

Використання моделей на основі механізму уваги для передбачення фінансових часових рядів

Attention-based models for finance time series prediction

Задохін Дмитро

керівник: д.ф.-м.н., проф. Глибовець М. М.

червень 2026

Механізм уваги — нова норма в глибокому навчанні

Архітектура **Transformer** досягла state-of-the-art у задачах із послідовною структурою — спершу в NLP, а потім і далеко за його межами.

Self-attention дозволяє моделі виявляти залежності між елементами послідовності незалежно від відстані між ними — фундаментальна перевага для часових рядів.

NLP

GPT · BERT · LLaMA

Генерація і розуміння тексту

Vision

ViT · DETR · Swin

Класифікація, детекція об'єктів

Audio

Whisper · Wav2Vec

Розпізнавання мовлення

Time series

Informer · PatchTST

Прогнозування рядів

Виклики при роботі з фінансовими часовими рядами

Природа даних

Фінансові часові ряди шумні, нестационарні і залежать від випадкових подій

Семантична бідність токена

Окрема часова точка майже не несе інформації порівняно із словом або областю зображення

Змішування різнотипних даних

Багатовимірні часові ряди мають атрибути різної природи, що змішуються в один токен

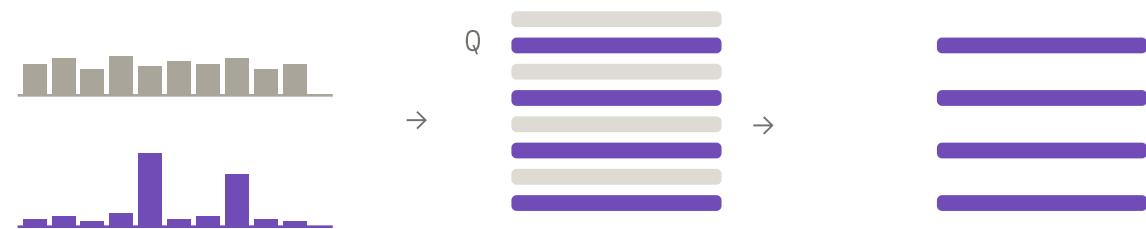
Лінива увага

Більшість активацій уваги — рівномірний шум, лише невелика частина пар Key-Query утворює інформативні активації

Розглянуті моделі

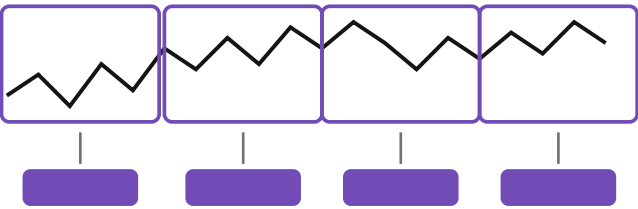
Informer

Розріджена увага: відбираються лише query з найбільшим відхиленням від рівномірного розподілу — ігноруються «ліниві» query.



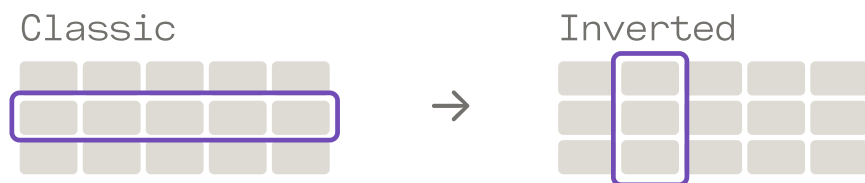
PatchTST

Часовий ряд розбивається на патчі — кожен стає одним токеном.
 $L^* = L/L_{\text{patch}}$, увага бачить семантично багатші одиниці.



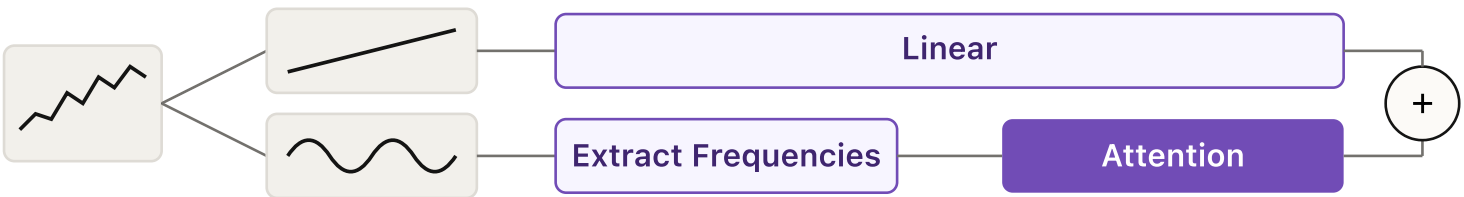
iTransformer

Замість токени на часовий крок — один токен на кожну змінну.
Увага виявляє кореляції між інструментами, а не між моментами.

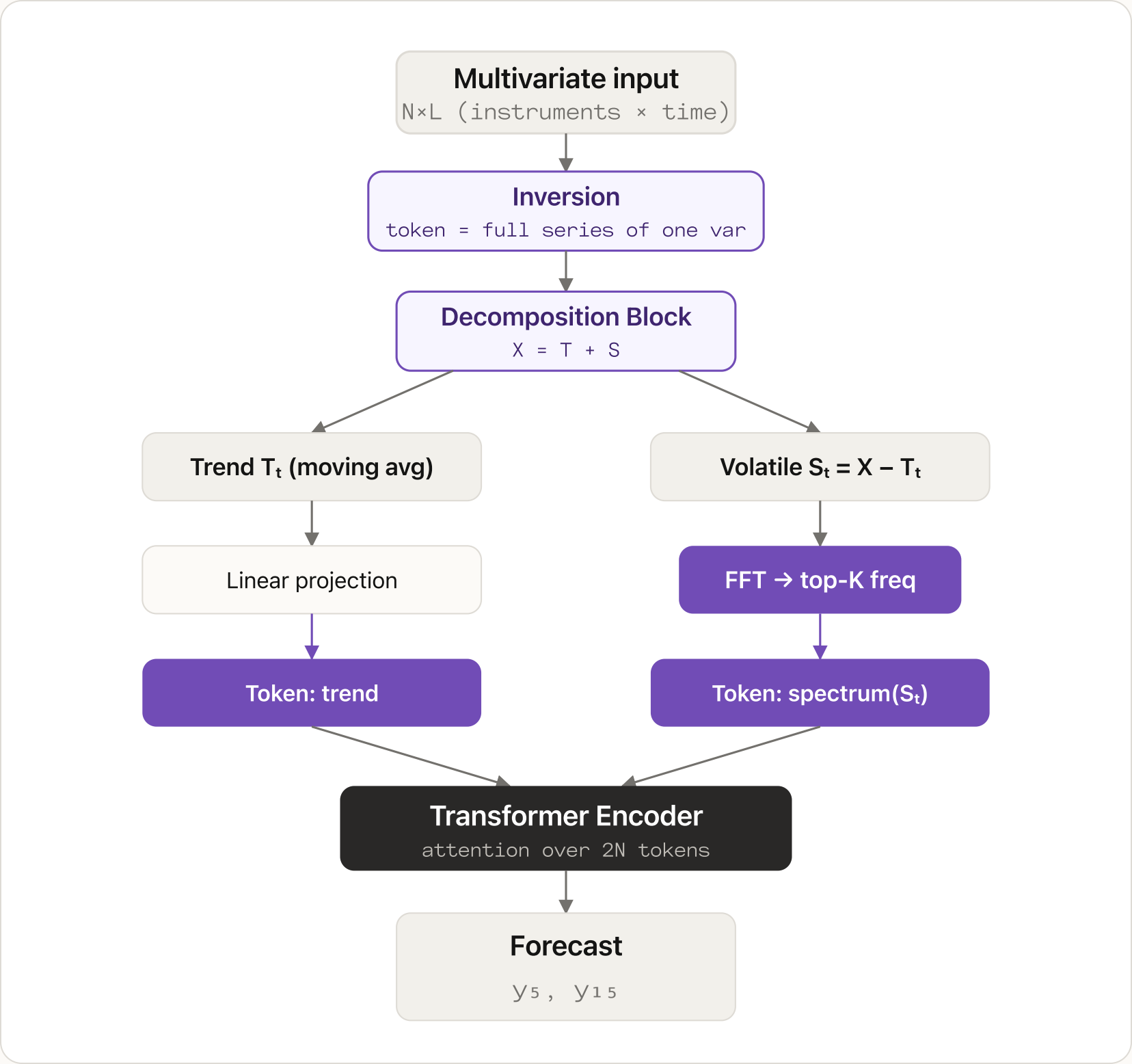
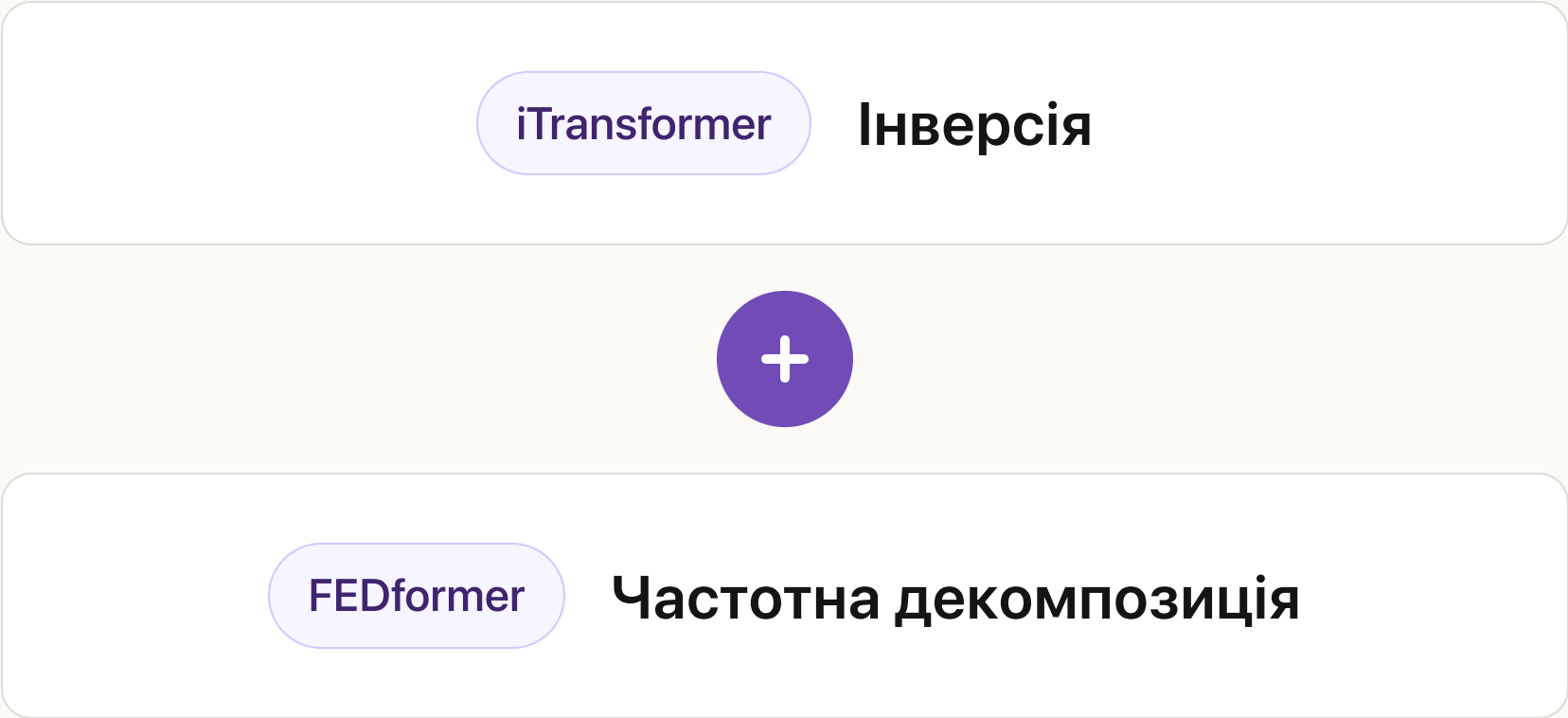


FEDformer

Ряд розкладається на тренд і волатильну складову; до останньої застосовується FFT / wavelet — увага лише на домінуючих частотах.



FEiTransformer — гібридна архітектура



Датасет 1: добові значення S&P500 (Yahoo Finance)

ЦІЛЬОВИЙ АКТИВ

Індекс S&P 500

Денні OHLCV, 2000 – 2025

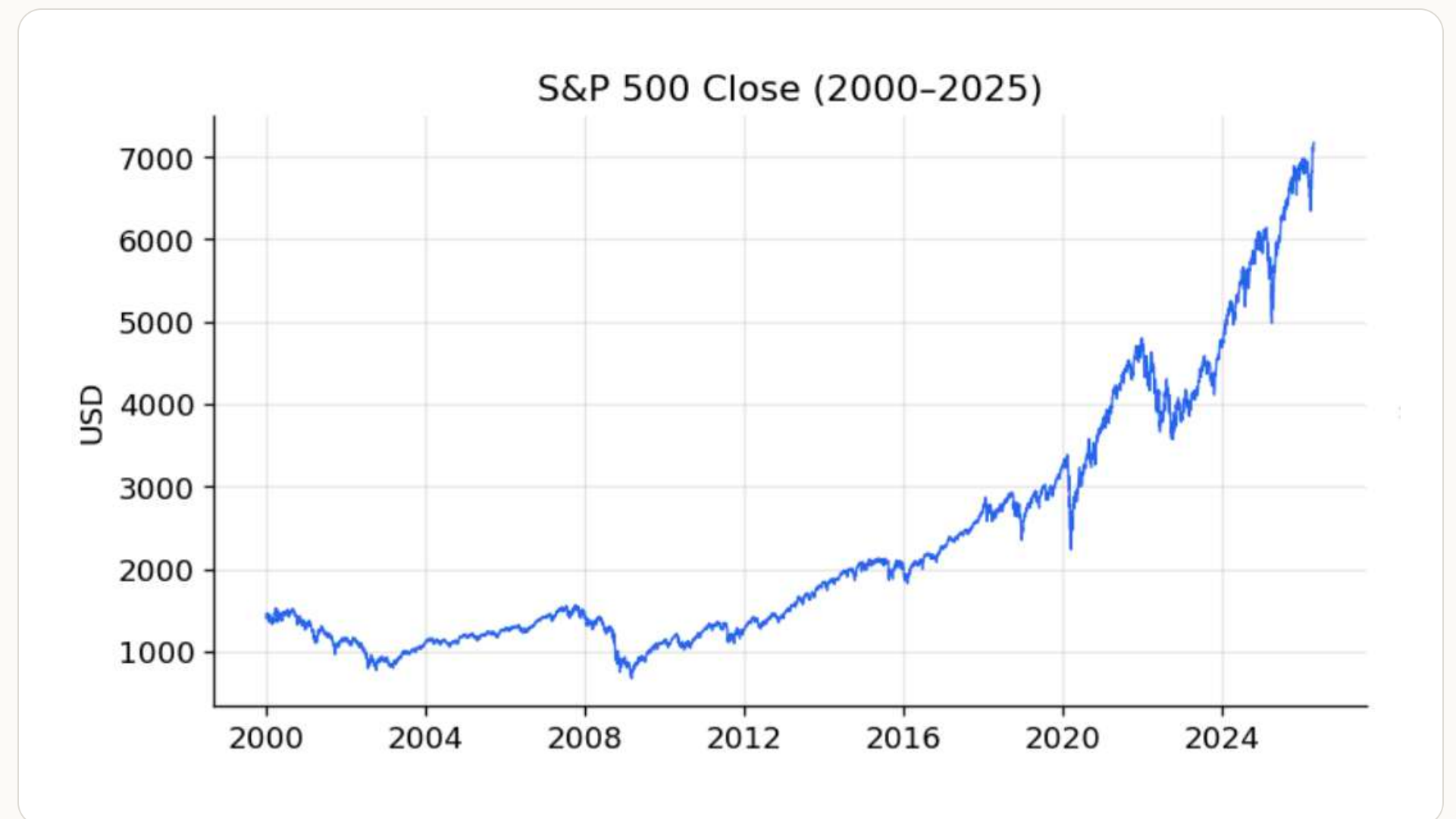
~10k

торгових днів

25

років історії

Мінімальні конфігурації моделей налічують десятки тисяч параметрів. Добових даних критично мало для моделей глибинного навчання.



Датасет 2: Kaggle — Binance Full History

ЦІЛЬОВИЙ АКТИВ

BTC / USDT

5-хвилинні OHLCV + об'єм торгів, 2017 – 2022

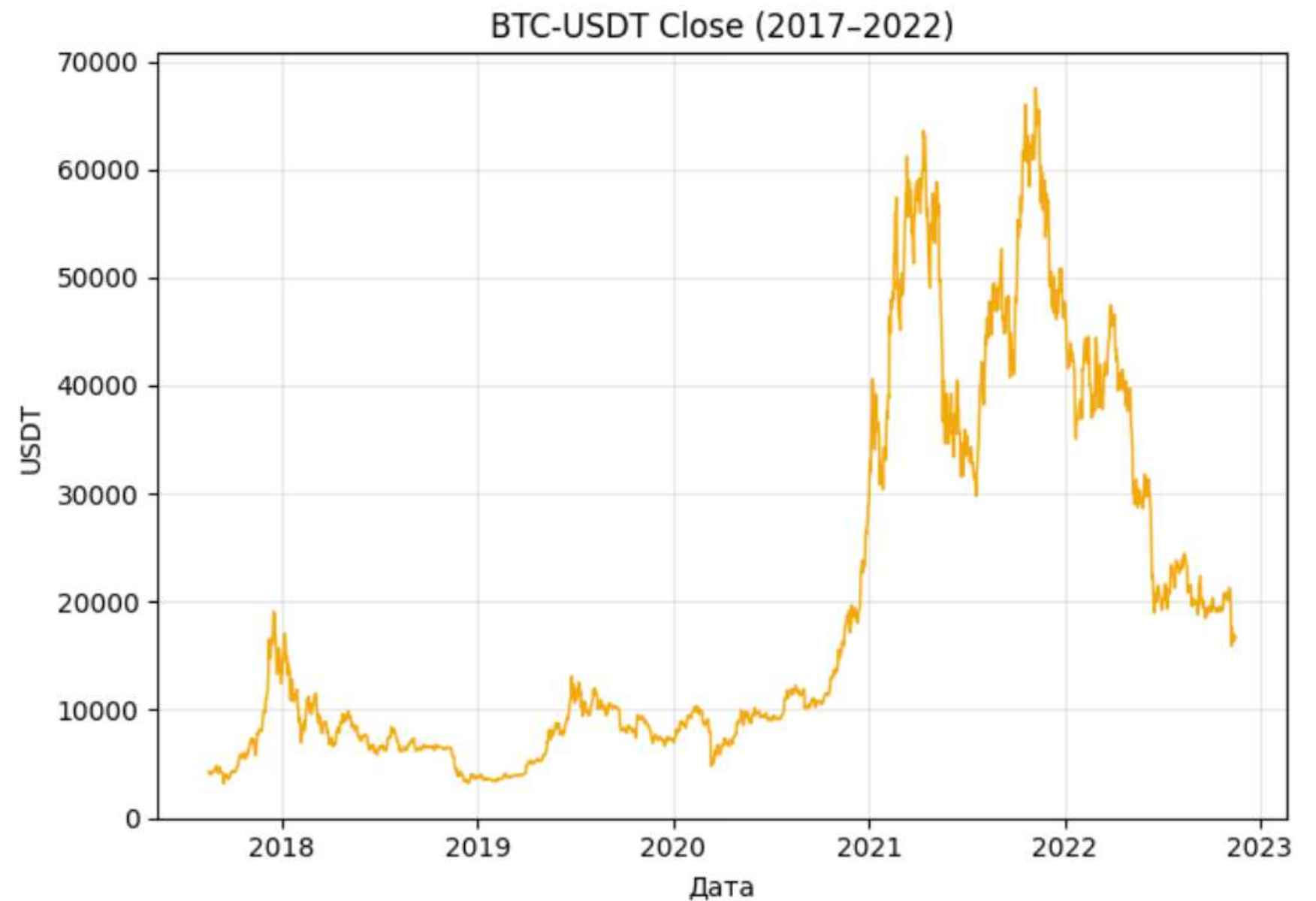
~500k

5-хв точок

× 50

за S&P 500

Достатня кількість точок для стабільного навчання трансформерів і повноцінного walk-forward тестування.



Постановка задачі

ГОРИЗОНТИ ПРОГНОЗУ

t+5 | t+15

LOOKBACK

L = 60

НАТРЕНОВАНІ ВАРІАЦІЇ МОДЕЛЕЙ

Багатовимірна

Цільова історія + історія інших фінансових інструментів і додаткові показники (об'єм, кількість угод).

Проста

Тільки історія ціни цільового активу.

ОБРОБКА ДАНИХ ПЕРЕД НАВЧАННЯМ

Ціни

$$r_t = \ln(P_t / P_{t-1})$$

Об'єми та кількість угод

$$\tilde{v}_t = \ln(1 + V_t) - \ln(1 + V_{t-1})$$

Пошук гіперпараметрів: random + neighbours

Простір пошуку занадто великий для повного перебору — реалізовано двофазний стохастичний пошук.

1 Exploration

RANDOM SAMPLING

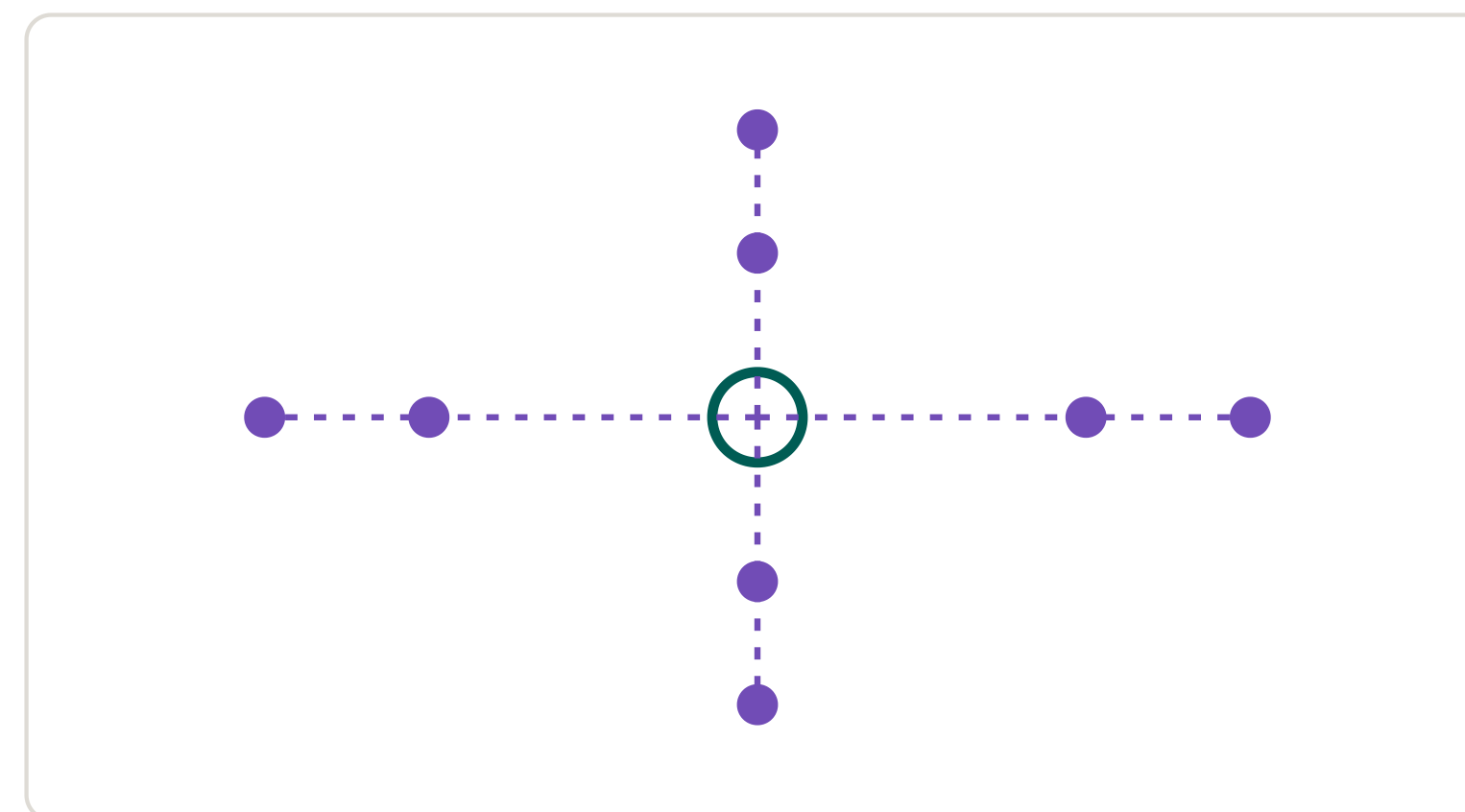
Випадкова вибірка валідних конфігурацій із фіксованим seed — рівномірне покриття простору.



2 Local refinement

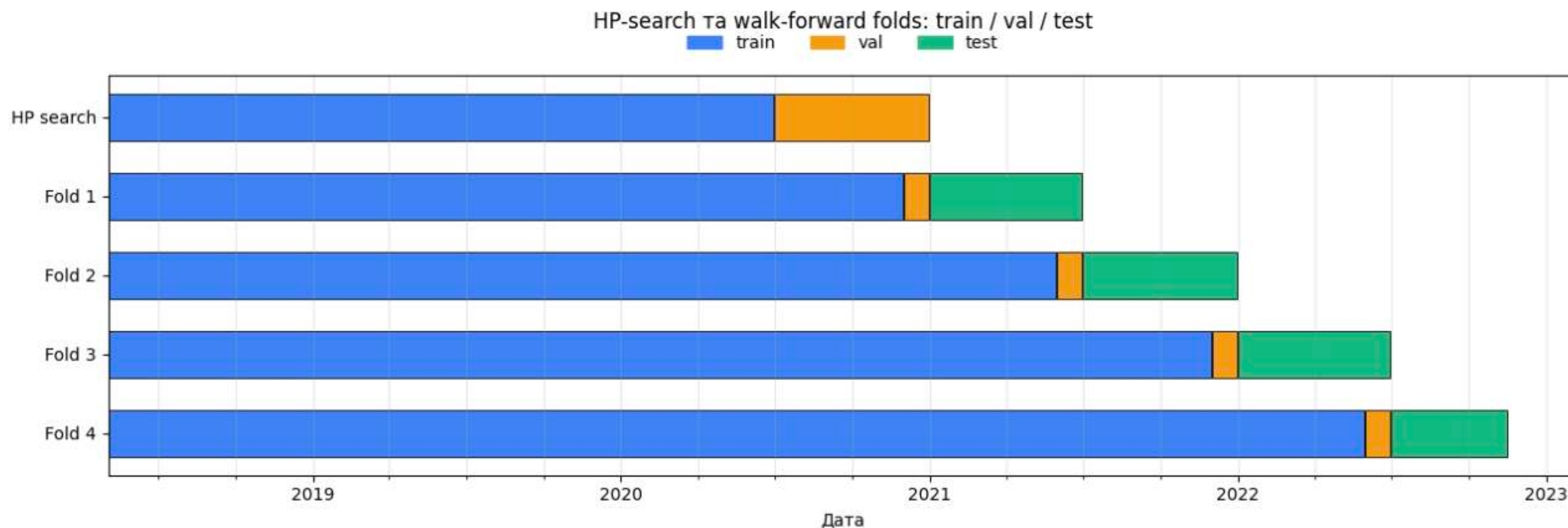
ORTHOGONAL NEIGHBOUR SEARCH

Для найкращої конфігурації — ортогональний пошук уздовж кожної осі гіперпараметрів.



Walk-forward оцінювання

Окремий зріз даних для **пошуку гіперпараметрів**; далі 4 послідовні фолди з розширенням train-вікна — кожна модель тестується на наступному, ще не баченому періоді.



Базові метрики

МЕТРИКИ ОЦІНЮВАННЯ

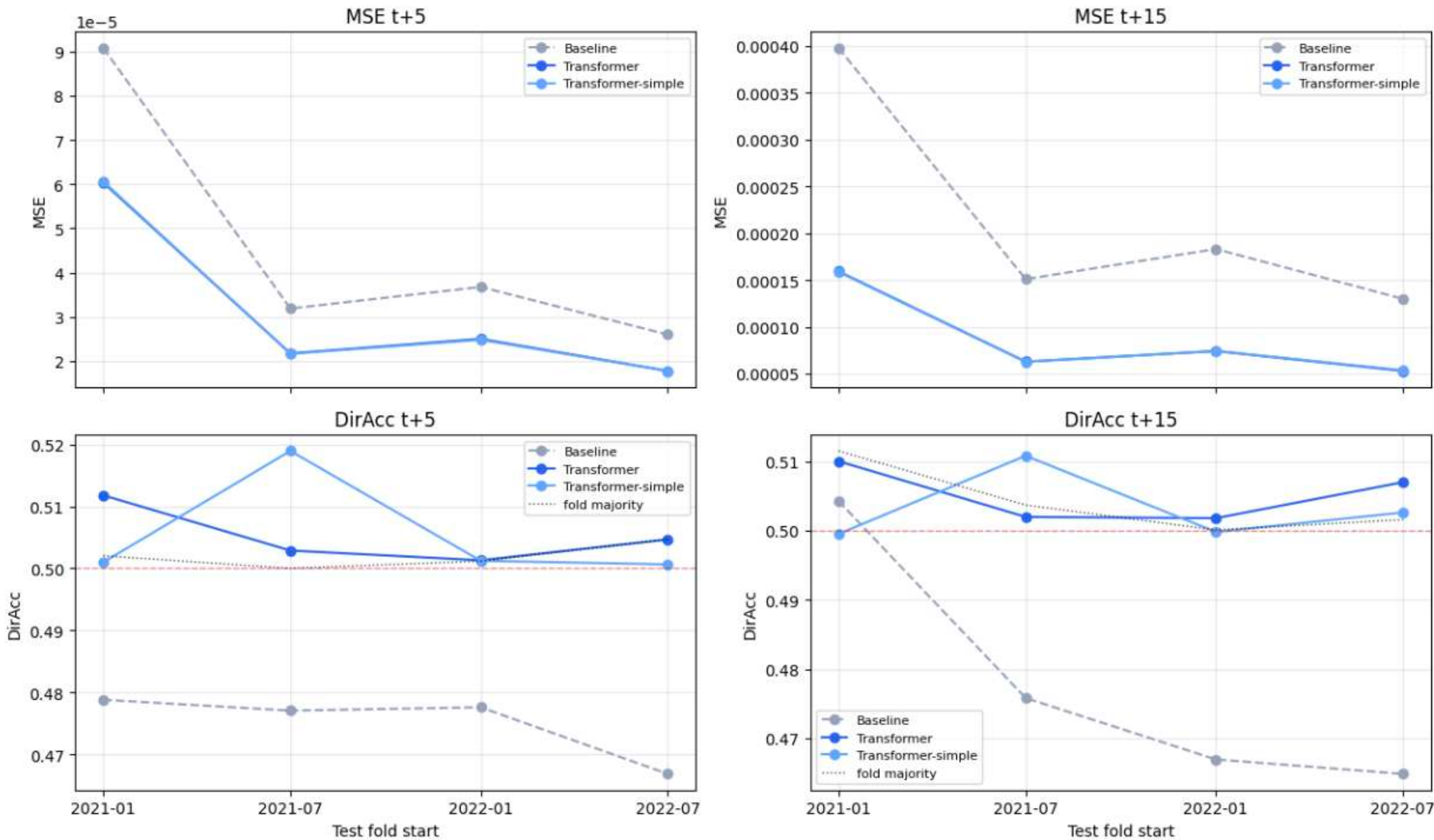
MSE

DirAcc

NAIVE BASELINE

Повертає **середній log-return** за період lookback, спроектований на горизонти 5 і 15 кроків.

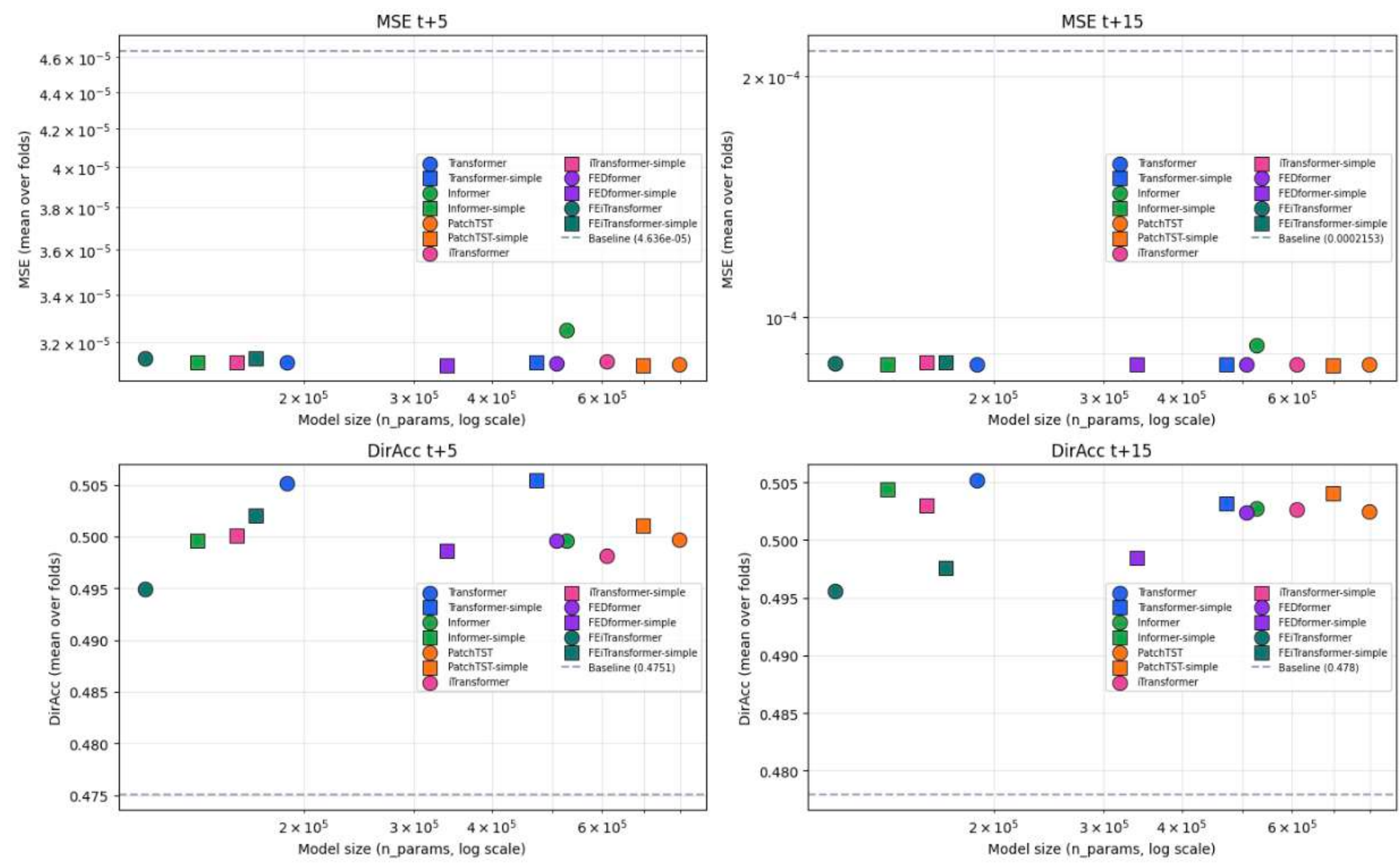
WALK-FORWARD РЕЗУЛЬТАТИ — 4 ФОЛДИ



Тестові фолди: 2021-01, 2021-07, 2022-01, 2022-07. Червона пунктирна лінія — DirAcc = 50%.

Результат vs Кількість параметрів

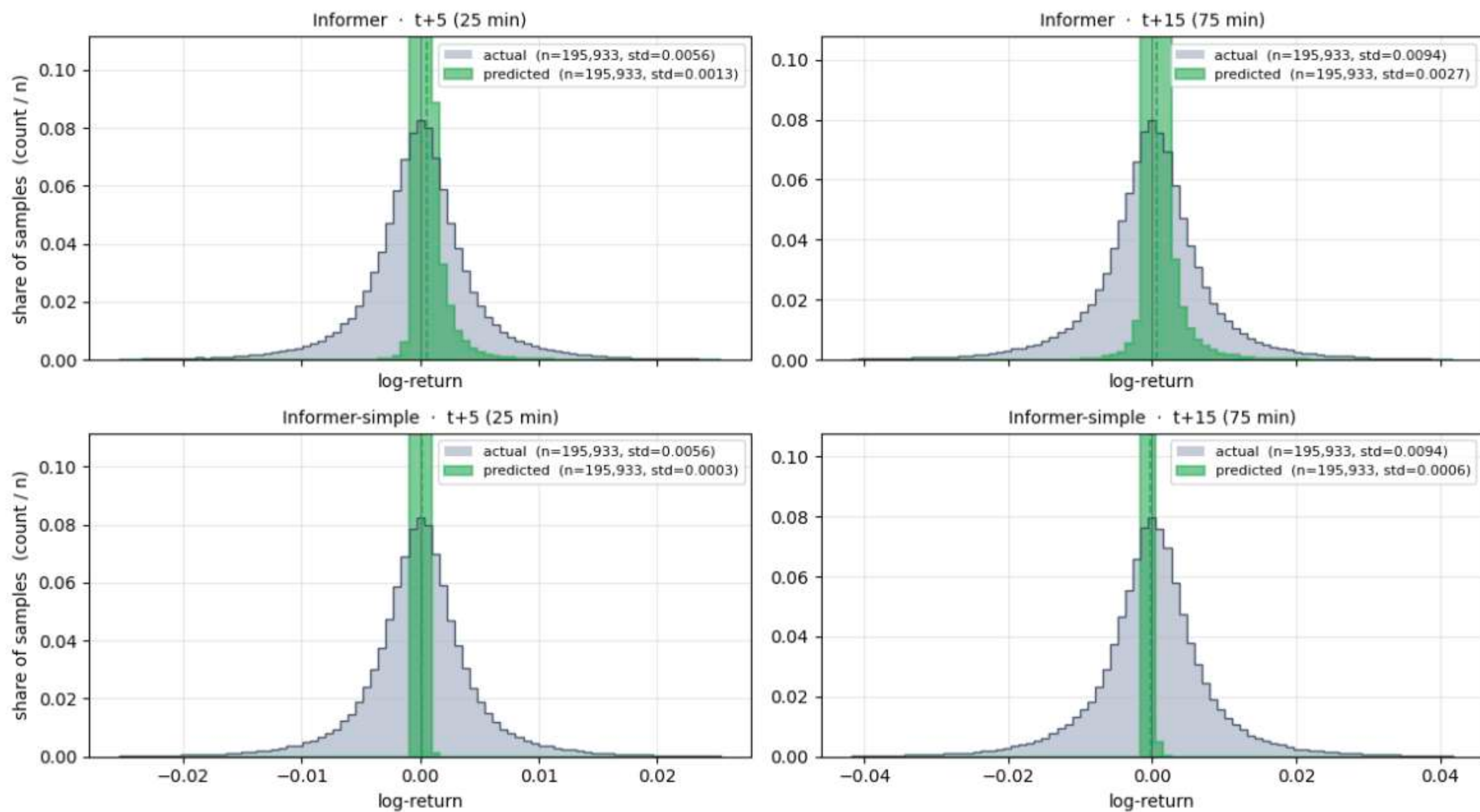
MSE TA DIRACC ПРОТИ РОЗМІРУ МОДЕЛІ — УСЕРЕДНЕНО ПО ФОЛДАХ



Пунктирна лінія — baseline. Розмір моделі від $\sim 1 \times 10^5$ до $\sim 7 \times 10^5$ параметрів (лог-шкала). Кожна точка — середнє по 4 walk-forward фолдах.

Аналіз розподілу прогнозів

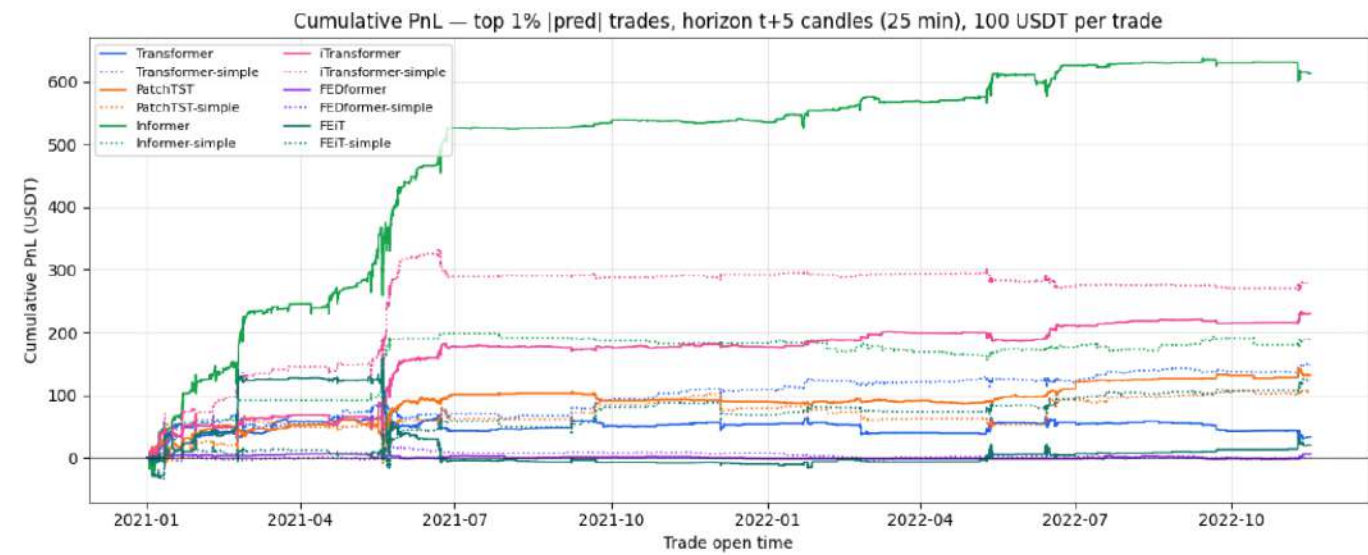
Класичні метрики не показують усієї картини — важлива не лише похибка, а й **амплітуда** сигналу.



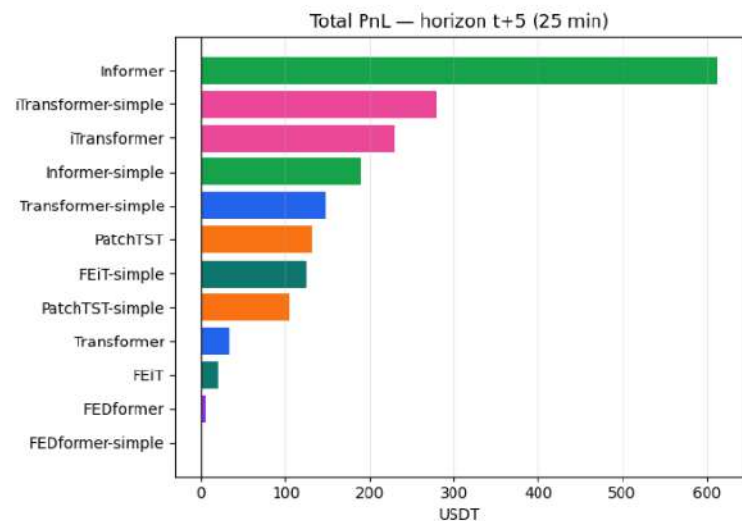
Ретроспективне тестування

1% найбільших за амплітудою прогнозів → long/short позиція на 100 USDT.

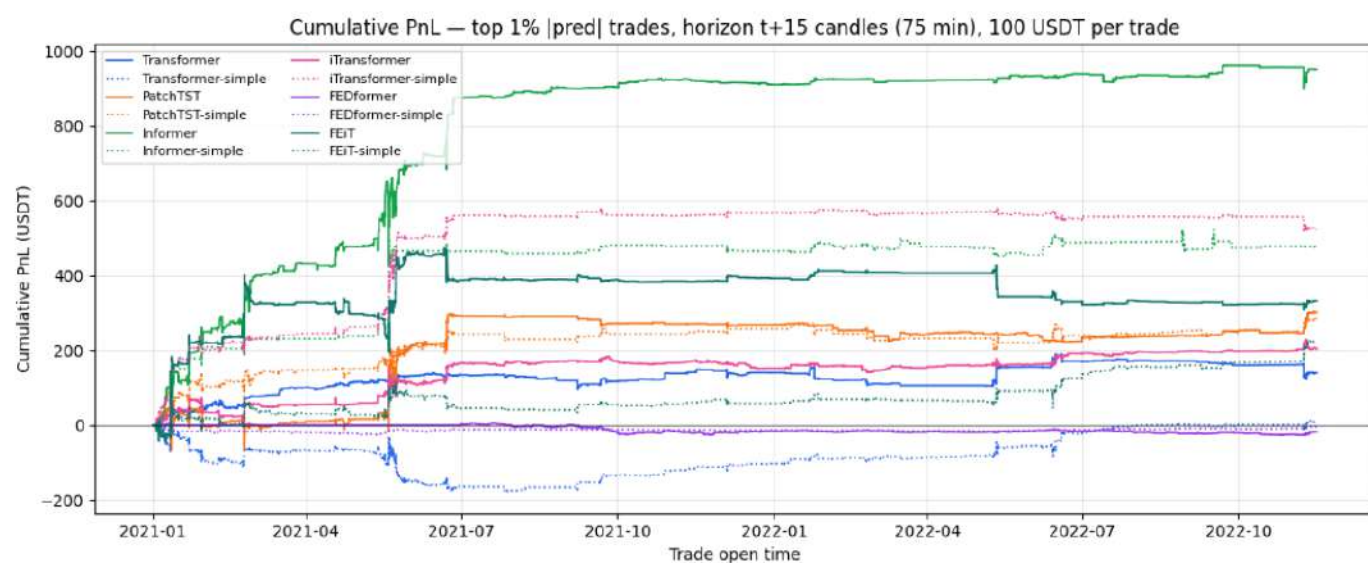
КУМУЛЯТИВНИЙ PNL · T+5



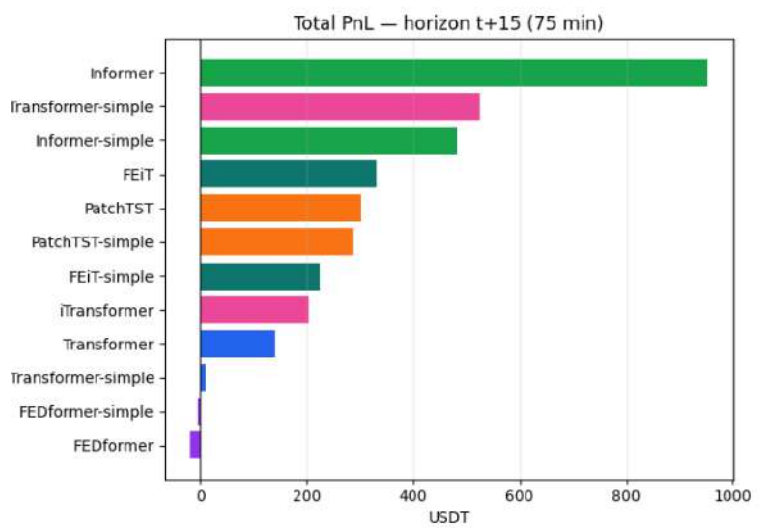
СУМАРНИЙ PNL · T+5



КУМУЛЯТИВНИЙ PNL · T+15



СУМАРНИЙ PNL · T+15



Висновки

ВОЛАТИЛЬНІСТЬ ДАНИХ

Висока невизначеність ринку змушує моделі давати **обережні, згладжені прогнози** — безпечніше передбачати близько до нуля, ніж ризикувати амплітудою.

MSE I DIRACC

За цими метриками моделі **відрізняються між собою мінімально**; усі перевищують базову статистичну модель.

ПРАКТИЧНА ДОЦІЛЬНІСТЬ

Більш сміливі моделі мають **гірші навчальні метрики**, але показують **кращі результати на практиці**.

ЧАСТОТНІ ПЕРЕТВОРЕННЯ

Моделі, що використовують **частотну декомпозицію**, показують погані результати на часових рядах з високою волатильністю.

Подальший розвиток

Покращення результату

- Адаптація функції втрат для збільшення амплітуди прогнозів.
- Перехід від точного передбачення до задачі класифікації.

Актуалізація датасету

Навчальні дані закінчуються на початку 2023 року та можуть бути неактуальними для практичного застосування.

Інтеграція

Побудова бота для алгоритмічної торгівлі, що інтегрований з реальною біржею.