

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

«До захисту допущено»

В. о. завідувача кафедри

_____ О. Л. Тимощук

« ___ » _____ 20__ р.

Дипломна робота

**на здобуття ступеня бакалавра
з напрямку підготовки 6.040303 «Системний аналіз»
на тему: «Модель часового ряду на основі
фрактального броунівського руху»**

Виконав:

студент ІV курсу, групи КА-51

Братусь Олександр Сергійович _____

Керівник:

професор, д.ф.-м.н. Бондаренко В. Г. _____

Консультант з економічного відділу:

доцент, к.е.н. Шевчук О.А. _____

Консультант з нормоконтролю:

доцент, к.т.н. Коваленко А. Є. _____

Рецензент:

професор, д.ф.-м.н. Іванов О. В. _____

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____

Київ – 2019 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу

Рівень вищої освіти – перший (бакалаврський)

Напрямок підготовки (програма професійного спрямування) – 6.040303

«Системний аналіз» («Системний аналіз та управління»)

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри

_____ О. Л. Тимощук

«__» _____ 20__ р.

ЗАВДАННЯ

на дипломну роботу студенту

Братусю Олександрю Сергійовичу

1. Тема роботи «Модель часового ряду на основі фрактального броунівського руху», керівник роботи професор кафедри ММСА, д.ф.-м.н. Бондаренко Віктор Григорович, затверджені наказом по університету від «25» травня 2019р. №1353с.

2. Термін подання студентом роботи _____

3. Вихідні дані до роботи _____

4. Зміст роботи _____

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо) _____

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	Шевчук О.А., доцент		

7. Дата видачі завдання _____

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка

Студент

_____ (підпис)

_____ (ініціали, прізвище)

Керівник роботи

_____ (підпис)

_____ (ініціали, прізвище)

РЕФЕРАТ

Дипломна робота: 78 с., 10 рис., 10 табл., 2 дод., 24 джерела.

АПРОКСИМАЦІЯ, МОДЕЛЬ, ПРОГНОЗУВАННЯ, СТАТИСТИКА, СТАЦІОНАРНІСТЬ, ФРАКТАЛЬНИЙ БРОУНІВСЬКИЙ РУХ.

Актуальність теми – моделювання часових рядів є базовим процесом при аналізі складних процесів у більшості галузей діяльності людини. З кожним роком складність процесів зростає, що спричиняє появу нових теоретичних та практичних підходів у дослідженні та моделюванні часових рядів.

Метою даної роботи є аналіз існуючих методів моделювання та обґрунтування нового методу, які використовуються для дослідження різноманітних процесів (як економічних чи фінансових явищ, так і природних явищ); порівняння запропонованої моделі з класичними.

Предмет дослідження – математична модель часового ряду: її властивості, особливості побудови, критерії ефективності, переваги та недоліки.

Об'єкт дослідження – нелінійні нестационарні часові ряди, що задовольняють певні вимоги, перш за все – стаціонарність приростів, а також відповідають характеристикам фрактальних моделей.

Програмний продукт реалізовано за допомогою мови програмування Python.

ABSTRACT

Thesis explanatory note: 78 p., 10 fig., 10 tabl., 2 append., 24 sources.

APPROXIMATION, FORECASTING, FRACTIONAL BROWNIAN MOTION, MODEL, STATIONARY, STATISTICS.

Actuality of the topic - the simulation of time series is the basic process in the analysis of complex processes in most areas of human activity. Every year, the complexity of processes is increasing, which leads to the emergence of new theoretical and practical approaches in the research and simulation of time series.

The purpose of this work is to analyze existing methods of modeling and justify the new method used to study various processes (both economic and financial phenomena and natural phenomena); comparison of the proposed model with the classic.

The subject of the research - mathematical model of time series: its properties, features of construction, criteria of efficiency, advantages and disadvantages.

The object of the research is nonlinear non-stationary time series satisfying certain requirements, first of all - stationary increments, and also corresponds to the characteristics of fractal models.

The software product is implemented using the Python programming language.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ.....	8
ВСТУП.....	9
РОЗДІЛ 1 ЕМПІРИЧНІ ЗАСАДИ МОДЕЛЮВАННЯ ЧАСОВИХ РЯДІВ.....	10
1.1 Часові ряди та їх моделі.....	10
1.2 Математичні засади моделювання на основі випадкових процесів.....	12
1.3 Фрактальний броунівський рух.....	15
Висновки до розділу.....	25
РОЗДІЛ 2 МАТЕМАТИЧНИЙ АЛГОРИТМ ПОБУДОВИ ФРАКТАЛЬНОЇ МОДЕЛІ.....	26
2.1 Методи оцінювання параметрів фрактального броунівського руху.....	26
2.2 Прогноз фрактального броунівського руху.....	33
2.3 Апроксимація часового ряду функціоналом від фрактального броунівського руху.....	37
2.3.1 Метод апроксимації.....	37
2.3.2 Перевірка якості апроксимації моделі.....	40
Висновки до розділу.....	48
РОЗДІЛ 3 ЗАСТОСУВАННЯ МАТЕМАТИЧНОЇ ФРАКТАЛЬНОЇ МОДЕЛІ.....	49
3.1 Опис даних.....	49
3.2 Апроксимація даних фрактальної моделлю.....	50
Висновки до розділу.....	52
РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ.....	54
4.1 Постановка задачі техніко-економічного аналізу.....	55
4.1.1 Обґрунтування функцій програмного продукту.....	55
4.1.2 Варіанти реалізації основних функцій.....	56
4.2 Обґрунтування системи параметрів ПП.....	58
4.2.1 Опис параметрів.....	58
4.2.2 Кількісна оцінка параметрів.....	59
4.2.3 Аналіз експертного оцінювання параметрів.....	62
4.3 Аналіз рівня якості варіантів реалізації функцій.....	66

4.4 Економічний аналіз варіантів розробки ПП	68
4.5 Вибір кращого варіанта ПП техніко-економічного рівня	73
4.6 Висновки до розділу.....	74
ВИСНОВКИ	76
ПЕРЕЛІК ПОСИЛАНЬ	77
ДОДАТОК А ЛІСТИНГ ПРОГРАМНОГО ПРОДУКТУ	79
ДОДАТОК Б ІЛЮСТРАТИВНІ МАТЕРІАЛИ ДО ДОПОВІДІ	84

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

fBm (fractional Brownian motion) – фрактальний броунівський рух

fGn (fractional Gaussian noise) – фрактальний гаусівський шум

ARMA (Autoregressive–moving-average) – модель авторегресії – ковзного середнього

ARIMA (Autoregressive integrated moving average) – інтегрована модель авторегресії – ковзного середнього

GARCH (Generalized autoregressive conditional heteroskedasticity) – авторегресійна умовна гетероскедастична модель

ВСТУП

На сьогоднішній день процес моделювання часових рядів досліджено глибоко та детально, свідченням чого є велика кількість спеціальної літератури. Залежно від того, в якій системі (соціально-економічній, фінансовій, технічній чи іншій) спостерігається досліджуваний процес, можна підібрати вже відомий метод моделювання, який оптимізований саме в межах потрібної предметної області. Разом з тим ми можемо спостерігати появу нових лінійних та нелінійних процесів, що стимулює розробку та появу нових моделей часових рядів.

Поточна робота зосереджена на моделі часового ряду, яка бере за основу фрактальний броунівський рух. Вона може описувати різноманітні фізичні та економічні явища, які можна представити випадковими процесами, яким властива нелінійність, нестационарність та немарковість.

У першому розділі представлено огляд існуючих як класичних так і нових моделей часових рядів. Розглянуто фрактальний броунівський рух та його основні показники, описано доцільність його практичного застосування при моделюванні.

Другий розділ присвячений алгоритму побудови запропонованої математичної моделі нелінійних нестационарних процесів, описано критерії якості побудови.

У третьому розділі представлено опис практичного застосування досліджуваної моделі для моделювання вибраних нелінійних процесів за допомогою створеного програмного продукту. Також проведено аналіз результатів.

Четвертий розділ присвячений економічній складовій математичної моделі.

РОЗДІЛ 1 ЕМПІРИЧНІ ЗАСАДИ МОДЕЛЮВАННЯ ЧАСОВИХ РЯДІВ

1.1 Часові ряди та їх моделі

Впродовж минулих десяти – п'ятнадцяти років відбувається поява великого обсягу нових прикладних статистичних моделей, ґрунтовне дослідження яких лише починається, що ознаменовує справжню революцію в теорії часових рядів. Однак саме практичні потреби є чинниками, що зумовлюють запровадження та вивчення нових моделей.

Останні роки характеризуються інформаційною революцією, появою нових технологій, належною зміною поведінки людини і всього суспільства, а окрім цього суттєвими якісними змінами економічних ринків. Враховуючи те, що старі класичні методи вже не здатні описати швидкозмінні нелінійні процеси, адже вони ґрунтувались на припущеннях стаціонарності часових рядів, нагальні зміни вимагають нових математичних засобів для їхнього адекватного опису і прогнозування.

Після глибокого дослідження класичних моделей, що базувалися на властивостях лінійності, стаціонарності, гаусовості часових рядів, відбувся природний перехід до моделей, які відмовилися від попередніх припущень, але узагальнили одержані результати. Перш за все, це моделі, які зберігають певні ознаки схожості вже вивчених моделей та адаптують розроблену математичну теорію та засоби. Дуже часто використовуються такі моделі як ARMA, ARIMA, GARCH та інші.

ARMA – модель авторегресії – ковзного середнього, яка визначена формулою:

$$X(t) = c + \varepsilon_t + \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \quad (1.1)$$

де c – стала;

ε_t – білий шум;

φ_i – параметри авторегресії;

p – порядок авторегресії;

θ_i – параметри ковзного середнього;

q – порядок ковзного середнього.

Дана модель широко застосовується в моделюванні фінансових ринків та економічних процесів.

ARIMA – інтегрована модель авторегресії – ковзного середнього, яка визначена формулою:

$$\Delta^d X(t) = c + \varepsilon_t + \sum_{i=1}^p \varphi_i \Delta^d X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i}, \quad (1.2)$$

де d – порядок оператора різниці часового ряду.

Ця модель є розширенням моделі ARMA для нестационарних часових рядів з метою досягти стаціонарності взяттям різниць деякого порядку вихідного ряду.

GARCH – узагальнена авторегресійна умовна гетероскедастичність, яка описується наступним рівнянням:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (1.3)$$

де α_i – параметри ARCH-моделі;

u_i – члени ARCH-моделі;

β_j – параметри GARCH-моделі;

σ_i – значення часового ряду.

На сьогодні все більше і більше з'являється якісно нових моделей та методів їхнього дослідження. Принципово іншим підходом є не використання результатів класичних моделей, а створення математичних моделей на основі

відомого випадкового процес із застосуванням певного перетворення. Є всі підстави вважати, що цей та інші розділи теорії часових рядів знаходяться на стадії становлення та розвитку.

1.2 Математичні засади моделювання на основі випадкових процесів

Неперервною математичною моделлю спостережуваної траєкторії $x(t)$ зветься пара (Φ, ξ) , для якої

$$x(t) = \Phi(X(\cdot))(t), \quad (1.4)$$

де $\xi(t)$ – випадковий процес з відомими характеристиками;

$X(t)$ – певна реалізація процесу $\xi(t)$;

Φ – оборотне перетворення в $C(0; T)$.

Якщо покласти $t = t_k$, $k = 1, \dots, n$, $x(t_k) = x_k$, то отримуємо неперервну математичну модель часового ряду.

Процес ξ назвемо основним процесом певної моделі.

У даній роботі при моделюванні часового ряду відповідно до наведених вище міркувань будується відображення Φ і за основний процес обирається фрактальний броунівський рух.

Випадковий процес $\xi(t)$ називається процесом з обмеженою варіацією, якщо

$$P \left\{ \sum_k |\xi(t_k) - \xi(t_{k-1})| < \infty \right\} = 1 \quad (1.5)$$

для довільного розбиття $0 = t_0 < t_1 < \dots < t_n = T$.

Якщо виконується «протилежна» нерівність

$$P\left\{\sum_k |\xi(t_k) - \xi(t_{k-1})| < \infty\right\} = 0, \quad (1.6)$$

то $\xi(t)$ – процес необмеженої варіації.

Історично першими моделями, які ґрунтувалися на певному перетворенні випадкового процесу $X(t)$ з відомими характеристиками, були процеси Леві з незалежними стаціонарними приростами, які як слідує, були ще й марківськими. Наприклад, у першому наближенні число клієнтів $N(t)$ в системі масового обслуговування можна представити процесом Пуассона з одновимірними розподілами:

$$P\{N(t) = k\} = \exp(-\lambda t) \frac{(\lambda t)^k}{k!}, \quad (1.7)$$

де λ – інтенсивність пуассонівського процесу;

k – число клієнтів.

Перехід до дифузійних процесів став наступним кроком в етапі розвитку математичних моделей. Так, дифузійна модель ціни акції $S(t)$ отримана з наступних міркувань:

$$S(t + dt) - S(t) = S(t)(\xi(t + dt) - \xi(t)) + \mu(t)S(t)dt, \quad (1.8)$$

де $\xi(t)$ – певний основний випадковий процес;

dt – приріст часу;

$\mu(t)$ – відсоткова ставка (надалі вважатимемо її за константу).

Якщо покласти:

$$\xi(t) = \sigma w(t), \quad (1.9)$$

де $w(t)$ – вінерівський процес,

σ – коефіцієнт волатильності, то приходимо до рівняння:

$$dS(t) = \sigma S(t)dw(t) + \mu S(t)dt, \quad (1.10)$$

яке можна трактувати як стохастичне рівняння Іто. Розв'язок цього рівняння тоді можна записати у вигляді геометричного (економічного) броунівського руху:

$$S(t) = S(0)\exp\left\{\sigma w(t) + \left(\mu - \frac{\sigma^2}{2}\right)t\right\}, \quad (1.11)$$

Варто зауважити, що знаменита формула Блека-Шоулза для раціональної (справедливої) ціни опціона була отримана в рамках моделі (1.11) ціни акції, з якої був отриманий ряд інших відомих результатів.

Що стосується процесів Леві як базових неперервних моделей для часових рядів (зокрема в економіці та фінансах), то існує величезна кількість робіт, що зумовило широке їх використання. Разом з тим, цим процесам (і дифузійним в тому числі) притаманна марковість, що часто є недоліком. Так, моделі, що мають марківську властивість:

$$P\{X(t_n) \in \Delta | X(t_1) = a_1, \dots, X(t_{n-1}) = a_{n-1}\} = P\{X(t_n) \in \Delta | X(t_{n-1}) = a_{n-1}\}, \quad (1.12)$$

можуть описувати порівняно прості фізичні явища на кшталт дифузії. Що стосується процесів у біології, економіці, фінансах, кліматології та інших сферах – відсутність післядії є непереконливою і зумовлює природний перехід до немарківських моделей.

Якщо часовий ряд пройшов процедуру деякого попереднього згладжування (фільтрації), то в якості моделі можна обрати процес з гладкими траєкторіями (тобто процес з необмеженими варіаціями). Найпростішим прикладом застосування такої моделі є проінтегрований дифузійний процес:

$$X(t) = \int_0^t f(\tau, w(\tau)) d\tau, \quad (1.13)$$

де функція f мусить бути оцінена. Згладжування отриманих даних широко використовується як в технічних системах, так при опрацюванні даних деяких фізичних явищ.

Процес із залежними стаціонарними приростами є ще одним прикладом немарківської моделі, скінченновимірні розподіли якої цілковито визначаються особливостями модельованого часового ряду. Монографія [1] наводить посилання на результати статистичної обробки фінансових часових рядів. В усіх розглянутих випадках гіпотеза про гаусовість одновимірних розподілів моделі відхиляється; одновимірний розподіл процесів, які моделюють прирости часового ряду в більшості досліджених випадків виявляються симетричними та мають «важкі» гіперболічні хвости:

$$P\{y_n > x\} \sim x^{-\alpha}, \quad x \rightarrow +\infty, \quad \alpha > 3. \quad (1.14)$$

1.3 Фрактальний броунівський рух

Однією з найбільш поширених моделей часових рядів є фрактальний броунівський рух (гаусівський випадковий процес), який позначається $B_H(t)$, де $0 < H < 1$ – деякий його параметр (параметр Харста), значення якого буде висвітлено нижче. Даний процес є дуже затребуваним при моделюванні, що пояснюється його наступними властивостями:

а) фрактальний броунівський рух можна задати в інтегральному вигляді, що був запропонований Вейлем:

$$B_H(t) = B_H(0) + \frac{1}{\Gamma(H+\frac{1}{2})} \left\{ \int_{-\infty}^0 [(t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}] dB(s) + \int_0^t (t-s)^{H-\frac{1}{2}} dB(s) \right\}, \quad (1.15)$$

Де $B(s)$ – випадковий гаусівський процес;

H – параметр Харста;

б) математичне сподівання дорівнює нулю:

$$E B_H(t) = 0; \quad (1.16)$$

в) у нульовій точці дорівнює нулю:

$$B_H(0) = 0; \quad (1.17)$$

г) коваріаційна функція процесу має вигляд:

$$R(t, s) = E B_H(t)B_H(s) = \frac{1}{2} (t^{2H} + s^{2H} - |t - s|^{2H}); \quad (1.18)$$

г) дисперсія має вигляд:

$$D B_H(t) = t^{2H}; \quad (1.19)$$

д) майже всі траєкторії процесу задовольняють умові Гельдера:

$$|X(t) - X(s)| \leq c|t - s|^\alpha, \quad \text{де } \alpha < H, \quad (1.20)$$

що підтверджує факт, що гладкість траєкторій процесу $B_H(t)$ теж визначається параметром Харста і узагальнює відомий результат Леві для вінерівського

процесу, а саме означає необмеженість варіації траєкторій фрактального броунівського руху;

е) процес має автомодельні властивості (самоподібності), що є корисним при аналізі часових рядів:

$$\text{Law}(B_H(at)) = \text{Law}(a^H B_H(t)), \quad (1.21)$$

тобто зміна часового масштабу еквівалентна зміні «амплітуди» процесу, а рівність в даній формулі означає співпадіння одновимірних розподілів процесів $B_H(at)$ та $a^H B_H(t)$;

є) прирости фрактального броунівського руху мають такі характеристики:

$$B_H(t) - B_H(s) \sim B_H(t - s) \sim N(0, (t - s)^{2H}), \quad (1.22)$$

що свідчить про їхню стаціонарність;

ж) прирости $B_H(t_2) - B_H(t_1)$ та $B_H(t_4) - B_H(t_3)$, $t_1 < t_2 < t_3 < t_4$ формують гаусівський випадковий вектор, коваріація між координатами якого дорівнює:

$$\frac{1}{2} ((t_4 - t_1)^{2H} + (t_3 - t_2)^{2H} - (t_4 - t_2)^{2H} - (t_3 - t_1)^{2H}). \quad (1.23)$$

Врахувавши дискретність часу:

$$y_k = B_H(k + 1) - B_H(k), \quad (1.24)$$

отримаємо коефіцієнт кореляції:

$$\rho(y_j, y_k) = \frac{1}{2} (|k - j + 1|^{2H} + |k - j - 1|^{2H} - 2|k - j|^{2H}), \quad (1.25)$$

що дозволяє стверджувати, що прирости процесу фрактального броунівського руху утворюють стаціонарну (у вузькому сенсі) послідовність.

Покладемо $j = k + n$. Тоді кореляційна функція приростів матиме вигляд:

$$\rho(y_k, y_{k+n}) = \frac{1}{2}((n+1)^{2H} + (n-1)^{2H} - 2n^{2H}). \quad (1.26)$$

Асимптотика даного результату при $n \rightarrow \infty$:

$$\rho_n = \rho(y_k, y_{k+n}) \sim H(2H-1)n^{2H-2}, \quad (1.27)$$

тобто при $H \neq \frac{1}{2}$ убутання пам'яті для приростів має гіперболічний характер, а при $H = \frac{1}{2}$ прирости утворюють гаусівську послідовність незалежних випадкових величин.

Так як ряд $\sum_{n=1}^{\infty} \rho_n \sim \sum_{n=1}^{\infty} \frac{1}{n^{2-2H}}$ збігається при $H < \frac{1}{2}$ і розбігається при $H > \frac{1}{2}$, то в першому випадку прирости утворюють послідовність з короткою пам'яттю, а в другому – з довгою.

Таким чином, саме параметр Харста H визначає вид пам'яті в межах моделі fBm.

Якщо $H < \frac{1}{2}$, то $\rho_n < 0$, і саме від'ємність коваріації означає швидку мінливість значень y_n . Послідовність таких y_n приростів фрактального броунівського руху прийнято називати рожевим шумом. Для такого процесу fBm прийнята також термінологія «антиперсистентний», що підтверджують статистичні дані (фінансові), які показують антиперсистентність відповідних часових рядів.

Якщо $H > \frac{1}{2}$, то $\rho_n > 0$, значення y_n характеризуються сильною післядією (довгою пам'яттю). Послідовність таких y_n приростів фрактального броунівського руху прийнято називати чорним шумом. Для такого процесу fBm

прийнята також термінологія «персистентний», що підтверджують дані, які описують певні фізичні процеси, наприклад, сонячну активність.

Наведені вище міркування дають зробити висновок, що лише для персистентного випадку ($H > \frac{1}{2}$) можливий прогноз значень фрактального броунівського руху.

з) при $H = \frac{1}{2}$ фрактальний броунівський рух є вінерівським процесом.

Графічне зображення траєкторій фрактального броунівського руху приведено на рис. 1.1 ($H=0,1$), рис. 1.2 ($H=0,5$) та рис. 1.3 ($H=0,9$).

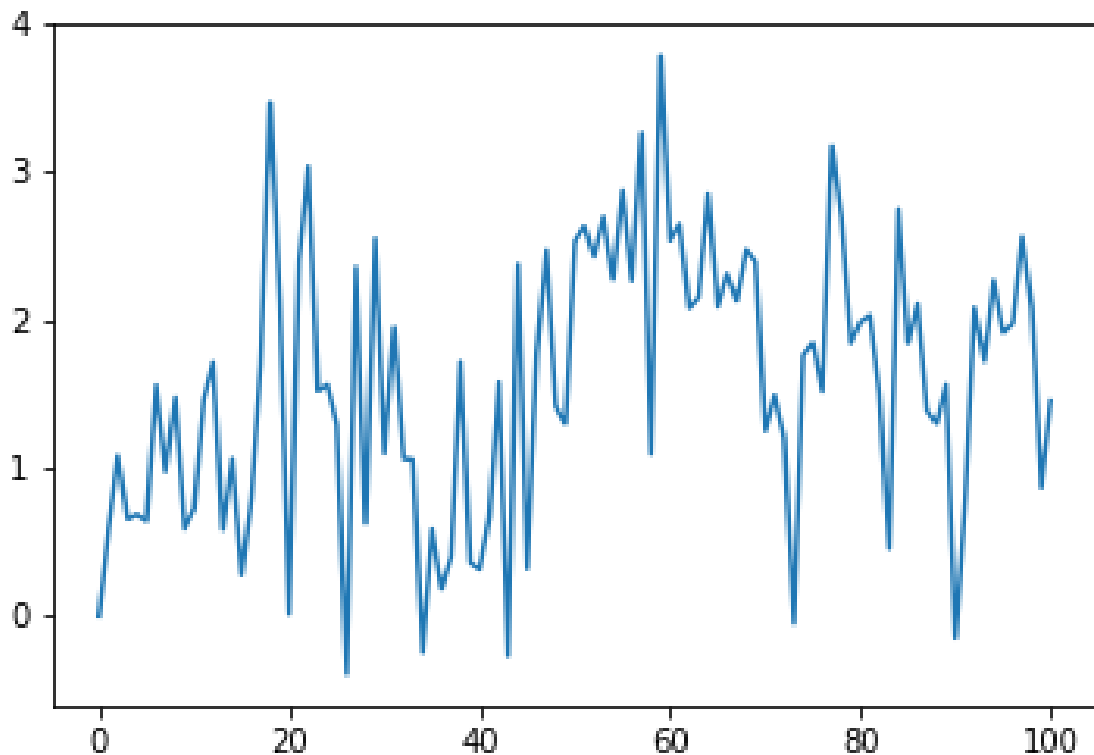
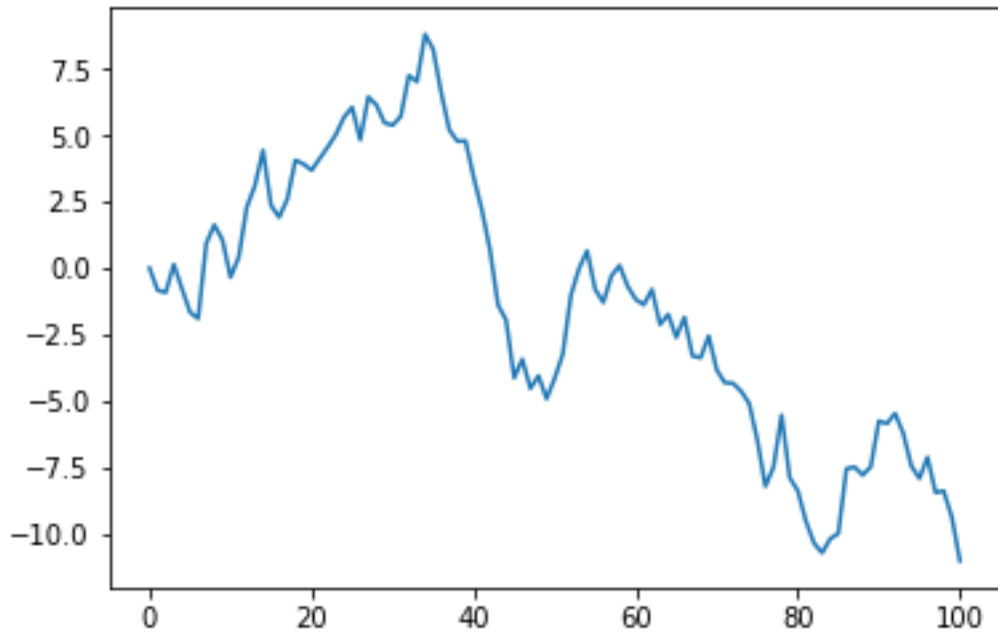
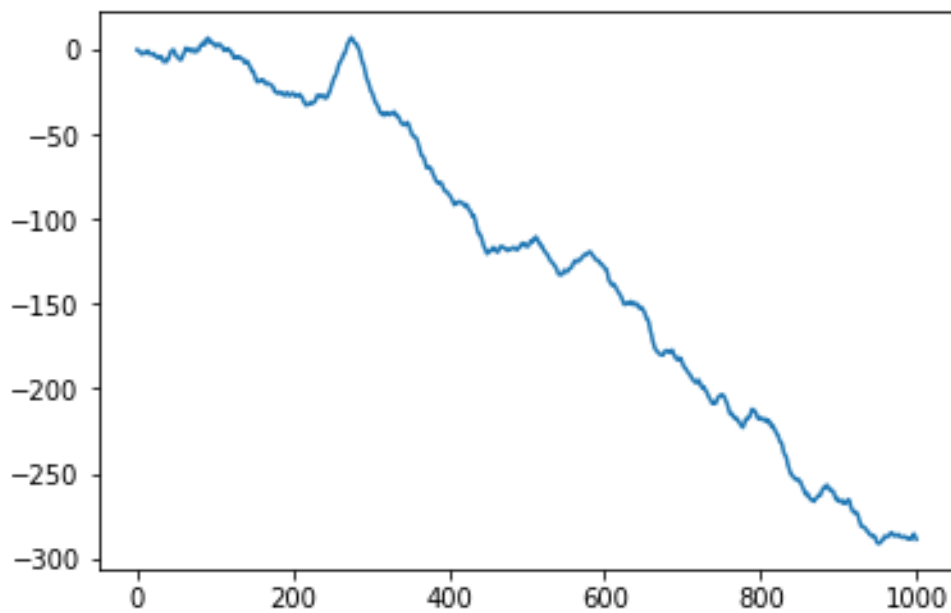


Рисунок 1.1 – Графік $V_H(t)$, $H = 0,1$.

Рисунок 1.2 – Графік $B_H(t)$, $H = 0,5$.Рисунок 1.3 – Графік $B_H(t)$, $H = 0,9$.

Наведені вище відомості щодо результатів обчислюваних експериментів дають підґрунтя судити про неадекватність гаусівських моделей для деяких часових рядів. При використанні фрактального броунівського руху вимога

гаусовості одновимірних розподілів можна задовольнити, якщо перейти до процесу:

$$X(t) = g(B_H(t)), \quad (1.28)$$

де g – непарна зростаюча функція.

При цьому одновимірні розподіли $X(t)$ мають цільність:

$$f_{X(t)}(y) = \frac{1}{\sqrt{2\pi t^H}} \exp\left\{-\frac{(g^{-1}(y))^2}{2t^{2H}}\right\} \frac{1}{g'(g^{-1}(y))}, \quad (1.29)$$

тобто вибір функції g може задовольнити вимоги, отримані при статистичному опрацюванні часового ряду.

Великий поштовх вивченню процесів з корельованими приростами дала робота Г. Харста [2], яка на сьогодні вже є хрестоматійною. Основний результат, приведений в цій роботі має наступні викладки.

Нехай $\{x_n\}$ - приріст спостережуваного часового ряду (Харст досліджував у своїй роботі рівень річки Ніл).

Покладемо $X_n = \sum_{k=1}^n x_k$ і розглянемо статистики $k \leq n$:

$$R_n = \max_k \left(X_k - \frac{k}{n} X_n \right) - \min_k \left(X_k - \frac{k}{n} X_n \right), \quad (1.30)$$

$$S_n = \left(\frac{1}{n} \sum_{k=1}^n x_k^2 - \left(\frac{1}{n} \sum_{k=1}^n x_k \right)^2 \right)^{\frac{1}{2}}, \quad (1.31)$$

та їхнє відношення $\frac{R_n}{S_n}$, яке характеризує нормоване відхилення X_k від середнього значення $\frac{k}{n} X_n$ за n років.

Спираючись на надзвичайно великий обсяг фактичних даних, Г. Харст вивів наступне асимптотичне співвідношення:

$$\frac{R_n}{S_n} \sim cn^H, \quad (1.32)$$

де $H \approx 0,7$.

Підійшовши до цієї залежності інакше, в роботі [3] пізніше був також доведений інший результат.

Нехай $\{x_n\}$ - послідовність однаково розподілених незалежних випадкових величин, $E x_n = 0$, $D x_n = 0$. Тоді має місце співвідношення:

$$\frac{R_n}{S_n} \sim c\sqrt{n}. \quad (1.33)$$

Отриманий Г. Харстом результат можна пояснити, припустивши незалежність приростів $\{x_k\}$.

Згадані вище результати стали базовими для досліджень Б. Мандельброта, який у своїх роботах [4-6] запропонував брати фрактальний броунівський рух за основу ймовірнісної моделі. Він розглянув теоретичні обґрунтування даної пропозиції, а також практичне застосування fBm в аналізі часових рядів, які описують різноманітні економічні та фізичні явища та процеси. Варто зауважити, що значна кількість дослідників наголошує, посилаючись на наведені вище факти, на універсальності даної моделі.

Що стосується статистики фрактального броунівського руху, то першим фундаментальним результатом стала гранична теорема для сум залежних випадкових величин – приростів y_k fBm, що була опублікована в роботі [7]. Для статистик:

$$R_{jn} = \frac{1}{n} \sum_{k=1}^n |y_k|^j, \quad (1.34)$$

де j – натуральне,

доведена наступна теорема:

$$\frac{R_{jn}}{E R_{jn}} \rightarrow 1, \quad n \rightarrow \infty, \quad \text{з ймовірністю } 1. \quad (1.35)$$

Доведене співвідношення не є тривіальним результатом, адже зазвичай граничні теореми доводяться лише для сум слабо залежних між собою величин. При доведенні використовуються досить спеціальні топологічні поняття – зокрема, фрактальна розмірність траєкторій процесу.

Також був отриманий результат у роботі [7] щодо обчислювального експерименту з імітаційним моделювання fBm, який дозволяє оцінити швидкість збіжності в граничних теоремах.

Наведений результат вплинув на розвиток досліджень зі статистики фрактального броунівського руху. В останні роки все більше результатів з'являється зі статистики фрактального броунівського руху. Цілий ряд робіт [8-21] присвячений доведенню низки граничних теорем, наслідком яких є можливість розрізнення гіпотез стосовно fBm. Загалом у цих роботах досліджується поведінка «інтегральних сум», зв'язаних з приростами fBm. Вводяться позначення:

$$x_k = B_H \left(\frac{k}{n} \right), \quad y_k = B_H \left(\frac{k+1}{n} \right) - B_H \left(\frac{k}{n} \right), \quad (1.36)$$

і досліджуються статистики вигляду:

$$I_n(q) = c_n \sum_{k=0}^{n-1} f(x_k)(n^H y_k)^q, \quad (1.37)$$

де c_n - нормуючий множник, $c_n \rightarrow 0$, $n \rightarrow \infty$;

f - довільна гладка функція.

Для $H = 0,5$, тобто для вінерівського процесу, $I_n(q)$ збігається (в певному сенсі) до випадкової величини, яку визначають як стохастичний інтеграл Іто. У згаданих раніше роботах були доведені результати збіжності для фрактального броунівського руху при різних значеннях параметру Харста H . Наприклад, для $H \in (0; \frac{3}{4})$:

$$\frac{1}{\sqrt{n}} \sum_{k=0}^n ((n^H y_k)^q - \mu_q) \rightarrow \eta, \quad (1.38)$$

де $\eta \sim \mathcal{N}(0, \sigma^2)$,

причому збіжність розуміється в сенсі розподілів:

$$\mu_q = E(\xi^q), \quad \xi \sim \mathcal{N}(0, 1), \quad (1.39)$$

Для двічі диференційованої функції доведено наступні граничні співвідношення:

$$\frac{1}{n^{1-H}} \sum_{k=0}^{n-1} f(x_k)(n^H y_k)^q \rightarrow -\frac{\mu_{q+1}}{2} \int_0^1 f'(B_H(s)) ds, \quad (1.40)$$

де q - непарне, $q \geq 3$;

$H \in (0; \frac{1}{2})$;

а збіжність розуміється в середньоквадратичному сенсі.

$$\frac{1}{n^{1-H}} \sum_{k=0}^{n-1} f(x_k) ((n^H y_k)^2 - 1) \rightarrow -\frac{1}{4} \int_0^1 f'(B_H(s)) ds, \quad (1.41)$$

де $H \in (0; \frac{1}{4})$.

Для $H > \frac{1}{2}$ має місце наступний результат:

$$\frac{1}{n^H} \sum_{k=0}^{n-1} f(x_k) (n^H y_k)^q \rightarrow q \mu_{q-1} \int_0^{B_H(1)} f(x) dx, \quad (1.42)$$

де q - непарне, $q \geq 3$.

Вибираючи в (1.37) функцію f з потрібними властивостями (наприклад, степеневу), можна отримати критерій перевірки параметричної гіпотези для перевірки значення параметра Харста, а також перевірити гіпотезу про відповідність досліджуваного ряду моделі, яка описується фрактальним броунівським рухом.

Висновки до розділу

В даному розділі було наведено теоретичні викладки щодо математичного моделювання часових рядів. Розглянути причини та передумови виникнення фрактального підходу. Проаналізовано теоретичне підґрунтя побудови моделі фрактального броунівського руху, відзначено її переваги та недоліки.

РОЗДІЛ 2 МАТЕМАТИЧНИЙ АЛГОРИТМ ПОБУДОВИ ФРАКТАЛЬНОЇ МОДЕЛІ

2.1 Методи оцінювання параметрів фрактального броунівського руху

Покладемо $X(t) = \sigma B_H(t)$. Розглянемо прирости:

$$y_k = X\left(\frac{k}{n}\right) - X\left(\frac{k-1}{n}\right) = \sigma(B_H\left(\frac{k}{n}\right) - B_H\left(\frac{k-1}{n}\right)), k = 1, 2, \dots, n, \quad (2.1)$$

Тоді ці прирости утворюють вектор $y = (y_1, y_2, \dots, y_n) \sim \mathcal{N}(0, V)$, що має нульове математичне сподівання, а кореляційна матриця визначається як:

$$V = \frac{\sigma^2}{n^{2H}} S, \quad (2.2)$$

де S – нормована кореляційна матриця, елементи якої визначаються рівністю:

$$s_{jk} = \frac{|k-j+1|^{2H} + |k-j-1|^{2H}}{2} - |k-j|^{2H}, \quad (2.3)$$

з чого слідує, що $\{y_k\}$ утворюють гаусівську стаціонарну послідовність.

Надалі саме прирости досліджуваного часового ряду y , як наслідок, прирости фрактального броунівського руху будуть об'єктом даного дослідження, тобто побудовані згодом оцінки будуть функціями від значень y_1, \dots, y_n .

В межах розглянутої моделі потрібно буде оцінити два невідомих параметри фрактального броунівського руху: коефіцієнт волатильності σ та показник Харста H . Буде запропоновано та обґрунтовано метод взаємозалежного оцінювання параметрів та метод одночасного оцінювання цих параметрів.

Оцінки $\hat{\sigma}$ та \hat{H} невідомих параметрів процесу $X(t) = \sigma B_H(t)$ мають задовольняти ряд вимог, основною з яких є конзистентність. Ця властивість гарантується, як правило, граничними теоремами, які, загалом, доведені для послідовностей незалежних (чи тих, що задовольняють умові сильного перемішування) випадкових величин, але прирости ΔB_m таким умовами не задовольняють. Тим не менше, справедливим є наступне твердження, яке було доведено в роботі Пельтьє [22].

Розглянемо статистики – абсолютні моменти приростів фрактального броунівського руху:

$$R_{jn} = \frac{1}{n} \sum_{k=1}^n |y_k|^j, \quad (2.4)$$

де j – натуральне число.

Для R_{jn} можна легко вирахувати математичне сподівання:

$$E_n(j) = ER_{jn} = \frac{\sigma^j}{n^{jH}} \frac{2^{\frac{j}{2}} \Gamma\left(\frac{j+1}{2}\right)}{\sqrt{\pi}}. \quad (2.5)$$

Також має місце наступний результат у вигляді теореми.

Теорема. З одиничною ймовірністю

$$\frac{R_{jn}}{ER_{jn}} \rightarrow 1, \quad n \rightarrow \infty. \quad (2.6)$$

Зокрема, при $j \rightarrow 1$ $\sqrt{\frac{\pi}{2}} \frac{n^H}{\sigma}$, звідки знаючи H , можна отримати конзистентну оцінку параметра σ :

$$\hat{\sigma}_{1n} = n^H \sqrt{\frac{\pi}{2}} R_{1n}, \quad (2.7)$$

і навпаки, знаючи σ , знаходимо оцінку H :

$$\hat{H}_n = \frac{\ln\left(\sqrt{\frac{2}{\pi}} \frac{\sigma}{R_{1n}}\right)}{\ln n}. \quad (2.8)$$

Варто зауважити, що наведений вище метод отримання консистентних оцінок не дозволяє оцінити незалежно H та σ . Приведений нижче метод забезпечує одночасне оцінювання оцінюваних параметрів. Візьмімо до уваги статистики, які не залежать від параметра σ . Розглянемо наступні:

$$\lambda_n = \frac{1}{n-1} \sum_{k=1}^{n-1} y_k y_{k+1}, \quad E\lambda_n = \frac{\sigma^2}{n^{2H}} (n^{2H-1} - 1) = \frac{\sigma^2}{n^{2H}} \rho, \quad (2.9)$$

де ρ – коефіцієнт кореляції між сусідніми приростами.

Розглянемо емпіричну оцінку цього параметру ρ :

$$\hat{\rho} = \frac{\lambda_n}{R_{2n}}. \quad (2.10)$$

Результати обчислювального експерименту свідчать про те, що збіжність $\hat{\rho}$ (якщо вона має місце, що є не завжди, так як доведення збіжності не є опублікованим) дуже повільна, тобто для об'єму вибірки $n \approx 500$ таке оцінювання не може бути застосованим.

Було виведене інше співвідношення між H та σ .

Нехай ε – канонічний гаусівський вектор, тобто:

$$E\varepsilon = 0, \quad E(\varepsilon, u)(\varepsilon, v) = (u, v), \quad \dim \varepsilon = n. \quad (2.11)$$

В такому разі вектор приростів fVn можна записати у вигляді:

$$y = V^{\frac{1}{2}}\varepsilon. \quad (2.12)$$

звідки:

$$n = E(\varepsilon, \varepsilon) = E(V^{-1}y, y) = \frac{\sigma^2}{n^{2H}} E(S^{-1}y, y), \quad (2.13)$$

а отже статистика $n^{2H-1}(S^{-1}y, y)$ є незміщеною оцінкою параметра σ^2 . Доведемо тепер, що запропонована оцінка є консистентною. Уведемо позначення:

$$\hat{\sigma}_{2n} = \sqrt{n^{2H-1}(S^{-1}y, y)}. \quad (2.14)$$

Нехай y – гаусівський вектор у евклідовому просторі X :

$$\begin{aligned} y \sim \mathcal{N}(0; V), \quad E(y, h) = 0, \quad E(y, h)(y, k) = (Vh, k), \\ E(Ay, y) = \text{tr}(AV), \quad h, k \in X \end{aligned} \quad (2.15)$$

де A – лінійний оператор в X .

Якщо f – диференційована на X функція, то формула інтегрування частинами має вигляд:

$$E(\text{grad } f(x), h) = Ef(y)(V^{-1}y, h). \quad (2.16)$$

Твердження. Якщо A, B – симетричні лінійні оператори, то має місце формула:

$$E(Ay, y)(By, y) = 2\text{tr}(VAVB) + \text{tr}(AV) \cdot \text{tr}(BV). \quad (2.17)$$

Доведення. Покладемо:

$$f(y) = (Ay, y)(By, k), \quad (2.18)$$

де k – фіксований вектор.

Тоді:

$$\text{grad } f(y) = 2(By, k)Ay + (Ay, y)Bk, \quad (2.19)$$

і тоді формулу (2.16) можна записати:

$$\begin{aligned} E(Ay, y)(By, k)(V^{-1}y, h) &= 2E(Ay, h)(By, k) + (By, k)E(Ay, y) = \\ &= 2(AVBh, k) + (Bh, k)\text{tr}(AV). \end{aligned} \quad (2.20)$$

Нехай $\{e_i\}$ – ортобазис в X . Якщо покласти $h = e_i, k = Ve_i$, і просумувати останню рівність за i , то прийдемо до потрібного твердження.

Твердження доведено.

Наслідок. $\hat{\sigma}_{2n}$ – консистентна оцінка параметра σ .

Доведення. У формулі (2.17) покладемо $A = B = S^{-1}$. Тоді:

$$E(S^{-1}y, y)^2 = \frac{2\sigma^4}{n^{4H-1}} + \frac{\sigma^4}{n^{4H-2}}, \quad (2.21)$$

а отже:

$$D\hat{\sigma}_{2n} = n^{4H-2}E(S^{-1}y, y)^2 - \sigma^4 = n^{4H-2} \left(\frac{2\sigma^4}{n^{4H-1}} + \frac{\sigma^4}{n^{4H-2}} \right) - \sigma^4 = \quad (2.22)$$

$$= \frac{2\sigma^4}{n} \rightarrow 0, \quad n \rightarrow \infty.$$

Доведено.

Рівності (2.7) та (2.14) утворюють систему, яку доцільно вирішувати ітераційно. Пропонується для певного значення $N \in (0; 1)$ вираховувати оцінку $\hat{\sigma}_{1n}$, матрицю S^{-1} та оцінку $\hat{\sigma}_{2n}$. Повторюємо таку ітерацію з перебором значень N з певним кроком, доки не досягаємо відношення

$$\frac{\hat{\sigma}_{2n}}{\hat{\sigma}_{1n}} = \frac{0,8}{R_{1n}} \sqrt{\frac{(S^{-1}y, y)}{n}} \approx 1, \quad \text{тобто} \quad \left| \frac{\hat{\sigma}_{2n}}{\hat{\sigma}_{1n}} - 1 \right| \rightarrow \min \quad (2.23)$$

Значення \hat{N} параметра N , для якого виконується умова (2.23), приймається за оцінку, і

$$\hat{\sigma} = \frac{\hat{\sigma}_{1n} + \hat{\sigma}_{2n}}{2}. \quad (2.24)$$

Експериментальним шляхом було проведено чисельну перевірку ефективності запропонованого методу, результати якої представлені в таблиці 2.1. У ній показані значення величини

$$q_{kj} = \frac{0,8}{r_1} \sqrt{\frac{(S_j^{-1}z_k, z_k)}{n}}, \quad (2.25)$$

де z_k – згенерований вектор приростів fBm з показником Харста H_k ;

S_j – нормована кореляційна матриця, яка відповідає fBm з показником Харста H_j .

Для кожного значення H_k (в певному рядку) значення q_{kj} обчислюються перебором параметра H_j з кроком $\Delta H_j = 0,1$. Результати обчислень наведені у таблиці 2.1, генерація z_k була здійснена при:

$$\begin{aligned} n = 256; H_k = 0,4; H_k = 0,7; \\ n = 1024; H_k = 0,3; H_k = 0,6; H_k = 0,7; \end{aligned} \quad (2.26)$$

Таблиця 2.1 – Ефективність методу оцінювання

$H_j \backslash H_k$	0,1	0,2	0,3	0,4	0,6	0,7	0,8	0,9
$n = 256$								
0,4	1,09	1,10	1,05	1,0	1,12	1,17	1,23	1,28
0,7	1,10	0,92	0,91	0,89	0,95	1,01	1,10	1,28
$n = 1024$								
0,3	0,85	0,89	1,04	1,20	1,4	1,37	1,42	1,48
0,6	1,18	1,12	1,20	1,23	1,04	1,15	1,23	1,25
0,7	1,20	1,18	1,15	1,14	1,13	1,02	0,93	0,89

Проведений аналіз результатів експерименту свідчить, що при кожному H_k (в певному рядку)

$$|q_{kj} - 1| \rightarrow \min, \quad \text{якщо } H_j = H_k, \quad \text{тобто } \hat{H}_k = H_j, \quad (2.27)$$

а отже запропонований алгоритм є ефективним.

2.2 Прогноз фрактального броунівського руху

Нехай спостерігаємо траєкторію (реалізацію) випадкового процесу $X(t)$, $0 \leq t \leq T$. Випадкова величина $\hat{X}(T + \tau)$ називається оптимальним прогнозом значень процесу в точці $T + \tau$, якщо

$$E(\hat{X}(T + \tau) - X(T + \tau))^2 = \min_{\xi} E(\xi - X(T + \tau))^2. \quad (2.28)$$

Оптимальний прогноз визначається формулою умовного середнього [23, 24]:

$$\hat{X}(T + \tau) = E(X(T + \tau) | D_{\tau}). \quad (2.29)$$

Так як процес обчислення такого умовно середнього є складною задачею, як приклад, розглянемо задачі, які допускають спрощення (2.29).

Приклад 1. Нехай $X(t) = w(t)$.

Тоді $E(w(T + \tau) | D_{\tau}) = E(w(T + \tau) - w(T) | D_{\tau}) + E(w(T) | D_{\tau}) = w(T)$ – в силу незалежності приростів вінерівського процесу.

Приклад 2. Нехай $x_k = X\left(\frac{k}{n}\right)$, $k = 1, \dots, n$, де послідовність x_k будується за схемою авторегресії $AR(p)$, тобто $x_k = \sum_{i=1}^p a_i x_{k-i} + b \varepsilon_k$, де $\varepsilon_k \sim \mathcal{N}(0; 1)$ – незалежні.

Тоді однокроковий прогноз в силу вимірності x_{k+1} щодо x_1, \dots, x_k задається рівністю

$$\hat{x}_{k+1} = E(x_{k+1} | x_1, \dots, x_k) = \sum_{j=1}^p a_j x_{k+j-1}. \quad (2.30)$$

Приклад 3. Розглянемо гаусівський вектор:

$$\begin{aligned} \xi &= (\xi_1, \dots, \xi_n), \xi = (\eta, \zeta), \dim \eta = m, \dim \zeta = n - m, \\ \xi &\approx \mathcal{N}(0; S), \eta \approx \mathcal{N}(0; A), \zeta \approx \mathcal{N}(0; D) \end{aligned} \quad (2.31)$$

тобто ξ – кореляційний оператор у вигляді блочної матриці:

$$S = \begin{bmatrix} A & B \\ C & D \end{bmatrix}, \quad (2.32)$$

де елементи матриць B, C визначають взаємну кореляцію координат η та ζ . Якщо η – спостережуваний вектор, ζ – оцінюваний вектор, то через гаусовість оптимальний прогноз співпадає з лінійною оцінкою:

$$\zeta = E(\zeta|\eta) = CA^{-1}\eta, \quad (2.33)$$

або в координатній формі:

$$\hat{\xi}_{m+j} = \sum_{k=1}^m \sum_{i=1}^m s_{m+j,k} a^{ki} \xi_i, \quad j = 1, \dots, n - m. \quad (2.34)$$

Для однокрокової екстраполяції ($m = n - 1$, B – вектор, ковектор $C = B^T$, $D = E\xi_n^2$) середньоквадратична абсолютна помилка прогнозу δ визначається таким співвідношенням:

$$\delta^2 = E(\hat{\xi}_n - \xi_n)^2 = D - (A^{-1}B, B), \quad (2.35)$$

а відносна похибка наступним:

$$\delta D^{-0,5}. \quad (2.36)$$

Прогноз фрактального броунівського руху можна будувати як для його приростів, поклавши $\xi_k = y_k = B_H\left(\frac{k}{n}\right) - B_H\left(\frac{k-1}{n}\right)$, так і для його значень, поклавши $\xi_k = B_H\left(\frac{k}{n}\right)$.

В першому випадку елементи матриці S визначається рівністю (2.3) і формула (2.34) приводиться до виду:

$$\hat{y}_{m+j} = \sum_{k=1}^m \sum_{i=1}^m \left(\frac{(m+j-k+1)^{2H} + (m+j-k-1)^{2H}}{2} - (m+j-k)^{2H} \right) a^{ki} y_i, \quad (2.37)$$

$$j = 1, \dots, r, \quad r = m - n.$$

Інакше, при екстраполяції значень фрактального броунівського руху елементи s_{jk} кореляційної матриці S визначаються рівнянням

$$s_{jk} = 0,5(j^{2H} + k^{2H} - (k-j)^{2H}). \quad (2.38)$$

Для визначення якості прогнозу на генерованих даних було проведено обчислювальний експеримент, в ході якого було перевірено обидва підходи. За навчальною вибіркою довжини m будується прогноз на 8 кроків.

Результати прогнозу за першим підходом (за формулою (2.37)) є незадовільними – абсолютна похибка $\delta_j = \left| \frac{\hat{y}_{m+j} - y_{m+j}}{y_{m+j}} \right|$, $j = 1, \dots, 8$ в середньому дорівнює 0,8 – 1,2 і не залежить від об'єму навчальної вибірки.

Обчислення прогнозованих значень \hat{x}_{m+j} фрактального броунівського руху за другим підходом (2.34) дає очікуваний результат. Так прогноз антиперсистентного ($H < 0,5$) процесу є незадовільним, а його помилка не залежить від m . Водночас для персистентного процесу якість прогнозу покращується зі зростанням об'єму навчальної вибірки. Результати експерименту

наведені в таблиці 2.2. Приведено значення відносних похибок $\delta_j = \left| \frac{\hat{y}_{m+j} - y_{m+j}}{y_{m+j}} \right|$ при прогнозі на 8 кроків ($j = 1, \dots, 8$) для $H = 0,3; H = 0,7; H = 0,9$ при $m = 100; m = 500; m = 1000$.

Таблиця 2.2 – Ефективність покрокового прогнозування

H		1	2	3	4	5	6	7	8
0.3	$m = 100$	0.67	0.71	0.45	0.07	2.32	1.49	1.58	0.11
	$m = 500$	0.25	0.16	0.06	0.17	1.68	0.61	0.71	0.69
	$m = 1000$	0.04	0.12	0.16	0.01	0.10	0.04	0.18	0.17
0.7	$m = 100$	0.02	0.08	0.17	0.27	0.29	0.35	0.50	0.58
	$m = 500$	0.015	0.012	0.009	0.003	0.027	0.001	0.005	0.007
	$m = 1000$	0.008	0.015	0.031	0.017	0.006	0.001	0.021	0.017
0.9	$m = 100$	0.02	0.07	0.10	0.18	0.21	0.24	0.26	0.31
	$m = 500$	0.001	0.001	0.001	0.001	0.01	0.01	0.01	0.01
	$m = 1000$	0.001	0.01	0.02	0.04	0.05	0.05	0.07	0.07

2.3 Апроксимація часового ряду функціоналом від фрактального броунівського руху

2.3.1 Метод апроксимації

Нехай x_1, \dots, x_n – спостережуваний часовий ряд з даними довільного походження, до того ж

$$\bar{x} = \frac{1}{n} \sum x_k = 0. \quad (2.39)$$

Побудуємо таке перетворення Φ , для якого

$$x_k = \Phi(\sigma B_H(\cdot))\left(\frac{k}{n}\right), \quad (2.40)$$

вибравши fBm за базовий процес моделювання. Рівність (2.40) визначає апроксимацію часового ряду фрактальним броунівським рухом.

На першому етапі апроксимації застосовується емпіричний метод перевірки гіпотези про нормальність моделі для приростів перетворених даних x_1, \dots, x_n , і в разі неприйняття гіпотези, апроксимацією часового ряду функцією від приростів fBm .

Нехай $y_k = x_k - x_{k-1}$, $k = 2, \dots, n$, $E x_k = 0$. Критерієм гаусовості може бути значення вибіркового «коефіцієнта ексцесу»:

$$d_n = \frac{\left(\frac{1}{n-1} \sum_{k=2}^n |y_k|\right)^2}{\frac{1}{n-1} \sum_{k=2}^n y_k^2}. \quad (2.41)$$

Якщо значення d_n значно відрізняється від $\frac{2}{\pi}$, то слід замінити часовий ряд y_1, \dots, y_{n-1} новою послідовністю z_1, \dots, z_{n-1} , побудова якої описана за наступною процедурою.

Загальна ідея апроксимації полягає у виборі такого одновимірного функціонального перетворення g кожного приросту y_k , де g – зростаюча непарна функція, тобто $z_k = g(y_k)$.

При виборі функції g необхідно дотриматися умови, що граничні значення деяких вибірових характеристик послідовності $\{z_k\}$ мають співпадати з аналогічними значеннями для гаусівської послідовності. Якщо за таку характеристику обрати d_n , то

$$\lim_{n \rightarrow \infty} \frac{\left(\frac{1}{n-1} \sum_{k=1}^{n-1} |g(y_k)| \right)^2}{\frac{1}{n-1} \sum_{k=1}^{n-1} (g(y_k))^2} = \frac{2}{\pi}, \quad (2.42)$$

де $z_k = g(y_k)$ – гаусівська випадкова величина.

Наведемо демонстрацію запропонованої схеми, обравши за g степеневу функцію.

Покладемо

$$z_k = \operatorname{sgn} y_k |y_k|^{\frac{1}{\lambda}} = \operatorname{sgn} y_k |z_k|^\lambda, \quad \lambda > 0. \quad (2.43)$$

Тоді

$$d_n = \frac{\left(\frac{1}{n-1} \sum_{k=1}^{n-1} |z_k|^\lambda \right)^2}{\frac{1}{n-1} \sum_{k=1}^{n-1} |z_k|^{2\lambda}}, \quad (2.44)$$

і границя $d = \lim_{n \rightarrow \infty} d_n$ дорівнює відношенню відповідних математичних сподівань.

$$\text{Так як } \xi \sim \mathfrak{N}(0; \sigma^2), \quad E|\xi|^\alpha = \frac{2^{\frac{\alpha}{2}}}{\sqrt{\pi}} \sigma^\alpha \Gamma\left(\frac{\alpha+1}{2}\right), \quad \text{то } d = \frac{1}{\sqrt{\pi}} \frac{\Gamma^2\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\lambda+\frac{1}{2}\right)}.$$

Прирівнявши d_n до правої частини отриманої рівності та вирішивши отримане рівняння, знайдемо відповідне λ (таке λ є єдиним, адже права частина рівності є спадною функцією від λ). Виконавши описані дії, запропонована апроксимація пропонує наступне представлення вихідного часового ряду:

$$x_k = \sum_{j=1}^k \text{sgn } y_j \cdot |z_j|^\lambda. \quad (2.45)$$

Припустивши стаціонарність (у вузькому сенсі) послідовності моделей для $\{y_k\}$, можна стверджувати про стаціонарність моделей $\{z_k\}$, тобто:

$$Ez_j z_k = Dz_j \cdot f\left(\frac{k-j}{n}\right), \quad (2.46)$$

де $f(s)$ – спадна функція, $f(0) = 1$.

Апроксимуючи f степеневою функцією $\frac{(s+\frac{1}{n})^{2H} + (s-\frac{1}{n})^{2H}}{2} - s^{2H}$, приходимо до апроксимації $\{x_k\}$ фрактальним броунівським рухом у вигляді моделі

$$z_k = \sigma^H \left(B_H\left(\frac{k}{n}\right) - B_H\left(\frac{k-1}{n}\right) \right), \quad x_k = \sigma^\lambda \sum_{j=1}^k \text{sgn } y_j \cdot \left| B_H\left(\frac{j}{n}\right) - B_H\left(\frac{j-1}{n}\right) \right|^\lambda, \quad (2.47)$$

що дозволяє провести дослідження отриманого часового ряду методами, розробленими для fBm.

2.3.2 Перевірка якості апроксимації моделі

Те, що розглянута процедура апроксимації реального часового ряду ґрунтується на значення єдиного параметра d_n , може взагалі не гарантувати гаусовості перетворених приростів z_k . Аби перевірити властивості цих приростів застосовуються граничні теореми для фрактального броунівського руху, які доведені в роботах [9-21].

Нехай $B(t)$, $0 \leq t \leq 1$ – фрактальний броунівський рух з показником Харста H , f – двічі диференційована функція, для якої $E(|f^{(k)}(B(t))|^p) < \infty$, $k = 1, 2$, p – натуральне.

Позначимо:

$$\xi_k = n^H \left(B\left(\frac{k+1}{n}\right) - B\left(\frac{k}{n}\right) \right) \sim \mathfrak{N}(0; 1). \quad (2.48)$$

У згаданих наукових роботах доведені наступні граничні теореми:

$$\frac{n^H}{n} \sum_{k=1}^n f\left(B\left(\frac{k}{n}\right)\right) \xi_k^3 \rightarrow -\frac{3}{2} \int_0^1 f'(B(s)) ds, \quad H \in \left(0; \frac{1}{2}\right). \quad (2.49)$$

$$\frac{n^{2H}}{n} \sum_{k=1}^n f\left(B\left(\frac{k}{n}\right)\right) (\xi_k^2 - 1) \rightarrow \frac{1}{4} \int_0^1 f''(B(s)) ds, \quad H \in \left(0; \frac{1}{4}\right); \quad (2.50)$$

$$\frac{1}{n^H} \sum_{k=1}^n f\left(B\left(\frac{k}{n}\right)\right) \xi_k^3 \rightarrow 3 \int_0^{B(1)} f(x) dx, \quad H \in \left(\frac{1}{2}; 1\right). \quad (2.51)$$

Збіжність у даному випадку розуміється в середньоквадратичному:

$$\eta_n \rightarrow \eta, \text{ якщо } E(\eta_n - \eta)^2 \rightarrow 0. \quad (2.52)$$

Позначивши:

$$\alpha_k = n^H B\left(\frac{k}{n}\right) = \sum_j^{k-1} \xi_j. \quad (2.53)$$

і взявши у першій формулі $f(x) = x$ та $f(x) = x^2$, прийдемо до результатів:

$$\frac{1}{n} \sum_{k=1}^n \alpha_k \xi_k^3 \rightarrow -\frac{3}{2}, \quad H \in \left(0; \frac{1}{2}\right), \quad (2.54)$$

$$\frac{1}{n^{H+1}} \sum_{k=1}^n \alpha_k^2 \xi_k^3 \rightarrow 3\eta, \quad H \in \left(0; \frac{1}{2}\right), \quad (2.55)$$

де $\eta \sim \mathcal{N}\left(0; \frac{1}{2H+2}\right)$.

Поклавши в другій формулі $f(x) = x^2$, дійдемо:

$$\frac{1}{n} \sum_{k=1}^n \alpha_k^2 (\xi_k^2 - 1) \rightarrow \frac{1}{2}, \quad H \in \left(0; \frac{1}{4}\right). \quad (2.56)$$

Взявши в третій формулі $f(x) = x$ та $f(x) = 1$, отримаємо:

$$\frac{1}{n^{2H}} \sum_{k=1}^n \alpha_k \xi_k^3 \rightarrow \frac{3}{2} B^2(1), \quad H \in \left(\frac{1}{2}; 1\right), \quad (2.57)$$

$$\frac{1}{n^H} \sum_{k=1}^n \xi_k^3 \rightarrow 3B(1), \quad H \in \left(\frac{1}{2}; 1\right). \quad (2.58)$$

Отримані співвідношення (2.54), (2.55), (2.56), (2.57), (2.58) дають можливість здійснити перевірку гіпотези $T = \{\text{статистики } z_1, \dots, z_n, \text{ отримані перетворенням (2.43) реальних даних, моделюються приростами фрактального броунівського руху}\}$.

Алгоритм описаної вище перевірки за відомого H наступний.

Позначивши:

$$c = \frac{1}{n} \sum_{k=1}^n z_k^2. \quad (2.59)$$

Припускаємо, що гіпотеза T виконана, тобто

$$z_k = \sqrt{c} \xi_k = \sqrt{c} n^H \left(B\left(\frac{k+1}{n}\right) - B\left(\frac{k}{n}\right) \right). \quad (2.60)$$

Покладемо

$$v_k = \sum_{j=1}^{k-1} z_j. \quad (2.61)$$

і побудуємо статистики:

$$\begin{aligned} A_n &= \frac{1}{n} \sum v_k z_k^3, & H \in \left(0; \frac{1}{2}\right); \\ B_n &= \frac{1}{n^{1+H}} \sum v_k^2 z_k^3, & H \in \left(0; \frac{1}{2}\right); \\ C_n &= \frac{1}{n} \sum v_k^2 \left(\frac{z_k^2}{c} - 1\right), & H \in \left(0; \frac{1}{4}\right); \\ D_n &= \frac{1}{n^{2H}} \sum v_k z_k^3, & H \in \left(\frac{1}{2}; 1\right); \end{aligned} \quad (2.62)$$

$$F_n = \frac{1}{n^H} \sum v_k z_k^3, \quad H \in \left(0; \frac{1}{2}\right).$$

Якщо гіпотеза Т справджується, то відповідно до (2.54), (2.55), (2.56), (2.57), (2.58) мають місце такі збіжності:

$$A_n \rightarrow -\frac{3}{2}c^2; \quad B_n \rightarrow 3c^{\frac{5}{2}}\eta; \quad C_n \rightarrow \frac{c}{2}; \quad D_n \rightarrow \frac{3}{2}c^2B^2(1); \quad F_n \rightarrow 3c^{\frac{3}{2}}B(1). \quad (2.63)$$

Аби прийняти рішення щодо гіпотези Т, потрібно порівняти реальні значення статистик з їхніми граничними теоретичними значеннями. Для статистики A_n визначимо відхилення від граничного значення таким способом:

$$\delta = \left| \frac{A_n - A}{A} \right|. \quad (2.64)$$

Граничні функції розподілів для статистик B_n та D_n :

$$F_1(x) = P\{3c^{2,5}\eta < x\} = \Phi\left(\frac{x}{3\sigma}c^{2,5}\right), \quad (2.65)$$

$$F_2(x) = 2\Phi\left(\frac{1}{c}\sqrt{\frac{2}{3}x}\right) - 1, x > 0, \quad (2.66)$$

де Φ – функція Лапласа;

$$\sigma = \frac{1}{\sqrt{2H+2}}.$$

Гіпотеза Т приймається, якщо

$$\delta < \beta_0, \quad |B_n| < \beta_1 \text{ для } H < 0,5, \quad 0 < D_n < \beta_2 \text{ для } H > 0,5, \quad (2.67)$$

де β_1 та β_2 – квантилі розподілів F_1 та F_2 , які відповідають обраному рівню значимості α (прийmemo $\alpha = 0,1$). Тоді

$$\beta_1 = \frac{4,95c^{2,5}}{\sqrt{2N+2}}, \quad \beta_2 = 4,08 c^2, \quad c = \overline{z^2}. \quad (2.68)$$

Для визначених раніше статистик A_n, B_n, D_n за формулами (2.62) швидкість збіжності до граничних значень лишається невідомою. Щоб її оцінити, розглянемо «ідеальний» випадок імітованого фрактального броунівського руху – тоді гіпотеза T апіорно виконана. Це дасть можливість надалі порівнювати поведінку цих статистик вже для перетвореного часового ряду, визначеного рівностями (2.43), (2.45).

Генеруючи значення фрактального броунівського руху за алгоритмом, описаним в раніше, і використовуючи позначення формул наступне, вирахуємо наступні статистики:

$$\tilde{A}_n = \frac{1}{n} \sum_{k=1}^{n-1} x_k z_{k+1}^3, \quad \tilde{B}_n = \frac{1}{n^{1+H}} \sum_{k=1}^{n-1} x_k^2 z_{k+1}^3, \quad \text{беручи } H = 0,2; 0,3; 0,4; \quad (2.69)$$

$$\tilde{C}_n = \frac{1}{n} \sum_{k=1}^{n-1} x_k^2 (z_k^2 - 1), \quad \text{беручи } H = 0,2; \quad (2.70)$$

$$\tilde{D}_n = \frac{1}{n^{2H}} \sum_{k=1}^{n-1} x_k z_{k+1}^3, \quad \text{беручи } H = 0,6; 0,7; 0,8. \quad (2.71)$$

Кожна зі статистик визначається за трьома вибірками об'ємів 256; 1024; 4096.

Також слід зауважити, що в статистиках A_n, B_n, C_n, D_n, z_k має місце функціональне перетворення приростів реального часового ряду, а в статистиках $\tilde{A}_n, \tilde{B}_n, \tilde{C}_n, \tilde{D}_n, z_k$ – нормовані прирости фрактального броунівського руху:

$$z_k = n^H \left(B \left(\frac{k}{n} \right) - B \left(\frac{k-1}{n} \right) \right). \quad (2.72)$$

тобто $c = \overline{z^2} = r_{2n} = 1$.

Значення контрольних статистик $\tilde{A}_n, \tilde{B}_n, \tilde{D}_n, \lim_{n \rightarrow \infty} \tilde{A}_n = A$ та квантилі β_1 та β_2 граничних для B_n та D_n розподілів наведені в таблиці 2.3. Для \tilde{B}_n такий розподіл є гаусівським:

$$P\{-\beta < 3\eta < \beta\} = \frac{2}{\sqrt{2\pi}} \int_0^{\frac{\beta}{3\sigma}} e^{-\frac{x^2}{2}} dx = 2\Phi\left(\frac{\beta}{3\sigma}\right) - 1 = \rho, \quad (2.73)$$

де $\sigma = \frac{1}{\sqrt{2H+2}}$;

Φ – функція Лапласа;

$H = 0,2; 0,3; 0,4$.

Взявши за рівень значимості $1 - \rho = 0,1$, отримаємо

$$\frac{\beta}{3\sigma} = 1,65, \quad (2.74)$$

звідки отримаємо значення квантиля:

$$\beta = 3,17; H = 0,2, \quad \beta = 3,07; H = 0,3, \quad \beta = 2,97; H = 0,4. \quad (2.75)$$

Окремо зазначимо, що гіпотеза T приймається, якщо $|\tilde{B}_n| < \beta$.

Граничний розподіл статистики \tilde{D}_n ($H > \frac{1}{2}$) не залежить від H та описується функцією розподілу

$$F(x) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{\frac{2}{3}x}} e^{-\frac{y^2}{2}} dy = 2\Phi\left(\sqrt{\frac{2}{3}x}\right) - 1, \quad x > 0. \quad (2.76)$$

Гіпотеза T приймається, якщо $0 < \tilde{D}_n < \beta$, де β – квантиля розподілу F , що відповідає рівню значимості $1 - \rho$. Якщо цей рівень вважати рівним $0,1$, то знайдемо β з рівняння

$$\sqrt{\frac{2}{3}\beta} = 1,65, \quad \beta = 4,08. \quad (2.77)$$

Результати обчислень представлені у таблиці 2.3.

Таблиця 2.3 – Значення контрольних статистик для генерованого fBm

H	n	\tilde{A}_n	\tilde{B}_n	\tilde{C}_n	\tilde{D}_n
0,2	256	-1,19	0,25	0,78	
	1024	-2,1	0,17	0,6	
	4096	-1,6	0,08	0,62	
0,3	256	-1,07	0,13		
	1024	-1,85	0,099		
	4096	-1,48	0,072		
0,4	256	-1,14	0,11		
	1024	-1,76	0,082		
	4096	-1,54	0,063		
0,6	256				0,21
	1024				0,13
	4096				0,09
0,7	256				0,14
	1024				0,072
	4096				0,044
0,8	256				0,14
	1024				0,019
	4096				0,031

За даними таблицями дійдемо висновку, що виконується гіпотеза T для всіх значень n для H. Дані дозволяють також визначити відносне відхилення $\delta = \left| \frac{A_n - A}{A} \right|$ статистики A_n щодо граничного значення $A = -1,5c^2$, що складає в даному «ідеальному» випадку 30 – 40% для $n \sim 300$ та 5 – 15% для $n \sim 1000 - 4000$.

Висновки до розділу

В даному розділі досліджено процес побудови фрактальної моделі, а саме: методи оцінювання невідомих параметрів, методику прогнозу. Розглянуто критерії якості даної моделі часового ряду, перевірено їхню швидкість збіжності.

РОЗДІЛ 3 ЗАСТОСУВАННЯ МАТЕМАТИЧНОЇ ФРАКТАЛЬНОЇ МОДЕЛІ

3.1 Опис даних

Приклад 1. Часовий ряд, що характеризує осциляції хвиль у Північній Атлантиці, і який охоплює часовий проміжок 10.1980 – 10.2014. Показники задані помісячно, загальний проміжок – 409 даних (409 місяців).

Приклад 2. Набір даних, який показує вміст вуглекислого газу в атмосфері в період з березня 1958 по червень 2016 з щомісячними вимірами (693 даних).

Візуальний аналіз графічного зображення даних, наведених на рис. 3.1, 3.2, дозволяє прийняти рішення про застосування до досліджуваних рядів моделі з необмеженою варіацією згідно з процедурою, описаною вище.

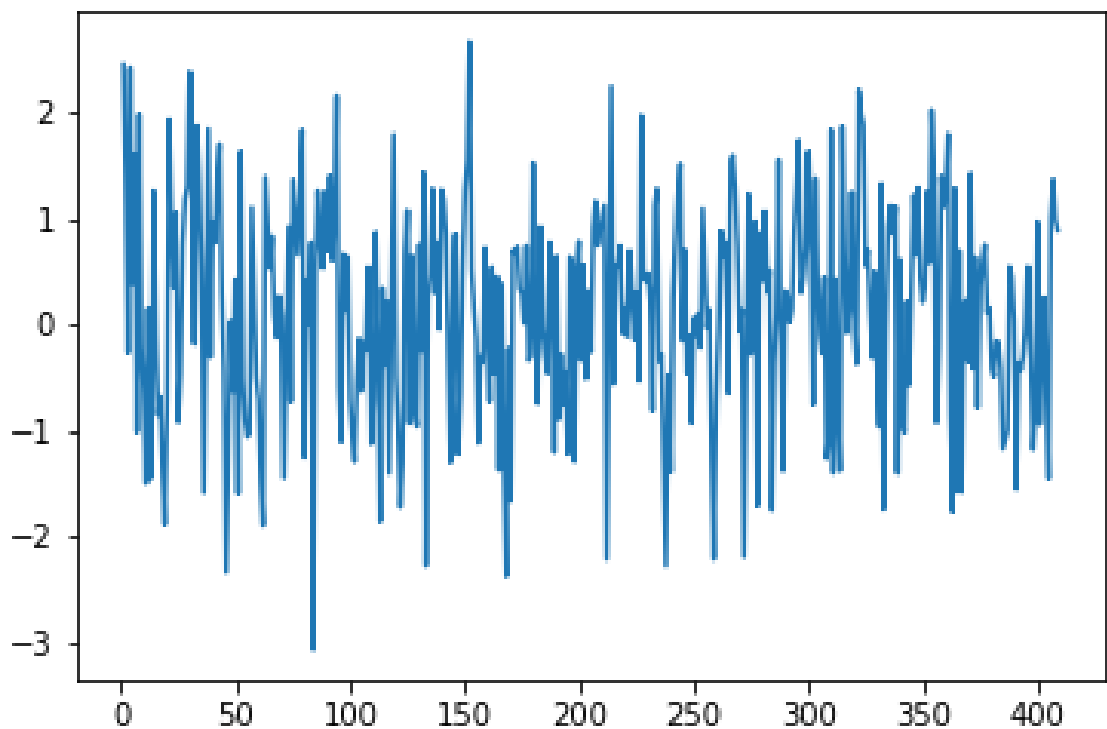


Рисунок 3.1 – Осциляція хвиль в Атлантиці

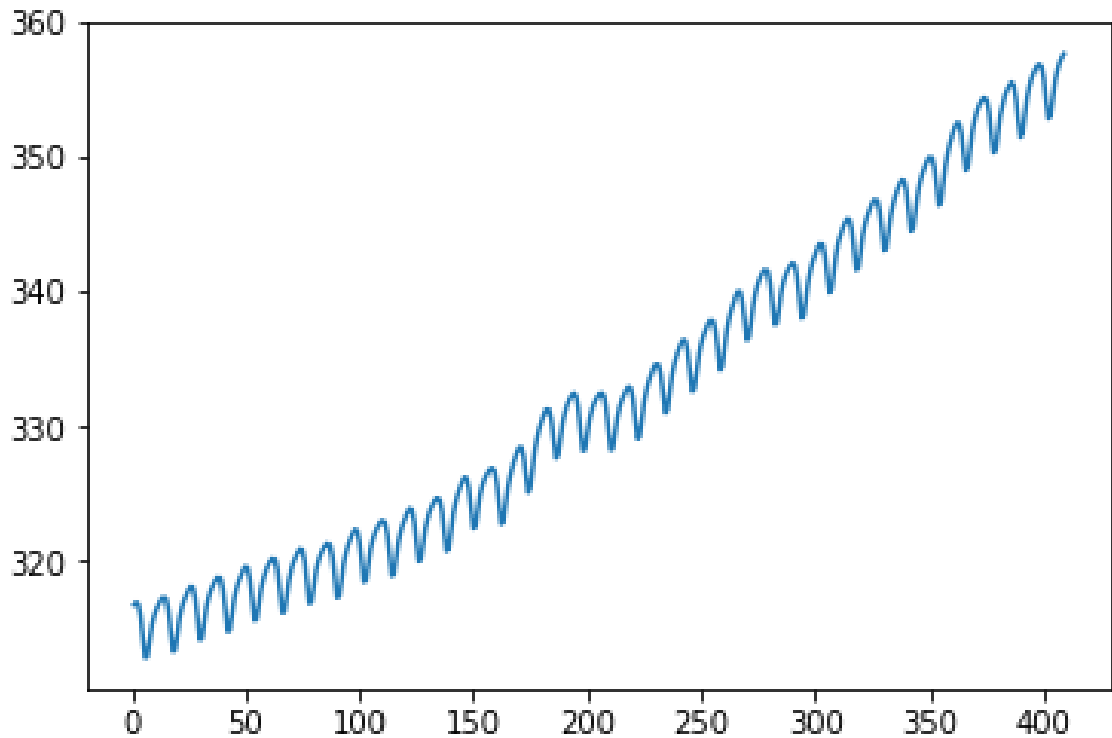


Рисунок 3.2 – Рівень вуглекислого газу в атмосфері

3.2 Апроксимація даних фрактальної моделлю

Спочатку усунемо лінійний тренд, віднявши від кожного значення їхнє середнє (математичне сподівання). На наступному етапі значення розбиваються на часові вікна, в кожному з яких будується лінійна апроксимація. Попереднє дослідження на стаціонарність приростів полягає в обчисленні вибіркового коефіцієнта кореляції для кожного часового вікна за формулою

$$\hat{\rho}_1 = \frac{\sum y_j y_{j+1}}{\sum y_j^2}.$$

Для першого прикладу $\hat{\rho}_1 = 0,44 - 0,49$.

Для першого прикладу $\hat{\rho}_1 = 0,73 - 0,77$.

Отже, прирости утворюють стаціонарну послідовність, так як значення $\hat{\rho}_1$ не залежать від номеру вікна.

Побудуємо модель для першого набору даних.

Обчислимо вибіркоче середнє:

$$\bar{x} = \frac{1}{409} \sum_{k=1}^{409} x_k \approx 0, \quad y_k = x_{k+1} - x_k.$$

Значення статистик:

$$R_{1n} = \frac{1}{409} \sum_{k=1}^{409} |y_k| = 1,03, \quad R_{2n} = \frac{1}{409} \sum_{k=1}^{409} y_k^2 = 1,72, \quad d_n = 0,62,$$

$$\text{звідки } \lambda = 1,04, \quad z_k = \text{sgn } y_k \cdot |y_k|^{0,96}, \quad k = 1, \dots, 408$$

Оцінювання параметра H за наведеним у попередньому розділі критерієм приводить нас до значення $\hat{H} = 0,1$ ($Q = 0,94$), і тоді $\hat{\sigma} = 158,57$.

Значення статистик: $A_n = -3,96$; $A = -4,03$; $B_n = 8,88 < \beta_1 = 11,24$.

Отже, моделювання є успішним.

$H < \frac{1}{2}$, а отже прогнозування неможливе.

Побудуємо модель для другого набору даних.

Обчислимо вибіркоче середнє:

$$\bar{x} = \frac{1}{682} \sum_{k=1}^{682} x_k \approx 0, \quad y_k = x_{k+1} - x_k.$$

Значення статистик:

$$R_{1n} = \frac{1}{682} \sum_{k=1}^{682} |y_k| = 0,79, \quad R_{2n} = \frac{1}{682} \sum_{k=1}^{682} y_k^2 = 0,87, \quad d_n = 0,71,$$

$$\text{звідки } \lambda = 0,81, \quad z_k = \text{sgn } y_k \cdot |y_k|^{1,23}, \quad k = 1, \dots, 681$$

Оцінювання параметра H за наведеним у попередньому розділі критерієм приводить нас до значення $\hat{H} = 0,8$ ($Q = 0,99$), і тоді $\hat{\sigma} = 181,91$.

Значення статистик: $D_n = 1,1 < \beta_2 = 4,08$. Отже, моделювання є успішним.

$H > \frac{1}{2}$, а отже прогнозування можливе.

Побудуємо прогноз на 4 кроки для об'ємів навчальних вибірок $m = 200; 400; 600$. У таблиці 3.1 наведено значення відносної похибки прогнозу

$$\delta_k = \frac{|\hat{s}_{m+k} - s_{m+k}|}{s_{m+k}}, \quad k = 1, 2, 3, 4.$$

Таблиця 3.1 – Значення похибки прогнозу

m	k=1	k=2	k=3	k=4
200	0,003	0,004	0,002	0,003
400	0,0006	0,002	0,005	0,005
600	0,001	0,006	0,01	0,01

Судячи з даних таблиці 3.1 можна робити висновок про задовільну якість прогнозування.

Висновки до розділу

Застосувавши досліджену в попередніх розділах модель для апроксимації реальних даних, було отримано задовільні результати математичного

модельовання. Підтверджено необхідність введення початкових обов'язкових необхідних умов і перевірено ефективність моделі на практиці.

РОЗДІЛ 4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

У даному розділі проводиться оцінка основних характеристик програмного продукту, призначеного для створення математичної моделі часового ряду на основі фрактального броунівського руху. Нижче наведено аналіз різних варіантів реалізації продукту з метою вибору оптимального, з огляду при цьому як на економічні фактори, так і на технічні характеристики продукту, що впливають на продуктивність роботи і на його сумісність з апаратним забезпеченням. Для цього було використано апарат функціонально-вартісного аналізу.

Функціонально-вартісний аналіз (ФВА) дозволяє оцінити реальну вартість продукту або послуги незалежно від організаційної структури компанії, розподіл прямих і побічних витрат по продуктам та послугам у залежності від потрібних на кожному етапі виробництва обсягів ресурсів. Дії, що виконуються на цих етапах, у контексті метода ФВА називаються функціями.

В основі ФВА лежить функціональний підхід, згідно з яким об'єктом аналізу є не сам виріб (програмний продукт, система), а функції, які він виконує. Функціонально-вартісний аналіз проводиться в два етапи: функціональний аналіз, вартісний аналіз.

Мета ФВА полягає у забезпеченні правильного розподілу ресурсів, виділених на виробництво продукції або надання послуг, на прями та непрямі витрати, а у даному випадку – аналізу функцій програмного продукту й виявлення усіх витрат на реалізацію цих функцій.

4.1 Постановка задачі техніко-економічного аналізу

У роботі застосовується метод ФВА для проведення техніко-економічного аналізу розробки.

Відповідно цьому варто обирати і систему показників якості програмного продукту.

Технічні вимоги до продукту наступні:

- програмний продукт повинен функціонувати на сучасних пристроях на базі актуальної версії ОС Windows або Unix-подібних ОС, які мають в наявності будь-який інтернет-браузер;
- забезпечувати високі показники висунутих у розділі 3 критеріїв;
- забезпечувати зручність і простоту взаємодії з користувачем програмного забезпечення;
- передбачати мінімальні витрати на впровадження програмного продукту.

4.1.1 Обґрунтування функцій програмного продукту

Головна функція F_0 – розробка програмного продукту, який аналізує процес за вхідними даними та будує його модель для подальшого прогнозування. Виходячи з конкретної мети, можна виділити наступні основні функції ПП:

F_1 – вибір мови програмування;

F_2 – вибір оптимальної бібліотеки для числових даних;

F_3 – джерело надходження даних часових рядів.

Кожна з основних функцій може мати декілька варіантів реалізації.

Функція F_1 :

а) мова програмування R;

б) мова програмування Python;

Функція F_2 :

а) dplyr;

б) purrr.

Функція F_3 :

а) зовнішні інтернет-ресурси;

б) вбудовані бібліотеки з даними, що постійно оновлюються.

4.1.2 Варіанти реалізації основних функцій

Варіанти реалізації основних функцій наведені у морфологічній карті системи (рисунок 4.1).

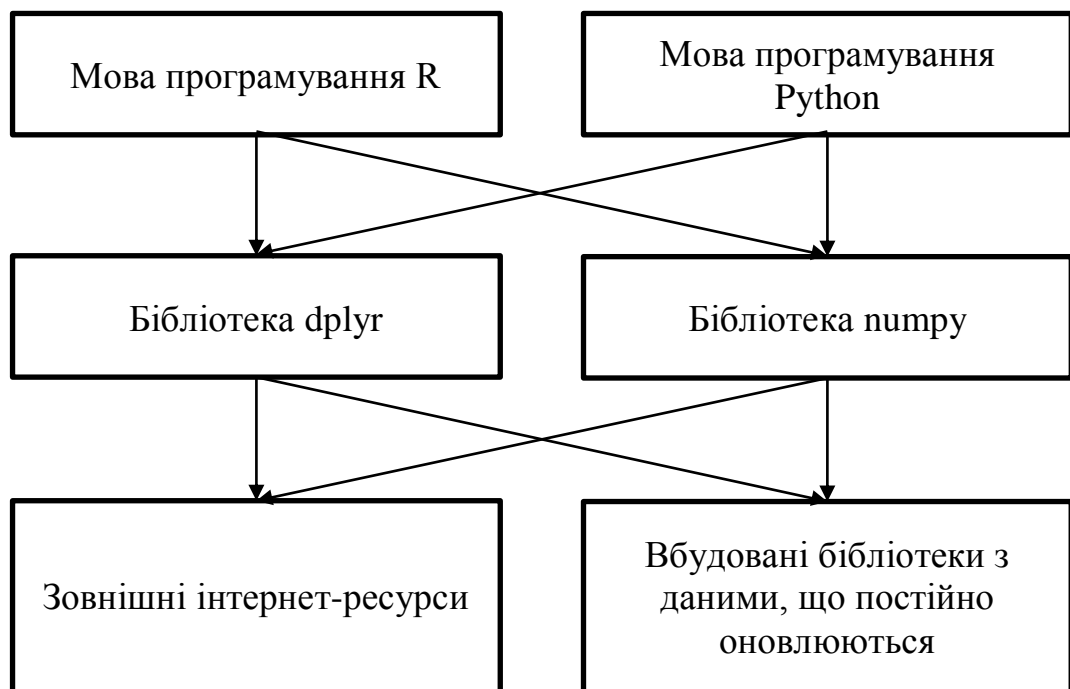


Рисунок 4.1 – Морфологічна карта

Морфологічна карта відображує всі можливі комбінації варіантів реалізації функцій, які складають повну множину варіантів ПП. На основі цієї карти

побудовано позитивно-негативну матрицю варіантів основних функцій (таблиця 4.1).

Таблиця 4.1 – Позитивно-негативна матриця

Основні функції	Варіанти реалізації	Переваги	Недоліки
F1	а	Багато функцій, необхідних для аналізу даних, є вбудованими функціями мови.	Низька продуктивність, порівняна складність розробки.
	б	Вища ефективність та нижча вартість розробки, адже використання можливе через веб-додаток.	Відсутність загального сховища та брак альтернатив.
F2	а	Підходить для великих об'ємів даних.	Необхідність в різноманітті функціоналу.
	б	Вища швидкість розробки, велика кількість готових алгоритмів, універсальність з різними типами даних.	Обмежений функціонал.
F3	а	Необмежена кількість даних, що є загальнодоступними і надіються спеціалізованими джерелами.	В багатьох випадках незручний формат даних.
	б	Універсальний формат даних, які є попередньо опрацьованими, легкодоступність.	Часто відсутнє різноманіття, велика кількість тривіальних даних.

На основі аналізу позитивно-негативної матриці робимо висновок, що при розробці програмного продукту деякі варіанти реалізації функцій варто відкинути, тому, що вони не відповідають поставленим перед програмним продуктом задачам. Ці варіанти відзначені у морфологічній карті.

Функція F1: оскільки обидві мови можуть бути використані для розробки і пропонують різні переваги, то слід розглянути обидва варіанти.

Функція F2: оскільки для даного продукту важливою можливістю опрацювати різні типи даних, відкидаємо варіант а).

Функція F3: оскільки для даного продукту важливим є доступ до різноманітних даних, використаємо варіант а) як єдиний можливий.

Таким чином, будемо розглядати такий варіант реалізації ПП:

F1a – F2б – F3а,

F1б – F2б – F3а.

Для оцінювання якості розглянутих функцій обрана система параметрів, описана нижче.

4.2 Обґрунтування системи параметрів ПП

4.2.1 Опис параметрів

На підставі даних про основні функції, що повинен реалізувати програмний продукт, вимог до нього, визначаються основні параметри виробу, що будуть використані для розрахунку коефіцієнта технічного рівня.

Для того, щоб охарактеризувати програмний продукт, будемо використовувати наступні параметри:

X1 – швидкодія мови програмування;

X2 – об'єм пам'яті для коректної роботи програми;

X3 – час опрацювання даних;

X4 – потенційний об'єм програмного коду.

X1 відображає швидкодію операцій залежно від обраної мови програмування.

X2 відображає необхідний для збереження та обробки даних об'єм оперативної пам'яті пристрою.

X3 відображає час, який витрачається на обробку вхідних даних та побудову математичної моделі, оцінивши якість моделювання.

X4 показує обсяг програмного коду, який необхідно створити розробнику.

4.2.2 Кількісна оцінка параметрів

Головна функція F_0 – розробка програмного продукту, який, приймаючи значень певного часового ряду має побудувати математичну модель, на основі якої буде здійснено прогнозування. На підставі даних про основні функції, що повинен реалізувати програмний продукт, вимог до нього, визначаються основні параметри виробу, що будуть використані для розрахунку коефіцієнта технічного рівня. На основі аналізу позитивно-негативної матриці робимо висновок, що при розробці програмного продукту деякі варіанти реалізації функцій варто відкинути, тому, що вони не відповідають поставленим перед програмним продуктом задачам. Гірші, середні і кращі значення параметрів вибираються на основі вимог замовника й умов, що характеризують експлуатацію ПП як показано у табл. 4.2.

Таблиця 4.2 – Основні параметри ПП

Назва Параметра	Умовні позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія мови програмування	X1	Оп/мс	1000	1500	2100
Необхідний об'єм пам'яті для стабільної роботи	X2	Мб	128	64	16
Час виконання алгоритму програмою	X3	мс	3000	2000	1200
Об'єм програмного коду	X4	кількість рядків коду	1000	700	500

За даними таблиці 4.2 будуються графічні характеристики параметрів, що продемонстровані на рисунках 4.2-4.5.

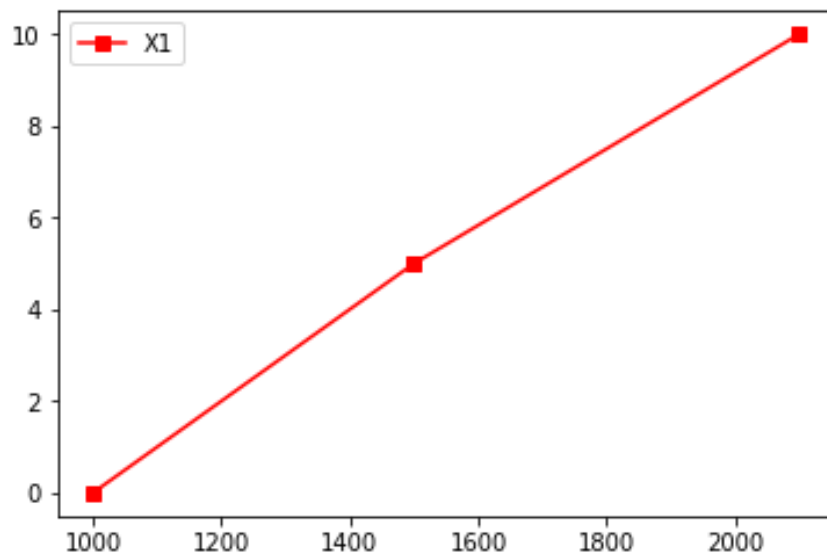


Рисунок 4.2 X1, швидкодія мови програмування

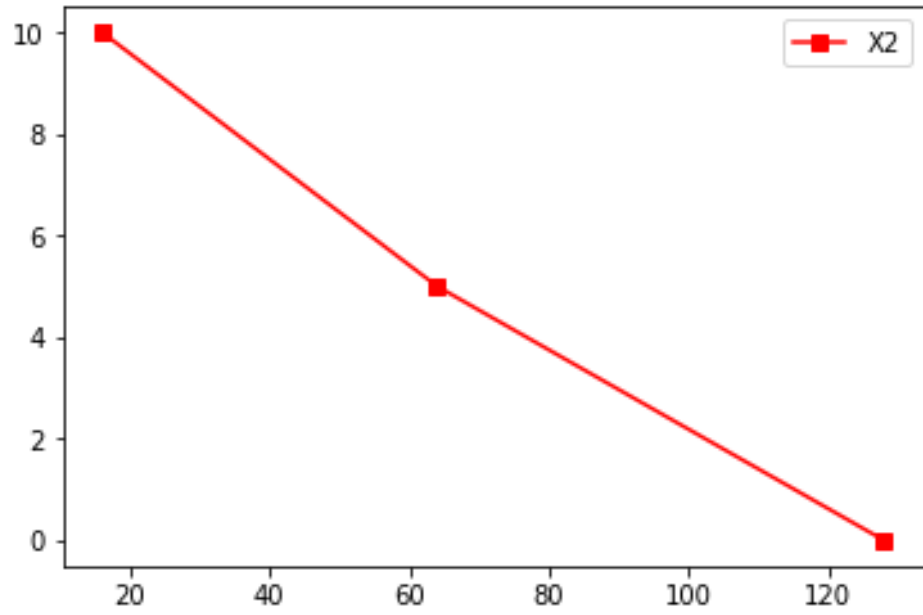


Рисунок 4.3 X2, об'єм пам'яті для коректної роботи

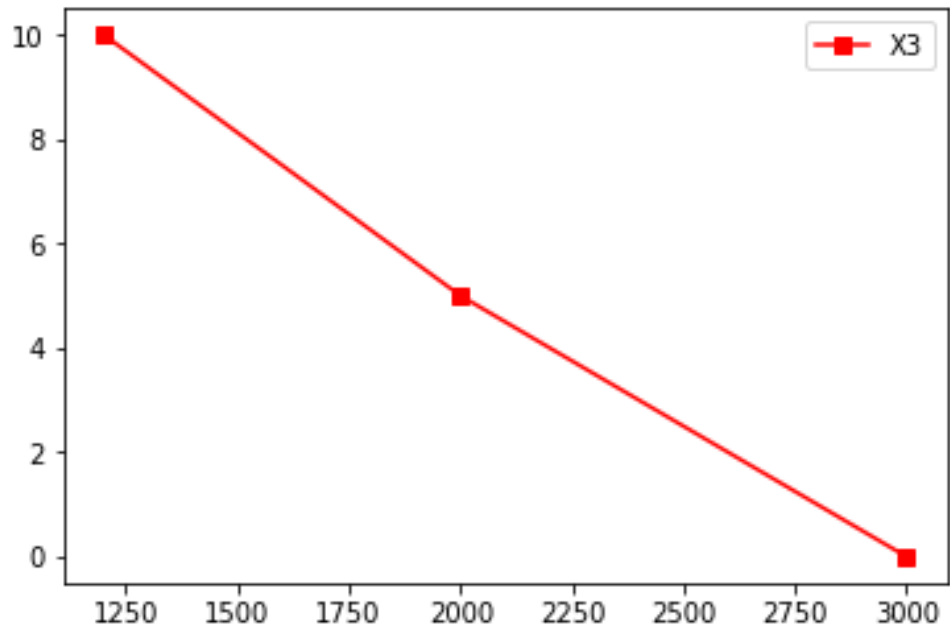


Рисунок 4.4 X3, час виконання алгоритму

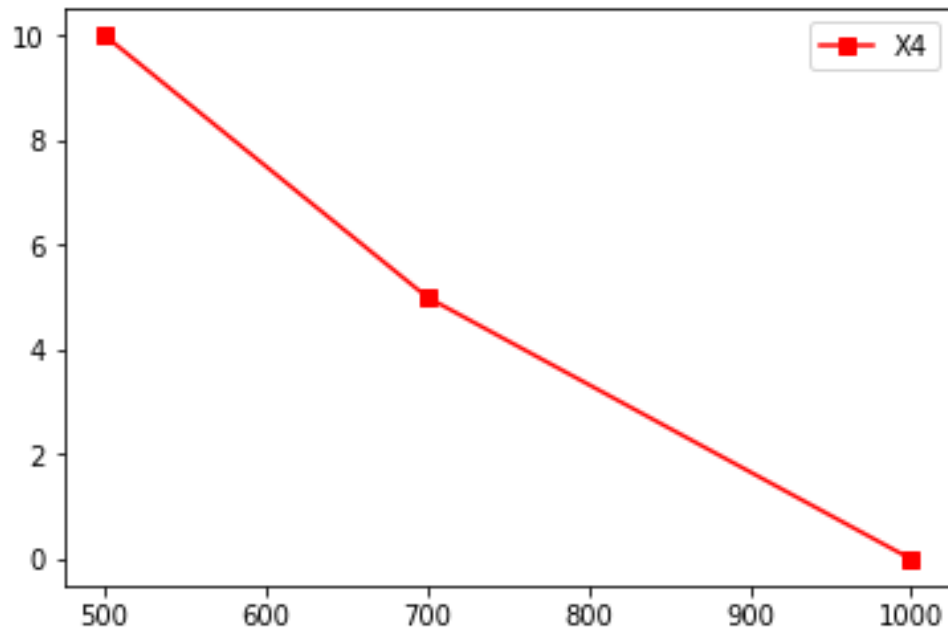


Рисунок 4.5 X4, потенційний об'єм програмного коду

4.2.3 Аналіз експертного оцінювання параметрів

Після детального обговорення й аналізу кожний експерт оцінює ступінь важливості кожного параметру для конкретно поставленої цілі – розробка програмного продукту, який дає найбільш точні результати при знаходженні параметрів моделей адаптивного прогнозування і обчислення прогнозних значень.

Значимість кожного параметра визначається методом попарного порівняння. Оцінку проводить експертна комісія із 7 людей. Визначення коефіцієнтів значимості передбачає:

- визначення рівня значимості параметра шляхом присвоєння різних рангів;
- перевірку придатності експертних оцінок для подальшого використання;
- визначення оцінки попарного пріоритету параметрів;
- обробку результатів та визначення коефіцієнту значимості.

Результати експертного ранжування наведені у таблиці 4.3.

Таблиця 4.3 – Результати ранжування параметрів

Параметр	Назва параметра	Одиниці виміру	Ранг параметра за оцінкою експерта							Сума рангів R _i	Відхилення Δ _i	Δ _i ²
			1	2	3	4	5	6	7			
X1	Швидкодія мови програмування	Оп/мс	1	1	2	1	2	1	1	9	-8,5	72,25
X2	Об'єм пам'яті для коректної роботи	Мб	3	2	3	4	4	3	2	21	-3,5	12,25
X3	Час виконання	мс	2	3	1	2	1	2	3	14	-3,5	12,25
X4	Потенційний об'єм програмного коду	кількість рядків коду	4	4	4	3	3	4	4	26	9,5	90,25
	Разом		10	10	10	10	10	10	10	70	0	201

Для перевірки ступеня достовірності експертних оцінок, визначимо наступні параметри:

а) сума рангів кожного з параметрів і загальна сума рангів:

$$R_i = \sum_{i=1}^N r_{ij} R_{ij} = \frac{Nn(n+1)}{2} = 70, \quad (4.1)$$

де N – число експертів;

n – кількість параметрів;

б) середня сума рангів:

$$T = \frac{R_{ij}}{n} = 17,5, \quad (4.2)$$

в) відхилення суми рангів кожного параметра від середньої суми рангів:

$$\Delta_i = R_i - T, \quad (4.3)$$

Сума відхилень по всіх параметрах повинна дорівнювати 0;

г) загальна сума квадратів відхилення:

$$S = \sum_{i=1}^N \Delta_i^2 = 187. \quad (4.4)$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{12S}{N^2(n^3 - n)} = \frac{12 \cdot 187}{7^2(4^3 - 4)} = 0,76 > W_k = 0,67. \quad (4.5)$$

Ранжування можна вважати достовірним, тому що знайдений коефіцієнт узгодженості перевищує нормативний, який дорівнює 0,67.

Скориставшись результатами ранжування, проведемо попарне порівняння всіх параметрів і результати занесемо у таблицю 4.4.

Таблиця 4.4 – Попарне порівняння параметрів

Параметри	Експерти							Кінцева оцінка	Числове значення
	1	2	3	4	5	6	7		
X1 і X2	<	<	<	<	<	<	<	<	1,5
X1 і X3	<	<	>	<	>	<	<	<	1,5
X1 і X4	<	<	<	<	<	<	<	<	1,5
X2 і X3	>	<	>	>	>	>	<	>	0,5
X2 і X4	<	<	<	>	>	<	<	<	1,5
X3 і X4	<	<	<	<	<	<	<	<	1,5

Числове значення, що визначає ступінь переваги і-го параметра над j-тим, a_{ij} визначається по формулі:

$$a_{ij} = \begin{cases} 1,5 \text{ при } X_i > X_j, \\ 1,0 \text{ при } X_i = X_j, \\ 0,5 \text{ при } X_i < X_j. \end{cases} \quad (4.6)$$

З отриманих числових оцінок переваги складемо матрицю: $A = [a_{ij}]$.

Для кожного параметра зробимо розрахунок вагомості K_{bi} за наступними формулами:

$$K_{bi} = \frac{b_i}{\sum_{i=1}^n b_i}, \text{ де } b_i = \sum_{j=1}^n a_{ij}, \quad (4.7)$$

Відносні оцінки розраховуються декілька разів доти, поки наступні значення не будуть незначно відрізнятися від попередніх (менше 2%). На другому і наступних кроках відносні оцінки розраховуються за наступними формулами:

$$K_{Bi} = \frac{b'_i}{\sum_{i=1}^n b'_i}, \text{ де } b'_i = \sum_{i=1}^N a_{ij} b_j, \quad (4.8)$$

Як видно з таблиці 4.5, в даному випадку знадобилося 3 ітерації.

Таблиця 4.5 – Розрахунок вагомості параметрів

X _i	X _j				Перша ітерація		Друга ітерація		Третя ітерація	
	X ₁	X ₂	X ₃	X ₄	b _i	K _{Bi}	b _i ¹	K _{Bi} ¹	b _i ²	K _{Bi} ²
X ₁	1,0	0,5	0,5	0,5	2,5	0,156	9,25	0,155	34,375	0,158
X ₂	1,5	1,0	1,5	0,5	4	0,25	16,5	0,278	59,75	0,274
X ₃	1,5	0,5	1,0	0,5	4	0,25	12,5	0,21	45,25	0,208
X ₄	1,5	1,5	1,5	1,0	5,5	0,344	21,25	0,357	78,625	0,361
Всього:					16	1	59,5	1	218	1

4.3 Аналіз рівня якості варіантів реалізації функцій

Визначаємо рівень якості кожного варіанту виконання основних функцій окремо.

Коефіцієнт технічного рівня для кожного варіанта реалізації ПП розраховується так (таблиця 4.6):

$$K_K(j) = \sum_{i=1}^n K_{Bi,j} B_{i,j}, \quad (4.9)$$

де n – кількість параметрів;

$K_{Bi,j}$ – коефіцієнт вагомості i -го параметра;

B_i – оцінка i -го параметра в балах.

Розрахунок показників рівня якості варіантів реалізації основних функцій ПП наведено у таблиці 4.6.

Таблиця 4.6 – Розрахунок показників рівня якості варіантів реалізації основних функцій ПП

Основні функції	Варіант реалізації функції	Параметри	Абсолютне значення параметра	Бальна оцінка параметра	Коефіцієнт вагомості параметра	Коефіцієнт рівня якості
F1	А	X1	1400	4	0,158	0,632
		X2	100	4	0,274	1,096
	Б	X1	1800	7,5	0,158	1,185
		X2	40	8	0,274	2,192
F2	Б	X3	1800	6	0,208	1,248
F3	А	X4	700	5	0,361	1,805

За даними з таблиці 4.6 за формулою:

$$K_K = K_{TY}[F_{1k}] + K_{TY}[F_{2k}] + \dots + K_{TY}[F_{zk}] \quad (4.10)$$

визначаємо рівень якості кожного з варіантів:

$$K_{K1} = 0,632 + 1,096 + 1,248 + 1,805 = 4,781, \quad (4.11)$$

$$K_{K2} = 1,185 + 2,192 + 1,248 + 1,805 = 6,43. \quad (4.12)$$

Як видно з розрахунків, кращим є другий варіант, для якого коефіцієнт технічного рівня має найбільше значення.

4.4 Економічний аналіз варіантів розробки ПП

Для визначення вартості розробки ПП спочатку проведемо розрахунок трудомісткості.

Всі варіанти включають в себе два окремих завдання:

1. Розробка проекту програмного продукту;
2. Розробка програмної оболонки;

Завдання 1 за ступенем новизни відноситься до групи А, завдання 2 – до групи Б. За складністю алгоритми, які використовуються в завданні 1 належать до групи 1; а в завданні 2 – до групи 3.

Для реалізації завдання 1 використовується довідкова інформація, а завдання 2 використовує інформацію у вигляді даних.

Проведемо розрахунок норм часу на розробку та програмування для кожного з завдань.

Проведемо розрахунок норм часу на розробку та програмування для кожного з завдань. Загальна трудомісткість обчислюється як

$$T_0 = T_P \cdot K_{\Pi} \cdot K_{СК} \cdot K_M \cdot K_{СТ} \cdot K_{СТ.М}, \quad (4.13)$$

де T_P – трудомісткість розробки ПП;

K_{Π} – поправочний коефіцієнт;

$K_{СК}$ – коефіцієнт на складність вхідної інформації;

K_M – коефіцієнт рівня мови програмування;

$K_{СТ}$ – коефіцієнт використання стандартних модулів і прикладних програм;

$K_{СТ.М}$ – коефіцієнт стандартного математичного забезпечення.

Для першого завдання, виходячи із норм часу для завдань розрахункового характеру ступеню новизни А та групи складності алгоритму 1, трудомісткість дорівнює: $T_P = 90$ людино-днів. Поправочний коефіцієнт, який враховує вид

нормативно-довідкової інформації для першого завдання: $K_{\Pi} = 1.6$. Поправочний коефіцієнт, який враховує складність контролю вхідної та вихідної інформації для всіх семи завдань рівний 1: $K_{СК} = 1$. Оскільки при розробці першого завдання використовуються стандартні модулі, врахуємо це за допомогою коефіцієнта $K_{СТ} = 0.8$. Тоді, за формулою (4.13), загальна трудомісткість програмування першого завдання дорівнює:

$$T_1 = 90 \cdot 1.6 \cdot 0.8 = 112,2 \text{ людино} - \text{днів.} \quad (4.14)$$

Проведемо аналогічні розрахунки для подальших завдань.

Для другого завдання (використовується алгоритм третьої групи складності, степінь новизни Б), тобто $T_p = 28$ людино-днів, $K_{\Pi} = 0.7$, $K_{СК} = 1$, $K_{СТ} = 0.8$:

$$T_2 = 28 \cdot 0.7 \cdot 0.8 = 15,68 \text{ людино} - \text{днів.} \quad (4.15)$$

Складаємо трудомісткість відповідних завдань для кожного з обраних варіантів реалізації програми, щоб отримати їх трудомісткість:

$$T_I = (112,2 + 15,68 + 4,8 + 15,68) \cdot 8 = 1186,88 \text{ людино} - \text{годин,} \quad (4.16)$$

$$T_{II} = (112,2 + 15,68 + 6,91 + 15,68) \cdot 8 = 1203,76 \text{ людино} - \text{годин.} \quad (4.17)$$

Найбільш високу трудомісткість має варіант II.

В розробці бере участь один інженер з моделювання складних систем з окладом 24000 грн. Визначимо зарплату за годину за формулою:

$$C_{ч} = \frac{M}{T_m \cdot t} \text{ грн.,} \quad (4.18)$$

де M – місячний оклад працівників;

T_m – кількість робочих днів тиждень;

t – кількість робочих годин в день.

$$C_q = \frac{24000}{1 * 21 * 8} = 142,88 \text{ грн.} \quad (4.19)$$

Тоді, розрахуємо заробітну плату за формулою:

$$C_{зп} = C_q \cdot T_i \cdot K_d, \quad (4.20)$$

де C_q – величина погодинної оплати праці робітника;

T_i – трудомісткість відповідного завдання;

K_d – норматив, який враховує додаткову заробітну плату.

Зарплата розробників становить:

$$C_{зп1} = 142,88 \cdot 1186,88 \cdot 1,2 = 203497,697 \text{ грн.}, \quad (4.21)$$

$$C_{зп2} = 142,88 \cdot 1203,76 \cdot 1,2 = 206391,874 \text{ грн.} \quad (4.22)$$

Відрахування на соціальний внесок становить 22,0%:

$$C_{від1} = C_{зп1} \cdot 0,22 = 203497,697 \cdot 0,22 = 44769,49 \text{ грн.}, \quad (4.23)$$

$$C_{від2} = C_{зп2} \cdot 0,22 = 206391,874 \cdot 0,22 = 45406,21 \text{ грн.} \quad (4.24)$$

Тепер визначимо витрати на оплату однієї машино-години (C_M). Так як одна ЕОМ обслуговує одного інженера з окладом 24000 грн., з коефіцієнтом зайнятості 0,3 то для однієї машини отримаємо:

$$C_{\Gamma} = 12 \cdot M \cdot K_3 = 12 \cdot 24000 \cdot 0,3 = 86400 \text{ грн.} \quad (4.25)$$

З урахуванням додаткової заробітної плати:

$$C_{3П} = C_{\Gamma} \cdot (1 + K_3) = 86400 \cdot (1 + 0,3) = 112320 \text{ грн.} \quad (4.26)$$

Відрахування на соціальний внесок:

$$C_{\text{ВІД}} = C_{3П} \cdot 0,22 = 51840 \cdot 0,22 = 24710,4 \text{ грн.} \quad (4.27)$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 30000 грн.:

$$C_A = K_{\text{ТМ}} \cdot K_A \cdot C_{\text{ПР}} = 1,15 \cdot 0,25 \cdot 30000 = 8625 \text{ грн.,} \quad (4.28)$$

де $K_{\text{ТМ}}$ – коефіцієнт, який враховує витрати на транспортування та монтаж приладу у користувача;

K_A – річна норма амортизації;

$C_{\text{ПР}}$ – договірна ціна приладу.

Витрати на ремонт та профілактику розраховуємо як:

$$C_{\text{Р}} = K_{\text{ТМ}} \cdot C_{\text{ПР}} \cdot K_{\text{Р}} = 1,15 \cdot 30000 \cdot 0,05 = 1725 \text{ грн.,} \quad (4.29)$$

де $K_{\text{Р}}$ – відсоток витрат на поточні ремонти.

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$\begin{aligned} T_{\text{ЕФ}} &= (D_{\text{К}} - D_{\text{В}} - D_{\text{С}} - D_{\text{Р}}) \cdot t_3 \cdot K_{\text{В}} = \\ &= (365 - 104 - 8 - 16) \cdot 8 \cdot 0,9 = 1706 \text{ годин,} \end{aligned} \quad (4.30)$$

де $D_{\text{К}}$ – календарна кількість днів у році;

D_B, D_C – відповідно кількість вихідних та святкових днів;

D_P – кількість днів планових ремонтів устаткування;

t – кількість робочих годин в день;

K_B – коефіцієнт використання приладу у часі протягом зміни.

Витрати на оплату електроенергії розраховуємо за формулою:

$$\begin{aligned} C_{\text{ЕЛ}} &= T_{\text{ЕФ}} \cdot N_C \cdot K_3 \cdot C_{\text{ЕН}} = \\ &= 1706,4 \cdot 0,22 \cdot 0,78 \cdot 2,7515 = 805,69 \text{ грн.} \end{aligned} \quad (4.31)$$

де N_C – середньо-споживча потужність приладу;

K_3 – коефіцієнтом зайнятості приладу;

$C_{\text{ЕН}}$ – тариф за 1 КВт-годин електроенергії.

Накладні витрати розраховуємо за формулою:

$$C_H = C_{\text{ПР}} \cdot 0,67 = 30000 \cdot 0,67 = 20100 \text{ грн.} \quad (4.32)$$

Тоді, річні експлуатаційні витрати будуть:

$$\begin{aligned} C_{\text{ЕКС}} &= C_{\text{ЗП}} + C_{\text{ВІД}} + C_A + C_P + C_{\text{ЕЛ}} + C_H = \\ &= 112320 + 24710,4 + 8625 + 1725 + 805,69 + 20100 = \\ &= 168285,79 \text{ грн.} \end{aligned} \quad (4.33)$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$C_{\text{М-Г}} = C_{\text{ЕКС}} / T_{\text{ЕФ}} = 168285,79 / 1706,4 = 98,62 \text{ грн/год.} \quad (4.34)$$

Оскільки в даному випадку всі роботи, які пов'язані з розробкою програмного продукту ведуться на ЕОМ, витрати на оплату машинного часу складають:

$$C_M = C_{M-\Gamma} \cdot T, \quad (4.35)$$

$$C_{M1} = 98,62 \cdot 1186,88 = 117050,88 \text{ грн.}, \quad (4.36)$$

$$C_{M2} = 98,62 \cdot 1203,76 = 118714,81 \text{ грн.} \quad (4.37)$$

Накладні витрати складають 67% від заробітної плати:

$$C_H = C_{ЗП} \cdot 0,67, \quad (4.38)$$

$$C_{H1} = 117050,88 \cdot 0,67 = 78424,09 \text{ грн.}, \quad (4.39)$$

$$C_{H2} = 118714,81 \cdot 0,67 = 79538,92 \text{ грн.} \quad (4.40)$$

Отже, вартість розробки ПП становить:

$$C_{ПП} = C_{ЗП} + C_{ВІД} + C_M + C_H, \quad (4.41)$$

$$\begin{aligned} C_{ПП1} &= 203497,7 + 44769,49 + 117050,88 + 78424,09 = \\ &= 443742,16 \text{ грн.}, \end{aligned} \quad (4.42)$$

$$\begin{aligned} C_{ПП2} &= 206391,87 + 45406,21 + 118714,81 + 79538,92 = \\ &= 450051,81 \text{ грн.} \end{aligned} \quad (4.43)$$

4.5 Вибір кращого варіанта ПП техніко-економічного рівня

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$K_{\text{TEP}j} = K_{\text{K}j} / C\Phi_j, \quad (4.44)$$

$$K_{\text{TEP}1} = \frac{4,781}{443742,16} = 1,08 \cdot 10^{-5}, \quad (4.45)$$

$$K_{\text{TEP}2} = \frac{6,43}{450051,81} = 1,43 \cdot 10^{-5}. \quad (4.46)$$

Як бачимо, найбільш ефективним є другий варіант реалізації програми з коефіцієнтом техніко-економічного рівня $K_{\text{TEP}2} = 1,43 \cdot 10^{-5}$.

4.6 Висновки до розділу

В даному розділі проведено повний функціонально-вартісний аналіз ПП, який було розроблено в рамках дипломного проекту. Процес аналізу можна умовно розділити на дві частини.

В першій з них проведено дослідження ПП з технічної точки зору: було визначено основні функції ПП та сформовано множину варіантів їх реалізації; на основі обчислених значень параметрів, а також експертних оцінок їх важливості було обчислено коефіцієнт технічного рівня, який і дав змогу визначити оптимальну з технічної точки зору альтернативу реалізації функцій ПП.

Другу частину ФВА присвячено вибору із альтернативних варіантів реалізації найбільш економічно обґрунтованого. Порівняння запропонованих варіантів реалізації в рамках даної частини виконувалось за коефіцієнтом ефективності, для обчислення якого були обчислені такі допоміжні параметри, як трудомісткість, витрати на заробітну плату, накладні витрати.

Після виконання функціонально-вартісного аналізу програмного комплексу що розроблюється, можна зробити висновок, що з альтернатив, що залишилися після першого відбору двох варіантів виконання програмного комплексу оптимальним є другий варіант реалізації програмного продукту. У нього виявився найкращий показник техніко-економічного рівня якості $K_{TEP2} = 1,43 \cdot 10^{-5}$.

Цей варіант реалізації програмного продукту має такі параметри:

мова програмування – python;

бібліотека – numpy;

використання даних, наданих інтернет-ресурсами.

Даний варіант виконання програмного комплексу дає користувачу можливість створювати математичні моделі часових рядів з достатньою точністю та швидкістю та для великого різноманіття вхідних даних.

ВИСНОВКИ

Дана робота присвячена аналізу, побудові та використанню математичної моделі часового ряду, в основі якої лежить процес фрактального броунівського руху.

Після ознайомлення з теоретичним матеріалом щодо понять випадкових процесів та математичних моделей, етапами розвитку методів моделювання в певній предметній області було запропоновано та досліджено модель часового ряду. За результатами дослідження можна постановити, що дана фрактальна модель має непогану прогнозовану здатність, зберігає інтерпретованість, подібно до класичних моделей.

В процесі дослідження було створено програмний додаток, який дозволяє побудувати модель та здійснити прогноз на нових даних. Перевірка додатку показала, що запропонована модель на основі фрактального броунівського руху є конкурентоспроможною по обчислювальній потужності.

Рекомендації щодо подальших досліджень – порівняння фрактальної моделі з іншими моделями (як класичними, так і вузькоспеціалізованими), що використовуються на практиці.

ПЕРЕЛІК ПОСИЛАНЬ

1. Ширяев А. Н. Основы стохастической финансовой математики: в 2 т. Москва: Фазис, 1998. Т. 1.: Факты и модели. 512 с.
2. H. Hurst. Long term storage capacity of reservation. *Transactions of the American Society of Civil Engineers*. 1951. Vol. 116, P. 770–799.
3. W. Feller. The asymptotic distribution of the range of sums of independent random variables. *Ann. Math. Stat.* 1951. Vol. 22, No. 3. P. 427–432.
4. Mandelbrot B. B. The Fractional Brownian motion, fractional noise sand. *SIAM Review*. 1968. Vol. 10, No. 4. P. 422–437.
5. B. Mandelbrot. The Fractal Geometry of Nature. *Freeman and Co.* San-Francisco, 1982. Vol. 89, 460 p.
6. Mandelbrot B. B. The Misbehavior of Market. *A Fractal View of Risk, Ruin, and Reward*. New York: Basic Books, 2004. 368 p.
7. Peltier R. F. A new method for estimating the parameter of fractional Brownian motion. *Rapport de recherché de l'INRIA*. 1994. P. 27.
8. Cheridito P. Stochastic integral of divergence type with respect to fractional Brownian motion with Hurst parameter $H \in (0, 1/2)$. *Ann. Inst. Henri Poincaré Probab. Stat.* 41, 2005. P. 1049–1081.
9. J. M. Corcuera, Woerner. Power variation of some integral long-memory processe. *Bernoulli* 12, 2006. P. 713–735.
10. J. Leónand. Limits for weighted p-variation sand likewise functional soff ractional diffusions with drift. *Stochastic Proc. Appl.* 117, 2006. P. 271–296.
11. A. Neuenkirch. Exactrate of convergence of some approximations chemes associated to SDE sdriven by a fractional Brownian. *J. Theoret. Probab.* 2007. No. 20. P. 871–899.
12. I. Nourdin. Weighted power variations of iterated Brownian motion. *Electron. J. Probab.* 2007. 13. P. 1229–1256.

13. I. Nourdin. Asymptotic behavior of weighted quadratic variations of fractional Brownian motion: The critical case $H=1/4$. *Ann. Probab.* 2008. 210 p.
14. Breton J-C. Error bounds on the non-normal approximation of Hermite power variations of fractional Brownian motion. *Electron. Comm. Probab.*, 2008. No. 13. P. 482–493.
15. Nourdin I. Asymptotic behavior of weighted quadratic and cubic variations of fractional Brownian motion. *Ann. Probab.*, 2008. Vol. 36, No. 6. P. 2159–2175.
16. I. Nourdin. Noncentral convergence of multiple integrals. *Ann. Probab.*, 2009. Vol. 37, No. 4. P. 1412–1426.
17. I. Nourdin. Density formula and concentration inequalities with Malliavin. *Electron. J. Probab.*, 2009. No. 14. P. 2287–2300.
18. Gradinaru M. Milstein's type schemes for fractional SDEs. *Ann. Inst H. Poincaré Probab Statist*, 2009. Vol. 45, No. 4. P. 1085–1098.
19. Breton J-C. Exact confidence intervals for the Hurst parameter of a fractional Brownian motion. *Electron. J. Statist.*, 2009. Vol. 3, P. 416–425.
20. Nourdin I. Central and non-central limit theorems for weighted power variations of fractional Brownian motion. *Ann. Inst H. Poincaré Probab Statist.* 2010. Vol. 46, No. 4. P. 1055–1079.
21. Nourdin I. Selected Aspects of fractional Brownian motion. Milano: Springer, 2012. 124 p.
22. S. Taqqu. Estimators for long-range dependence: An empirical study. *Fractals*, 1995. No. 3. P. 785–798.
23. Ширяев А. Н. Вероятность. Москва: Наука, 1989. 639 с.
24. Гихман И. И., Скороход А. В. Введение в теорию случайных процессов. Москва: Наука, 1977. 570 с.

ДОДАТОК А ЛІСТИНГ ПРОГРАМНОГО ПРОДУКТУ

Фрагмент 1: завантаження даних та виведення часового ряду

```
import numpy as np
import matplotlib.pyplot as plt
import math
from stochastic.continuous import FractionalBrownianMotion
from math import gamma

k = 3

files = ['Data_co2.txt', 'Data_waves_2.txt', 'Data_eu_1.txt', 'Data_eu_2.txt',
'Data_oil_2.txt']
file = files[k-1]
f = open(file, 'r')

nn = [682, 409, 1259, 1008, 381]
# co2 - 682, oil_2 - 381, waves_2 - 409

data = [list([float(x) for x in f.readline().split()]) for i in range(nn[k-1])]
data = np.matrix(data)
N = data.shape[0]
dat = np.zeros((N, 2))
for i in range(N):
    dat[i, 0] = i+1
    dat[i, 1] = round(data[i, 0], 6)
plt.plot(dat[:,0], dat[:,1])
```

Фрагмент 2: первинна обробка даних, перевірка на стаціонарність приростів

```
E = sum(dat[:,1])/N
print('Математичне сподівання до усунення тренду = ', E)
norm = dat
norm[:,1] = dat[:,1]-E
E_0 = sum(norm[:,1])/N
print('Математичне сподівання після усунення тренду = ', E_0)
y = norm
for i in range(1, N):
```

```

    y[i-1, 1] = norm[i, 1] - norm[i-1, 1]
norm_y = y[:-1, :]
plt.plot(norm_y[:,0], norm_y[:,1])

N = N-1
n = 2
r = [0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]
for k in range(n):
    s_1 = 0
    s_2 = 0
    for i in range(round(N*k/n), round(N*(k+1)/n-1)):
        s_1 = s_1 + norm_y[i, 1]*norm_y[i+1, 1]
        s_2 = s_2 + norm_y[i, 1]*norm_y[i, 1]
    r[k] = s_1/s_2
print(r)
N = N+1

```

Фрагмент 3: Знаходження значення лямбда

```

R_1 = 0
R_2 = 0
for i in range(N-1):
    R_1 = R_1 + abs(norm_y[i,1])
    R_2 = R_2 + norm_y[i,1]*norm_y[i,1]
R_1 = R_1/(N-1)
R_2 = R_2/(N-1)
d = R_1*R_1/R_2
print(' Абсолютні моменти приростів')
print('R1 = ', R_1)
print('R2 = ', R_2)
print("'Коефіцієнт ексцесу" = ', d)

def f(g):
    return gamma(g/2+1/2)*gamma(g/2+1/2)/gamma(g+1/2)/pow(math.pi, 0.5)
def MPD(f, a, b):
    eps = 10** -5
    while (b-a) > eps:
        c = (a+b)/2.0
        if f(c) > d:
            a=c
        elif f(c) < d:
            b=c
    else:

```

```

        b=c
    return(c)
lamda = MPD(f, 0, 10)
print('Лямбда = ', lamda)

```

Фрагмент 4: перехід до нової послідовності, оцінка параметрів моделі

```

z = np.zeros((N-1, 2))
for i in range (0, N-1):
    z[i, 0] = i+1
    z[i, 1] = np.sign(norm_y[i,1])*np.power(abs(norm_y[i,1]), 1/lamda)
    if i % 100 == 0:
        print(i)

from numpy.linalg import inv

def matr_S(H):
    S = np.zeros((N-1, N-1))
    for i in range(N-1):
        for j in range(N-1):
            S[i, j] = (np.power(abs(j-i+1), 2*H) + np.power(abs(j-i-1), 2*H) - 2 *
np.power(abs(j-i), 2*H))/2
    return S
n = 10
H = np.zeros((1, n-1))
Q = np.zeros((1, n-1))
for i in range(n-1):
    H[0, i] = (i+1)/n
    if H[0, i] == 0.5:
        k = i
H = np.delete(H, k, 1)
print(H)
for i in range(n-2):
    Q[0, i] = 0.8/R_1 * np.power(np.dot((np.dot(inv(matr_S(H[0, i])), z)).T, z)[1,
1]/(N-1), 0.5)
    print('H = ', H[0, i], '; Q = ', Q[0, i])
print(H)
print(Q)
Q = abs(Q - 1)
print(Q)
H_est = H[0, np.argmin(Q, axis = 1)[0]]
print("\nH_est = ', H_est)
s_est = np.power(N-1, H_est)*1.25*R_1
print("Оцінка волатильності: ", s_est)

```

Фрагмент 5: обчислення статистичних критеріїв

```

v = np.zeros((N-1, 1))
v_2 = np.zeros((N-1, 1))
z_2 = np.zeros((N-1, 1))
z_3 = np.zeros((N-1, 1))
v[0, 0] = 0
for i in range (1, N-1):
    v[i, 0] = z[i-1, 1] + v[i-1, 0]
for i in range (0, N-1):
    z_2[i, 0] = z[i, 1] * z[i, 1]
    z_3[i, 0] = z[i, 1] * z[i, 1] * z[i, 1]
    v_2[i, 0] = v[i, 0] * v[i, 0]
c = 1/N*np.sum(z_2[:, 0])

a_n = -3/2*c*c
beta_1 = 4.95*np.power(c, 2.5)/np.power(2*H_est+2, 0.5)
c_n = c/2
beta_2 = 4.08*c*c
f_n = np.power(4.08/1.5, 0.5)*3*np.power(c, 1.5)

An = np.dot(v.T, z_3)[0,0]/N
Bn = np.dot(v_2.T, z_3)[0, 0]/np.power(N, 1+H_est)
Cn = (np.dot(v_2.T, z_2)[0,0]/c - np.sum(v_2[:, 0]))/N
Dn = np.dot(v.T, z_3)[0,0]/np.power(N, 2*H_est)
Fn = np.sum(z_3[:, 0])/np.power(N, H_est)
if H_est < 0.5:
    print("An = ", round(An, 4), " a_n = ", round(a_n, 4))
    print("Bn = ", round(Bn, 4), " beta_1 = ", round(beta_1, 4))
    if H_est < 0.25:
        print("Cn = ", round(Cn, 4), " c_n = ", round(c_n, 4))
if H_est > 0.5:
    print("Dn = ", round(Dn, 4), " beta_2 = ", round(beta_2, 4))
    print("Fn = ", round(Fn, 4), " f_n = ", round(f_n, 4))

```

Фрагмент 6: побудова прогнозу

```

S = matr_S(H_est)
m = [200, 400, 600]
u = [0, 0, 0, 0]
uu = 0
j = 1

```

```
for k in range (m[0]):  
    for i in range (m[0]):  
        uu = uu + z[i, 1] * S[m[0]+j][k] * S[k, i]  
print(uu)  
print(z[m[0]+j, 1])
```

ДОДАТОК Б ІЛЮСТРАТИВНІ МАТЕРІАЛИ ДО ДОПОВІДІ

Модель часового ряду на основі фрактального броунівського руху

Виконав:

Студент 4-го курсу групи КА-51

Братусь Олександр

Науковий керівник:

професор кафедри ММСА, д.ф.-м.н. Бондаренко В. Г.

■ **Об'єкт дослідження:**

нелінійні нестационарні процеси зі стаціонарними приростами

■ **Предмет дослідження:**

математична модель часового ряду

■ **Мета роботи:**

дослідити запропоновану модель часового ряду, обґрунтувати доцільність її застосування в зазначених випадках, довести ефективність досліджуваної моделі, в тому числі експериментальним шляхом

Актуальність обраної теми

Моделювання часових рядів є невід'ємним процесом у більшості галузей діяльності людини. Випадкові процеси будь-якого походження (економічні, фінансові, природні тощо) можуть бути проаналізовані з можливістю подальшої апроксимації певною моделлю задля можливості робити короткострокові та довгострокові прогнози.

Варіанти побудови моделі

- Використовуючи існуючі (класичні) моделі (ARMA, ARIMA, GARCH тощо)
- Обравши відомий випадковий процес, побудувати відображення

Неперервна математична модель

$$x(t) = \Phi(X(\cdot))(t)$$

- де $\xi(t)$ – випадковий процес з відомими характеристиками;
- $X(t)$ – певна реалізація процесу $\xi(t)$;
- Φ – оборотне перетворення в $C(0; T)$.

- $P\{N(t) = k\} = \exp(-\lambda t) \frac{(\lambda t)^k}{k!}$

- $S(t + dt) - S(t) = S(t)(\xi(t + dt) - \xi(t)) + \mu(t)S(t)dt$

- $S(t) = S(0)\exp\left\{\sigma w(t) + \left(\mu - \frac{\sigma^2}{2}\right)t\right\}$

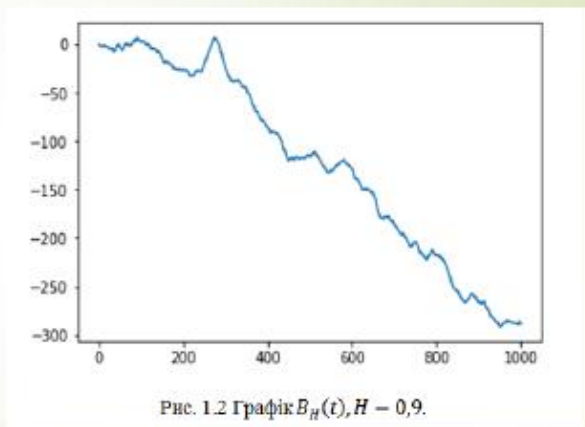
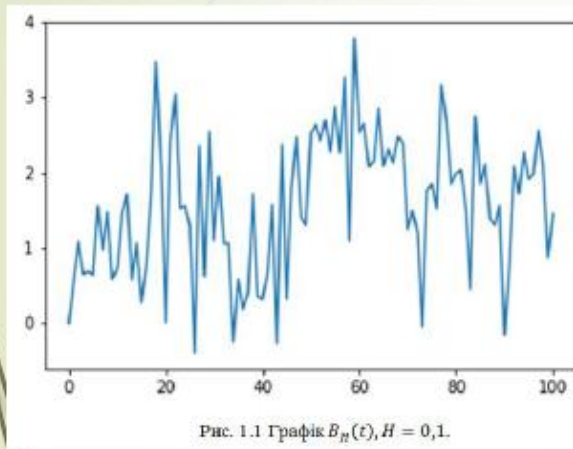
Фрактальний броунівський рух

$$B_H(t), \quad 0 < H < 1$$

- $E B_H(t) = 0$
- $B_H(0) = 0$
- $R(t, s) = E B_H(t)B_H(s) = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H})$
- $D B_H(t) = t^{2H}$
- $B_H(t) - B_H(s) \sim B_H(t - s) \sim N(0, (t - s)^{2H})$

Пам'ять фрактального броунівського руху

- $H < \frac{1}{2}$ - прирости утворюють послідовність з короткою пам'яттю – антиперсистентний процес fBm;
- $H > \frac{1}{2}$ - прирости утворюють послідовність з довгою пам'яттю – персистентний процес fBm (сильна післядія);
- при $H = \frac{1}{2}$ фрактальний броунівський рух є вінерівським процесом.



Гранична теорема для сум незалежних випадкових величин

Для статистик:

$$R_{jn} = \frac{1}{n} \sum_{k=1}^n |y_k|^j, \text{ де } j - \text{натуральне,}$$

доведена наступна теорема:

$$\frac{R_{jn}}{E R_{jn}} \rightarrow 1, \text{ } n \rightarrow \infty, \text{ з ймовірністю } 1.$$

Прирости часового ряду

- Покладемо $X(t) = \sigma B_H(t)$.
- Розглянемо прирости:

$$y_k = X\left(\frac{k}{n}\right) - X\left(\frac{k-1}{n}\right) = \sigma(B_H\left(\frac{k}{n}\right) - B_H\left(\frac{k-1}{n}\right)), k = 1, 2, \dots, n.$$

- $\{y_k\}$ утворюють гаусівську стаціонарну послідовність.

Оцінка параметрів

- $\hat{H}_n = \frac{\ln\left(\sqrt{\frac{2}{\pi}} \frac{\sigma}{R_{1n}}\right)}{\ln n}$
- $\hat{\sigma}_{1n} = n^H \sqrt{\frac{\pi}{2}} R_{1n}$
- Статистика $n^{2H-1}(S^{-1}y, y)$ є незміщеною оцінкою параметра σ^2
- $\hat{\sigma}_{2n} = \sqrt{n^{2H-1}(S^{-1}y, y)}$ - консистентна оцінка параметра σ
- $\frac{\hat{\sigma}_{2n}}{\hat{\sigma}_{1n}} = \frac{0,8}{R_{1n}} \sqrt{\frac{(S^{-1}y, y)}{n}} \approx 1$, тобто $\left| \frac{\hat{\sigma}_{2n}}{\hat{\sigma}_{1n}} - 1 \right| \rightarrow \min$
- $\hat{\sigma} = \frac{\hat{\sigma}_{1n} + \hat{\sigma}_{2n}}{2}$

Прогноз фрактального броунівського руху

- $$\hat{\xi}_{m+j} = \sum_{k=1}^m \sum_{i=1}^m s_{m+j,k} a^{ki} \xi_i, \quad j = 1, \dots, n - m$$
- $$\xi_k = y_k = B_H\left(\frac{k}{n}\right) - B_H\left(\frac{k-1}{n}\right)$$
- $$\hat{y}_{m+j} = \sum_{k=1}^m \sum_{i=1}^m \left(\frac{(m+j-k+1)^{2H} + (m+j-k-1)^{2H}}{2} - (m+j-k)^{2H} \right) a^{ki} y_i,$$

$$j = 1, \dots, r, \quad r = m - n.$$
- $$\delta_j = \left| \frac{\hat{y}_{m+j} - y_{m+j}}{y_{m+j}} \right|$$

Апроксимація часового ряду

- $$\bar{x} = \frac{1}{n} \sum x_k = 0$$
- $$y_k = x_k - x_{k-1}, \quad k = 2, \dots, n$$
- $$d = \frac{1}{\sqrt{n}} \frac{\Gamma^2\left(\frac{\lambda+1}{2}\right)}{\Gamma\left(\lambda+\frac{1}{2}\right)}$$
- $$z_k = \operatorname{sgn} y_k |y_k|^{\frac{1}{\lambda}} = \operatorname{sgn} y_k |z_k|^\lambda, \quad \lambda > 0$$
- $$z_k = \sigma^H \left(B_H\left(\frac{k}{n}\right) - B_H\left(\frac{k-1}{n}\right) \right), \quad x_k = \sigma^\lambda \sum_{j=1}^k \operatorname{sgn} y_j \cdot \left| B_H\left(\frac{j}{n}\right) - B_H\left(\frac{j-1}{n}\right) \right|^\lambda$$

Перевірка апроксимації

- $A_n = \frac{1}{n} \sum v_k z_k^3, H \in (0; \frac{1}{2});$
- $B_n = \frac{1}{n^{1+H}} \sum v_k^2 z_k^3, H \in (0; \frac{1}{2});$
- $C_n = \frac{1}{n} \sum v_k^2 \left(\frac{z_k^2}{c} - 1 \right), H \in (0; \frac{1}{4});$
- $D_n = \frac{1}{n^{2H}} \sum v_k z_k^3, H \in (\frac{1}{2}; 1);$
- $F_n = \frac{1}{n^H} \sum v_k z_k^3, H \in (0; \frac{1}{2}).$

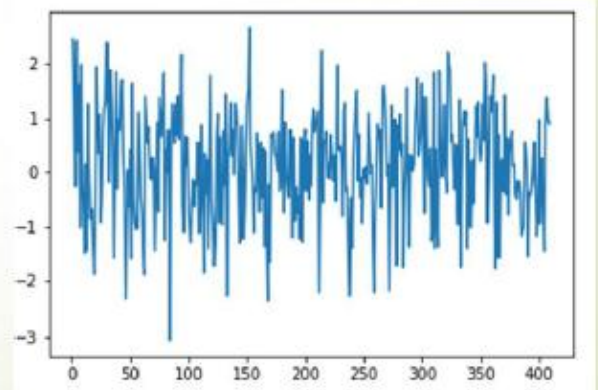
$$A_n \rightarrow -\frac{3}{2}c^2; \quad B_n \rightarrow 3c^{\frac{5}{2}}\eta; \quad C_n \rightarrow \frac{c}{2}; \quad D_n \rightarrow \frac{3}{2}c^2B^2(1); \quad F_n \rightarrow 3c^{\frac{3}{2}}B(1).$$

Приклад роботи програмного продукту

Часовий ряд, що характеризує осциляції хвиль у Північній Атлантиці, і який охоплює часовий проміжок 10.1980 – 10.2014

$$\begin{aligned} \hat{\rho}_1 &= 0,44 - 0,49 \\ \hat{d}_n &= 0,62 \\ \lambda &= 1,04 \\ \hat{H} &= 0,1 \quad (Q = 0,94) \\ \hat{\sigma} &= 158,57 \end{aligned}$$

$$\begin{aligned} A_n &= -3,96; \quad A = -4,03 \\ B_n &= 8,88 < \beta_1 = 11,24 \end{aligned}$$

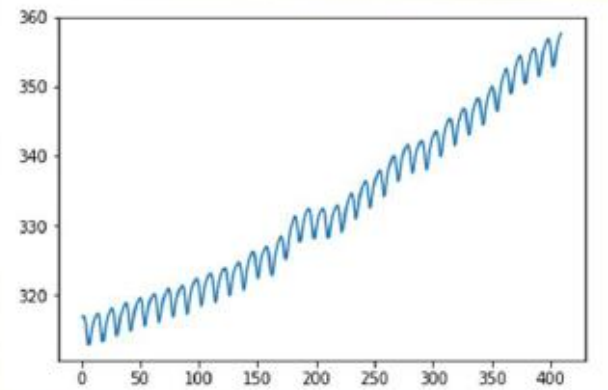


Приклад роботи програмного продукту

Набір даних, який показує вміст вуглекислого газу в атмосфері в період з березня 1958 по червень 2016

$$\begin{aligned}\hat{\rho}_1 &= 0,73 - 0,77 \\ \hat{d}_n &= 0,71 \\ \hat{\lambda} &= 0,81 \\ \hat{H} &= 0,8 (Q = 0,99) \\ \hat{\sigma} &= 181,91\end{aligned}$$

$$D_n = 1,1 < \beta_2 = 4,08$$



Приклад роботи програмного продукту

Набір даних, який показує вміст вуглекислого газу в атмосфері в період з березня 1958 по червень 2016

Прогнозування

$$\delta_k = \frac{|\hat{s}_{m+k} - s_{m+k}|}{s_{m+k}}, \quad k = 1, 2, 3, 4$$


m	k=1	k=2	k=3	k=4
200	0,003	0,004	0,002	0,003
400	0,0006	0,002	0,005	0,005
600	0,001	0,006	0,01	0,01

Висновки

- Дипломна робота присвячена аналізу, побудові та використанню математичної моделі часового ряду, в основі якої лежить процес фрактального броунівського руху.
- Постановлено, що дана фрактальна модель має непогану прогнозовану здатність, зберігає інтерпретованість, подібно до класичних моделей.
- Створено програмний додаток, який дозволяє побудувати модель та здійснити прогноз на нових даних. Перевірка додатку показала, що запропонована модель на основі фрактального броунівського руху є конкурентоспроможною по обчислювальній потужності.

Майбутні дослідження

- Підвищення швидкодії при роботі з масивними об'ємами даних
- Створення багатокрокових прогнозів
- Проведення порівняльного аналізу з іншими моделями в даній предметній області



Дякую за увагу!