Харківський національний університет імені В. Н. Каразіна
Факультет математики і інформатики
Кафедра прикладної математики

Кваліфікаційна робота

бакалавра

на тему

«Застосування часових рядів для аналізу криптовалютного ринку»

Виконав: студент групи МП41 IV курсу
(перший бакалаврський рівень)
спеціальності 113
«Прикладна математика»
освітньої програми
«Прикладна математика»
Куцин А. С.
Керівник: кандидат фіз.-мат. наук ,
доцент кафедри
прикладної математики
Сморцова Т. І.
Рецензент: кандидат пед. наук ,
доцент кафедри
вищої математики та
інформатики
Жовтоніжко І. М.

Харків – 2024 рік

Анотація

Куцин А.С. Застосування часових рядів для аналізу криптовалютного ринку. Кваліфікаційна (бакалаврська) робота на здобуття вищої освіти спеціальності 113 «Прикладна математика» Освітньо-професійної програми «Прикладна математика» – Харківський національний університет імені В.Н Каразіна. Харків. 2024.

Розглянуто основні означення щодо теорії часових рядів та використання моделей ARMA. Наведено моделі розвинення часових рядів, методи зведення часових рядів до стаціонарних, а саме за допомогою дискретного диференціювання та поліноміального згладження та методи перевірки часових рядів на стаціонарність, у тому числі за допомогою візуальної оцінки та за допомогою статистичних тестів. Також застосовано наведені методи та підходи до реальних даних ціни криптовалюти Bitcoin за березень 2024-го року.

Проведена оцінка параметрів та підбір порядків моделі ARMA на вказаних даних. Зроблений підрахунок похибок прогнозування для різних порядків моделі, а також аналіз отриманих прогнозів за допомогою цих моделей. У якості аналізу також було проведено додаткове тестування зведених рядів для виявлення причин отриманих прогнозів.

Ключові слова: Криптовалюти, часові ряди, ARMA-моделі, поліноміальне згладження, дискретне диференціювання, тестування на незалежність, тестування на стаціонарність.

Abstract

Kutsyn A.S. The Use of Time Series for the Analysis of the Cryptocurrency Market. Bachelor's Thesis of higher education in the specialty 113 «Applied Mathematics» of the educational and professional program «Applied Mathematics» - V.N. Karazin Kharkiv National University. Kharkiv. 2024.

The main definitions related to time series theory and the use of ARMA models are considered. Models for the development of time series, methods for reducing time series to stationary, namely by means of discrete differentiation and polynomial smoothing, and methods for checking time series for stationarity, including visual assessment and statistical tests, are presented. The methods and approaches discussed are also applied to real data of Bitcoin prices for March 2024.

An evaluation of parameters and selection of ARMA model orders on the specified data has been conducted. The calculation of forecasting errors for different model orders, as well as the analysis of the obtained forecasts using these models, has been carried out. As part of the analysis, additional testing of the reduced series was also performed to identify the reasons for the obtained forecasts.

Keywords: Cryptocurrencies, time series, ARMA models, polynomial smoothing, discrete differentiation, independence testing, stationarity testing.

Зміст

Вступ

В останні роки криптовалюти здобули значне зацікавлення як у світі фінансів, так і в академічних колах. Зрослий інтерес до цього нового класу активів викликаний їх потенційною роллю як альтернативної форми інвестицій, а також можливістю використання технології блокчейн для різних цілей, включаючи фінансові транзакції, управління ланцюжками поставок та забезпечення безпеки даних.

Однак, незважаючи на широкий спектр досліджень, присвячених криптовалютам, є небагато робіт, присвячених аналізу та прогнозуванню їх часових рядів. Часові ряди криптовалют – це послідовність спостережень цін і обсягів торгів, які змінюються з часом. Аналіз цих часових рядів має важливе значення для розуміння динаміки цін на криптовалюти, виявлення внутрішніх закономірностей та розробки стратегій інвестування.

Мета цієї дипломної роботи полягає в проведенні аналізу часових рядів криптовалют з використанням методів та інструментів аналізу даних. Зокрема, будуть розглянуті методи візуалізації, дескриптивного аналізу, моделювання та прогнозування часових рядів, застосовані до даних криптовалютних ринків.

Для прогнозування та моделювання буде взятий клас моделей авторегресії та ковзного середнього (ARMA). Основною ідеєю роботи є порівняння результату використання цієї моделі у комбінації з різним методами виділення трендової складової ряду.

У кінцевому підсумку, результати цієї роботи можуть бути корисні для інвесторів, трейдерів, аналітиків ринку та всіх зацікавлених осіб, які прагнуть краще зрозуміти та успішно оперувати на ринку криптовалют.

1. Вибір даних дослідження

Для аналізу та прогнозування візьмемо записи щодо ціни криптовалюти Bitcoin за березень 2024 року у доларах США з архіву цін відомого онлайн-сервісу обміну цифрових валют Binance [3].

Ці дані доступні у відомій криптовалютній системі Binance і є записами за відповідними часовими інтервалами. Для кожного часового інтервалу $[t_i, t_{i+1}]$ є набір показників, які характеризують криптовалюту на ньому. До таких показників, зокрема, належать:

- Ціна відкриття (ціна криптовалюти в момент часу t_i).
- Ціна закриття (ціна криптовалюти в момент часу t_{i+1}). Ця ціна є ціною відкриття наступного часового інтервалу.
- Найвища ціна, яка була на проміжку.
- Найнижча ціна, яка була на проміжку.

Можна розглядати низку інших показників, але ми зосередимося на ціні закриття, бо цей показник є найцікавішим для прогнозування, а саме, прогнозування ціни наприкінці теперішнього часового проміжку.

Для аналізу взяті часові проміжки довжиною у 15 хвилин.

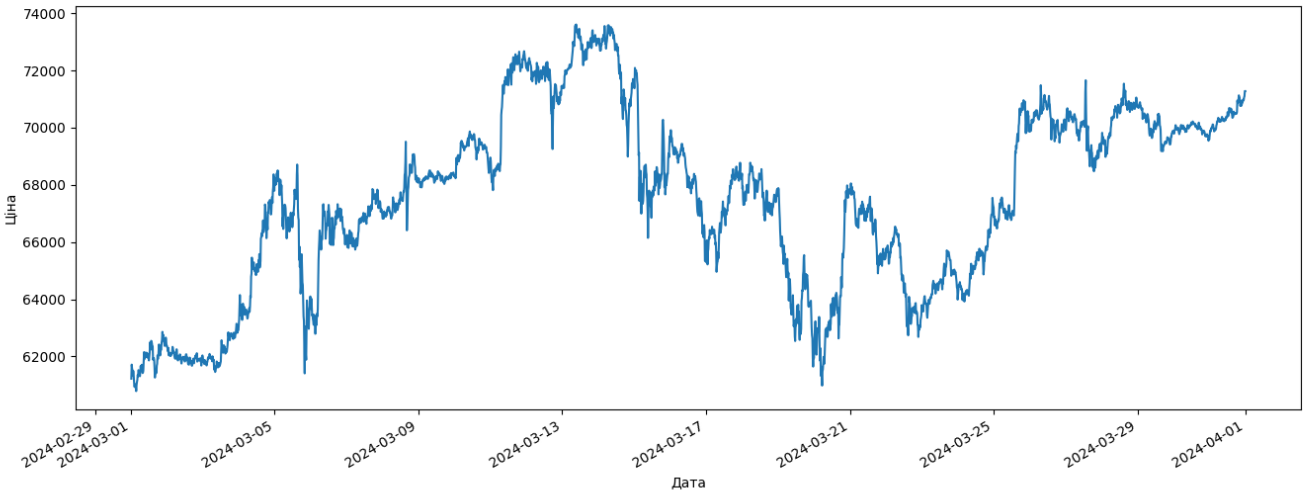

Рис. 1. Графік ціни Bitcoin за березень 2024-го року.

2. Основні означення та перетворення часового ряду

У цьому розділі наведемо основні означення та факти, які будуть використані для дослідження.

2.1. Стаціонарні часові ряди

Одним з найважливіших класів часових рядів є клас стаціонарних часових рядів. Для того, щоб надати означення стаціонарності, наведемо декілька початкових термінів.

Означення 2.1. (ст. 8 з [2]) Стохастичним процесом називають множину випадкових величин $\{X_t, \quad t \in T\}$ визначених на деякому ймовірнісному просторі $\{\Omega, \mathcal{F}, \mathrm{P}\}$.

У свою чергу стохастичний процес ще називають часовим рядом. Також часовим рядом можуть називати не сам процес, а його конкретну реалізацію. Якщо множина індексів T часового ряду є дискретною, то часовий ряд, відповідно, теж називається дискретним. У випадку, що розглядається, будемо працювати саме з дискретним часовим рядом.

Із визначенням стохастичного процесу та часового ряду тісно пов’язане поняття автоковаріаційної функції.

Означення 2.2. (ст. 11 з [2]) Якщо $\{X_t,\ t \in T\}$ – стохастичний процес зі скінченною дисперсією $Var(X_t) < \infty$ для кожного $t \in T$, то автоковаріаційною функцією (або АКВФ) $\gamma_X(r, s)$ називають

$$\gamma_X(r, s) = Cov(X_r, X_s) = E\left[(X_r - EX_r)(X_s - EX_s)\right].$$

Ця функція є продовженням поняття коваріаційної матриці на випадок нескінченого вектору випадкових величин. Вона показує коваріацію між випадковими величинами часового ряду у різні моменти часу.

Дуже важливим класом часових рядів, для яких побудована змістовна математична теорія, є клас стаціонарних часових рядів.

Означення 2.3. (ст. 12 з [2]) Часовий ряд $\{X_t,\ t \in Z\}$ з множиною індексів $Z = \{0,\ \pm 1,\ \pm 2,\ \pm 3, \ldots\}$ називають стаціонарним, якщо

1) $E|X_t|^2 < \infty$ для кожного $t \in Z$,

2) $EX_t = m$ для кожного $t \in Z$,

3) $\gamma_X(r, s) = \gamma_X(r + t, s + t)$ для будь-яких $r, s, t \in Z$.

Іншими словами, стаціонарний ряд є таким рядом, у якому математичне сподівання не залежить від часу та коваріація між випадковими величинами залежить тільки від відстані між ними у часі, а не від того факту, в який саме момент часу ми їх узяли. За такої логіки АКВФ можна розглядати як функцію від однієї змінної, а не від двох.

Зауваження: Якщо $\{X_t,\ t \in Z\}$ є стаціонарним часовим рядом, то $\gamma_X(r,\ s) = \gamma_X(r - s,\ 0)$, тому досить зручно перевизначити АКВФ від стаціонарного часового ряду так

$$\gamma_X(h) = \gamma_X(h, 0) = Cov(X_{t+h}, X_t) \text{ для всіх } t, h \in Z.$$

За тим же принципом визначається поняття автокореляційної функції або АКФ, яка показує кореляцію між величинами ряду у різні моменти часу.

Означення 2.4. (ст. 12 з [2]) Автокореляційною функцією стаціонарного ряду $\{X_t,\ t \in Z\}$ називається функція

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr\left(X_{t+h},\ X_t\right).$$

АКФ може бути дуже корисним індикатором нестаціонарності ряду. Наприклад для ряду, який має тренд, АКФ показує плавний спад, а для стаціонарного ряду АКФ проявляє різкий спад відразу до невеликих значень кореляції.

2.2. Класична модель розвинення ряду

Зрозуміло, що далеко не кожен ряд з реального життя буде стаціонарним. Зокрема, далеко не кожен ряд буде мати стале математичне сподівання. Задля вирішення цієї проблеми з часом виникла ідея розвинення ряду на декілька компонент, кожна з яких відповідає за окремий аспект поведінки ряду. Найчастіше використовують модель розвинення, яку називають «класичною» (ст. 15 з [2]), а саме,

$$X_t = m_t + s_t + Y_t.$$

У свою чергу, компоненту m_t називають трендом, компоненту s_t називають сезоном або сезонною компонентою, а Y_t називають залишком розвинення. Неформально кажучи, тренд показує рух ряду «у середньому» та є невипадковою функцією. Сезонна компонента є також невипадковою функцією, але періодичною з деяким періодом d та показує вплив сезонних факторів на ряд (наприклад, ціна на авіабілети буде різною в залежності від пори року). А ось залишок Y_t є деяким випадковим залишком, який отримується у результаті видалення m_t та s_t з ряду.

Ідея такого розвинення полягає в тому, щоб спробувати видалити невипадкові компоненти m_t та s_t з нашого ряду, сподіваючись, що залишок Y_t виявиться стаціонарним, щоб потім можна було використовувати теорію стаціонарних часових рядів.

У свою чергу ми будемо використовувати ще простішу модель розвинення, яка буде складатися тільки з трендової частини та залишку. Для обґрунтування такого підходу, поділимо початковий часовий ряд на невеликі часові проміжки, припускаючи, що на цих проміжках відсутня періодичність/сезонність.

Розбиття було зроблено рівномірно, на 32 рівні за розмірами інтервали, кожен по 93 елемента. Саме таке розбиття було обрано суто через рівномірність.

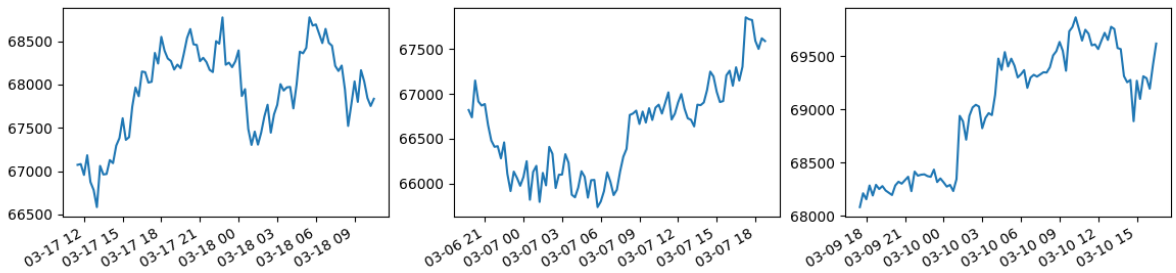

Рис. 2. Графік цін на Bitcoin на трьох з 32-х проміжках.

2.3. Методи видалення трендової частини ряду

Після введення моделі розвинення ряду на тренд/сезон та залишок, який ми сподіваємось отримати стаціонарним, наступним питанням є прибирання невипадкових компонент, щоб отримати наш «стаціонарний» залишок. У нашому випадку прибирати треба тільки тренд. Існують різні методи для цього. Одні намагаються спочатку побудувати оцінку на тренд, а потім

віднімати цю оцінку від початкового ряду, а інші намагаються прибирати його без попереднього оцінювання. Ми розглянемо два різні методи, для розв’язання цієї задачі.

2.3.1. *Метод 1. Оцінка тренду за МНК*

Згідно з цим методом, m_t намагаються шукати у вигляді заздалегідь визначеного сімейства функцій, оцінюючи при цьому параметри функції за допомогою методу найменших квадратів. На практиці дуже часто використовують оцінки за допомогою поліномів невеликої степені, оскільки поліноми є досить гарним інструментом наближення функцій. Цей метод ще називають методом згладжування початкового ряду.

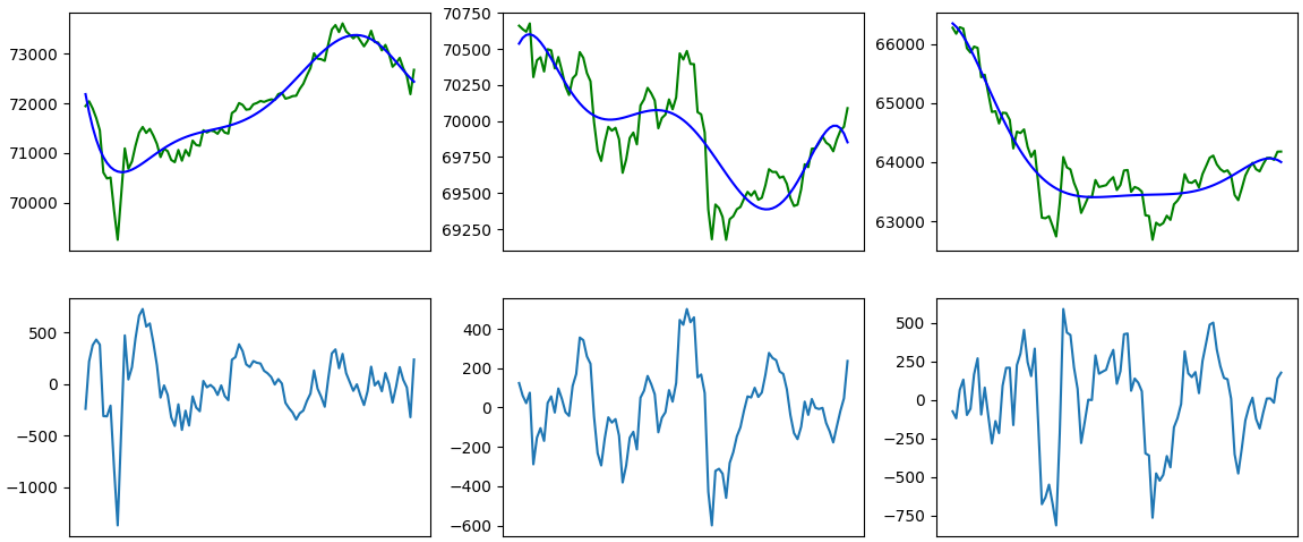

Рис. 3. Приклад згладження ряду поліномом 6-го ступеню.

2.3.2. *Метод 2. Видалення тренду за допомогою дискретного диференціювання*

Цей метод не робить ніякої оцінки на тренд, а намагається напряму видалити тренд. Для цього визначимо оператор ∇ як

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t,$$

де B – оператор зсуву назад

$$BX_t = X_{t-1}.$$

Степені операторів B та ∇ визначаються так:

$$B^j\left(X_t\right) = X_{t-j} \text{ а } \nabla^j\left(X_t\right) = \nabla\left(\nabla^{j-1}(X_t)\right) \text{ з } \nabla^0\left(X_t\right) = X_t.$$

Досить гарною властивістю цього методу є той факт, що якщо заздалегідь відомо, що тренд ряду має вигляд поліному степені p, то оператор диференціювання степені $p+1$ гарантовано прибирає цей тренд. Більш детально, для моделі $X_t = m_t + Y_t$, де $m_t = \sum_{j=0}^{k} a_j t^j$ та Y_t є стаціонарним рядом з нульовим середнім,

$$\nabla^k X_t = k! a_k + \nabla^k Y_t.$$

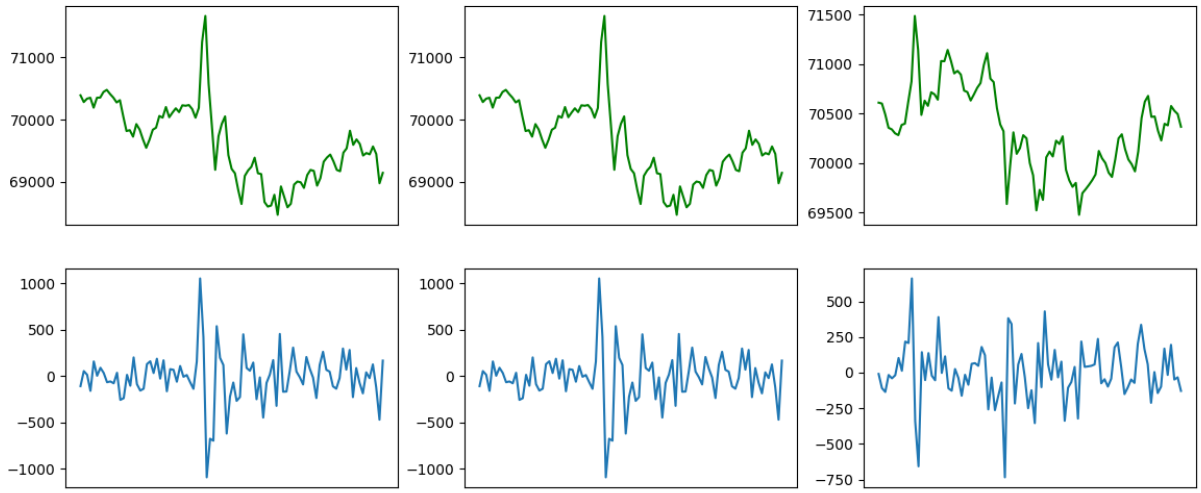

Рис. 4. Приклад диференціювання першого порядку ряду.

2.4. Методи перевірки стаціонарності часового ряду

Після проведення етапу прибирання тренду потрібен деякий механізм, щоб впевнитись у тому, що отриманий залишок Y_t дійсно є стаціонарним. Для початку можна це зробити й візуально (наприклад, впевнитись у тому, що ряд не має очевидних ознак тренду).

Також можна подивитися на АКФ від перетвореного ряду. Для АКФ від стаціонарного ряду характерною є поведінка, за якої АКФ різко зменшується до невеликих значень, на відміну від для нестаціонарніх рядів, для яких АКФ показує плавний рух.

Однак, існують ще статистичні тести для перевірки наявності стаціонарності, одним з яких є розширений тест Дікі-Фуллера. Його ще називають «тестом на одиничні корені». Щоб пояснити як він працює, дамо таке означення.

Означення 2.5. (ст. 74 з [1]) Часовий ряд $\{X_t,\ t \in Z\}$ називаються $AR(p)$ -процесом, якщо він має вигляд

$$X_t = \mu + \sum_{i=1}^{p} a_i X_{t-i} + \epsilon_t.$$

У свою чергу, якщо використовувати оператор зсуву B^j, то останню рівність можна записати так:

$$X_t = \mu + \left(\sum_{i=1}^{p} a_i B^i\right) X_t + \epsilon_t \quad \rightarrow \quad \left(1 - \sum_{i=1}^{p} a_i B^i\right) X_t = \epsilon_t + \mu.$$

Вираз $\left(1 - \sum_{i=1}^{p} a_i z^i\right)$ називають характеристичним поліномом цього процесу. Виявляється, що коли характеристичний поліном має одиничні корені, то процес $AR(p)$ не є стаціонарним. Тест Дікі-Фуллера припускає, що ряд підпорядковується такій моделі:

$$\nabla X_t = \mu + \alpha X_{t-1} + \sum_{i=1}^{p-1} a_i \nabla X_{t-i}.$$

Нульовою гіпотезою є те, що у цій моделі коефіцієнт α дорівнює нулю. У такому випадку перші різниці ряду є $AR(p-1)$ процесом, що еквівалентно умові одиничного кореню в рівнянні.

Альтернативною гіпотезою є те, що коефіцієнт α менше нуля, тобто що початковий ряд не має одиничного кореню та є стаціонарним. Метою дослідження є прибирання тренду таким чином, щоб відхилити нульову гіпотезу.

2.5. Перевірка стаціонарності часового ряду

При зведенні наших рядів до стаціонарних методом диференціювання виявилося, що вже навіть однократне диференціювання призводить до ситуації, коли ряди виглядають та поводять себе як стаціонарні.

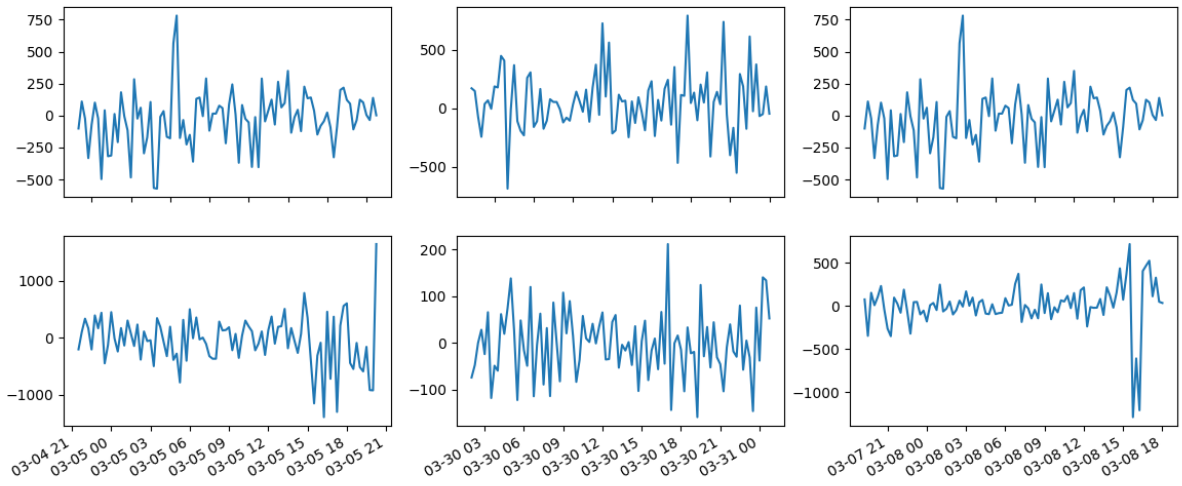

Рис. 5. Приклад диференційованих рядів на 6-ти різних проміжках з 32-х

Але за візуальною складовою можна тільки чітко побачити наявність сталого математичного очікування, що ще не гарантує стаціонарність. Подивимось на АКФ від деяких рядів.

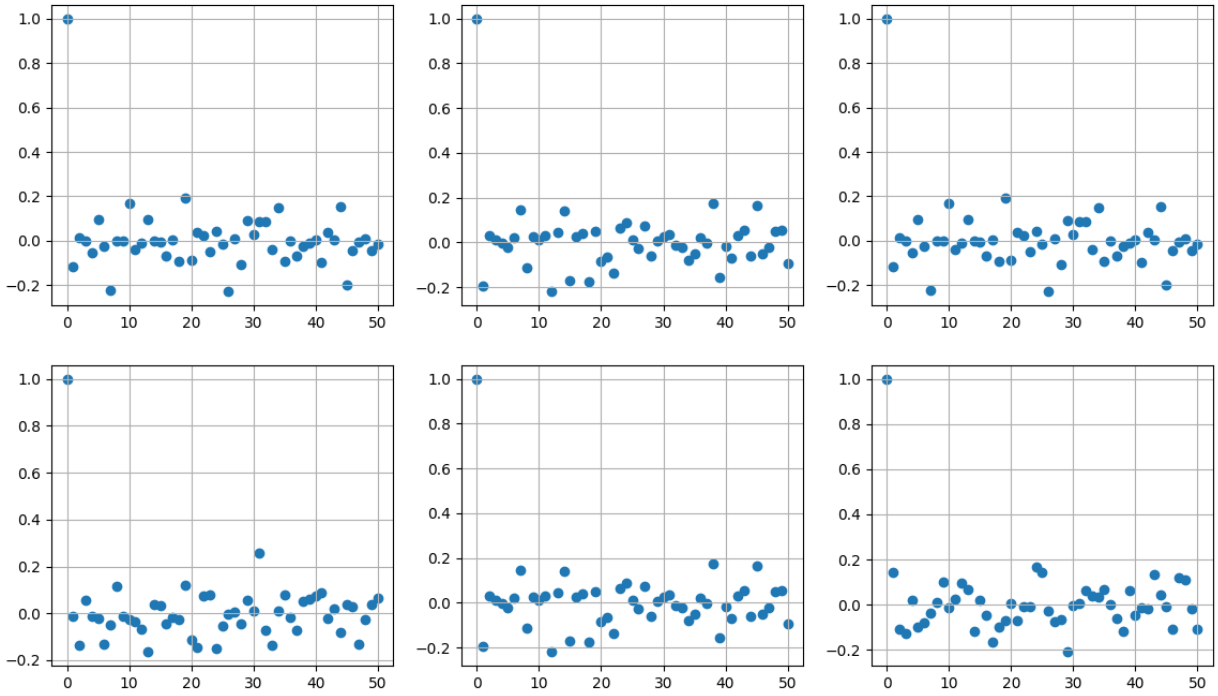

Рис. 6. Приклад 6-ти випадкових АКФ від диференційованих рядів

Бачимо, що АКФ поводить себе так, як повинна поводити себе за наявності стаціонарності (різкий спад до невеликих значень).

У свою чергу, тест Дікі-Фуллера відхилив нульову гіпотезу про нестаціонарність у всіх 32 випадках. При тестуванні використовувався рівень значущості $\alpha = 0.05$, але він відхиляє нестаціонарність навіть за меншими рівнями значущості, тому що дійсно досягнені рівні значущості (p-значення) при тестуванні вийшли дуже малими. Наприклад максимальне за усіма 32-ма рядами p-значення вийшло менше ніж 0.03.

Як висновок, ми приймаємо гіпотезу стаціонарності при однократному диференціюванні.

Розглянемо тепер поліноміальне згладжування. Якщо у випадку дискретного диференціювання параметром, який можна змінювати, був порядок диференціювання, то у цьому випадку треба розумно підібрати степінь поліному, за допомогою якого буде проводитись згладжування. Для

цього можна використати наступний підхід: перебрати степені у деякому діапазоні та для кожної степені оцінити поліном методом МНК, після чого відняти цей поліном від початкового ряду та застосувати до того, що залишиться, тест Дікі-Фуллера.

В наступній таблиці показано, на якій кількості рядів тест Дікі-Фуллера відхилив нульову гіпотезу про нестаціонарність та максимальне з *p*-значень за всіма рядами для степенів поліному від 1 до 10.

Степінь поліному	Рядів, на яких відхилилась гіпотеза про нестаціонарність	Максимальне з *p*-значень
1	5 з 32	0.74076
2	12 з 32	0.59974
3	21 з 32	0.36622
4	28 з 32	0.17312
5	31 з 32	0.05708
6	32 з 32	0.04362
7	32 з 32	0.01347
8	32 з 32	0.00179
9	32 з 32	0.00878
10	32 з 32	0.00353

Табл. 1. Таблиця кількості рядів на яких нульова гіпотеза тесту Дікі-Фуллера була відхилена, а також максимальне з p-значень тесту по всім рядам.

Бачимо, що починаючи з 6-ї степені при рівні значущості $\alpha = 0.05$ гіпотеза нестаціонарності відхиляється на усіх рядах.

Подивимось на результуючі залишки після видалення поліномів 6-ї степені.

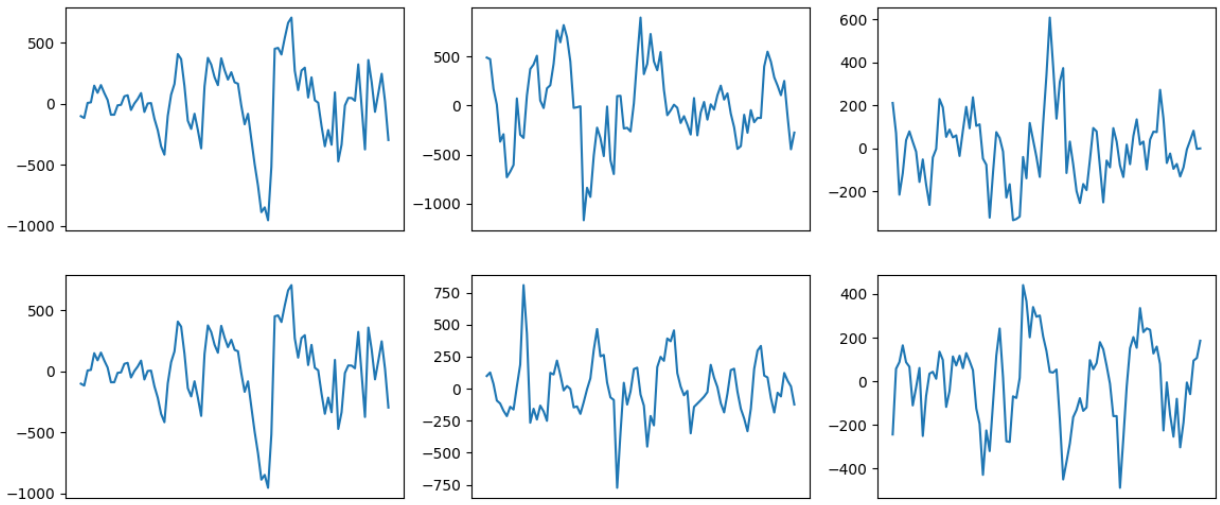

Рис. 7. Приклад 6-ти рядів після видалення трендової частини через поліноміальне згладження поліномами 6-го ступеню

Розглянемо також їхні АКФ.

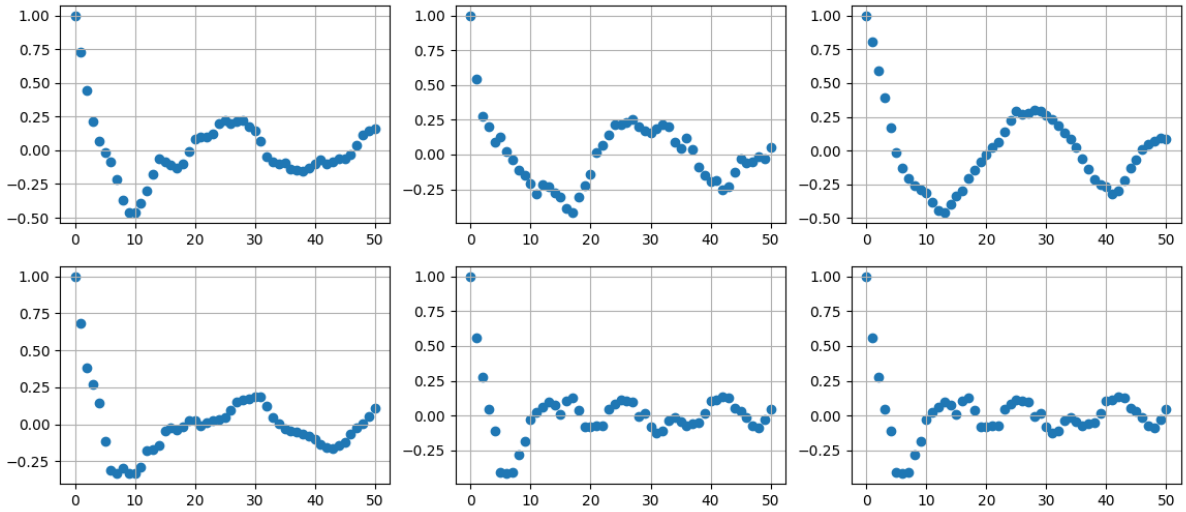

Рис. 8. Приклад 6-ти АКФ від рядів після поліноміального згладження Поліномами 6-го ступеню

Бачимо суттєву різницю між поведінкою АКФ в диференційованих рядах та в рядах після видалення поліноміальної оцінки тренду. Це може свідчити про те, що під час побудови моделі ми будемо мати різні результати для висновків.

3. ARMA-моделі

Одним із потужних класів для моделювання стаціонарних часових рядів є клас ARMA моделей. ARMA-модель визначається так.

Означення 3.1. (ст. 74 з [2]) Процес $\{Z_t\}$ з нульовим середнім та дисперсією σ^2, величини якого некорельовані між собою, називають білим шумом.

Означення 3.2. (ст. 74 з [2]) Процес $\{X_t,\ t \in Z\}$ називається $ARMA(p, q)$ процесом, якщо $\{X_t\}$ стаціонарний та якщо для кожного t

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q},$$

де Z_t – білий шум. $\{X_t\}$ називають $ARMA(p,\ q)$ процесом з середнім μ якщо $\{X_t - \mu\}$ є $ARMA(p, q)$ процесом.

Як і у випадку з $AR(p)$ процесом $ARMA(p,\ q)$ процес може бути записаний більш компактно у вигляді

$$\phi(B)X_t = \theta(B)Z_t \qquad t = 0,\ \pm 1,\ \pm 2, \ldots,$$

де ϕ та θ є поліномами степені p та q відповідно.

Логіка, яка стоїть за цією моделлю є досить природньою. Припускається, що теперішнє значення ряду X_t залежить, по перше, від p попередніх значень ряду, при чому лінійно, а, по друге, залежить від q попередніх значень похибок Z_{t-i} та від похибки у теперішній момент часу Z_t.

3.1. Оцінка параметрів моделі

Звісно одним з найважливіших кроків моделювання є оцінка параметрів обраної моделі. Для оцінки параметрів моделі ARMA існує декілька способів. У цій роботі будемо оцінювати параметри моделі за допомогою метода максимальної вірогідності. Щоб зрозуміти як він працює нам потрібно буде узяти декілька фактів з [2].

Зосередимось поки що на оцінці коефіцієнтів $\phi_1,\phi_2,\ldots,\phi_p$ та $\theta_1, \theta_2,\ldots,\theta_q$, σ, вважаючи, що порядки p та q моделі, що розглядається, задані.

Відомо, що для незміщених випадкових величин їх коваріація дорівнює

$$Cov\left(X_i, X_j\right) = E(X_i X_j)$$

Нехай $\hat{X}_{n+1}$ – оцінка для X_{n+1} яка дає найменшу середньоквадратичну похибку.

Твердження 3.1. (Алгоритм інновацій, ст. 172 з [2]). *Нехай процес* $\{X_t\}$ *має нульове середнє та АКВФ* $\kappa(i, j) = E(X_i X_j)$, *причому матриця* $\left[\kappa(i, j)\right]_{i,j=1}^{n}$ *невироджена для кожного* $n = 1, 2, 3\ldots$

Тоді оцінки на один крок вперед $\hat{X}_{n+1}$, $n \geq 0$ *та їх середньоквадратичні похибки* v_n, $n \geq 1$ *можуть бути знайдені так:*

$$\hat{X}_{n+1} = \begin{cases} 0, & якщо\ n = 0 \\ \sum_{j=1}^{n} \theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & якщо\ n \geq 1, \end{cases}$$

Середньоквадратичні похибки v_n *знаходяться з рівнянь*

$$\begin{cases} v_0 = \kappa(1, 1) \\ \theta_{n,n-k} = v_k^{-1}\left(\kappa\left(n+1, k+1\right) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j} v_j\right), \quad k = 0, 1, \ldots n-1 \\ v_n = \kappa\left(n+1, k+1\right) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 v_j. \end{cases}$$

Зауважимо, що цю систему можна розв’язати рекурсивно.

Тепер розглянемо модель ARMA

$$\phi(B)X_t = \theta(B)Z_t,$$

де Z_t є процесом білого шуму з нульовим середнім та дисперсією σ^2. Зробимо заміну (ст. 175-176 з [2])

$$\begin{cases} W_t = \sigma^{-1}X_t & t = 1,\dots,m, \\ W_t = \sigma^{-1}\phi(B)X_t & t > m, \end{cases}$$

де $m = max(p, q)$. Тоді отримаємо новий процес з АКВФ $\kappa(i,j) = E(W_i W_j)$, яка має вигляд

$$\kappa(i,j) = \begin{cases} \sigma^{-2}\gamma_X(i-j) & 1 \le i,j \le m \\ \sigma^{-2}\left[\gamma_X(i-j) - \sum_{r=1}^{p}\phi_r\gamma_X(r-|i-j|)\right] & min(i,j) \le m < \max(i,j) \\ \sum_{r=0}^{q}\theta_r\theta_{r+|i-j|} & \min(i,j) > m \\ 0 & інакше \end{cases}$$

де $\gamma_X(h)$ – АКВФ процесу $\{X_t\}$, а θ_j, ϕ_i – коефіцієнти ARMA моделі.

Після цього можна застосовуючи твердження 3.1 до нового процесу та після зворотної заміни отримати оцінку на один крок вперед саме для ARMA процесу:

$$\begin{cases} \hat{X}_{n+1} = \sum_{j=1}^{n}\theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & для \quad 1 \le n < m, \\ \hat{X}_{n+1} = \phi_1 X_n + \dots + \phi_p X_{n+1-p} + \sum_{j=1}^{q}\theta_{nj}\left(X_{n+1-j} - \hat{X}_{n+1-j}\right), & для\ n \ge m, \end{cases}$$

а також середньоквадратичні похибки цих оцінок:

$$E\left(X_{n+1} - \hat{X}_{n+1}\right)^2 = \sigma^2 E\left(W_{n+1} - \hat{W}_{n+1}\right)^2 = \sigma^2 r_n,$$

де θ_{nj} та r_n знаходяться з АКВФ $\kappa(i,j)$ за допомогою твердження 3.1.

Тепер перейдемо до оцінки параметрів за допомогою методу максимальної вірогідності. У моделі ARMA, що розглядається, потрібно оцінити два вектора параметрів $\phi = (\phi_1, \ldots, \phi_p)'$ та $\theta = (\theta_1, \ldots, \theta_q)'$, а також дисперсію білого шуму σ^2 величин $\{Z_i\}$.

Для моделі ARMA функція вірогідності має вигляд: (ст. 256 з [2])

$$L(\phi, \theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} (r_0 r_1 \ldots r_{n-1})^{-\frac{1}{2}} exp\left[-\frac{1}{2}\sigma^{-2} \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}} \right]$$

Диференціюючи логарифм цієї функції відносно σ^2 та враховуючи, що $\hat{X}_j$ та r_j не залежать від σ^2, отримуємо, що оцінки $\hat{\phi}, \hat{\theta}$ та $\hat{\sigma}$ задовольняють співвідношення

$$\hat{\sigma}^2 = n^{-1} S(\hat{\phi}, \hat{\theta}),$$

де

$$S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$$

та $\hat{\phi}, \hat{\theta}$ є величинами, які мінімізують величину

$$l(\phi, \theta) = ln\left(n^{-1} S(\phi, \theta)\right) + n^{-1} \sum_{j=1}^{n} ln\ (r_{j-1}).$$

3.2. Підбір порядків моделі. Показники AIC та BIC

Тепер зосередимось на підборі порядків моделі. Цей етап є дуже важливим через те, що дуже маленький порядок може призвести до нездатності моделі вловлювати закономірності у даних, а дуже великий порядок може призвести до перенавчання моделі.

Для підбору порядків використаємо так званий інформаційний критерій Акаіке (AIC) який визначається за формулою:

$$AIC = -2lnL + 2k,$$

де L – максимізоване значення функції правдоподібності, а $k-$ число параметрів моделі.

Підставивши $n^{-1}S(\hat{\phi},\theta)$ замість $\hat{}\sigma$ та врахувавши, що кількість параметрів дорівнює $(p+q+1)$, отримаємо, що для ARMA моделі цей інформаційний критерій набуде вигляду

$$AIC(p,q) = -2ln\left(L\left(\hat{\phi},\hat{}\theta,\frac{S(\hat{\phi},\hat{}\theta)}{n}\right)\right) + 2(p+q+1).$$

Існує також Баєсовський інформаційний критерій (BIC), який для ARMA моделі обчислюється так:

$$BIC = (n-p-q)ln\left[\frac{\hat{}\hat{n}\sigma^2}{n-p-q}\right] + n\left(1+ln\sqrt{2\pi}\right) + (p+q)ln\left[\frac{\sum_{t=1}^{n} X_t^2 - \hat{n}\sigma^2}{p+q}\right],$$

де додатковий параметр $n-$ кількість елементів в даних.

3.3 Підрахунок та візуалізація AIC та BIC

Розглянемо практичне застосування викладених фактів. Для підбору параметрів p та q для кожного ряду обчислимо два інформаційні критерії щодо тренованої моделі. Після цього візьмемо середній показник AIC та BIC за всіма 32-ма рядами.

На рисунках 9 зображений середній показник AIC та BIC за всіма 32-ма рядами, обчислений для ARMA моделі, застосованої до даних, в яких трендова частина була видалена за допомогою диференціювання. При моделюванні були перебрані параметри p та q в діапазоні від 1 до 10.

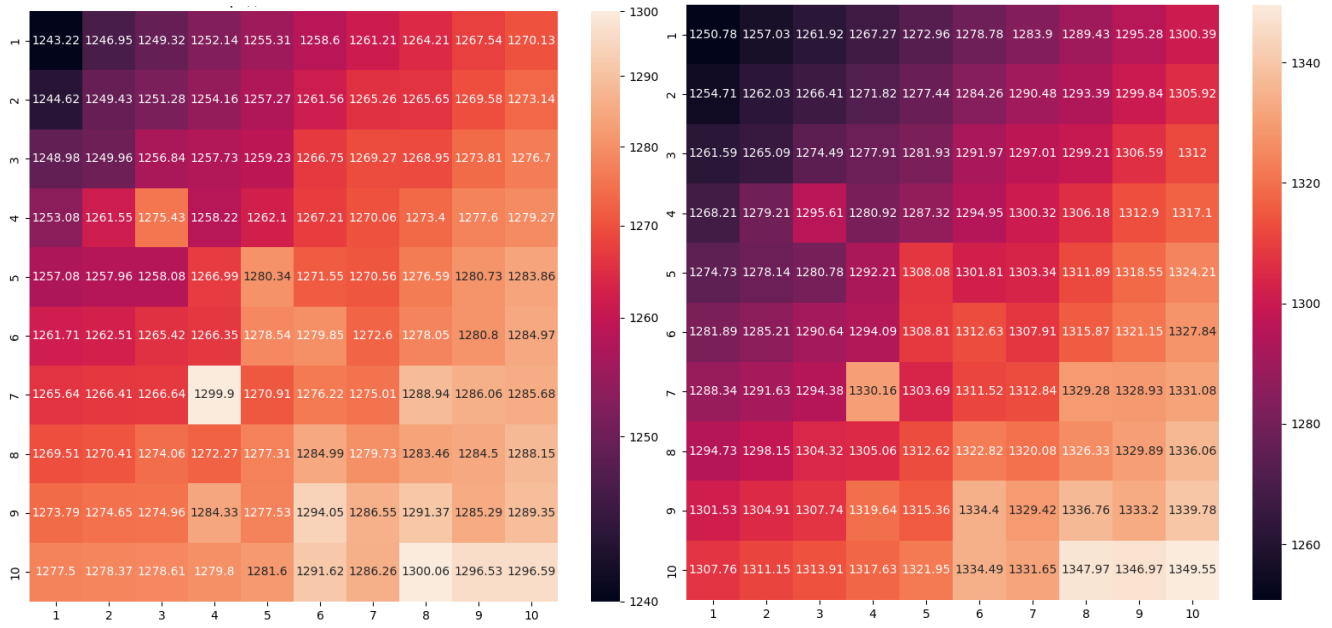

Рис. 9. Значення AIC (зліва) та BIC (справа) для різних комбінацій значень *p* (по осі Y) та *q* (по осі X) для випадку диференціювання

На рисунку 10 зображені середні показники AIC та BIC для моделі ARMA на даних, в яких трендова частина була видалена за допомогою поліноміального згладжування.

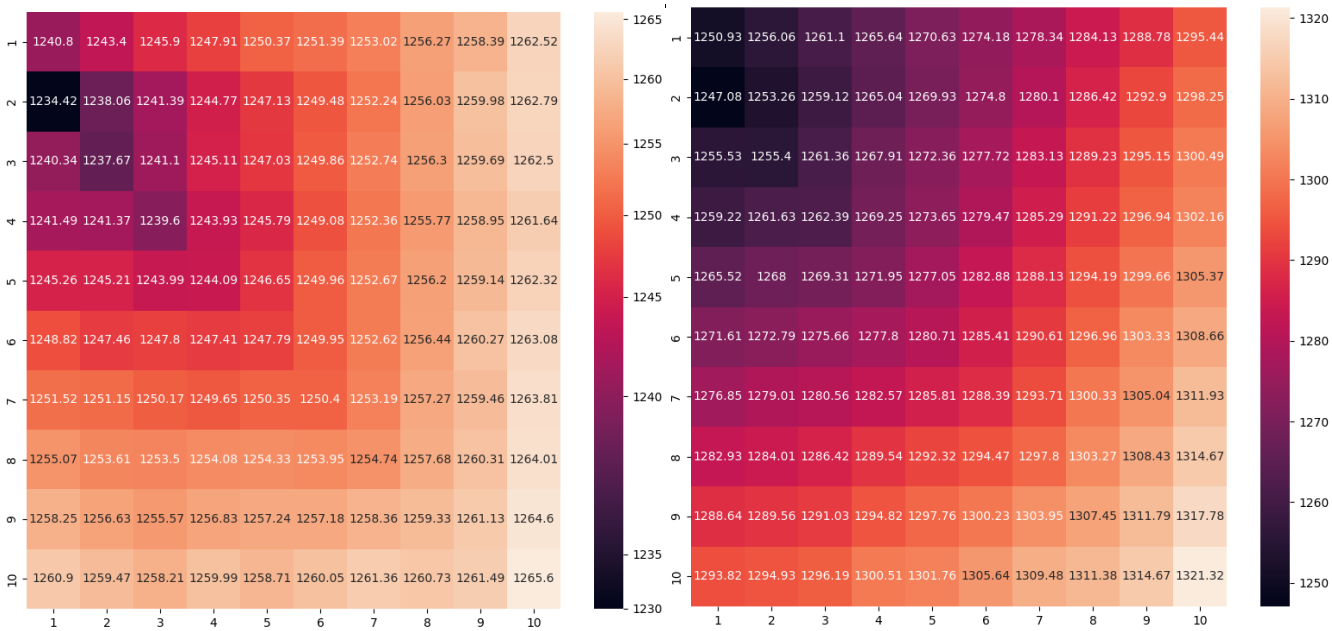

Рис. 10. Значення AIC (зліва) та BIC (справа) для різних комбінацій значень p (по осі Y) та q (по осі X) для випадку поліноміального згладження

Проаналізуємо рисунки. Як бачимо, у випадку з дискретним диференціюванням найменші середні значення AIC та BIC вийшли для параметрів $p=1$ ma $q=1$. Це може свідчити про те, що при збільшенні параметрів ми не отримуємо суттєвого виграшу з точки зору функції правдоподібності. Складова, пов'язана з функцією правдоподібності не здатна перекрити складову, пов'язану з кількістю параметрів, через яку AIC та BIC і збільшуються. Наперед кажучи, це може свідчити про те, що модель не здатна отримати виграш у прогнозуванні за рахунок збільшення параметрів, тобто що вона не навчається на даних.

Щодо випадку з поліноміальним виділенням тренду бачимо, що невеликий приріст є тільки для випадку $p=2$ та $q=1$ Це теж не дуже добре. Для того, щоб не робити висновки про якість моделі за одним критерієм, розглянемо інші її параметри.

Для кожного ряду та прогнозованих значень моделі обчислимо корінь із середньоквадратичної похибки (RMSE) та усереднимо її за всіма рядами. У випадку дискретного диференціювання отримуємо наступні значення:

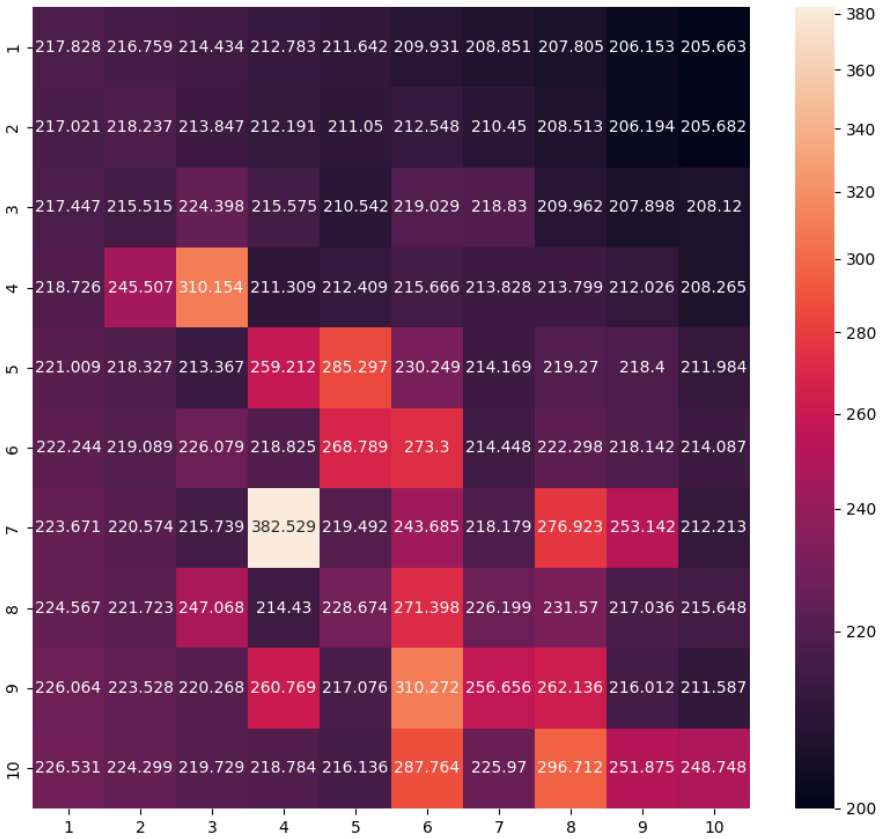

Рис. 11. Значення RMSE для різних комбінацій значень *p* (по осі Y) та *q* (по осі X) для випадку диференціювання

Серед цих значень немає значення, що сильно відрізняється від інших в кращу сторону. Значення нижче 200 одиниць отримати не вдалось.

Для поліноміального згладжування отримуємо інші результати:

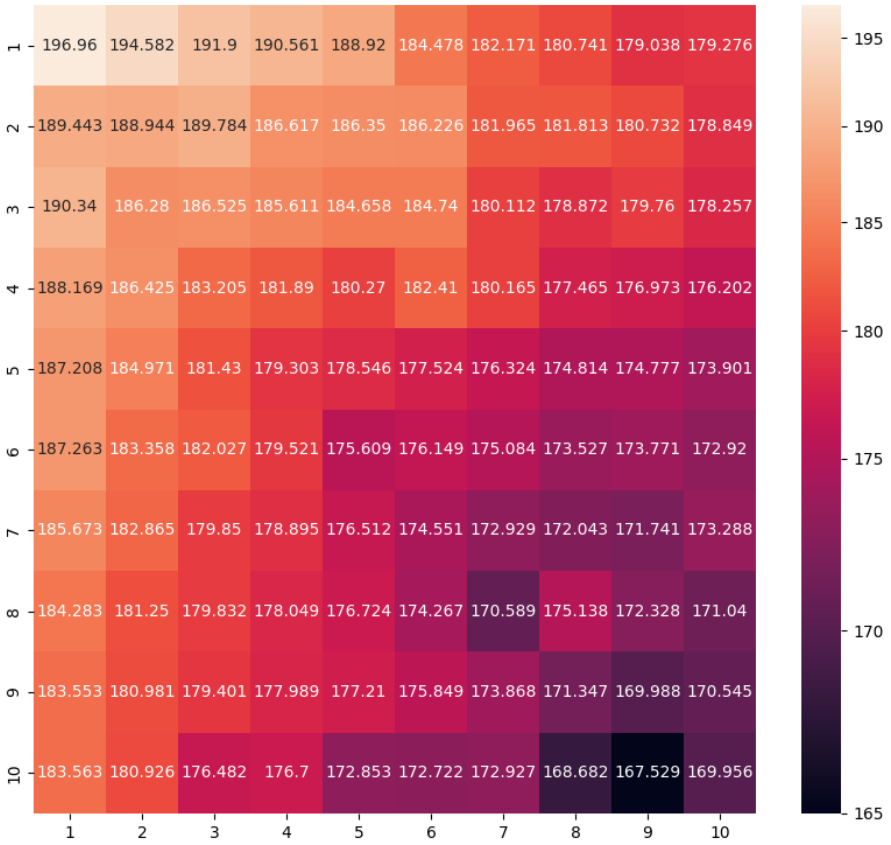

Рис. 12. Значення RMSE для різних комбінацій значень p (по осі Y) та q (по осі X) для випадку поліноміального згладження

Бачимо, що з деяких причин з-поміж усіх значень RMSE виділяється значення для параметрів $p = 9$ та $q = 10$. Також бачимо суттєву відмінність у характері змін цих значень від випадку диференціювання. Зміна значень відбувається не хаотично та зменшується зі збільшенням кількості параметрів.

Розглянемо також графіки прогнозованих значень на деяких рядах у порівнянні з істинними значеннями ряду. Спочатку розглянемо їх для випадку видалення трендової частини ряду за допомогою диференціювання.

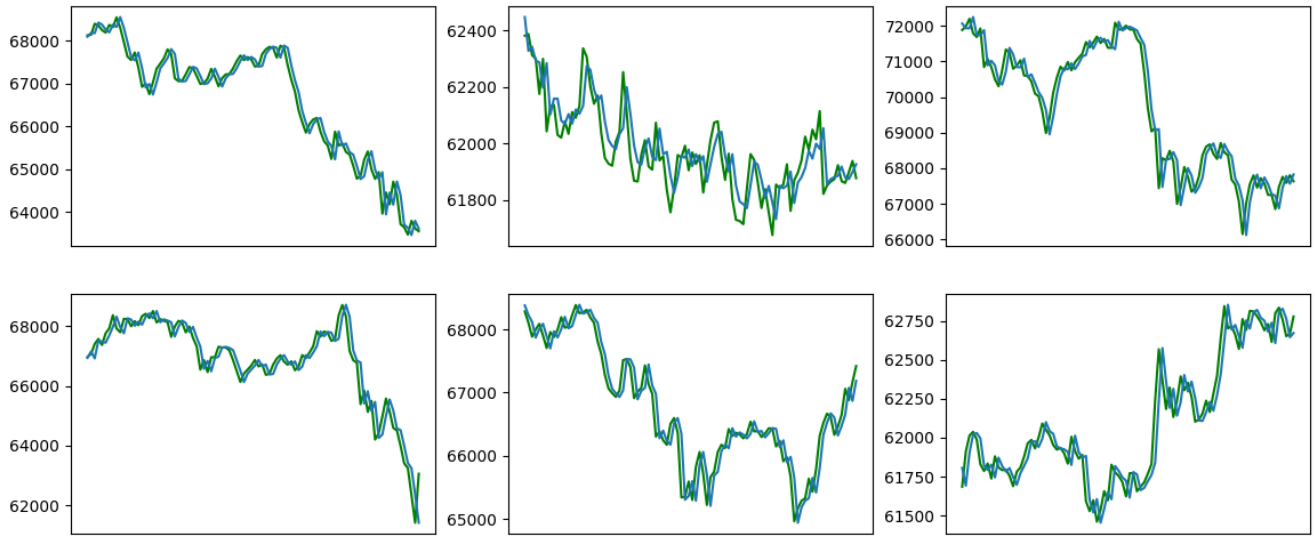

Рис. 12. 6 прикладів прогнозованих рядів (синім) у порівнянні з істинним (зеленим) у випадку диференціювання

Майже в усіх випадках, незалежно від обраних порядків моделі, отримуємо наступний результат: прогноз виявляється близьким до наївного, через що рисунок виглядає так, ніби ряд істинних значень відрізняється від ряду прогнозованих значень лише зсувом на одиницю. Це підтверджують конкретні рисунки зі збільшеним масштабом.

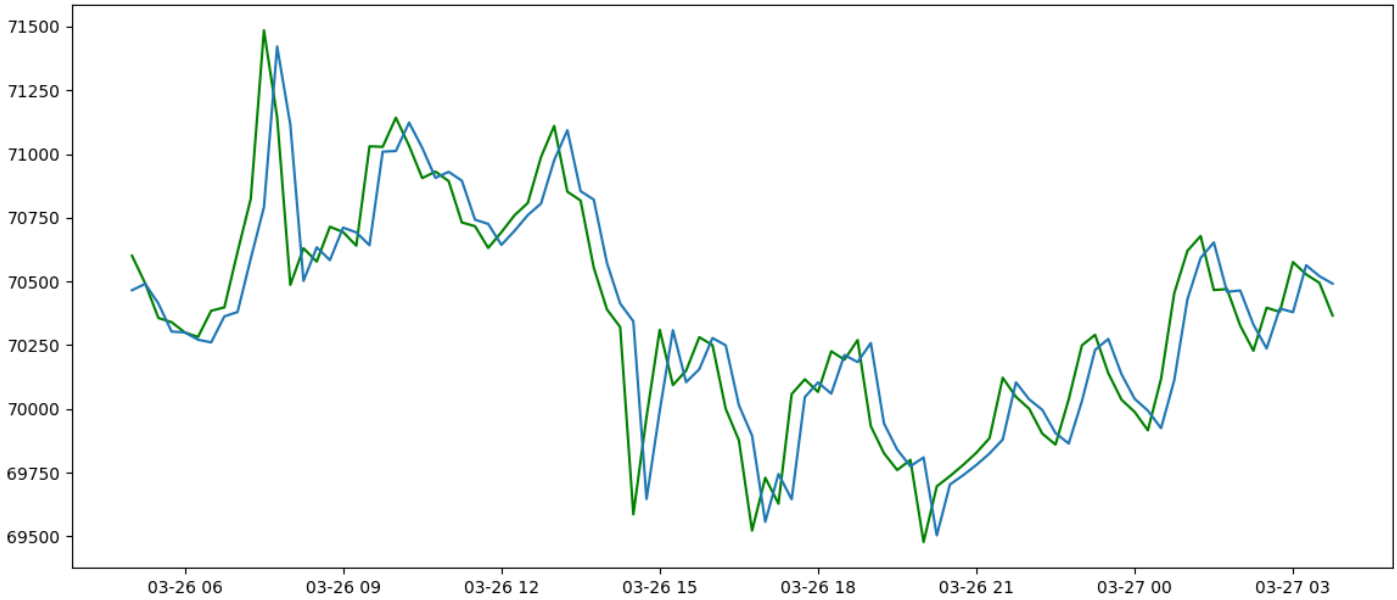

Рис. 13. Приклад прогнозованого ряду (синім) у порівнянні з істинним (зеленим) у випадку диференціювання

Чому рисунки в цьому випадку мають саме такий характер розглянемо в підрозділі 3.4.

Для випадку з видаленням трендової частини ряду за допомогою поліноміального згладжування проблема наївного прогнозування відсутня.

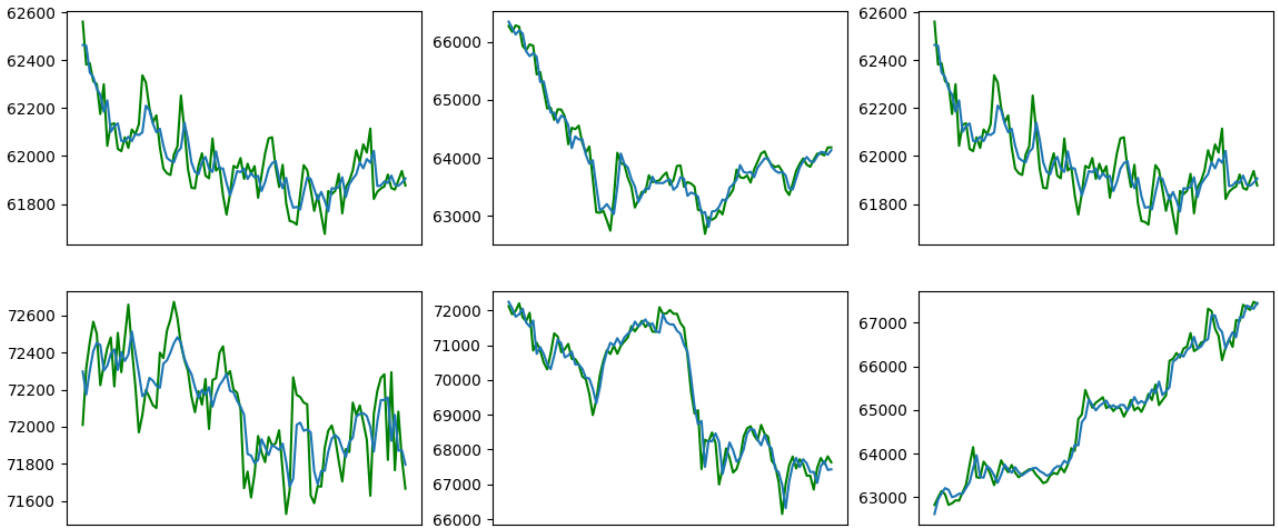

Рис. 14. 6 прикладів прогнозованих рядів (синім) у порівнянні з істинним (зеленим) у випадку поліноміального згладження

Але якщо придивитись детальніше до багатьох результатів, можна побачити особливість, яка полягає у тому, що прогнозований ряд частково копіює поведінку оригінального.

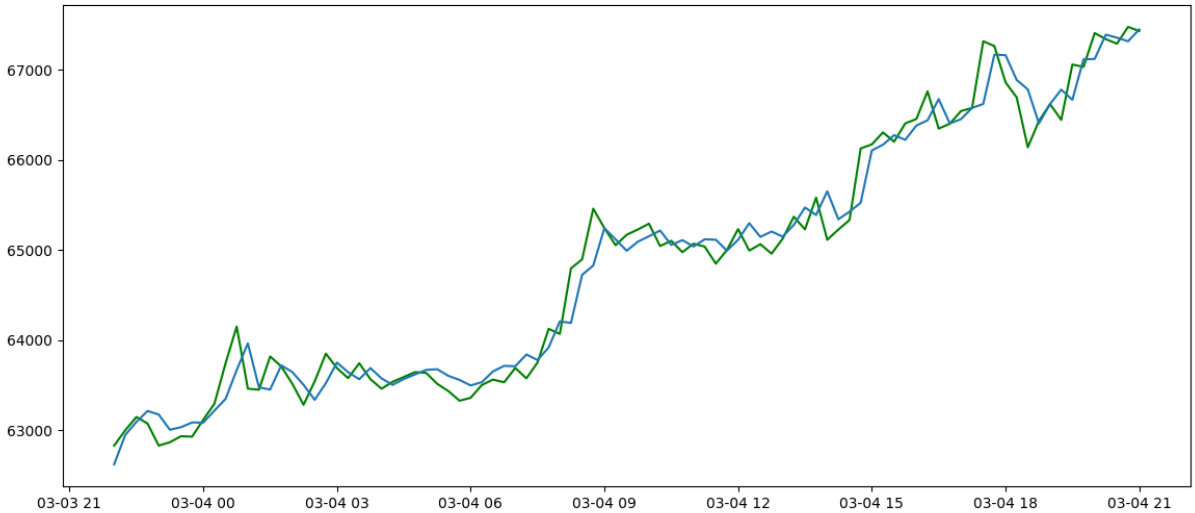

Рис. 15. Приклад прогнозованого ряду (синім) у порівнянні з істинним (зеленим) у випадку поліноміального згладження

3.4. Тестування згладжених рядів на незалежність

Розглянемо тепер питання про те, чому у випадку з диференціюванням прогноз моделі найчастіше виявився наївним. Наївний прогноз може бути результатом того, що модель не бачить кращого варіанту, ніж прогнозувати наступну ціну виключно на базі ціни у попередній момент часу незважаючи на те, що було ще раніше. Це означає, що модель не бачить зв’язку між цінами у попередні моменти часу та у наступні. Тоді постає питання про наявність або відсутність зв’язку між цінами у попередні моменти часу та у наступні. Дослідимо це питання.

Протестуємо ряди, які були отримані у результаті видалення трендової частини, на незалежність. Зробимо це двома способами.

3.4.1. *Метод 1. Використання АКФ*

Нехай ряд $\{x_1, x_2 \ldots, x_n\}$ є реалізацією незалежних та однаково розподілених випадкових величин, тоді (ст. 30 з [1]) вибіркова АКФ є випадковою величиною із розподілом $N(0, \frac{1}{n})$. Тому в такому випадку приблизно 95% усіх значень АКФ повинні лежати у діапазоні $\frac{\pm 1.96}{\sqrt{n}}$. На основі цього факту можна зробити статистичний тест для перевірки ряду на незалежність. Якщо більше 5% значень АКФ виходять за вказаний діапазон, то ми можемо відхилити гіпотезу щодо незалежності.

У таблиці 2 наведений результат такого тестування для випадків диференціювання та поліноміального згладжування.

Метод видалення трендової частини ряду	Кількість рядів, на яких приймається гіпотеза про незалежність
Диференціювання	27 з 32 (84.375%)
Поліноміальне згладжування	0 з 32 (0%)

Табл. 2. Кількість рядів на яких приймається гіпотеза про незалежність за допомогою тестування через АКФ

Бачимо, що у випадку поліноміального згладжування достатня кількість значень АКФ виявляється поза вказаними межами, в силу чого ряди, які отримуються у результаті такого методу прибирання тренду, з великою ймовірністю не можуть бути незалежними. А для випадку диференціювання значення АКФ поводять себе так ніби ряд дійсно є реалізацією незалежних величин.

3.4.2. *Метод 2. Тест Люнга-Бокса*

Іншим методом перевірки незалежності є так званий тест Люнга-Бокса. Він також базується на значеннях вибіркової АКФ, але обчислює статистику:

$$Q_{LB} = n(n+2)\sum_{j=1}^{h}\frac{\hat{\rho}^2(j)}{n-j},$$

яка у випадку незалежності елементів ряду є величиною, що розподілена за законом χ^2. Далі встановлюється деякий рівень значущості та виконується тестування, як і в багатьох тестах.

Із застосуванням тесту Люнга-Бокса отримані такі результати.

Метод видалення трендової частини ряду	Кількість рядів, на яких приймається гіпотеза про незалежність
Диференціювання	24 з 32 (75%)
Поліноміальне згладження	0 з 32 (0%)

Табл. 3. Кількість рядів на яких приймається гіпотеза про незалежність за використовуючи тест Люнга-Бокса

Бачимо такі ж результати, як і в попередній перевірці. Це наводить на думку, що при диференціюванні виходять лишки, які є просто реалізаціями незалежних випадкових величин.

Ймовірно, операція диференціювання видаляє занадто багато важливої інформації, пов’язаної із зв’язком між різними елементами ряду.

Можна сказати, що знайдено можливу причину наївного прогнозу нашого ряду у випадку дискретного диференціювання.

4. Висновки

Під час дипломної роботи була досліджена можливість прогнозування криптовалютних часових рядів за допомогою моделі ARMA та порівняння її ефективності у комбінації з різними методами зведення часових рядів до стаціонарних, а саме з методами поліноміального згладжування та дискретного диференціювання.

На етапі зведення до стаціонарності було виявлено, що на вказаних даних для зведення рядів до стаціонарних достатнім є диференціювання першого порядку або поліноміальне згладжування поліномами 6-ї степені.

Також у процесі виявилось, що у випадку дискретного диференціювання прогноз моделі ARMA виявляється у багатьох випадках просто наївним та дуже неефективним. Також була встановлена можлива причина такого прогнозування, а саме те, що після дискретного диференціювання ряд втрачає значну кількість інформації, в результаті чого зведений ряд виявляється реалізацією незалежних випадкових величин.

У випадку поліноміального згладжування проблема наївного прогнозу була відсутня, але саме прогнозування все одно виявилось досить невдалим з точки зору занадто великої середньоквадратичної похибки. Тому можна сказати, що розглянуті у роботі підходи або зовсім не можна використовувати для прогнозування у реальному житті, або вони потребують подальшого аналізу та покращення, можливо за допомогою спеціальних додаткових перетворень рядів. Також зазначимо, що, можливо, вказані методи не здатні впоратись із задачею прогнозування саме в одновимірному випадку, і можна розглянути ті самі моделі, але багатовимірні, використовуючи деякі додаткові дані для криптовалютних часових рядів.

Список використаних джерел

[1] Peter J. Brockwell Richard A. Davis – Introduction to Time Series and Forecasting Peter J. Brockwell Richard A. Davis. Third Edition

[2] Peter J. Brockwell Richard A. Davis – Time Series: Theory and Methods. Second Edition

[3] https://data.binance.vision/?prefix=data/spot/monthly/klines/BTCUSDT/15m/