оброблення даних, який передбачає декілька логічно пов'язаних етапів: збір і підготовку фінансових даних, побудову системи ознак, навчання ML-моделей, генерацію alpha-сигналів, формування портфельних рішень і проведення портфельного бектестингу. Така модульна структура дає змогу забезпечити прозорість дослідження, контроль потенційних методологічних помилок і надає можливість адаптації окремих компонентів без порушення цілісності всієї системи.

Особливістю підходу є орієнтація на середньостроковий інвестиційний горизонт, що відповідає практиці квантитативного управління портфелями. Частота прийняття інвестиційних рішень є істотно нижчою за частоту оновлення ринкових даних, що дає змогу використовувати щоденну інформацію та агреговані внутрішньоденні показники як джерело сигналів, уникаючи при цьому специфічних обмежень і витрат, характерних для високочастотної або внутрішньоденної торгівлі.

Важливим концептуальним положенням є твердження, що якість alpha-сигналів не може бути адекватно оцінена виключно за допомогою аналітичних метрик прогнозної точності. Навіть статистично стабільні сигнали можуть не мати практичної інвестиційної цінності чи втрачати ефективність у портфельному застосуванні. Тому портфельний бектестинг розглядається як необхідний етап модельного процесу, що доповнює аналітичну оцінку ML-моделей і дає змогу перевірити їх здатність генерувати стійкі економічні результати з урахуванням динаміки ринку та взаємодії між активами.

### Загальна логіка алгоритмічного процесу

Алгоритмічний процес реалізується у вигляді послідовності етапів обробки (pipeline), що наведено на **рис. 1**.

Кожен етап виконує визначену функцію та інтегрований у єдину часово узгоджену систему

прийняття рішень. Наукова новизна полягає у формалізованій інтеграції часово-залежного (time-aware) ML-навчання та портфельного бектестингу в єдину квантитативну експериментальну систему.

Розглянемо структурну схему модельного комплексу у вигляді функціональних блоків і потоків даних між ними (**рис. 2**).

У **табл. 1** наведена інформація щодо інформаційних потоків, якими обмінюються блоки системи. Функціональне навантаження компонентів системи наведено в **табл. 2**.

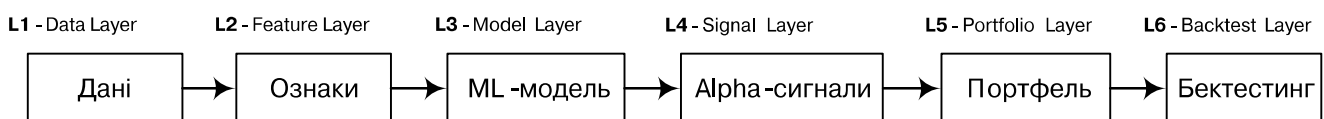
*Вхідні дані та простір ознак (Data Description & Feature Space).* Емпіричною базою дослідження є щоденні фінансові часові ряди ліквідних інструментів фондового ринку США (акції та індексні активи). Аналіз здійснюється на середньостроковому горизонті 1–12 тижнів із використанням щоденних даних (Open, High, Low, Close). Внутрішньоденні значення застосовуються винятково для побудови агрегованих характеристик.

Простір ознак формується як структурована система кількісних показників динаміки активів і охоплює три групи:

- price-based: лагові дохідності та кумулятивні зміни цін;
- volatility-based: історична волатильність і дисперсійні характеристики;
- intraday-aggregated: діапазон найвищої та найнижчої цін (High — Low) та рух відкриття до закриття (Open — Close).

Усі ознаки на момент  $t$  формуються виключно на основі інформації  $\leq t$ , що забезпечує часову коректність моделювання.

*Побудова ML-based alpha-сигналів.* У сучасній квантитативній літературі з ML в asset pricing домінує постановка задачі прогнозування очікуваної (excess) дохідності як регресії, а не як бінарної класифікації напрямку руху ціни. У канонічній праці S. Gu, B. Kelly та D. Xiu очікувані премії за ризик/надлишкові дохідності оцінюються саме як регресійна задача (expected returns forecasting), що дає змогу безпосередньо інтерпретувати вихід моделі як кількісне оцінювання очікуваної дохідності та використовувати її для формування портфельних рішень [1]. Аналогічно, сучасні дослідження з прогнозування дохідностей на горизонтах 1–3–12 місяців (типових



**Рис. 1.** Алгоритмічний процес роботи з моделлю

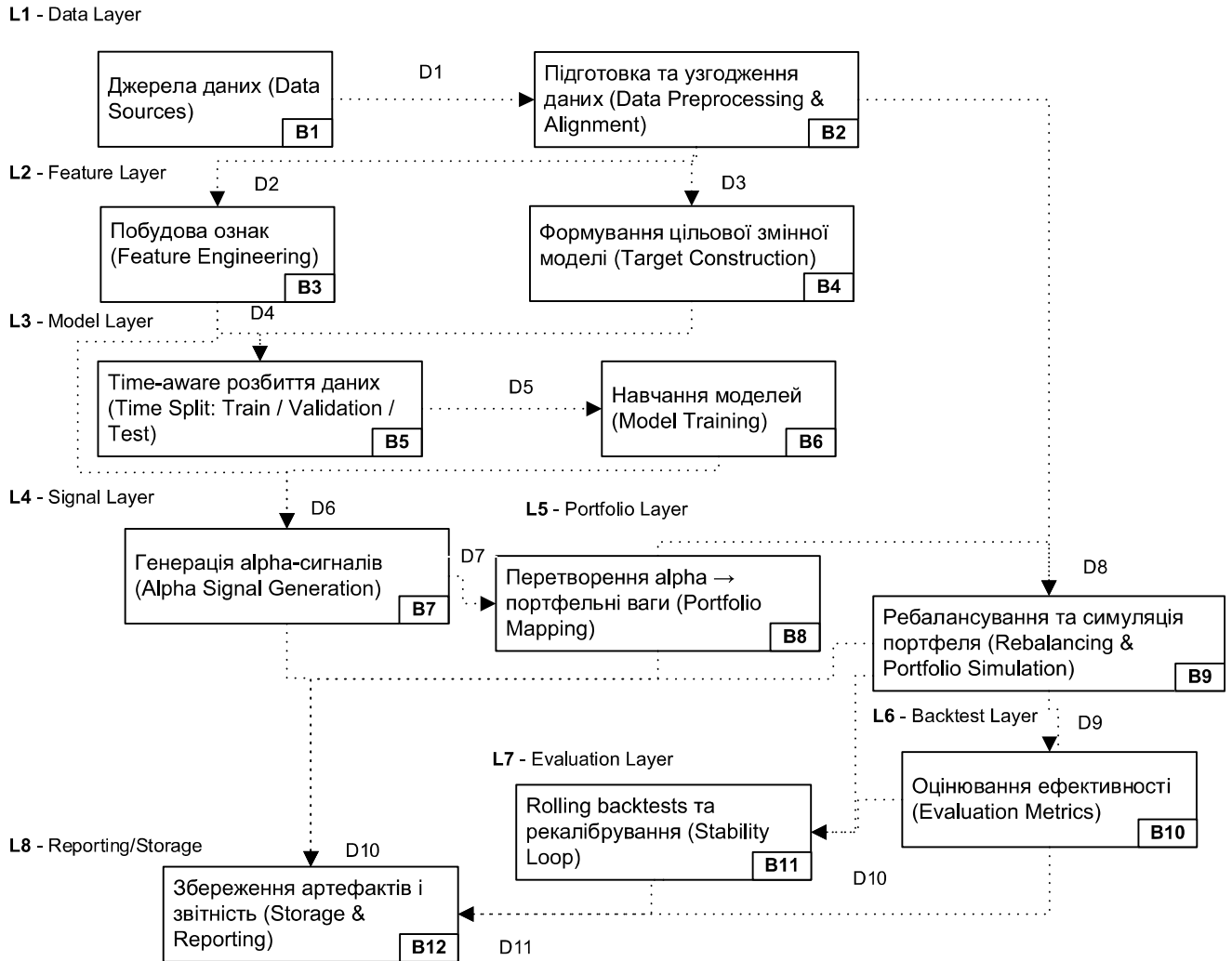


Рис. 2. Структурна схема модельного комплексу

Таблиця 1

Інформаційні потоки даних

Інформаційні потоки	Перетворення потоку
D1	$\mathcal{D}_0 \rightarrow \mathcal{D}_1$ — первинні дані перетворюються на очищені та узгоджені часові ряди
D2	$\mathcal{D}_1 \rightarrow X(t)$ — очищені ціни формують простір ознак
D3	$\mathcal{D}_1 \rightarrow y(t, H)$ — очищені ціни задають цільову змінну прогнозування
D4	$\{X, y\} \rightarrow \mathcal{S}$ — ознаки та цільова змінна формують time-aware спліти
D5	$\mathcal{S} \rightarrow M(t)$ — на основі сплітів формується навчена модель
D6	$\{M(t), X(t)\} \rightarrow \alpha(t)$ — модель та актуальні ознаки генерують alpha-сигнал
D7	$\alpha(t) \rightarrow w(t)$ — alpha-сигнал трансформується в портфельні ваги
D8	$\{w(t), \mathcal{D}_1\} \rightarrow \{R_p(t), \mathcal{V}(t)\}$ — ваги та реалізовані дохідності формують динаміку портфеля
D9	$\{R_p(t), \mathcal{V}(t)\} \rightarrow Met$ — портфельні результати перетворюються на метрики ефективності
D10	$\{R_p, Met\} \rightarrow$ тригер рекалібрування (recalibration trigger) – результати та метрики ініціюють перевірку стабільності та рекалібрування
D11	$\{\alpha, w, R_p, Met\} \rightarrow \mathcal{A}$ — ключові артефакти зберігаються та формують звітність

## Компоненти системи та їх функціональне навантаження

Блок	Функціональне навантаження
<b>B1</b>	Агрегація та первинна уніфікація ринкових даних за активами
<b>B2</b>	Синхронізація дат; очищення пропусків; перевірка аномалій; формування базових дохідностей
<b>B3</b>	Обчислення price-based, volatility-based та intraday-aggregated ознак із дотриманням правила $\leq t$
<b>B4</b>	Обчислення майбутньої дохідності $y(t, H)$ ; часовий зсув таргету; узгодження $X$ та $y$
<b>B5</b>	Часовий поділ вибірки; формування схем ковзного прогнозу (walk-forward схем); визначення ковзних/розширюваних (rolling/expanding) вікон
<b>B6</b>	Навчання ML-моделі; підбір гіперпараметрів; фіксація версії моделі для моменту $t$
<b>B7</b>	Прогноз очікуваної дохідності $\alpha_{i,t}(H)$ для кожного активу
<b>B8</b>	Ranking активів; нормалізація; застосування обмежень (концентрація, тільки довгі позиції (long-only) тощо)
<b>B9</b>	Застосування правил ребалансування; контроль turnover; урахування транзакційних витрат; обчислення портфельної дохідності
<b>B10</b>	Розрахунок кумулятивної дохідності (cumulative return), волатильності, коефіцієнта Шарпа (Sharpe ratio), максимального просідання (max drawdown)
<b>B11</b>	Серія ковзних бектестів (rolling-бектестів); моніторинг деградації; тригер оновлення моделі
<b>B12</b>	Логування; збереження конфігурацій; формування таблиць і графіків

для середньострокових квант-стратегій) також формулюють задачу як прогнозування величини дохідності (point forecasts) і оцінюють моделі за регресійними метриками (MSE) [7].

#### Постановка задачі як регресії очікуваної дохідності

У межах даного дослідження alpha-сигнал трактується як безперервна (continuous) прогнозна величина, що відповідає оцінці очікуваної дохідності активу на фіксованому горизонті  $H$  (1–12 тижнів). Типова формалізація:

$$\alpha_{i,t}(H) = E[r_{i,t \rightarrow t+H} | x_{i,t}], \quad (1)$$

де  $x_{i,t}$  — вектор ознак (price-based, volatility-based, intraday-aggregated), а  $r_{i,t \rightarrow t+H}$  — майбутня (excess) дохідність за горизонтом  $H$ . Така постановка є узгодженою з підходами емпіричного asset pricing, де ML застосовується для оцінювання очікуваних премій за ризик у крос-секції активів і в часі [1], а також із сучасними ML-підходами до прогнозування дохідностей на різних середньострокових горизонтах [7].

Популярна постановка “вгору/вниз” (класифікація) у фінансових задачах є інтуїтивною, однак методологічно слабшою для задач портфельного управління, де ключовою змінною є очікувана величина дохідності (expected return magnitude), а не лише знак. Класифікацію можна описати так:

- втрачає інформацію про *масштаб* очікуваної дохідності (сильні та слабкі сигнали потрапляють в одну категорію);
- потребує довільного вибору порогу (threshold) та/або перетворення прогнозу на торгове рішення, що підвищує нестабільність результатів між вибірками;
- складніше інтегрується в портфельні процедури, де ваги активів природно пов’язані з безперервними очікуваними дохідностями та ризиком.

Натомість регресійна постановка безпосередньо підтримує центральну ідею статті: ML використовується для побудови безперервних alpha-сигналів, а їх практична інвестиційна цінність оцінюється лише в портфельному бектестингу. Такий акцент є типовим для сучасної критичної дискусії щодо коректного використання ML в asset pricing: важливо не лише спрогнозувати, а й методологічно правильно вбудувати прогноз у процес побудови портфеля та перевірки стійкості результатів [8].

*Клас моделей для побудови alpha-сигналів.* З урахуванням вимог відтворюваності та інтерпретованості (для статті фахового рівня та бакалаврського диплома) доцільно розглядати декілька класів моделей, які в літературі використовуються як базові та/або конкурентні для прогнозування очікуваних дохідностей:

1) лінійні та регуляризовані моделі (baseline/benchmark) –регуляризовані регресії (Ridge/Lasso/Elastic Net) широко застосовуються як сильний базовий рівень у задачах прогнозування дохідностей у високовимірному просторі ознак, оскільки вони контролюють перенавчання та частково вирішують проблему мультиколінеарності предикторів. У підході S. Gu, B. Kelly та D. Xiu лінійні/регуляризовані моделі виступають важливою частиною порівняльного аналізу методів [1], а сучасні дослідження із прогнозування дохідностей також використовують їх як стандартний benchmark [7];

2) tree-based моделі (Random Forest/Gradient Boosting) — деревні методи (tree-based methods) та градієнтний бустинг є стандартним інструментом для виявлення нелінійностей та взаємодій між предикторами без ручного задання функціональної форми. У емпіричному asset pricing показано, що саме моделі, здатні захоплювати нелінійні взаємодії (зокрема tree-based і NN), забезпечують суттєві прирости прогнозу та економічної ефективності порівняно з суто лінійними підходами [1]. У сучасних прикладних роботах з прогнозування дохідностей також типово порівнюють tree-based ансамблі з лінійними моделями та нейромережами [7];

3) ensemble-підхід — у прикладному quant-процесі ансамблювання використовується переважно як механізм підвищення стабільності прогнозів (зменшення дисперсії моделі) та зниження чутливості до специфічного вибору алгоритму. Важливо, що в asset pricing методологічні рішення (зокрема дизайн експерименту, схема валідації, період навчання та перевірки) можуть суттєво впливати на результати навіть за фіксованого класу моделей. З огляду на це, ансамбль доцільно розглядати як інструмент стійкості (robustness), а не як самоціль [8].

*Оркестрація моделей та алгоритмічний pipeline (Model Orchestration & Workflow).* Центральним елементом запропонованого підходу є формування узгодженого алгоритмічного pipeline, що забезпечує коректну інтеграцію методів машинного навчання в процес квантитативного портфельного аналізу. На відміну від фрагментарного використання ML-моделей для прогнозування окремих показників, у цьому дослідженні реалізується послідовний і замкнений робочий цикл, що імітує реальні умови прийняття інвестиційних рішень.

*Навчання моделей у часовому контексті.* З огляду на нестационарність фінансових рядів навчання організовується в послідовному часовому режимі. Застосовуються схеми rolling window або expanding window, у межах яких модель на кожному кроці часу використовує лише

інформацію, доступну на момент прийняття рішення.

Rolling window підвищує адаптивність до зміни ринкових умов, тоді як expanding window забезпечує стабільність оцінок завдяки накопиченню довгострокової інформації. Вибір схеми розглядається як частина експериментального дизайну.

Train-вибірка застосовується для оцінювання параметрів, validation — для вибору гіперпараметрів, test — для фінальної портфельної оцінки. Такий підхід забезпечує коректну оцінку позавивіркової (out-of-sample) ефективності.

*Генерація alpha-сигналів.* Alpha-сигнали формуються послідовно в часі: для кожного моменту  $t$  модель генерує прогноз очікуваної дохідності на заданому горизонті, використовуючи лише інформацію  $\leq t$ . Це усуває look-ahead bias та забезпечує інформаційну узгодженість між етапами ML і портфельного формування.

*Інтеграція з портфельним бектестингом.* Бектестинг інтегрований у pipeline та виконується для кожного циклу навчання і генерації сигналів. Оцінювання здійснюється на рівні портфельної динаміки, що дає змогу аналізувати не лише точність прогнозів, а й їхню реальну інвестиційну ефективність у часі.

*Формування портфельних рішень на основі alpha-сигналів (Portfolio Construction).* Ключовим етапом запропонованого підходу є перехід від прогнозних ML-based alpha-сигналів до конкретних портфельних рішень. Саме на цьому етапі відбувається інтеграція результатів ML з інвестиційною логікою, що дає змогу уникнути використання ML як самоцілі та оцінювати його практичну цінність у реальному портфельному контексті.

У межах дослідження alpha-сигнали інтерпретуються як відносні оцінки очікуваної дохідності активів, а не як готові торговельні рекомендації. Тому формування портфеля розглядається як окрема задача, що базується на ранжуванні активів за величиною alpha та подальшому перетворенні цього ранжування в систему портфельних ваг. Такий підхід дає змогу зберегти інтерпретованість рішень і водночас уникнути надмірної складності, що властива оптимізаційним моделям із великою кількістю припущень.

*Ранжування активів за alpha-сигналами.* Першим кроком формування портфеля є ранжування активів на основі згенерованих alpha-сигналів. Для кожного моменту часу активи впорядковуються відповідно до прогнозованої очікуваної дохідності на заданому горизонті. Таке ранжування відображає відносну привабливість активів з точки зору ML-моделі та

створює підґрунтя для подальшого розподілу капіталу.

Використання ранжування, а не абсолютних значень  $\alpha$ , дає змогу зменшити чутливість портфельних рішень до масштабу прогнозів і потенційних зсувів у рівні сигналів між різними періодами. Це особливо важливо в умовах нестационарності фінансових ринків, коли абсолютні значення прогнозованої дохідності можуть змінюватися разом із загальним ринковим фоном.

*Нормалізація та побудова портфельних ваг.* Після ранжування  $\alpha$ -сигнали перетворюються в портфельні ваги за допомогою процедур нормалізації. У найпростішому випадку це може бути лінійна нормалізація, за якої ваги активів пропорційні відношенню активу в  $\alpha$ -ранжуванні. Такий механізм забезпечує прозорий зв'язок між прогнозованою інформацією та портфельним розподілом капіталу.

Нормалізація виконує подвійну функцію: по-перше, вона гарантує дотримання бюджетного обмеження портфеля, а по-друге, знижує ризик надмірної концентрації капіталу в окремих активах унаслідок екстремальних значень  $\alpha$ -сигналів. Завдяки цьому портфельні рішення залишаються стабільними навіть у періоди підвищеної волатильності прогнозів.

*Обмеження та контроль портфельного ризику.* Для забезпечення практичної реалізованості портфельних рішень застосовуються прості та інтерпретовані обмеження: максимальні та мінімальні ваги активів, контроль концентрації капіталу та, за потреби, симетричні умови для довгих і коротких позицій. Запропонований підхід не потребує розв'язання складних оптимізаційних задач і не спирається на припущення щодо стаціонарності коваріацій чи нормальності розподілів дохідностей.

У межах дослідження не використовується класична оптимізація Марковіца або її модифікації, що дає змогу уникнути залежності результатів від нестійких оцінок матриці коваріацій. Портфель розглядається як механізм агрегування  $\alpha$ -сигналів, а не як результат формальної оптимізації.

Такий підхід забезпечує чітке розмежування між задачею прогнозування та задачею управління капіталом і дає змогу інтерпретувати портфель як інструмент трансформації ML-прогнозів у практичні інвестиційні рішення з урахуванням заданих обмежень [10].

*Портфельний бектестинг і оцінювання ефективності (Portfolio Backtesting & Evaluation).* Портфельний бектестинг у межах цього дослідження розглядається як ключовий інструмент перевірки практичної інвестиційної цінності ML-

based  $\alpha$ -сигналів. На відміну від оцінювання моделей лише за аналітичними метриками прогнозованої точності, бектестинг надає можливість відтворити послідовність реальних рішень інвестора у часі та оцінити, чи трансформуються прогнозовані сигнали в стабільний економічний результат на рівні портфеля [12].

Методологічно бектестинг реалізується як замкнений цикл, у якому  $\alpha$ -сигнали генеруються на основі інформації, доступної на момент прийняття рішення, після чого на їх основі формується портфель, виконується ребалансування та фіксується портфельна дохідність за наступний інтервал часу. Отриманий часовий ряд портфельних дохідностей використовується для розрахунку показників ефективності та ризику.

*Загальна схема проведення бектестингу.* Бектестинг організовується як послідовність кроків, що повторюються для кожного моменту ребалансування. На кроці  $t$  виконуються такі операції:

1) оновлення даних і ознак — формується набір ознак  $x_{i,t}$  для кожного активу на основі даних до моменту  $t$  включно;

2) навчання/оновлення ML-моделі (за rolling або expanding схемою): модель оцінюється на історичному вікні даних, що закінчується моментом  $t$ ;

3) генерація  $\alpha$ -сигналів — для кожного активу отримується  $\alpha_{i,t}(H)$  — прогноз очікуваної дохідності на горизонті  $H$ ;

4) побудова портфельних ваг  $w_{i,t}$  на основі  $\alpha$  (ranking  $\rightarrow$  normalization  $\rightarrow$  constraints);

5) ребалансування портфеля — портфель переходить до нових ваг із урахуванням обраної частоти ребалансування;

6) обчислення реалізованої портфельної дохідності на інтервалі  $(t, t+\Delta)$  та оновлення історії портфельних результатів.

Завдяки такій схемі бектестинг забезпечує причинно-часову коректність: рішення у момент  $t$  не можуть залежати від інформації, що з'явиться після  $t$ . Це є критично важливим для уникнення завищених результатів і забезпечення відтворюваності дослідження.

*Розрахунок портфельних дохідностей у бектесті.* Головним вихідним результатом бектестингу є часовий ряд портфельних дохідностей  $R_{p,t}$ , що формується на кожному інтервалі між ребалансуваннями. У базовому випадку портфельна дохідність на періоді  $(t, t+\Delta)$  визначається як зважена сума дохідностей активів за цей період:

$$R_{p,t \rightarrow t+\Delta} = \sum_i w_{i,t} \cdot r_{i,t \rightarrow t+\Delta}, \quad (2)$$

де  $w_{i,t}$  — ваги портфеля після ребалансування у момент  $t$ , а  $r_{i,t \rightarrow t+\Delta}$  — реалізована дохідність активу на відповідному інтервалі.

Для забезпечення практичності моделювання портфельного процесу допускається врахування простих і інтерпретованих елементів, зокрема:

- обмежень на ваги (щоб уникати надмірної концентрації);
- обмежень на зміну ваг між періодами (як механізм контролю turnover);
- умовного моделювання транзакційних витрат як параметра чутливості (за потреби).

Причому ключовим є не деталізація торговельної інфраструктури, а методологічно коректне відтворення послідовності “сигнал → рішення → результат” у часі.

**Метрики оцінювання ефективності та ризику.** Оцінювання результатів бектестингу здійснюється на основі набору метрик, що відображають як прибутковість, так і ризикові характеристики портфеля. Вибір метрик обмежується базовими показниками, які легко інтерпретуються та є достатніми для порівняння альтернативних моделей і варіантів pipeline.

**Кумулятивна дохідність (Cumulative Return).** Кумулятивна дохідність характеризує накопичений результат стратегії за весь період бектесту. Вона визначається як нарощення капіталу при послідовному реінвестуванні:

$$V_t = V_0 \cdot \prod_k (1 + R_{p,k}), \quad (3)$$

де  $V_t$  — вартість портфеля у момент  $t$ , а  $R_{p,k}$  — портфельні дохідності на інтервалах ребалансування. Кумулятивна крива капіталу дає змогу візуально оцінити стабільність зростання та наявність кризових періодів.

**Волатильність (Volatility).** Волатильність використовується як базовий вимір ризику, що характеризує мінливість портфельних дохідностей. Вона визначається як стандартне відхилення  $R_p$  на відповідній частоті (тижневій/місячній), з можливістю приведення до річної шкали для порівнянності:

$$\sigma_p = std(R_p). \quad (4)$$

Волатильність важлива для розуміння того, якою “ціною” (у вигляді коливань капіталу) досягається кумулятивний результат.

**Коефіцієнт Шарпа (Sharpe ratio).** Sharpe ratio застосовується як компактний показник співвідношення “дохідність/ризик”, але без надмірної інтерпретації як універсального критерію. У базовому вигляді:

$$SR = mean(R_p - r_f) / std(R_p - r_f), \quad (5)$$

де  $r_f$  — безризикова ставка на відповідній частоті. У практичному експериментальному дизайні допускається використання спрощеної постановки з  $r_f \approx 0$  для коротких інтервалів або для порівняння моделей за однакових умов.

Важливо, що Sharpe ratio розглядається як один із декількох показників, а не як єдина метрика якості.

**Максимальне просідання (Maximum Drawdown).** Максимальне просідання (drawdown) вимірює найбільше падіння капіталу від локального максимуму до наступного мінімуму:

$$DD_t = (Peak_t - V_t) / Peak_t, \quad (6)$$

$$MaxDD = max(DD_t).$$

Цей показник є критично важливим у прикладному контексті, оскільки відображає найгірший сценарій втрати капіталу в межах бектесту та дає змогу оцінити “психологічну” та ризик-менеджерську прийнятність стратегії.

**Інтерпретація результатів і критерії порівняння.** Результати бектестингу інтерпретуються як комплексна характеристика інвестиційної доцільності alpha-сигналів. Порівняння альтернативних підходів (різні моделі, різні набори ознак, різні схеми rolling/expanding) здійснюється не лише за рівнем прибутковості, а й за профілем ризику:

- вища кумулятивна дохідність при контрольованій волатильності;
- вищий Sharpe ratio за схожої стабільності;
- менше максимальне просідання або більш “плавна” крива капіталу.

Окрім того, у межах середньострокового підходу важливим є аналіз чутливості результатів до частоти ребалансування та до вибору прогнозного горизонту. Це дає змогу оцінити, чи є ефективність сигналів стійкою властивістю підходу, а не випадковим результатом конкретної конфігурації експерименту.

Таким чином, портфельний бектестинг у запропонованому дослідженні виконує роль ключового методологічного механізму, який забезпечує перехід від статистичного оцінювання ML-моделей до оцінювання їхньої реальної інвестиційної цінності.

### **Стійкість сигналів та адаптація до ринкових режимів (Stability & Recalibration)**

Однією з ключових характеристик фінансових ринків є їх динамічність і наявність різних ринкових режимів, що відрізняються рівнем волатильності, кореляційною структурою активів, ліквідністю та доміантними драйверами дохідності [11]. У таких умовах навіть статистично стабільні та економічно обґрунтовані alpha-сигнали можуть із часом втрачати ефективність або демонструвати суттєве погіршення портфельних результатів. Це зумовлює необхідність розглядати прогнозні сигнали не як стаціонарні характеристики активів, а як умовні оцінки, чутливі до поточного ринкового середовища.

Порівняння результатів моделей (2015–2023)

Модель	Середня річна дохідність	Волатильність	Sharpe ratio	Max Drawdown
Ridge	11,2 %	16,5 %	0,68	–18,4 %
Random Forest	13,6 %	17,8 %	0,76	–19,7 %
Equal-weight benchmark	8,9 %	15,9 %	0,56	–22,1 %

У межах запропонованого підходу визнається, що ефективність ML-based alpha-сигналів не є постійною в часі та може змінюватися внаслідок переходу між ринковими режимами. Такі зміни можуть бути пов'язані з макроекономічними шоками, фазами ринкових циклів, періодами підвищеної або зниженої волатильності, а також зі структурними змінами в поведінці учасників ринку. Відповідно, оцінювання сигналів лише на фіксованому історичному інтервалі не є достатнім для висновків щодо їх довгострокової інвестиційної придатності.

*Регулярна переоцінка та повторне калібрування моделей.* Для забезпечення адаптивності до змін ринкових умов у дослідженні передбачається регулярна переоцінка та повторне калібрування ML-моделей [9], що використовуються для генерації alpha-сигналів. У рамках алгоритмічного pipeline це реалізується через послідовне оновлення моделей у rolling або expanding window режимі, де параметри моделі оцінюються на актуальному підборі історичних даних.

Такий підхід дає змогу, з одного боку, врахувати нову інформацію та зміну статистичних властивостей ринку, а з іншого — уникати надмірної реакції на короткострокові флуктуації. Повторне калібрування моделей розглядається не як виняткова процедура у відповідь на погіршення результатів, а як невіддільний елемент системного квантитативного процесу.

*Rolling backtests як інструмент аналізу стійкості.* Для оцінювання стабільності alpha-сигналів і портфельних результатів у часі застосовується підхід rolling backtests, за якого бектестинг виконується на послідовних перекривних часових інтервалах. Кожен такий інтервал дає змогу оцінити поведінку стратегії в різних ринкових умовах і зменшити ризик узагальнення результатів, отриманих на окремому сприятливому періоді.

Rolling backtests дають змогу аналізувати динаміку ключових метрик ефективності — кумулятивної дохідності, волатильності, Sharpe ratio та максимальних просідань — у різних часових підвбірках. Це допомагає виявляти

періоди деградації сигналів, оцінювати тривалість таких фаз і перевіряти, чи відновлюється ефективність після повторного калібрування моделей.

Емпіричні результати квантитативного моделювання.

Для демонстрації практичної придатності запропонованого підходу було проведено емпіричне тестування на вибірці з 20 ліквідних інструментів фондового ринку США за період 2015–2023 років. Прогнозний горизонт становив чотири тижні, частота ребалансування — щотижнева. Навчання моделей здійснювалося у режимі rolling window із використанням щоденних даних.

У межах експерименту було порівняно дві специфікації моделей:

- регуляризована лінійна регресія (Ridge) як базова модель;
- Random Forest як нелінійний ансамблевий підхід.

Оцінювання ефективності здійснювалося на основі портфельного бектестингу із застосуванням стандартних ризик-орієнтованих метрик, результати якого наведено в **табл. 3**.

Обидві ML-моделі демонструють покращення ризик-скоригованої ефективності в порівнянні з рівноваговим бенчмарком. Random Forest забезпечує вищу середню дохідність та Sharpe ratio, що свідчить про здатність враховувати нелінійні взаємозв'язки між ознаками. Водночас регуляризована лінійна модель демонструє стабільніші результати та нижчу чутливість до ринкових коливань.

Отримані результати підтверджують практичну доцільність інтеграції ML-based alpha-сигналів у портфельні стратегії в межах середньострокового квантитативного підходу.

## ВИСНОВКИ

У межах дослідження розроблено та реалізовано інтегрований квантитативний підхід до формування ML-based alpha-сигналів та їхньої перевірки в портфельному контексті. На відміну від робіт, що обмежуються оцінюванням прогнозової точності моделей, запропонований

підхід поєднує регресійну постановку задачі прогнозування очікуваної дохідності з процедурою послідовного портфельного бектестингу.

Емпіричні результати на вибірці ліквідних акцій американського ринку підтверджують, що використання ML-based alpha-сигналів дає змогу підвищити ризик-скориговану дохідність портфеля в порівнянні з наївною стратегією рівновагового розподілу. Зокрема alpha-орієнтовані портфелі демонструють вищу кумулятивну дохідність і Sharpe ratio за порівнянного рівня волатильності та контрольованого максимального просідання.

Отримані результати свідчать про те, що економічна цінність ML-моделей проявляється не стільки в покращенні статистичних метрик прогнозу, скільки у їхній здатності генерувати стійкий портфельний ефект у середньостроковому горизонті за умов регулярного оновлення моделей і контролю ризику.

Водночас дослідження підтверджує, що ефективність alpha-сигналів є часово-змінною величиною, що потребує систематичного моніторингу та recalібрування моделей. Таким чином, запропонований підхід може розглядатися як методологічне підґрунтя для побудови адаптивних квантитативних стратегій, у яких машинне навчання інтегрується в контрольований портфельний процес.

#### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Gu S. Empirical Asset Pricing via Machine Learning / S. Gu, B. Kelly, D. Xiu // *The Review of Financial Studies*. — 2020. — Vol. 33. — Issue 5. — P. 2223–2273. DOI: <https://doi.org/10.1093/rfs/hhaa009>.
2. Bagnara M. Asset Pricing and Machine Learning: A critical review / Matteo Bagnara // *Journal of Economic Surveys*. — 2024. — Vol. 38. — P. 27–56. DOI: [10.1111/joes.12532](https://doi.org/10.1111/joes.12532).
3. Lopez de Prado M. *Advances in Financial Machine Learning* / M. Lopez de Prado. — Hoboken, NJ : John Wiley & Sons, 2018. — 400 p.
4. Harvey C. R. Backtesting / C. R. Harvey, Y. Liu. — Available at SSRN 2345489. — 2015. — Jul 28.
5. Bailey D. H. The Sharpe ratio efficient frontier / D. H. Bailey, M. López de Prado // *The Journal of Risk*. — 2012. — Vol. 15. — No. 2. — P. 3–44. DOI: <https://doi.org/10.21314/JOR.2012.255>.
6. Nadler P. Empirical Asset Pricing with Functional Factors / Philip Nadler, Alessio Sancetta // *Journal of Financial Econometrics*. — 2023. — Vol. 21. — Issue 5. — P. 1258–1281. DOI: <https://doi.org/10.1093/jffinec/nbac003>.
7. Wang Z. Machine learning for stock return prediction: Transformers or simple neural networks? / Z. Wang // *Finance Research Letters*. — 2025. — Vol. 86. — Part F. — 108783. DOI: <https://doi.org/10.1016/j.frl.2025.108783>.
8. Giglio S. Asset pricing with omitted factors / S. Giglio, D. Xiu // *Journal of Political Economy*. — 2021. — Vol. 129. — No. 7. — P. 1947–1990. — Access mode: <https://www.journals.uchicago.edu/doi/10.1086/714090>.
9. Identifying the system of value factors of green consumer choice / L. Potrashkova, V. Zaruba, D. Raiko, O. Yevsyeyev // *Innovative Marketing*. — 2024. — Vol. 20. — Issue 1. — P. 199–211. DOI: [http://dx.doi.org/10.21511/im.20\(1\).2024.17](http://dx.doi.org/10.21511/im.20(1).2024.17).
10. Modern Approach To Modeling Of Efficiency Of Financial Market Based On Methods Of Dynamic Programming / O. Maslyhan, T. Shabelnyk, O. Korolovych, N. Liba // *Ефективна економіка*. — 2022. — No. 9. DOI: <http://doi.org/10.32702/2307-2105.2022.9.7>.
11. Influence of regional financial market models on the structure of global financial assets / O. Bulatova, T. Shabelnyk, T. Marena, N. Reznikova // *Advances in Economics, Business and Management Research (AEBMR): 6th International Conference on Strategies, Models and Technologies of Economic Systems Management (SMTESM 2019)*. Proceedings of the International Scientific Conference. — 2019. — Vol. 95. — P. 339–345. DOI: <https://doi.org/10.2991/smtesm-19.2019.55>.
12. Shabelnyk T. V. Approaches to optimize investment risks / T. V. Shabelnyk // *Problems of Decision Making Under Uncertainties: Proceedings of the International Conference..* — 2018.
13. Wang Y. Predicting Stock Prices Based on Machine Learning to Build Self-adaptive Trading Strategy / Y. Wang, P. Huang, J. Luo // *Computational Economics*. — 2025. DOI: [10.1007/s10614-025-11054-4](https://doi.org/10.1007/s10614-025-11054-4).
14. Forecasting high-frequency excess stock returns via data analytics and machine learning / E. Akyildirim, D. Khuong Nguyen, A. Sensoy, M. Šikić // *International Review of Financial Analysis*. — 2023. — Vol. 29. — Issue 1. — P. 22–75. DOI: <https://doi.org/10.1111/eufm.12345>.

#### REFERENCES

1. Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33 (5), 2223-2273. DOI: <https://doi.org/10.1093/rfs/hhaa009>.
2. Bagnara, M. (2024). Asset Pricing and Machine Learning: A critical review. *Journal of Economic Surveys*, 38, 27-56. DOI: [10.1111/joes.12532](https://doi.org/10.1111/joes.12532).
3. Lopez de Prado, M. (2018). *Advances in Financial Machine Learning*. Hoboken, NJ: John Wiley & Sons, 400 p.
4. Harvey, C. R., & Liu, Y. (2015). *Backtesting*. Available at SSRN 2345489.
5. Bailey, D. H., & Lopez de Prado, M. (2012). The Sharpe ratio efficient frontier. *The Journal of Risk*, 15 (2), 3-44. DOI: <https://doi.org/10.21314/JOR.2012.255>.
6. Nadler, P., & Sancetta, A. (2023). Empirical Asset Pricing with Functional Factors. *Journal of Financial Econometrics*, 21 (5), 1258-1281. DOI: <https://doi.org/10.1093/jffinec/nbac003>.
7. Wang, Z. (2025). Machine learning for stock return prediction: Transformers or simple neural networks? *Finance Research Letters*, 86 (Part F), 108783. DOI: <https://doi.org/10.1016/j.frl.2025.108783>.
8. Giglio, S., & Xiu, D. (2021). Asset pricing with omitted factors. *Journal of Political Economy*, 129 (7), 1947-1990. Retrieved from: <https://www.journals.uchicago.edu/doi/10.1086/714090>.
9. Potrashkova, L., Zaruba, V., Raiko, D., & Yevsyeyev, O. (2024). Identifying the system of value factors of green consumer choice. *Innovative Marketing*, 20 (1), 199-211. DOI: [http://dx.doi.org/10.21511/im.20\(1\).2024.17](http://dx.doi.org/10.21511/im.20(1).2024.17).
10. Maslyhan, O., Shabelnyk, T., Korolovych, O., & Liba, N. (2022). Modern Approach to Modeling Of Efficiency

- ciency Of Financial Market Based On Methods Of Dynamic Programming. *Elektronnyi zhurnal "Efektivna ekonomika"* [Electronic magazine "Effective Economy"], 9. DOI: <http://doi.org/10.32702/2307-2105.2022.9.7>.
11. Bulatova, O., Shabelnyk, T., Marena, T., & Reznikova, N. (2019). Influence of regional financial market models on the structure of global financial assets. *Advances in Economics, Business and Management Research (AEBMR)*, 95. 6th International Conference on Strategies, Models and Technologies of Economic Systems Management (SMTESM 2019). Proceedings of the International Scientific Conference, 95. P. 339-345. DOI: <https://doi.org/10.2991/smtesm-19.2019.55>.
  12. Shabelnyk, T. V. (2018). Approaches to optimize investment risks. Proceedings of the International Conference "Problems of Decision Making Under Uncertainties" (Conference Materials).
  13. Wang, Y., Huang, P., & Luo, J. (2025). Predicting Stock Prices Based on Machine Learning to Build Self-adaptive Trading Strategy. *Computational Economics*. DOI: 10.1007/s10614-025-11054-4.
  14. Akyildirim, E., Nguyen, D. K., Sensoy, A., & Šikić, M. (2023). Forecasting high-frequency excess stock returns via data analytics and machine learning. *International Review of Financial Analysis*, 29 (1), 22-75. DOI: <https://doi.org/10.1111/eufm.12345>.

**T. V. SHABELNYK**, D. Sc. in Economics, Professor  
**S. O. YEVSYEYeva**, Student

### QUANTITATIVE MODELING OF PORTFOLIO STRATEGIES USING MACHINE LEARNING-BASED ALPHA SIGNALS AND BACKTESTING

**Abstract.** *This paper proposes a quantitative framework for constructing ML-based alpha signals and integrating them into a portfolio backtesting procedure over a medium-term investment horizon. The empirical analysis relies on daily financial time series of liquid equity market instruments, providing statistically consistent input data for quantitative modeling. A structured feature space is constructed, incorporating price-based, volatility-based, and intraday-aggregated characteristics. Model training is implemented in a time-aware setting using rolling or expanding window schemes and strict chronological separation of the training, validation, and test sets, ensuring robust quantitative out-of-sample evaluation and minimizing information leakage.*

*Alpha signals are generated at discrete rebalancing dates as numerical forecasts of expected returns over a predefined horizon and are directly embedded into portfolio construction. Performance is assessed through quantitative portfolio backtesting using standard risk-adjusted metrics, including returns, volatility, the Sharpe ratio, and maximum drawdown. The proposed algorithmic pipeline formalizes the interaction between machine learning and portfolio analysis and provides a methodological basis for the quantitative evaluation of the investment performance of ML-based strategies.*

**Keywords:** *quantitative modeling, machine learning, alpha signals, portfolio backtesting, equity market, medium-term strategies, risk-adjusted performance metrics.*

#### ІНФОРМАЦІЯ ПРО АВТОРІВ

**Шабельник Тетяна Володимирівна** — д-р екон. наук, проф., завкафедри Економічної кібернетики і системного аналізу, Харківський національний економічний університет ім. С. Кузнеця, просп. Науки, 9А, м. Харків, Україна, 61001; +38 (050) 176-84-78; Tanya.Shabelnik17@gmail.com; ORCID: 0000-0001-9798-391X

**Євсєєва Софія Олексіївна** — студентка, Харківський національний економічний університет імені Семена Кузнеця, просп. Науки, 9А, м. Харків, Україна, 61001; +38 (068)805-58-00; Sofiiia.Yevsieieva@hneu.net; ORCID: 0009-0001-9496-6804

#### INFORMATION ABOUT THE AUTHORS

**Shabelnyk T. V.** — D. Sc. in Economics, Professor, Chair of the Department of Economic Cybernetics and Systems Analysis Simon Kuznets Kharkiv National University of Economics, 9A, Nauky Ave, Kharkiv, Ukraine, 61001; +38 (050) 176-84-78; Tanya.shabelnik17@gmail.com; ORCID: 0000-0001-9798-391X

**Yevsyeyeva S. O.** — Student, Simon Kuznets Kharkiv National University of Economics, 9A, Nauky Ave., Kharkiv, Ukraine, 61001; +38 (050) 176-84-78; Sofiiia.Yevsieieva@hneu.net; ORCID: 0009-0001-9496-6804

Надійшла до редакції 21.02.2026

Прийнята до друку 09.03.2026

