

УДК 004.02, 004.67, 004.891.3  
DOI: 10.20535/SRIT.2308-8893.2023.4.04

**МОДЕЛЮВАННЯ ДИНАМІКИ РИНКУ КРИПТОВАЛЮТ  
З ВИКОРИСТАННЯМ ІНСТРУМЕНТІВ  
МАШИННОГО НАВЧАННЯ**

**Д. МАРТЪЯНОВ, Я. ВИКЛЮК, М. ФЛЕЙЧУК**

**Анотація.** Проаналізовано динаміку кон'юнктури ринку криптовалют (Bitcoin) з використанням інструментарію економетричного оцінювання на основі моделей машинного навчання. Удосконалено метод прогнозування на основі декомпозиції часових рядів та лагових зміщень фінансових індикаторів. Побудовано ансамбль моделей короткочасного прогнозу курсу Bitcoin та проаналізовано його точність порівняно з окремими складовими моделями. Використано моделі часових рядів на основі розрахованих фінансових індикаторів (ADODS, NATR, TRANGE, ATR, OBV, RSI, ADTV). Абсолютне відхилення короткочасного прогнозу склало 9,5\$ що становить 0,06% від абсолютного значення.

**Ключові слова:** ансамблі моделей, машинне навчання, часовий ряд, криптовалюта.

**АКТУАЛЬНІСТЬ**

На сучасному етапі фінансові ринки розвиваються надзвичайно високими темпами, зокрема ринок криптовалют; економетричне моделювання може стати основою для визначення факторів впливу на ціни та обсяги торгів криптовалютами. Таке моделювання передбачає збирання відповідних даних, визначення структури моделі, оцінювання її параметрів і проведення відповідного статистичного аналізу. Економетричні моделі можуть ґрунтуватися на різних методологіях, таких як аналіз часових рядів, регресійний аналіз чи аналіз панельних даних.

**Ключові етапи моделювання ринку криптовалют:** *Збирання даних.* З цією метою можливе використання ретроспективних даних про ціни на криптовалюту, обсяги торгів, ринкові та фінансові показники та інші змінні. Ці дані можна отримати з різних джерел (біржі криптовалют, постачальників фінансових даних і публічних баз даних навіть з можливістю їх комерційного використання).

*Специфікація моделі.* Визначення структури моделі, що окреслює зв'язок між залежною змінною (наприклад, ціною криптовалюти) і незалежними змінними (наприклад, ринковими та фінансовими показниками, іншими економічними факторами). Цей етап може передбачати вибір відповід-

них економетричних методів, таких, як моделі авторегресійної інтегрованої ковзної середньої (ARIMA) [1], моделі векторної авторегресії (VAR) [2] або алгоритми машинного навчання [3].

*Оцінювання параметрів.* Цей крок передбачає адаптацію моделі до ретроспективних даних та оцінювання відповідних коефіцієнтів, які окреслюють зв'язки між змінними.

*Оцінювання моделі.* Оцінювання відповідності оцінюваної моделі. Для оцінювання продуктивності моделі можна використовувати різноманітні статистичні тести, такі як перевірка гіпотез, діагностика моделі та перевірка поза вибіркою (наприклад, F-критерій Фішера [4], рівень статистичної похибки [5], T-критерій Ст'юдента [6], Тест Гренджера [7], Тест Дарбіна–Уотсона [8] тощо).

*Прогнозування та аналіз.* Після успішно застосованих описаних вище етапів модель можна використовувати для прогнозування та аналізу сценаріїв або ж відсіяти.

Указані кроки та побудова дієвої моделі допомагають зрозуміти потенційні майбутні тенденції та поведінку ринку криптовалют на основі оцінених взаємозв'язків.

Важливо зазначити, що економетричні моделі ґрунтуються на історичних даних і припущеннях, а їх точність залежить від обмежень і невизначеностей. Ринок криптовалют відомий своєю нестабільністю та чутливістю до зовнішніх факторів, таких як нормативні зміни, технологічний прогрес, інфляційні або девальваційні очікування (фіктивні змінні), інші тенденції ринку, які можуть створювати проблеми для економіко-математичного моделювання.

Для застосування економетричного моделювання з метою прогнозування тенденцій ринку криптовалют необхідно мати глибинне розуміння специфічних методів статистичного аналізу та особливостей динаміки досліджуваного ринку.

Аналіз останніх досліджень та розробок у даній галузі може допомогти підвищити точність і надійність моделей.

## АНАЛІЗ ОСТАННІХ ПУБЛІКАЦІЙ

З активізацією ринку криптовалют значна кількість вітчизняних та зарубіжних учених зацікавилися питаннями аналізу та прогнозування цього ринку з використанням сучасних методів та інструментів. З одного боку, масштабне глобальне зацікавлення цим ринком фінансовими трейдерами, з огляду на можливість одержання значної маржі у короткостроковому періоді та доступ до масштабних баз даних (динамічних датасетів з поділом на численні часові лаги) створює істотні перспективи для економетричного моделювання; на такий тип дослідження існує значний попит в аналітичних колах. Із другого боку, високий рівень волатильності та вплив багатьох суб'єктивних факторів, що важко піддаються екстраполяції, потребує від аналітиків даних використання комплексного підходу та застосування різноманітних методів і інструментів для об'єктивного обґрунтування прикладних моделей прогнозування динаміки ринку криптовалют та миттєвого прийняття рішень трейдерами, що подекуди буває непростим завданням щодо вирішення.

Питанням прогнозування курсу криптовалют, зокрема Bitcoin, з урахуванням низки факторів, що впливають на його вартість, а також окресленню щоденних тенденцій на ринку Bitcoin присвячено працю G. Gurupradeep, M. Narishvaran, K. Amsavalli [9]. У згаданій праці для прогнозування ціни закриття наступного дня, враховано такі фактори, як ціна відкриття, найвища ціна, найнижча ціна, ціна закриття, обсяг Bitcoin, обсяг інших валют і зважена ціна. При цьому автори використовують інструменти Scikit-Learn і моделі «випадкового лісу» і прогнозування.

Л. Кібальник у своїх дослідженнях [10] підтверджує, що ринок криптовалют характеризується значною волатильністю, курсовими коливаннями, складністю застосування адміністративних методів регулювання та кризовими явищами. Тому, як вважають автори, використання традиційних методів моделювання є неефективним, оскільки з використанням класичних методів аналізу досить важко отримати адекватні прогнози стосовно розвитку цього ринку. Дослідники пропонують застосування інструментарію фрактального аналізу та аналізу динаміки волатильності, що дозволяє здійснювати постійний моніторинг стану ринку та прогнозування динаміки криптовалют різного ступеня капіталізації.

Цікавим видається науковий підхід L. Catania та S. Grassi [11]. Науковці також відзначають, що дослідження фінансових часових рядів криптовалют досить складно піддається моделюванню, демонструючи екстремальні спостереження, асиметрії та часто нелінійні характеристики, які важко прогнозувати. Автори розробляють динамічну модель нового типу, здатну врахувати довгострокову пам'ять і асиметрію в процесі волатильності, а також наявність змінних у часі асиметрії та ексцесу. Емпіричне дослідження, виконане на великому масиві реальних даних щодо наборів криптовалют, засвідчує докази наявності тривалої пам'яті та ефекту кредитного плеча, що можна вважати вагомим внеском у теорію динаміки волатильності. Такі результати є важливими для управління інвестиційними активами та ризиками, пов'язаними з цим процесом.

Таким чином, аналіз та моделювання ринку криптовалют розкриває важливі наукові горизонти для застосування сучасного інструментарію економетричного прогнозування на фінансових ринках у прикладній площині.

## ПОСТАНОВКА ЗАВДАННЯ

З огляду на викладене вище, сформулюємо **мету** дослідження, як побудову ансамблевої моделі короткочасного прогнозу курсу Bitcoin на базі історичних даних та на фінансових індикаторах цієї криптовалюти. У роботі використано історичні дані з розбиттям по одній хвилині за 11,5 днів. Усього набір даних містив 18 056 записів і складався з таких змінних: «price» (курс криптовалюти), «volume» (обсяг трансакції), «count» (кількість трансакцій), «open» (курс на початок торгів), «high» (рівень максимального курсу), «close» (курс криптовалюти на завершення певного періоду), «low» (рівень мінімального курсу). Слід зазначити, що у наявному реальному ряду даних окремі з них були пропущені, проте період відсутності даних не перевищував трьох хвилин. Тому для подальшого аналізу відсутні дані були заповнені з використанням методу інтерполяції з використанням лінійної регресії. Результат інтерполяції наведено на рис. 1.

Як відомо, моделі прогнозування курсу валют можна поділити на дві категорії.

1. *Модель часових рядів.* Досліджується значення криптовалюти з урахуванням її значень за попередні періоди часу без урахування інших факторів [12].

2. *Модель на основі фінансових індикаторів.* Курс криптовалюти прогнозується на основі фінансових індикаторів, які у своїй природі враховують часові затримки та характеристики криптовалюти за попередні моменти часу [13].

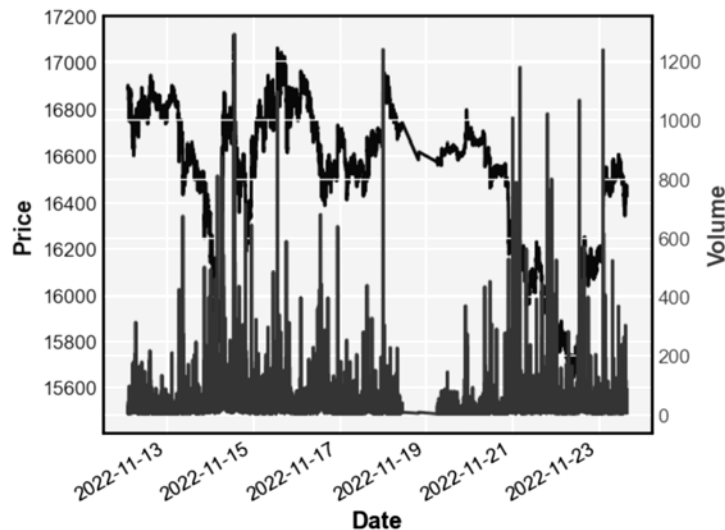


Рис. 1. Динаміка курсу та обсягів продажу Bitcoin, 2022-11-13 – 2022-11-23  
Розраховано авторами.

У роботі поєднано ці два підходи та використано різноманітні методи машинного навчання, а також поєднано найбільш дієві з них в ансамбль.

Алгоритм розрахунку:

- 1) розрахунок фінансових індикаторів;
- 2) часова декомпозиція цільового поля;
- 3) визначення лагових затримок та формалізація моделі;
- 4) побудова та аналіз множини моделей машинного навчання;
- 5) побудова прогнозу на основі ансамблю моделей.

## РОЗРАХУНОК ФІНАНСОВИХ ІНДИКАТОРІВ

Як досліджувані індикатори обрано найбільш популярні та вживані характеристики кон'юнктури ринку криптовалют: ADODS, NATR, TRANGE, ATR, OBV, RSI, ADTV.

Коротко викладемо основні підходи до розрахунку цих фінансових індикаторів.

ADODS — Chaikin A/D Oscillator [14].

Цей індикатор на основі обсягу для вимірювання сукупного грошового потоку. Індикатор припускає, що ступінь тиску купівлі або продажу можна визначити за розташуванням закриття з урахуванням значень максимуму та мінімуму обмінного курсу за досліджуваний період.

Крива ADODS — це загальна сума кожного періоду обсягу грошових потоків (MFV):

$$MFV = \frac{(Close - Low) - (High - Close)}{High - Low} \times Volume;$$

$$ADODS_{p+1} = ADODS_p + MFV,$$

де  $p$  — індикатор періоду.

ATR Normalized (NATR) [15].

Цей індикатор використовується в технічному аналізі для вимірювання рівня волатильності і визначається за рівняннями:

$$NATR = \frac{100 \times ATR}{Close};$$

$$ATR_{p+1} = \frac{ATR_p (n-1) + TR}{n}, \quad (1)$$

де  $p$  — індикатор періоду;  $n$  — кількість періодів;

$$ATR_p = \frac{1}{n} \sum_{i=1}^n TR_i;$$

$$TR = \max[(High_p - Low_p), |High_p - Close_{p-1}|, |Low_p - Close_{p-1}|], \quad (2)$$

де  $n$  — аналізована кількість періодів часу;  $p$  — індикатор періоду;  $TR$  — діапазон торгівлі.

- *True Range (TRANCE)* [16] — це технічний індикатор, що вимірює денний діапазон курсу плюс будь-який розрив від ціни закриття попереднього дня ((2)).

- *Середній справжній діапазон (ATR)* [17] — це індикатор технічного аналізу, який вимірює волатильність ринку шляхом розкладання всього діапазону ціни активу за цей період (рівняння (1)).

- *Балансовий обсяг (OBV)* [18] — це технічний індикатор імпульсу торгівлі, який використовує потік обсягу для прогнозування змін курсу акцій:

$$OBV_{p+1} = OBV_p + \begin{cases} Volume : Close_{p+1} > Close_p; \\ 0 : Close_{p+1} = Close_p; \\ Volume : Close_{p+1} < Close_p, \end{cases}$$

де  $p$  — індикатор періоду.

- *Індекс відносної сили (RSI)* — широко використовуваний індикатор осцилятора. Для трейдерів із середньою реверсією він може генерувати сигнали для визначення рівня перекупленої чи перепроданої ціни. RSI також можна використовувати для визначення сили руху тренду:

$$RSI_{step\ one} = 100 - \left[ \frac{100}{1 + \frac{Avg\ Gain}{Avg\ Loss}} \right],$$

де  $Avg\ Gain$  — середній прибуток (в абсолютному вираженні);  $Avg\ Loss$  — середній збиток в абсолютному вираженні.

- *Середній прибуток* ( $Avg\ Gain$ ) або *збиток* ( $Avg\ Loss$ ) [19], що використовується у цьому методі, є середнім відсотковим приростом або збитком відповідно впродовж досліджуваного періоду. У формулі використовується абсолютне значення обсягу середніх втрат. Періоди із втратою ціни враховуються як нуль у розрахунках середнього прибутку. Лаги з тенденцією до підвищення цін враховуються як нуль у процесі розрахунку середнього збитку. Стандартна кількість періодів, які використовуються для розрахунку початкового значення RSI, становить 14.

За умови коли буде розраховано значення 14 періодів, можна переходити до виконання наступного етапу, метою якого є згладжування результатів таким чином, щоб RSI лише наближався до 100 чи до нуля на чітко окресленому трендовому ринку:

$$RSI_{step\ two} = 100 - \left[ \frac{100}{1 + \frac{(Avg\ Gain)_p \times 13 + (Avg\ Gain)_{p+1}}{(Avg\ Loss)_p \times 13 + (Avg\ Loss)_{p+1}}} \right].$$

- *Середній щоденний обсяг торгів* ( $ADTV$ ) — це середній обсяг продажу певної валюти впродовж дня. Середній щоденний обсяг торгів є важливим показником, оскільки високий або ж низький обсяг торгів приваблює різні типи трейдерів та інвесторів. Багато трейдерів та інвесторів віддають перевагу вищому рівню середнього щоденного обсягу торгів порівняно з низьким, оскільки за умови великого обсягу легше відкривати та виходити з позицій. Активи з невеликим обсягом мають менше покупців і продавців, і тому, в цьому випадку, може бути важче увійти або вийти на торги на рівні бажаної ціни:

$$ADTV = \frac{Volume_{daily}}{\sum trades}.$$

## ЧАСОВА ДЕКОМПОЗИЦІЯ ЦІЛЬОВОГО ПОЛЯ

Як декомпозицію цільового поля у дослідженні обрано адитивну модель (11) [20]:

$$price(t) = Trend(t) + Seasonal(t) + Residual(t).$$

Трендову залежність  $Trend(t)$  [21] розраховано як лінійну згортку методом найменших квадратів, екстрапольованим на обох кінцях [22].

Сезонна компонента  $Seasonal(t)$  визначалась методом ковзного середнього із періодом один день (1440 хв) [23].

Результат до композиції подано на рис. 2. Як видно з рисунка, на графіку взаємозв'язку курсу Bitcoin з досліджуваними фінансовими індикаторами спостерігаються чіткі трендові та сезонні компоненти залежності з невеликими розривами у період вихідних днів. У результаті декомпозиції компонента  $Residual(t)$  виступатиме в ролі цільового поля.

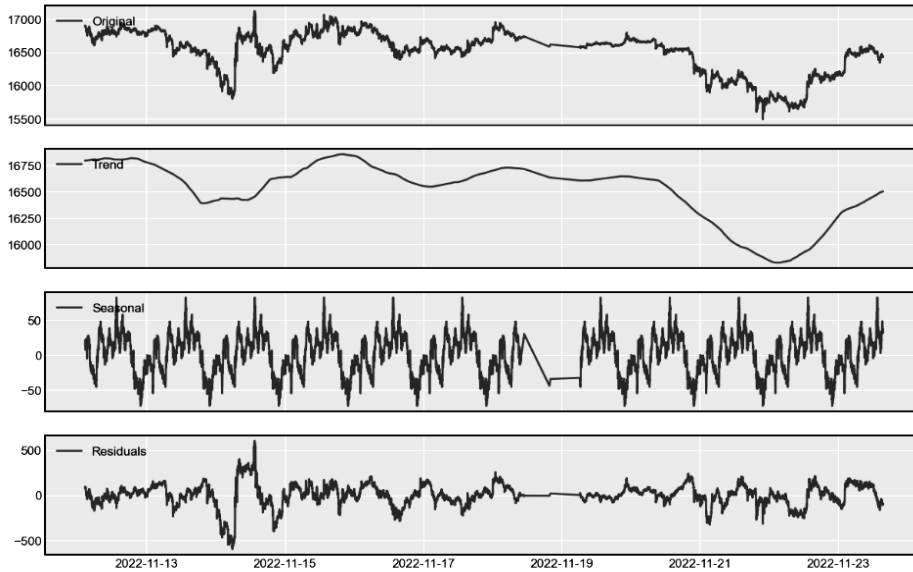


Рис. 2. Декомпозиція часового ряду курсу Bitcoin, 2022-11-13 – 2022-11-23. Розраховано авторами.

Порівняння динаміки фінансових індикаторів з отриманою випадковою компонентою показано на рис. 3.

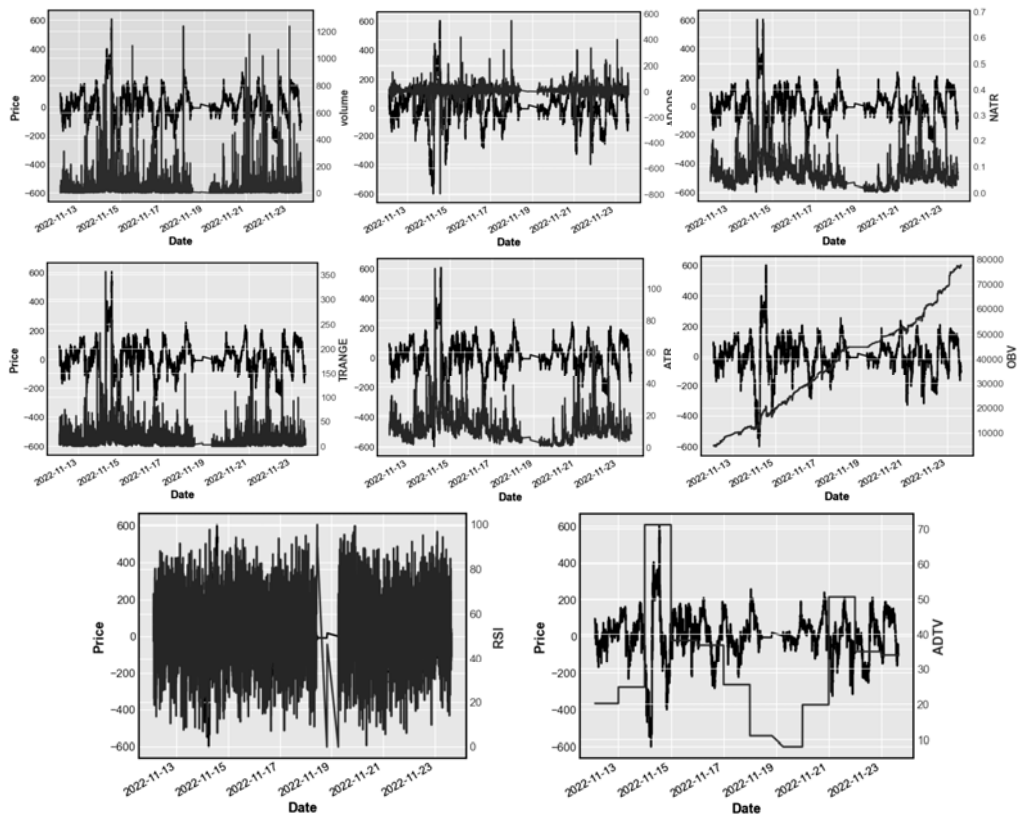


Рис. 3. Динаміка залежності курсу Bitcoin від фінансових індикаторів, 2022-11-13 – 2022-11-23  
Розраховано авторами

Як видно з рис. 3, динаміка досліджуваної залежної змінної від фінансових індикаторів суттєво розрізняються і потребують подальшого аналізу та встановлення наявності природи і типу функціональної залежності.

### ВИЗНАЧЕННЯ ЛАГОВИХ ЗАТРИМОК ТА ФОРМАЛІЗАЦІЯ МОДЕЛІ

Додатково, для аналізу наявності факту і напряму зв'язку між фінансовими індикаторами та цільовим полем, виконано кореляційний аналіз (рис. 4).

Як видно з аналізу, між полями ATR та NATR існує сильний лінійний зв'язок, тому для подальшого аналізу залишено поле NATR, оскільки воно являє собою нормалізоване ATR. Також існує зв'язок між полем Volume та TRANGE. Цей зв'язок є на межі лінійного, тому прийнято рішення залишити ці два поля для подальшого детальнішого аналізу. Окрім того, можна побачити, що випадкова компонента майже не корелює лінійно з жодним із зазначених фінансових індикаторів. Це свідчить або про відсутність лінійного зв'язку та необхідність використання нелінійних моделей, або про необхідність побудови комплексної лінійної моделі, що включає в себе всі зазначені фінансові індикатори (багатофакторної регресійної моделі). Проте, як показує практика трейдингу на ринку криптовалют, саме визначені нами індикатори найчастіше враховуються трейдерами для прийняття швидких біржових рішень.

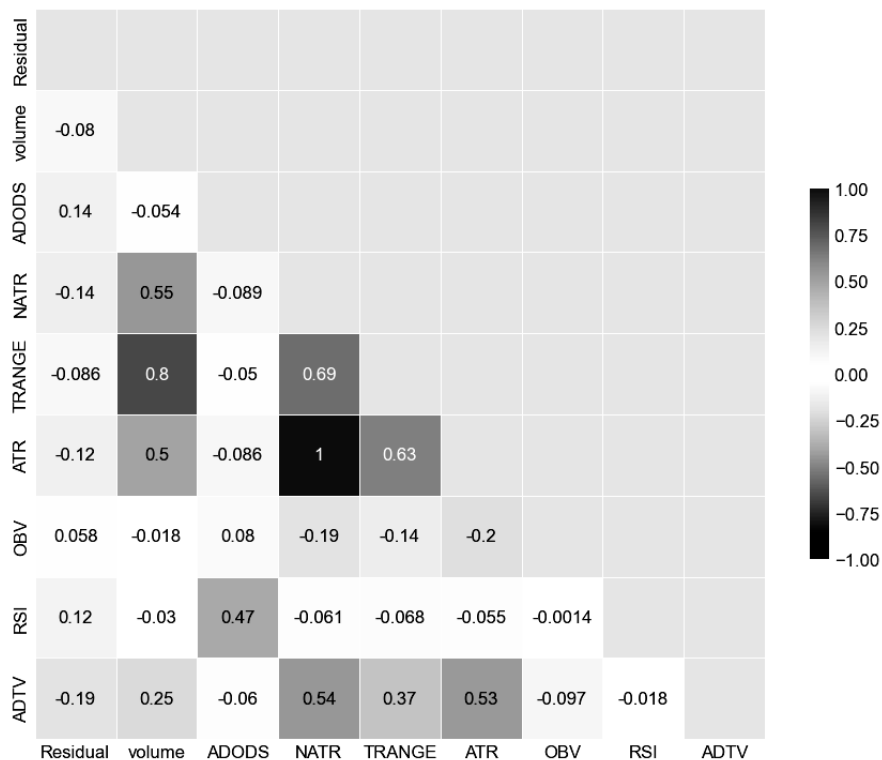


Рис. 4. Матриця кореляційного зв'язку між фінансовими індикаторами динаміки ринку криптовалют та курсом Bitcoin, 2022-11-13 – 2022-11-23  
Розраховано авторами



Для врахування впливу лагової затримки випадкової компоненти ( $Residual(t)$ ) виконано автокореляційний аналіз у межах сезонної компоненти, тобто на 1440 хв. Результат аналізу подано на рис. 5.

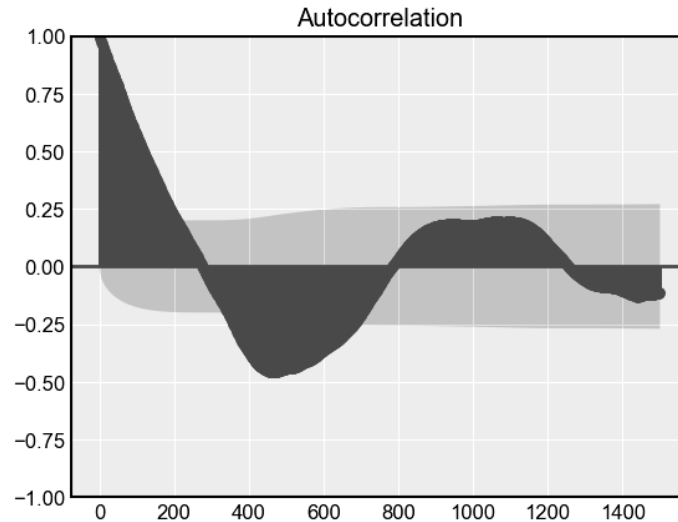


Рис. 5. Автокореляційний аналіз для випадкової компоненти ( $Residual(t)$ ), 2022-11-13 – 2022-11-23  
Розраховано авторами

Як видно з рисунка, лагова кореляція істотно знижується до нуля і обернено зростає в межах лагу 465 хв (7 год 45 хв), що відповідає сценарію класичних торгів на ринку криптовалют. Для уточнення необхідного лагу та усунення взаємної кореляції виконано частковий автокореляційний аналіз. Результат ілюструє рис. 6.

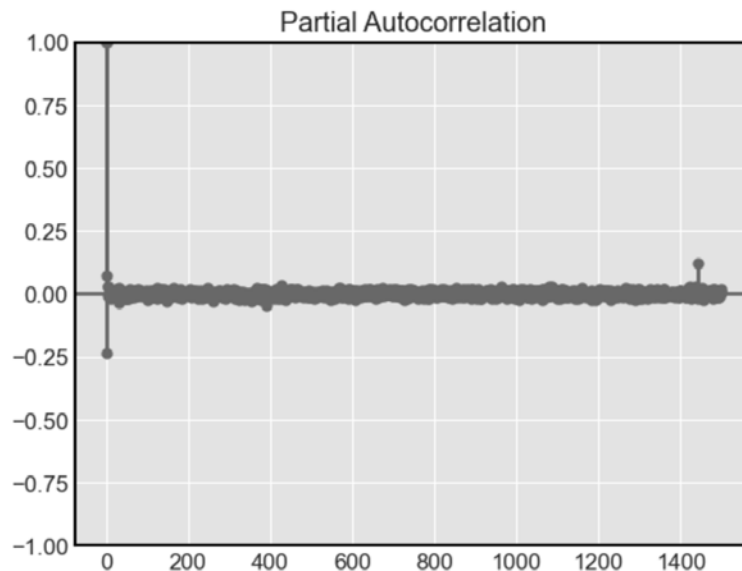


Рис. 6. Частковий автокореляційний аналіз для випадкової компоненти ( $Residual(t)$ ), 2022-11-13 – 2022-11-23  
Розраховано авторами.

Як видно з рис. 6, найвищого рівня кореляційний зв'язок притаманний для компоненти з лаговою затримкою 1, 2 та 1440. Для встановлення наявності лагових затримок між іншими фінансовими індикаторами та випадковою компонентною також виконано відповідний автокореляційний аналіз (рис. 7). Як видно з рисунка, всі коефіцієнти кореляції є дуже низькими, тому немає сенсу враховувати лагову затримку під час побудови моделей.

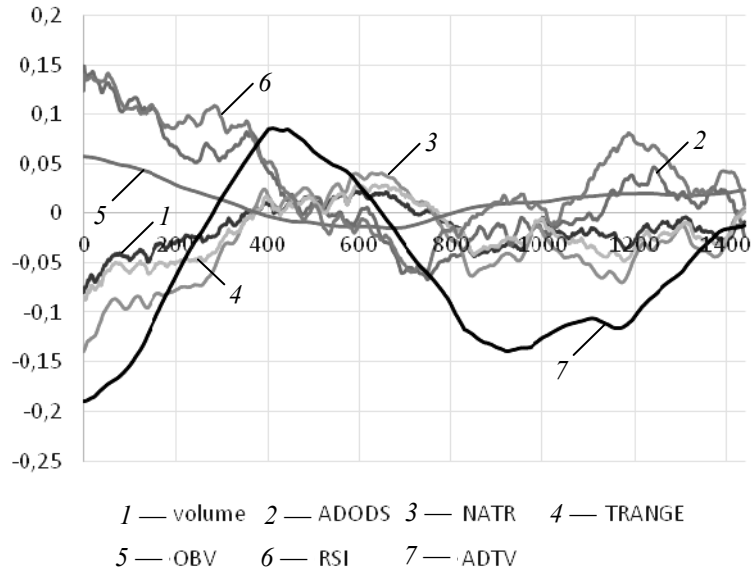


Рис. 7. Залежність коефіцієнта кореляції між випадковою компонентною ( $Residual(t)$ ) та фінансовими індикаторами з відповідним часовим лагом, 2022-11-13 – 2022-11-23  
Розраховано авторами

Це пояснюється специфікою виконаного аналізу. Оскільки йдеться про розбиття даних по одній хвилині, то логічним є відсутність такого впливу, адже кон'юнктура ринку криптовалют у таких часових лагах дуже швидко змінюється і часто трейдери враховують більші часові лаги.

Таким чином, у результаті проведеного аналізу формалізовану модель прогнозу курсу Bitcoin можна подати у такому вигляді:

$$BIT_{USD}(t+1) = Trend(t) + Seasonal(t) + F(Residual(t), Residual(t-1), Residual(t-2), Residual(t-465), Residual(t-1440), volume(t), ADODS(t), NATR(t), TRANGE(t), OBV(t), RSI(t), ADTV(t)). \quad (3)$$

Як видно з рівняння (3), курс Bitcoin можна розраховувати як суму трендової, сезонної та випадкової компонент, яка прогнозується за допомогою ансамблю моделей.

## ПОБУДОВА ТА АНАЛІЗ МНОЖИНИ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ

Одним із завдань дослідження є побудова моделі для прогнозу курсу валют на майбутні періоди. Модель, подана у вигляді рівняння (3), дає змогу прогнозувати лише на один період наперед. Для побудови прогнозу на декілька майбутніх періодів можна застосувати два підходи:

1. Для кожного із факторів побудувати прогнозу модель на один період наперед і відповідно використовувати прогнозні значення як вхідні для прогнозування на наступний період часу (ланцюговий метод).

Як вихідне поле використовувати значення  $Residual(t + lag)$  зміщеного на необхідні періоди наперед.

Перший підхід потребує побудови великої кількості моделей і зумовить накопичення помилки за ітерацій на кожний наступний період. Другий підхід дозволить зменшити накопичувальну помилку, але потребує побудови окремої моделі для кожного лагу прогнозування. Тому для розрахунків обрано саме другий підхід.

Як тестові моделі виокремлено: лінійну множинну регресію [24] (Linear), регресію Губара [25] (Huber) та нейронну мережу зворотного поширення помилки [26] (MLP), що складається із двох прошарків по 100 нейронів кожен. Для тестування точності та адекватності моделей використано кросвалідацію з величиною розбиття 3. Результати оцінювання точності прогнозування наведено в табл. 1.

Як видно з таблиці, точність прогнозу у випадку тестових даних є досить високою. Як і слід було очікувати, зі збільшенням часу прогнозу точність моделей поступово зменшується. Однак навіть на 10 періодів наперед вона є досить висока.

**Таблиця 1.** Коефіцієнти кореляції моделі прогнозу ( $R^2$ ) для різних лагів для  $Residual(t + lag)$ , 2022-11-13 – 2022-11-23

Lag	Linear	MLP	Huber	VotingRegressor
0	0,99	0,99	0,99	0,99
1	0,98	0,98	0,98	0,98
2	0,97	0,96	0,97	0,97
3	0,96	0,96	0,96	0,96
4	0,96	0,94	0,95	0,95
5	0,95	0,92	0,95	0,95
6	0,94	0,93	0,94	0,94
7	0,93	0,92	0,93	0,93
8	0,92	0,91	0,92	0,92
9	0,91	0,90	0,91	0,91
10	0,90	0,89	0,90	0,90

*Розраховано авторами.*

Для уникнення випадкових флуктуацій моделей їх об'єднано в ансамбль за допомогою VotingRegressor, який усереднює результати окремих моделей.

## ПОБУДОВА ПРОГНОЗУ НА ОСНОВІ АНСАМБЛЮ МОДЕЛЕЙ

Для побудови прогнозу набір даних було поділено на навчальний та тестовий в пропорції 90/10. Як тестовий набір обрано дані за останній період. Результати навчання і тестування наведено в табл. 2.

**Таблиця 2.** Точність навчання ( $R^2$ ) на навчальному та тестовому наборі даних для  $Residual(t + lag)$

Lag	Linear		MLP		Huber		VotingRegressor	
	Train	Test	Train	Test	Train	Test	Train	Test
0	0,99	0,99	0,99	0,95	0,99	0,99	0,99	0,99
1	0,99	0,98	0,99	0,98	0,98	0,98	0,99	0,98
2	0,98	0,97	0,98	0,97	0,97	0,97	0,98	0,97
3	0,97	0,96	0,97	0,94	0,97	0,96	0,97	0,96
4	0,96	0,95	0,94	0,85	0,96	0,95	0,96	0,95
5	0,95	0,94	0,93	0,88	0,95	0,94	0,95	0,90
6	0,94	0,93	0,94	0,89	0,94	0,93	0,94	0,92
7	0,93	0,92	0,93	0,92	0,93	0,92	0,93	0,92
8	0,92	0,91	0,92	0,89	0,92	0,91	0,92	0,91
9	0,91	0,90	0,91	0,89	0,91	0,90	0,91	0,89
10	0,90	0,89	0,91	0,86	0,90	0,89	0,90	0,87

Розраховано авторами.

Як видно з таблиці, точність навчання поступово спадає з лагом прогнозу. Чим менша затримка в прогнозі, тим точніший результат. У розглянутому випадку, кожен лаг становить 1 хв. Отже, чим швидше буде прийнято рішення, тим менша помилка в його прийнятті. Це точність прогнозу для випадкової компоненти числового ряду. Для визначення точності моделі (3) необхідно до цих результатів додати сезонну та трендову компоненти і побудувати прогноз реального курсу валют. Прогноз курсу Bitcoin для тестових значень подано на рис. 8.

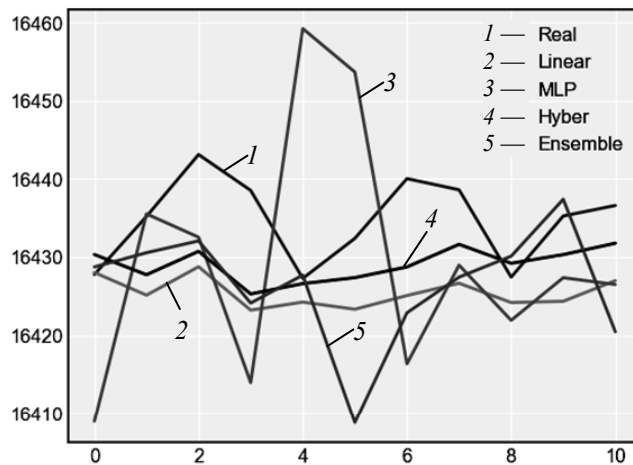


Рис. 8. Прогноз курсу Bitcoin згідно з розробленим ансамблем моделей на 10 хв наперед порівняно з тестовими даними

Розраховано авторами

Як видно з рисунка, значення мають досить сильне відхилення від реальних даних, однак розрахунок середнього абсолютного відхилення становить 9,5, що становить 0,06% від абсолютного значення. Це може свідчити про достатній рівень точності моделі.

## ВИСНОВКИ

1. На сучасному етапі фінансові ринки розвиваються надзвичайно високими темпами, зокрема і ринок криптовалют, для якого економетричне моделювання може стати основою для визначення факторів впливу на його кон'юнктуру. Попри істотні можливості застосування економетричного інструментарію для аналізу цього ринку (генерування об'ємних даних у режимі реального часу), слід вказати й на певні обмеження такого аналізу, зокрема істотна волатильність ринку та часто суб'єктивний чи інтуїтивний характер прийняття рішень трейдерами криптовалют.

2. Для оперативного прийняття рішень щодо купівлі-продажу валютних активів на фондових біржах запропоновано застосування ансамблевої моделі короткочасного прогнозу курсу валюти, ґрунтуючись на історичних даних основних характеристик ринкової кон'юнктури (курс криптовалюти, обсяг транзакції, кількість транзакцій, курс на початок торгів, рівень максимального курсу, курс криптовалюти на завершення певного періоду, рівень мінімального курсу та фінансових індикаторів цієї криптовалюти (ADODS, NATR, TRANGE, ATR, OBV, RSI, ADTV).

3. У процесі аналізу моделі прогнозування курсу криптовалют можна використати дві категорії моделей: модель часових рядів; модель на основі фінансових індикаторів.

4. Визначено ключові етапи моделювання ринку криптовалют: 1) збирання даних; 2) специфікація моделі; 3) оцінювання параметрів; 4) оцінювання моделі; 5) прогнозування та аналіз.

5. Серед основних кроків для побудови моделі для аналізу ринку криптовалют рекомендовано застосовувати: часову декомпозицію цільового поля; визначення лагових затримок та формалізацію моделі; побудову та аналіз множини моделей машинного навчання; побудову прогнозу на основі ансамблю моделей.

6. Основні висновки з виконаного аналізу:

- у взаємозв'язку курсу Bitcoin з досліджуваними фінансовими індикаторами спостерігаються чіткі трендові та сезонні компонентні залежності з невеликими розривами у період вихідних днів;

- найміцніший зв'язок визначено між такими змінними, як ATR і NATR, а також Volume та TRANGE. Цей зв'язок є на межі лінійного, тому прийнято рішення залишити ці два поля для подальшого детальнішого аналізу. Відзначено, що випадкова компонента майже не корелює лінійно з жодним із зазначених фінансових індикаторів. Це свідчить або про відсутність лінійного зв'язку та необхідність використання нелінійних моделей, або про потребу побудови комплексної лінійної моделі, що включає в себе основні фінансові індикатори (багатофакторної регресійної моделі);

- встановлено, що лагова кореляція істотно знижується до нуля і обернено зростає в межах лагу 465 хв (7 год 45 хв), що відповідає сценарію класичних торгів на ринку криптовалют; найвищий рівень кореляційного зв'язку притаманний для компоненти з лаговою затримкою 1, 2 та 1440; оскільки всі коефіцієнти кореляції між випадковою компонентою та фінансовими індикаторами є надто низькими, тому немає сенсу враховувати лагову затримку у побудові такого типу моделей; курс Bitcoin можна розраховувати як суму трендової, сезонної та випадкової компонент, яка прогнозується за допомогою ансамблю моделей.

## ЛІТЕРАТУРА

1. Gandhi Pratik, *7 Statistical Tests to validate and help to fit ARIMA model*. Available: <https://towardsdatascience.com/7-statistical-tests-to-validate-and-help-to-fit-arima-model-33c5853e2e93>
2. Kotzé Kevin, *Vector autoregression models*. Available: <https://kevinkotze.github.io/ts-7-var/>.
3. Taiwo Oladipupo Ayodele, *Types of Machine Learning P.19-28. Algorithms*. Available: [https://cdn.intechopen.com/pdfs/10694/InTech-Types\\_of\\_machine\\_learning\\_algorithms.pdf](https://cdn.intechopen.com/pdfs/10694/InTech-Types_of_machine_learning_algorithms.pdf)
4. Onchiri Sureiman and Callen Moraa Mangera, “F-test of overall significance in regression analysis simplified,” *Journal of the Practice of Cardiovascular Sciences*, vol. 6, issue 2, pp. 116–122, May-August 2020. doi: 10.4103/jpcs.jpcs\_18\_20.
5. Sander Greenland et al., “Statistical Tests, P-values, Confidence Intervals, and Power: A Guide to Misinterpretations,” *The American Statistician, Online Supplement*, pp. 1–12, 2016. Available: [https://events.gwdg.de/event/482/attachments/391/580/anwer\\_to\\_the\\_ASA\\_statement\\_misinterpretations.pdf](https://events.gwdg.de/event/482/attachments/391/580/anwer_to_the_ASA_statement_misinterpretations.pdf)
6. L. Brown, “The conditional level of Student’s – test,” *The Annals of Mathematical Statistics*, vol. 38, no. 4, pp. 1068–1071, Aug., 1967. Available: <https://faculty.wharton.upenn.edu/wp-content/uploads/2012/04/Conditional-level-of-students-t-test.pdf>
7. Xiaojun Song and Abderrahim Taamouti, “A better understanding of Granger causality analysis: A big data environment,” *Oxford Bulletin of Economics & Statistics*, pp. 2–25, August 2019. Available: [https://www.researchgate.net/publication/329803300\\_A\\_Better\\_Understanding\\_of\\_Granger\\_Causality\\_Analysis\\_A\\_Big\\_Data\\_Environment](https://www.researchgate.net/publication/329803300_A_Better_Understanding_of_Granger_Causality_Analysis_A_Big_Data_Environment)
8. Champion Robert and Mills M. Terence, “Demonstrating the Durbin-Watson Statistic,” *Journal of the Royal Statistical Society Series D (The Statistician)*, 47(4), pp. 643–644, December 1998. doi: 10.1111/1467-9884.00161.
9. G. Guruprdeep, M. Harishvaran, and K. Amsavalli, “Cryptocurrency Price Prediction using Machine Learning, 2023,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 12, issue 4, pp. 808–814, April 2023. doi: 10.17148/IJARCCCE.2023.124140.
10. L.O. Kibalnyk, O.A. Kovtun, and G.B. Danylchuk, “Modeling and analysis of the current state of the cryptocurrency market,” *Economics and management in the period of digital transformation of business, society and the state: materials of the Jubilee International Scientific and Practical Conference (May 28–29, 2020, Zaporizhzhia)*. Zaporizhzhia: ZNU Engineering Institute, 2020, pp. 112–117.
11. Leopoldo Catania and Stefano Grassi, “Modelling Crypto-Currencies Financial Time-Series,” *SSRN Electronic Journal*, pp. 2–38, August 2017. doi: 10.2139/ssrn.3028486.
12. George E.P. Box, *Time series analysis: forecasting and control*; fifth edition, 2016, 668 p. Available: [http://repo.darmajaya.ac.id/4781/1/Time%20Series%20Analysis\\_%20Forecasting%20and%20Control%20%28%20PDFDrive%20%29.pdf](http://repo.darmajaya.ac.id/4781/1/Time%20Series%20Analysis_%20Forecasting%20and%20Control%20%28%20PDFDrive%20%29.pdf)
13. Andrea Majdánková, Blanka Giertliová, and Iveta Hajdúchová, “Prediction by financial and economic analysis in the conditions of forest enterprises,” *Journal of Forest Science*, 66, 2020 (1), pp. 1–8. Available: <https://www.agriculturejournals.cz/pdfs/jfs/2020/01/01.pdf>
14. *Chaikin A/D Oscillator*. Available: <https://www.cryptohopper.com/resources/technical-indicators/283-chaikin-a-d-oscillator>.
15. *Normalized Average True Range*. Available: <https://taapi.io/indicators/normalized-average-true-range/>
16. *Average True Range*. Available: <https://www.fidelity.com/learning-center/trading-investing/technical-analysis/technical-indicator-guide/atr>.

17. *Average True Range*. Available: <https://www.wallstreetmojo.com/average-true-range/>
18. William Wai Him Tsang and Terence Tai Leung Chong, "Profitability of the On-Balance Volume Indicator," *Economics Bulletin*, 29(3), pp. 2424–2431, January 2009.
19. *Average Rate Of Return: Meaning, How To Calculate And Uses*. Available: <https://in.indeed.com/career-advice/career-development/average-rate-of-retur>
20. Kramar Vadim and Alchakov Vasiliy, "Time-Series Forecasting of Seasonal Data Using Machine Learning Methods," *Algorithms*, vol. 16, issue 5, pp. 2–16, 2023. doi: 10.3390/a16050248.
21. Andrius Buteikis, *Time series with trend and seasonality components*. 66 p. Available: [http://web.vu.lt/mif/a.buteikis/wp-content/uploads/2019/02/Lecture\\_03.pdf](http://web.vu.lt/mif/a.buteikis/wp-content/uploads/2019/02/Lecture_03.pdf)
22. Harding Ben and Clayton V. Deutsch, *Trend Modeling and Modeling with a Trend*. May 12, 2021, 14 p. Available: <https://geostatisticslessons.com/lessons/trendmodeling>
23. R. Shumway and D. Stoffer, *Time series analyses and its applications*; 3rd ed. 2011, 576 p. Available: <http://pzs.dstu.dp.ua/DataMining/times/bibl/TimeSeries.pdf>
24. Hariaji Joko, *Simple Linear Regression (SLR) Model and Multiple Linear Regression (MLR) Model*, May 2021, pp. 1–34. doi:10.13140/RG.2.2.17237.35044.
25. Qiang Sun and Wenxin Zhou, "Adaptive Huber Regression," *Journal of the American Statistical Association*. October 2018, 56 p. doi: 10.1080/01621459.2018.154312. [https://www.researchgate.net/publication/317732614\\_Adaptive\\_Huber\\_Regression](https://www.researchgate.net/publication/317732614_Adaptive_Huber_Regression)
26. Dastres Roza and Soori Mohsen, "Artificial Neural Network Systems," *International Journal of Imaging and Robotics*, 21(2), pp. 13–25, March 2021. Available: [https://www.researchgate.net/publication/350486076\\_Artificial\\_Neural\\_Network\\_Systems](https://www.researchgate.net/publication/350486076_Artificial_Neural_Network_Systems)

Надійшла 30.06.2023

#### INFORMATION ON THE ARTICLE

**Dmytro I. Martjanov**, ORCID: 0009-0003-3919-4412, Lviv Polytechnic National University, Ukraine, e-mail: [d.martjnoff@gmail.com](mailto:d.martjnoff@gmail.com)

**Yaroslav I. Vyklyuk**, ORCID: 0000-0003-4766-4659, Lviv Polytechnic National University, Ukraine, e-mail: [vyklyuk@ukr.net](mailto:vyklyuk@ukr.net)

**Mariya I. Fleychuk**, ORCID: 0000-0002-0417-9066, Stepan Gzhytskyi National University of Veterinary Medicine and Biotechnologies, Ukraine, e-mail: [fleychukm@gmail.com](mailto:fleychukm@gmail.com)

#### MODELING CRYPTOCURRENCY MARKET DYNAMICS USING MACHINE LEARNING TOOLS / D.I. Martjanov, Ya.I. Vyklyuk, M.I. Fleychuk

**Abstract.** The article analyzes the dynamics of the cryptocurrency market (Bitcoin) using econometric estimation tools based on machine learning models. The forecasting method is improved based on time series decomposition and lagged shifts of financial indicators. An ensemble of short-term forecast models for the Bitcoin exchange rate is built, and its accuracy is analyzed and compared to individual component models. Time series models are used along with calculated financial indicators (ADODS, NATR, TRANGE, ATR, OBV, RSI, ADTV). The absolute deviation of the short-term forecast amounted to \$9.5, which is 0.06% of the absolute value.

**Keywords:** ensemble models, machine learning, time series, cryptocurrency.