

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»**

**НН Інститут прикладного системного аналізу
Кафедра математичних методів системного аналізу**

До захисту допущено:

Завідувач кафедри

_____ Оксана ТИМОЩУК

«__» _____ 2023 р.

Дипломна робота

на здобуття ступеня бакалавра

за освітньо-професійною програмою «Системний аналіз і управління»

спеціальності 124 «Системний аналіз»

**на тему: «Моделювання вартості акцій та міжнародних індексів
розподілами з важкими хвостами»**

Виконав:

студент IV курсу, групи КА-91

Грішин Костянтин Дмитрович

Керівник:

доц., к.ф.-м.н. Каніовська І.Ю.



Консультант з економічного розділу:

доцент, к.е.н. Рощина Н.В.

Консультант з нормоконтролю:

к.ф.-м.н. Статкевич В.М.

Рецензент:

доцент, к.ф.-м.н. Ільєнко А.Б.



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

Студент _____

Київ – 2023 року

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

Рівень вищої освіти – перший (бакалаврський)

Спеціальність – 124 «Системний аналіз»

Освітня програма «Системний аналіз і управління»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Оксана ТИМОЩУК

«__» травня 2023 р.

ЗАВДАННЯ
на дипломну роботу студенту
Грішину Костянтину Дмитровичу

1. Тема роботи «Моделювання вартості акцій та міжнародних індексів розподілами з важкими хвостами», керівник роботи Каніовська І.Ю., к.ф.-м.н., доцент затверджені наказом по університету від «__» травня 2023 р. № _____
2. Термін подання студентом роботи 12.06.2023.
3. Вихідні дані до роботи: ДСТУ 3008:2015, статистичні дані Yahoo Finance
4. Зміст роботи: 1) дослідження предметної області; 2) розподіли з важкими хвостами, їх властивості, оцінки параметрів та показники ризику інвестування; 3) реалізація власного експерименту; 4) функціонально-вартісний аналіз програмного продукту.
5. Перелік ілюстративного матеріалу: 1) розподіл лог-дохідностей акцій, індексів; 2) основні етапи розвитку застосування розподілів; 3) розподіли із важкими хвостами 4) результати власного експерименту з індексом США Standard & Poor 500; 5) порівняння лог-дохідностей та VaR для різних методів.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Економічний	доцент, канд. екон. наук Рощина Н.В.		

7. Дата видачі завдання: 17.01.2023 

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Затвердження теми ДР	17.04.2023-23.04.2023	виконано
2	Ознайомлення зі структурою БДР згідно з Положенням про державну атестацію студентів НТУУ «КПІ ім. І. Сікорського»	17.04.2023-23.04.2023	виконано
3	Ознайомлення з ДСТУ 3008-95 та стандарти ЄСПД	17.04.2023-30.04.2023	виконано
4	Проведення дослідження за темою БДР під керівництвом керівника	17.04.2023-30.04.2023	виконано
5	Завершення роботи над першим варіантом частини БДР	24.04.2023-14.05.2023	виконано
6	Проведення роботи над експериментальною частиною БДР	01.05.2023-14.05.2023	виконано
7	Проведення роботи над програмним продуктом	01.05.2023-14.05.2023	виконано
8	Оформлення БДР та аналіз отриманих результатів	15.05.2023-21.05.2023	виконано

Студент

Керівник

_____ Костянтин ГРІШИН

 _____ Ірина КАНІОВСЬКА

РЕФЕРАТ

Дипломна робота: 132 с., 26 рис., 11 табл., 5 додатків, 47 джерел.

Ключові слова: РОЗПОДІЛ ІЗ ВАЖКИМИ ХВОСТАМИ, СТІЙКИЙ РОЗПОДІЛ, УЗАГАЛЬНЕНИЙ СКОШЕНИЙ РОЗПОДІЛ СТЬЮДЕНТА, УЗАГАЛЬНЕНИЙ ГІПЕРБОЛІЧНИЙ РОЗПОДІЛ, МІЖНАРОДНИЙ ФОНДОВИЙ ІНДЕКС, ПРОСТА АКЦІЯ, ЛОГАРИФМІЧНА ДОХІДНІСТЬ.

Об'єкт дослідження: моделі вартості акцій та міжнародних фондових індексів.

Мета дослідження: проаналізувати переваги застосування розподілів із важкими хвостами в моделях цін акцій та міжнародних фондових індексів.

Використані моделі: стійкого розподілу, узагальненого скошеного розподілу Стюдента, узагальненого гіперболічного розподілу, модель GARCH.

Отриманні результати: створено програмний продукт на мові програмування Python та R, побудовано моделі лог-дохідностей фондового індексу США Standard & Poor 500 з 3 січня 2000 по 30 грудня 2022 року розподілами з важкими та напівважкими хвостами, зроблено висновок про їх значну перевагу над моделлю з нормальним розподілом. Дані для аналізу було взято з Yahoo Finance.

В рамках подальшого дослідження пропонується підвищити точність моделі шляхом удосконалення механізму попередньої обробки даних, розширення множини розподілів. Також пропонується застосувати більш новітні моделі оцінки ризиків інвестиційного портфеля.

ABSTRACT

Thesis 132 p., 26 fig., 11 tables, 5 appendices, 47 sources.

Keywords: HEAVY TAILED DISTRIBUTION, STABLE DISTRIBUTION, SKEWED GENERALIZED T DISTRIBUTION, GENERALIZED HYPERBOLIC DISTRIBUTION, STOCK INDEX, COMMON STOCK, LOG RETURN, VALUE AT RISK.

Object of study: price models of stocks and stock indexes.

Purpose: to investigate the advantages of using heavy tailed distributions in price models of stocks and stock indices.

Used models: stable distribution, skewed generalized t distribution, generalized hyperbolic distribution, model GARCH.

Results: a software product was created in the Python and R programming languages, models of log returns of USA stock index Standard & Poor 500 from January 3, 2000 to December 30, 2022 are built via heavy tailed and semi-heavy tailed distributions, their significant superiority over model which utilize normal distribution is concluded. The data for the analysis was taken from Yahoo Finance.

As part of further research, it is proposed to increase model accuracy by data preprocessing system improvement and enlarging set of distributions. Also, it is proposed to apply more novel models of investment portfolio risk evaluation.

ЗМІСТ

ВСТУП	8
1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ	10
1.1 Актуальність проблеми	10
1.2 Основні поняття та етапи розвитку моделей ціноутворення акцій	11
1.3 Огляд публікацій, які присвячені дослідженню лог-дохідностей акцій	15
1.4 Міра ризику інвестування	24
Висновки до розділу	27
2 РОЗПОДІЛИ З ВАЖКИМИ ХВОСТАМИ, ЇХ ВЛАСТИВОСТІ, ОЦІНКИ ПАРАМЕТРІВ ТА ПОКАЗНИКИ РИЗИКУ ІНВЕСТИВАННЯ	29
2.1 Розподіли з важкими хвостами	29
2.1.1 Стійкий розподіл	29
2.1.2 Узагальнений скошений розподіл Стюдента	38
2.2 Розподіли з напівважкими хвостами	40
2.3 Методи оцінки показника капіталу під ризиком	45
Висновки до розділу	50
3 РЕАЛІЗАЦІЯ ВЛАСНОГО ЕКСПЕРИМЕНТУ	52
3.1 Огляд бібліотек Python та R для роботи з розподілами та тестування їх функцій	52
3.2 Моделювання лог-дохідностей індексу S&P 500 розподілами	56
3.3 Обчислення показника VaR для лог-дохідностей різними методами	61
Висновки до розділу	65
4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ	67
4.1 Постановка задачі проектування	67
4.2 Обґрунтування функцій програмного продукту	68
4.3 Обґрунтування системи параметрів ПП	70
4.4 Аналіз експертного оцінювання параметрів	72
4.5 Аналіз рівня якості варіантів реалізації функцій	75

4.6 Економічний аналіз варіантів розробки ПП	77
4.7 Вибір кращого варіанту ПП техніко-економічного рівня	82
Висновки до розділу	83
ВИСНОВКИ	84
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	85
ДОДАТОК А Лістинг програми на мові R для оцінки параметрів узагальненого скошеного розподілу Стюдента та побудови відповідної щільності	90
ДОДАТОК Б Лістинг програми на мові Python для оцінки параметрів розподілів та візуалізації результатів	92
ДОДАТОК В Лістинг програми на мові R для оцінки VAR ковзним вікном для стійкого та узагальненого скошеного розподілу Стюдента	105
ДОДАТОК Г Лістинг програми на мові Python для оцінки VAR ковзним вікном розподілами та візуалізації результатів	108
ДОДАТОК Д Презентація	123

ВСТУП

Інвестування в акції стала нерозривною частиною будь-якої економіки. Цей механізм дозволяє компаніям отримати гроші для розвитку, приносячи вигоду інвестору в довгостроковій перспективі, коли ціна паперів зростає. За останнє сторіччя вкладення у вигідні акції перетворилося з бізнесу в науку. Цим займаються банки та компанії. Тому інвестування у наші дні вимагає не лише фінансової грамотності, а й знань сучасних математичних моделей поведінки акцій.

До задачі моделювання поведінки цін акцій було розроблено багато різних підходів: фундаментальний аналіз, технічний аналіз, стохастичні моделі, а в останні десятиліття також машинне навчання. Але пошук моделей поведінки цін акцій ускладнюється тим, що ринок, немов «чинить супротив» спробам гравця отримати вигоду на біржі – адже це означає збитки для інших гравців. Існує навіть гіпотеза про ефективність ринку, один із наслідків якої в тому, що передбачити поведінку акцій за рахунок пошуку закономірностей неможливо. Її підтвердження знецінило би більшість підходів.

Тому часто дослідники віддають перевагу моделі випадкового блукання, яка зберігає свою адекватність навіть якщо гіпотеза про ефективність ринку справджується. У такій моделі прирости акцій покладаються певною випадковою величиною.

Визначення розподілу приросту цін акцій у моделі випадкового блукання є критичним для портфельних інвесторів та банків. До 1960-их років приріст цін акцій покладався розподіленням за нормальним законом. Але з часом все більше спеціалістів відхиляли цю гіпотезу, серед яких лауреат Нобелівської премії з економіки 2013 року Юджин Фама та відомий математик, основоположник теорії фракталів, Бенуа Мандельброт.

У сучасних дослідженнях моделей випадкового блукання перейшли від приросту цін акцій до приросту їх логарифмів (логарифмічні дохідності), а на

заміну нормальному розподілу прийшли так звані розподіли з важкими хвостами. Щільності цих розподілів спадають на нескінченості повільніше, ніж для гаусівського, це означає, що вони не нехтують ймовірністю реалізації великих числових значень, які у менеджменті ризиків асоціюються зі аномальними прибутками або збитками. Більш того, серед цих розподілів є такі, що природньо лягають на сучасне уявлення про ціноутворення акцій.

Тому об'єктом цієї роботи стало дослідження моделей ймовірнісного розподілу логарифмічних дохідностей (скор. лог-дохідностей) акцій та міжнародних індексів за допомогою розподілів з важкими хвостами. Обрані розподіли будуть використані як наближення лог-дохідностей міжнародного індексу США Standard & Poor 500 (скор. S&P 500), одного з головних показників економіки цієї країни. Також для цього індексу буде різними методами розраховано капітал під ризиком (англ. Value-at-Risk) – показник ризику інвестування, який після виходу другої базельської угоди, став загальноприйнятим.

У першому розділі проводиться аналіз предметної області: розкривається актуальність проблеми, ретельно проводиться огляд публікацій, які присвячені наближенню лог-дохідностей акцій та індексів різними розподілами, аналізується парадигма оцінки капіталу під ризиком.

У другому розділі наводиться математична теорія, необхідна для проведення дослідження ймовірнісного розподілу лог-дохідностей акцій та міжнародних індексів за допомогою розподілів з важкими хвостами: дається означення цих розподілів, описуються їх характеристики, властивості, методи оцінки параметрів, алгоритми генерації випадкових величин. Також дається формалізація техніки обчислення капіталу під ризиком.

У третьому розділі проводиться пошук найкращої моделі поведінки міжнародного індексу США Standard & Poor 500, із застосуванням розподілів із важкими та напівважкими хвостами, результати порівнюються, робляться висновки. Також обчислюється капітал під ризиком різними методами.

1 ДОСЛІДЖЕННЯ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Актуальність проблеми

У моделюванні поведінки цін акцій одними з найбільш популярних моделей є стохастичні, зокрема, модель випадкового блукання. Навіть нещодавні прориви в сфері машинного навчання не витіснили цей підхід. Тим не менш, питання, які розподіли мають використовуватись в моделі випадкового блукання, залишається відкритим. Неправильний вибір розподілу матиме серйозні наслідки. Найбільш небезпечним може бути недооцінка ймовірності значних відхилень від середніх значень – в економіці вони асоціюються з рідкісними, але руйнівними подіями.

Наприклад, така недооцінка призвела до банкрутства одного з найбільших банків США Lehman Brothers у 2008 році. Інший приклад, у 2007 році в серпні сталася черга подій, які не могли бути поясненими загальноприйнятими моделями. Головний стратег банку Lehman Brothers дав коментар «Події, які моделі передбачали лише раз на 10 000 років, відбувалися три дні поспіль» [1]. Фінансовий стратег Goldman Sachs, однієї з найбільших інвестиційних компаній світу, казав про ці події: «Спостерігалися значення, віддалені на 25 стандартних відхилень кілька днів поспіль» [2].

Перші моделі випадкового блукання покладали прирости цін акцій нормальним розподілом. Це узгоджувалося з уявленнями про приріст ціни як суперпозиції великої кількості факторів, і як наслідок, виконання центральної граничної теореми. До 60-их років минулого сторіччя в дослідженнях нормальний розподіл приростів цін акцій не піддавався сумніву [3, С. 420]. Згодом виявилось, що це не є адекватною моделлю, так як вона недооцінювала ймовірності великих відхилень від середніх значень [4, С. 49].

Для вирішення цієї проблеми було запропоновано замінити різницю цін акцій на різницю логарифмів (лог-дохідності, підрозділ 1.2) та використовувати в розподілі з важкими хвостами, перевага яких у тому, що вони не нехтують малоймовірними катастрофічними подіями. Причому, як показали теоретичні дослідження, серед цих розподілів є такі, що узгоджуються з природнім уявленням про приріст ціни акцій як результату суми дії великої кількості факторів, наприклад стійкі розподіли [5, С. 331]. Це робить їх найкращими «кандидатами» для моделей поведінки цін акцій та міжнародних індексів.

1.2 Основні поняття та етапи розвитку моделей ціноутворення акцій

Означення 1. Прості акції – один із видів цінних паперів, володіння якими надає право власнику отримувати дивіденди від компанії. Вони продаються та купуються на біржі в торгові дні (наприклад в США 252 торгових днів). Отже, можна отримати прибуток від приросту ціни акцій та від дивідендів. У роботі розглядається саме прибуток від зростання цін акцій.

Означення 2. Портфель – набір цінних паперів, який може включати акції, облігації, валюту, векселі, нерухомість, об'єкти мистецтва тощо. В роботі будемо розглядати лише портфелі, які складаються з акцій. У портфелі може бути декілька акцій однієї й тої ж компанії (можлива дробова частина, наприклад, 2,5 акції компанії Boeing).

Означення 3. Міжнародний індекс (фондовий індекс) – портфель, який складається з акцій крупних компаній деякої країни чи регіону. Ціна індексу є показником стану економіки. Наприклад, у США таким індекс Standard & Poor 500, який складається з акцій 500 провідних американських компаній. Він є найбільш авторитетним для економістів.

Означення 4. Ціна закриття акції в день t – ціна, за якою останній раз акція була продана на біржі в день t [6].

Один із найбільш популярних показників зміни ціни акції є логарифмічні дохідності (лог-дохідності). Саме їх будемо використовувати в роботі.

Означення 5. Лог-дохідність (англ. Log-return) в день t – значення, яке розраховується для портфелю за формулою (1.1)

$$r_t = \ln(\sum_j P_t^j n_j) - \ln(\sum_j P_{t-1}^j n_j) \quad (1.1)$$

де P_t^j – ціна закриття акції j в день t

n_j – кількість акцій j у портфелі

Лог-дохідність можна визначити для акції компанії – тоді портфель буде складатись з однієї акції. Однією з найбільш поширених моделей поведінки часового ряду $\{r_t\}$ є модель випадкового блукання.

Означення 6. Модель випадкового блукання – модель, яка покладає прирости цін акцій реалізацією деяких випадкових величин (в загальному випадку, не незалежних, не однаково розподілених). Цими моделями займався, зокрема, лауреат Нобелівської премії з економіки 2013 року Юджин Фама [7]. Багато публікацій підтверджують адекватність даного припущення. Наприклад у роботі [8], де було проведено аналіз акцій США, Франції, Італії, Бельгії та інших країн з 1967-1988 рік, гіпотезу не було відхилено.

Аналіз публікацій, які присвячено моделям випадкового блукання для лог-дохідностей, показав, що їх часто вважають незалежними однаково розподіленими. Більш просунуті дослідження ще враховують волатильність.

Означення 7. Волатильність – міра того наскільки за фіксований час акція може змінити своє значення. Часто вимірюється за допомогою стандартного відхилення.

Якщо в дослідженні враховується зміна волатильності, тоді від часового ряду лог-дохідностей $\{r_t\}$ переходять до деяких стандартизованих приростів

$\{e_t\}$, які вважають реалізацією незалежних однаково розподілених випадкових величин. Часто для цього застосовують модель GARCH (Generalized AutoRegressive Conditional Heteroskedasticity). Ця модель враховує зміну дисперсії випадкових доданків у часовому ряду. Із них отримують оцінку волатильності в кожен момент часу $\{\sigma_t\}$, із яких потім знаходять стандартизовані прирости $\{e_t\}$.

Найперша модель випадкового блукання для ціноутворення акцій була розроблена математиком Луї Башельє у 1900 році [9]. У ній приріст ціни акції покладался нормально розподіленим – так-звана гаусівська гіпотеза. Вона була обґрунтована: на ціну акцій діє багато випадкових факторів, якщо їх вплив представити як арифметичну суму, тоді (при обмеженій дисперсії) виконується центральна гранична теорема [4, С. 49]. Публікації, присвячені аналізу акцій, до 1960-их років не піддавали сумнівам цю гіпотезу [10, С. 3] або вважали що вона виконується для лог-дохідностей [11].

З часом виявилось, що гаусівська гіпотеза не є правдоподібною, тому що швидкість спадання щільності нормального розподілу на $x \rightarrow \infty$ занадто висока порівняно з побудованими гістограмами – в економіці це можна трактувати як недооцінка моделлю ймовірності великих прибутків та збитків. Розподіли, у яких щільність спадає повільніше стали називати розподіли з важкими хвостами (англ. Heavy-Tailed distribution) [4, С. 49]. Нижче у цьому пункті буде їм дано строге математичне означення.

У 1963 році видатний вчений Бенуа Мандельброт опублікував статтю, де вперше поставив під сумніви адекватність гаусівської гіпотези [12] та запропонував замінити нормальний розподіл стійким, його ідею підтримав і розвинув вчений Юджин Фама у статті 1965 року [3, С. 1]. Застосування стійкого розподілу виявилось логічним узагальненням моделі оскільки нормальний розподіл був його частковим випадком. Причому, як довели вчені Колмогоров та Гнеденко, стійкий розподіл і лише він є граничним для суми незалежних, однаково розподілених випадкових величин із нескінченною дисперсією чи середнім [13, С. 162]. Тому ці розподіли дають можливість не

відхиляти гіпотезу про ціну акції як результату дії великої кількості незалежних випадкових факторів.

Після цих двох публікацій розподіли з важкими хвостами почали отримувати все більшу популярність, зокрема, стійкі розподіли [14]-[15]. У сучасних публікаціях застосовуються й інші розподіли, наприклад, розподіл Стюдента, узагальнений гіперболічний (підрозділ 2.2). Їх використання не протирічить згаданій гіпотезі ціноутворення акцій, оскільки вони відносяться до так званих нескінченно подільних розподілів.

Як знову ж було показано Колмогоровим та Гнеденко, у граничних розподілів сум незалежних випадкових величин існує певна «ієрархія». Якщо маємо суму однаково розподілених величин зі скінченими моментами другого порядку – граничним розподілом буде нормальний. Якщо опустити умову скінченності моментів – граничним є стійкий (підрозділ 2.1). Якщо додатково допустити в сумі неоднакові випадкові величини граничним будуть нескінченно подільні розподіли [13, С. 116]. Тобто їх можна уявити як результат дії великої кількості незалежних випадкових факторів, як і уявляється ціноутворення акції. Згадані класи розподілів зображено на рис 1.1.



Рисунок 1.1 – Класи розподілів, які застосовуються в моделюванні цін акцій, які є граничним для послідовності сум незалежних випадкових величин

Як було зазначено, серед стійких та нескінченно подільних розподілів є розподіли з важкими хвостами. Дамо цьому поняттю строге математичне означення.

Означення 8. [16,С. 5]. Розподіл із важкими хвостами – такий що

$$\forall \mu > 0, \lim_{x \rightarrow +\infty} \sup_x (1 - F(x))e^{\mu x} = \infty \quad (1.2)$$

де $F(x)$ – функція розподілу

Умова (1.2) визначає швидкість спадання «хвосту» функції розподілу. Означення 8 еквівалентне існуванню скінченної генератриси [16,С. 6]. Її можна записати в ряд Тейлора [17,С. 75]

$$E(e^{tX}) = 1 + \frac{t}{1!} E(X) + \frac{t^2}{2!} E(X^2) + \frac{t^3}{3!} E(X^3) + \dots + \frac{t^k}{k!} E(X^k) + \dots \quad (1.3)$$

З формули (1.3) слідує, що існування генератриси еквівалентно існуванню скінчених моментів усіх порядків. Тоді одразу можна зробити висновок, що на рис 1.1 блакитна множина містить лише розподіли з важкими хвостами.

1.3 Огляд публікацій, які присвячені дослідженню лог-дохідностей акцій

Основною проблемою в дослідженні цін акцій та індексів є пошук розподілу, який найточніше наближував емпіричний розподіл лог-дохідностей. Було проведено огляд публікацій, де цін акцій та індексів моделювались різними розподілами. На його основі складено таблицю 1.1. Показником узгодженості в більшості досліджень була статистика Колмогорова-Смірнова (чим точніше наближення, тим вона менше, мінімальна нуль). В усіх публікаціях було зроблено висновок, що гіпотеза про

нормальність розподілу лог-дохідностей відхиляється. Також можна помітити, що різні розподіли виявлялись кращою моделлю залежно від акції чи фондового індексу.

Таблиця 1.1 – Статистика Колмогорова-Смірнова для лог-дохідностей акцій та міжнародних індексів залежно від розподілу в різних публікаціях

Дослідження			Значення статистики Колмогорова-Смірнова, яка використовувалась в критерії узгодженості					
Джерело	Акції/індекси	Час	N	S	SGT	NIG	Нур	GH
[18]	DAX	2000-2009	1.858	1.389		0.796	0.779	
	WIG20	2000-2009	1.584	0.881		0.682	0.681	
[19]	XTRZM	2007-2017	Відхилено	Краще				
[20]	Nikkei225	1978-2017	Відхилено	Краще				
	S&P 500	1975-2017	Відхилено	Краще				
	Dow 30	1986-2017	Відхилено	Краще				
	SSEC	1998-2017	Відхилено.	Краще				
[21]	S&P 500	1957-2017	0.018		0.006			0.007
[22]	DJIA	1987-1994	4.6383	0.5583				
	Boeing	1997-2003	2.1361	0.4522				
[23]	ISPIX	1998-2005	Відхилено	0.02584				
	AMEX	1983-2005	Відхилено	0.01824				
	FCHI	1990-2005	Відхилено	0.01198				
	NASDAQ	1999-2005	Відхилено	0.02928				
[24]	S&P 500	1998-2008	0.0541					0.0183
	Nikkei225	1998-2008	0.0417					0.0191
	HangSeng	1998-2008	0.0691					0.0254
	DAX	1998-2008	0.0617					0.0335
[25]	AGL	2006-2011			0.0467	0.0414	0.024	0.0214
	AMS	2006-2011			0.03	0.0267	0.032	0.036
	ANG	2006-2011			0.0274	0.018	0.4746	0.028
[26]	Vale5	1994-2001	0.0751			0.0121	0.0099	0.0108
	Ibvsp	1994-2001	0.0661			0.0166	0.0208	0.0093
	Petr4	1994-2001	0.0640			0.0133	0.0142	0.0126
	Elet6	1994-2001	0.0598			0.0123	0.0150	0.0103
[27]	DJIA	1987-1992	3996.41*		4092.31*			
	Nikkei225	1987-1992	3323.08*		3468.71*			
	FT30	1987-1992	3564.29*		3727.05*			

Примітка 1. Складено автором на основі [21]-[27]

Примітка 2. Скорочені позначення такі:

N – нормальний розподіл,

S – стійкий розподіл,

SGT – скошений розподіл Стюдента,

NIG – нормальний обернений гаусівський,

Нур – гіперболічний,

GH – узагальнений гіперболічний

Абревіатури в стовпці «Акції/Індекси» – загальноприйняті скорочення акцій та індексів

Зеленим виділено найточніші моделі

«Краще», «Відхилено» – для публікацій, де було наведено лише якісні результати

* – замість статистики Колмогорова-Смірнова застосовувалась статистика likelihood ratio

Опишемо як проводилися дослідження в деяких публікаціях більш детально, щоб зрозуміти парадигму роботи з лог-дохідностями.

У роботі [23] досліджувались розподіли 27 акцій та міжнародних індексів. Для них методом максимальної правдоподібності було оцінено параметри для нормального розподілу та стійкого розподілу (який задається параметрами α, β, δ, c , підрозділ 2.1). За допомогою критеріїв узгодженості Колмогорова-Смірнова (скорочено К-С) та Андерсона-Дарлінга (скорочено А-Д) було перевірено гіпотези про нормальний та стійкий розподіли. Результати тестів наступні:

а) в усіх 27 випадках гіпотеза про нормальний розподіл відхилена на рівні значущості $p = 0.05$ як для К-С, так і для А-Д;

б) в 15 з 27 випадків гіпотеза про стійкий розподіл відхилена на $p = 0.05$ тестом А-Д, та відхилена у 23 з 27 випадків для К-С на тому ж рівні.

Для всіх акцій та індексів статистики критерію К-С та А-Д давали кращі значення для стійкого розподілу, ніж для нормального. Автори зауважують, що К-С більше чутливий до відхилень у центральній частині графіку функції розподілу, а тест А-Д – до відхилень у її кінцях [23, С. 45]. Тому дослідники зробили висновок, що стійкий розподіл краще наближує у хвостах, ніж у центральній частині розподілу.

Наступним пунктом дослідження була перевірка наявності дисперсії у розподілі лог-дохідностей. Якщо вона скінчена, тоді виконується

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \xrightarrow{\text{м.н.}} c, n \rightarrow \infty \quad (1.4)$$

де x_i – елементи вибірки $\{x_1, x_2, \dots, x_n\}$

$\bar{x}_n$ – середнє арифметичне

c – деяке дійсне число, яке є дисперсією розподілу

Для акцій Microsoft значення оцінки дисперсії необмежено зростала і сильно коливалась з чого автори зробили висновок, що розподіл немає скінченного другого моменту, що характерно, зокрема, для стійкого розподілу (Рис 1.2).

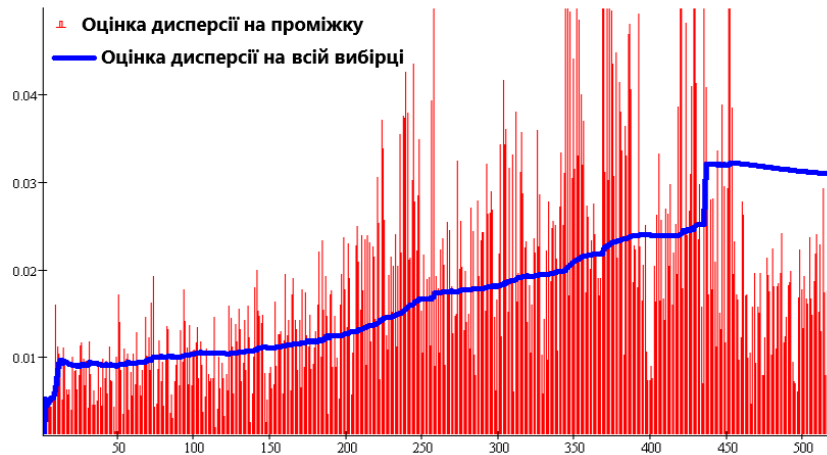


Рисунок 1.2 – Залежність оцінки дисперсії лог-дохідностей акцій Microsoft на проміжку від 13-03-1986 до 27-05-2005 року від кількості даних [23]

Перевірка гіпотези про стійкий розподіл у дослідженні [23] здійснювалась на додачу до тесту К-С та А-Д ще власним тестом. В його основі лежала властивість, що якщо незалежні однаково розподілені випадкові величини $\xi_1, \dots, \xi_n$ мають стійкий розподіл, то при додаванні $\xi_1 + \dots + \xi_n$ теж є стійким (підрозділ 2.1, означення 11). Дослідники запропонували такий тест.

Нехай є вибірка $\{X_1, X_2, \dots, X_n\}$. Можна покласти вибірку $\{Y_1, Y_2, \dots, Y_{\lfloor \frac{n}{d} \rfloor}\}$

$$Y_k = \sum_{i=(k-1) \cdot d + 1}^{k \cdot d} X_i, k = 1, \dots, \left\lfloor \frac{n}{d} \right\rfloor \quad (1.5)$$

де n – розмір загальної вибірки вибірки

d – деяке натуральне число

$\left\lfloor \frac{n}{d} \right\rfloor$ – ціла частина від ділення n на d

Середнє лог-дохідностей припускалось нуль, тоді в припущенні про стійкий розподіл виконується рівність (1.6) (підрозділ 2.1, властивість 2).

$$Y_i d^{-\frac{1}{\alpha}} = X_i \quad (1.6)$$

Гіпотеза про однаковий розподіл вибірок $\{X_i\}, \{Y_i\}$ перевірявся тестами А-Д та К-С. Дослідники проводили такий тест для $d = 10$ та $d = 15$. Результатом було прийняття гіпотези про стійкість для 20 акції з 27 на рівні $p = 0.05$.

В дослідженні [23] оцінювався також показник Херста H , міра довготривалої пам'яті часового ряду. Зокрема $H > 0.5$ відповідає персистентній поведінці – тренд у теперішньому робить більш ймовірним збереження такого ж тренду в майбутньому [28, С. 4]. Усі 27 акцій та індексів мали показник H в інтервалі від 0.51 до 0.599. Автори [23] також показали, що в моделі випадкового блукання, де випадкові прирости мають стійкий розподіл, виконується (1.7)

$$H^{-1} = \alpha \quad (1.7)$$

де H – показник Херста для часового ряду моделі випадкового блукання
 α – параметр стійкого розподілу випадкових приростів (підрозділ 2.1)

Відповідно, для лог-дохідностей акцій, де прийнята гіпотеза про стійкий розподіл, H^{-1} та α мали бути близькими. Дослідники отримали похибку $|H^{-1} - \alpha|$ від 0.004 (мінімальна) до 0.27 (максимальна).

У статті [22] розглянуто наближення стійким розподілом лог-дохідностей індексу DJIA (Dow Jones Industrial Average) в період 1987 по 1994 та акцій Boeing з 1997 по 2003 роки. Параметри стійкого розподілу $S(\alpha, \beta, c, \delta)$ оцінювались методом Кутрувелліса (підрозділ 2.1). Перевірка гіпотези про стійкий розподіл здійснювалась критеріями А-Д та К-С. Оцінка рівня p здійснювалась методом Монте-Карло [22, С. 19]. Для цього у вибірці розміром

n оцінювались параметри $\hat{\alpha}, \hat{\beta}, \hat{c}, \hat{\delta}$, далі генерувалось багато вибірок стійкого розподілу з параметрами $\hat{\alpha}, \hat{\beta}, \hat{c}, \hat{\delta}$ розміру n . Рівень p оцінювався за формулою

$$p = \frac{1}{N} \sum_{k=1}^N 1\{p_i \geq p_0\} \quad (1.8)$$

де $1\{\cdot\}$ – індикатор події

p_i – значення статистики критерію К-С/А-Д для згенерованої вибірки i

p_0 – значення статистики К-С/А-Д для початкової вибірки

N – кількість згенерованих вибірок

Для DJIA метод дав для А-Д $p = 0.02$, для К-С $p = 0.5$. Отже принаймні для критерію узгодженості К-С дані не суперечать гіпотезі про стійкий розподіл на рівні $p > 0.05$. Гіпотеза про гаусівський розподіл була відхилена на рівні $p < 0.005$ для А-Д та К-С. Для акцій Boeing застосування описаної техніки дало прийняття гіпотези про стійкість розподілу на рівні $p = 0.18$ для А-Д та $p = 0.8$ для К-С, що означало прийняття гіпотези. Гіпотеза про гаусівський розподіл була відхилена на рівні $p < 0.005$ (для А-Д та К-С). Результати наближення розподілами для DJIA зображено на рис 1.3 та рис 1.4.

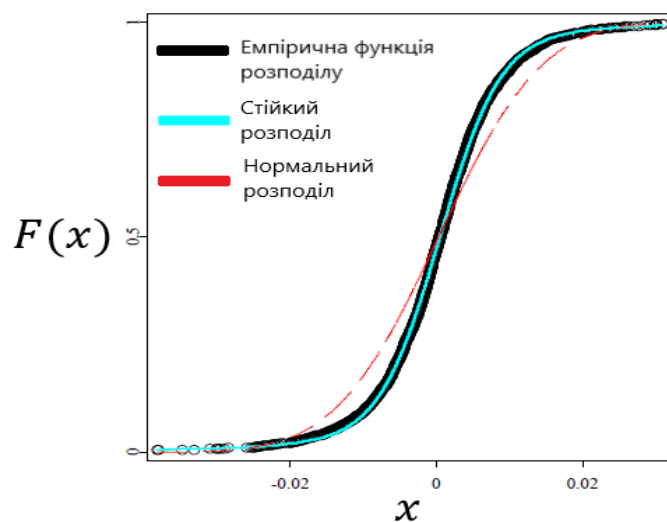


Рисунок 1.3 – Порівняння емпіричної функції розподілу лог-дохідностей індексу США DJIA з теоретичними функціями розподілу [22]

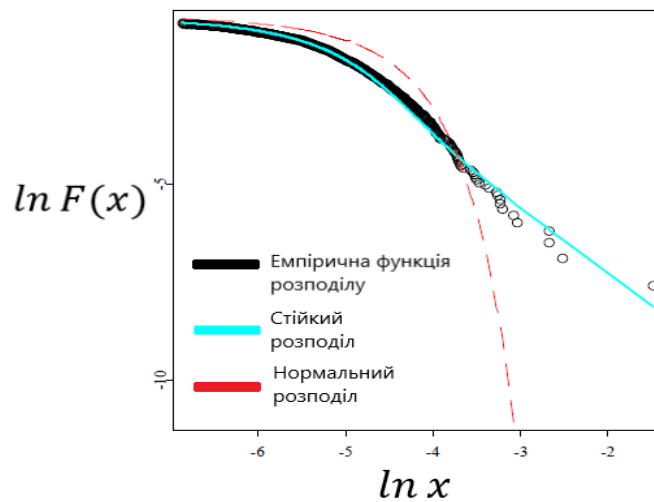


Рисунок 1.4 – Порівняння емпіричної функції розподілу лог-дохідностей індексу DJIA з теоретичними функціями розподілу на $(-\infty, 0)$ у подвійній логарифмічній шкалі [22]

У публікації [29] стійкий розподіл застосовується не лише як модель розподілу лог-дохідностей акцій та індексів, його параметри використовуються як передвісник криз і краху бірж. Також у цьому дослідженні було побудовано емпіричну функцію щільності для лог-дохідностей індексу DJIA з 1900 по 2019 рік та порівняно з кривою гаусівського розподілу з оціненими параметрами. Результат дозволяє зробити висновок, що в жодні роки нормальний розподіл не був адекватною моделлю.

У роботі [29] також лог-дохідності 27 індексів (дані за весь доступний час) було наближено стійким розподілом $S(\alpha, \beta, c, \delta)$ і виявлено, що великі α та β (для β за модулем) характерні для міцних економік. Наприклад, для Nikkei 225 (Японія) $\alpha = 1.71, \beta = -0.24$; для Swiss Market Index (Швейцарія) $\alpha = 1.69, \beta = -0.21$; для S&P Merval (Аргентина) $\alpha = 0.74, \beta = -0.055$.

Автори досліджували зміну параметрів стійкого розподілу з часом для лог-дохідностей акцій DJIA з 2000 до 2019 року [29]. Було використане ковзне вікно 500 днів, в якому оцінювались параметри α, β , потім відбувався зсув на день. На рис 1.5 видно, що перед крахами біржі α, β часто спадав. Було зроблено висновок, що існує залежність між крахами та параметрами α, β .

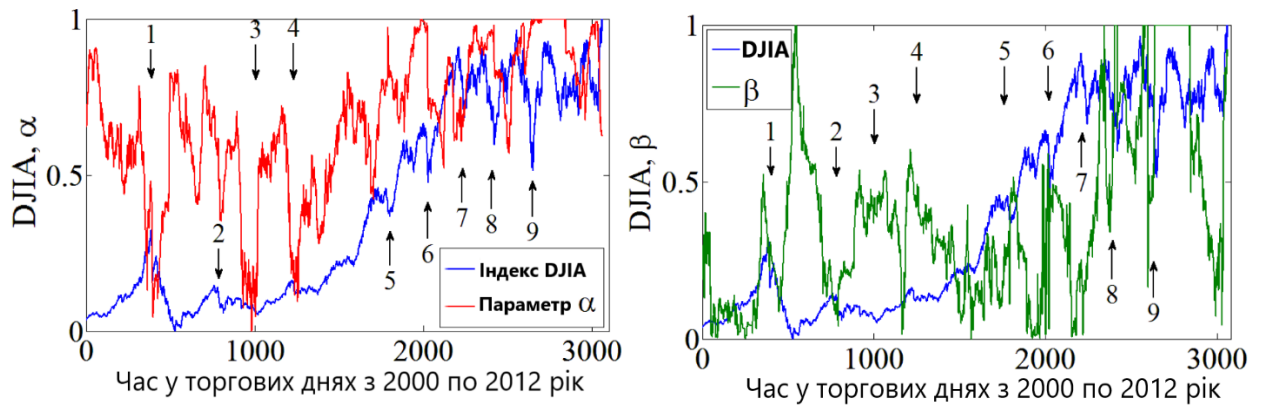


Рисунок 1.5 – Порівняння ціни індексу DJIA та оцінені за допомогою ковзного вікна параметри α, β . Цифрами показано екстремальні ситуації на біржі [29]

Такі передвісники мають математичне обґрунтування – у стійкому розподілі параметр α регулює «важкість хвосту», тобто ймовірність реалізацій значень далеких від нуля, β регулює зміщення моди від нуля вправо (підрозділ 2.1). Отже малий α означає більший ризик великих доходів та збитків, малий β означає більшу ймовірність від’ємних значень (тобто фінансових втрат).

Аналогічні результати отримали інші дослідники [20], які наближували лог-дохідності Nikkei 225 та S&P 500 стійкими розподілами через ковзне вікно. Перед банкрутством Lehman Brothers 15 вересня 2008 року можна помітити суттєве зниження α та β для обох індексів (рисунок 1.6).

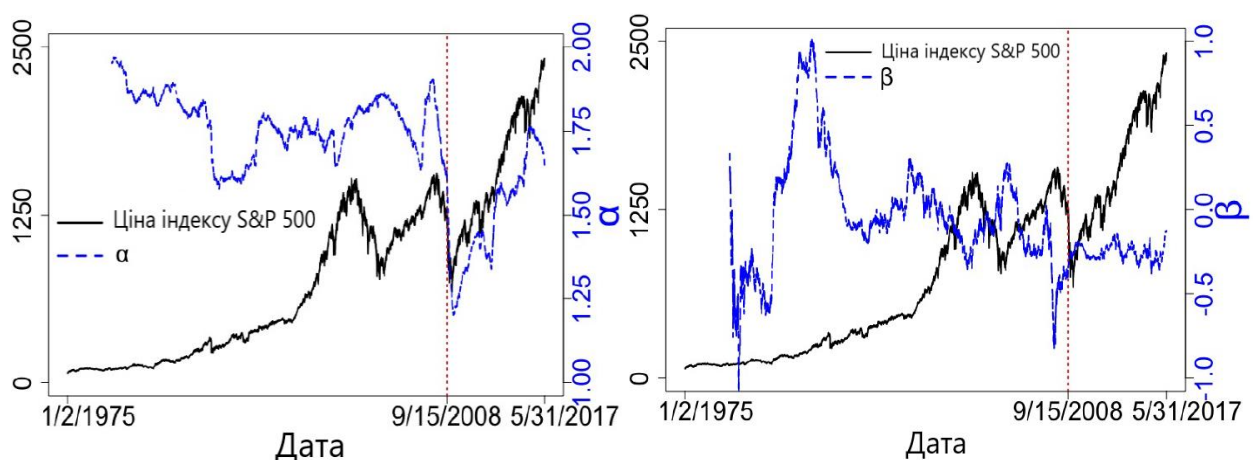


Рисунок 1.6 – Порівняння індексу S&P 500 та параметрів стійкого розподілу: індексу важкості хвоста α (зліва) та асиметрії β (справа). 09.15.2008 – банкрутство Lehman Brothers, початок світової фінансової кризи [20]

У іншому дослідженні [21] лог-дохідності S&P 500 з 1957 до 2017 року покладались однаково розподіленими та наближувались розподілами з важкими хвостами, зокрема, скошеним t-розподілом, узагальненим гіперболічним. Скошений t-розподіл є узагальненням t-розподілу – його асиметрія залежить від параметра β , при його нульовому значенні він стає звичайним розподілом Стюдента. Саме він дав найкращі наближення.

Інше дослідження розглядало лог-дохідності індексів DAX та WIG20 з 2000 по 2009 рік [18]. Особливістю цієї роботи в тому, що у лог-дохідностях акції була врахована волатильність. Щоб оцінити волатильність в кожен момент часу t використовувалась модель GARCH(1,1). Особливістю цієї моделі є те, що на відміну від звичайної оцінки стандартного відхилення вона миттєво реагує на останні зміни цін. Вона задається формулами (1.9) та (1.10)

$$r_t = \sigma_t \varepsilon_t \quad (1.9)$$

$$\sigma_t^2 = c_0 + c_1 r_{t-1}^2 + d_1 \sigma_{t-1}^2 \quad (1.10)$$

де r_t – лог-дохідностей в момент часу t

σ_t – волатильність в момент часу t , яка визначається за формулою

c_0, c_1, d_1 – константи

ε_t – нормально розподілена випадкова величина, $\varepsilon_t \sim N(0,1)$

Після оцінки параметрів моделі ставала відомою волатильність $\{\sigma_t\}$. Робився перехід від $\{r_t\}$ до $\{\varepsilon_t\}$ поелементним діленням r_t на σ_t . Це проілюстровано на рисунку 1.7. Елементи $\{\varepsilon_t\}$ покладались незалежними, однаково розподіленими та наближувались різними розподілами: стійким (підрозділ 2.1), гаусівським, нормальним оберненим гаусівським та гіперболічним (підрозділ 2.2). Для індексів DAX та WIG20 найточнішим розподілом виявився гіперболічний [18]. Порівняння розподілів у подвійній логарифмічній шкалі показано на рисунку 1.8.

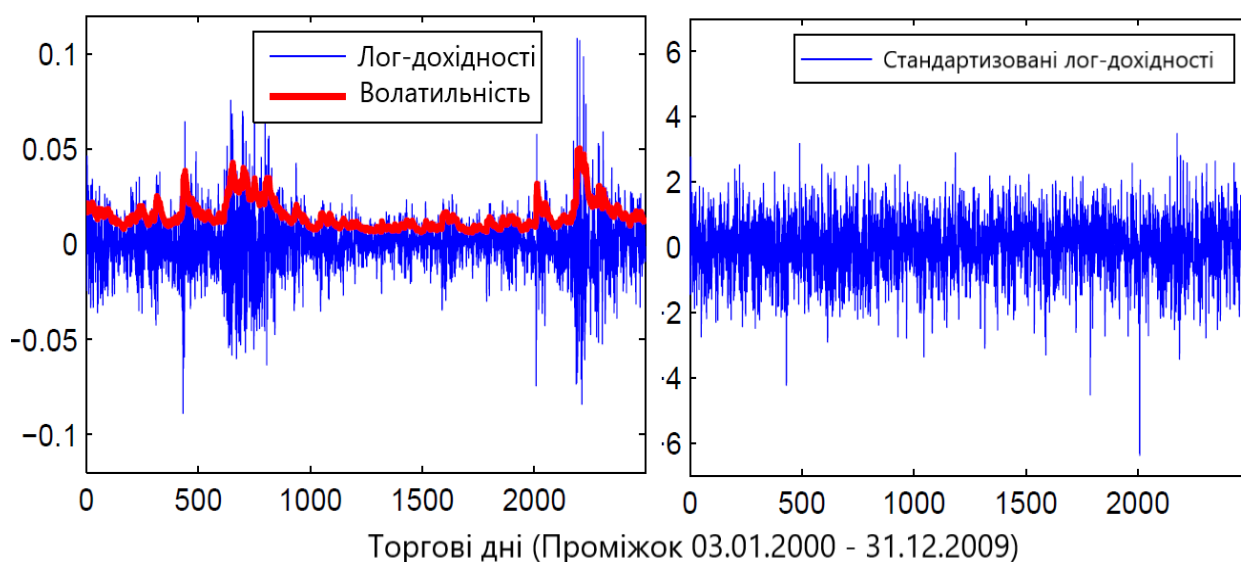


Рисунок 1.7 – Лог-дохідності індексу DAX $\{r_t\}$, оцінка волатильності (графік зліва) та відповідний стандартизований ряд $\{\varepsilon_t\}$ (графік справа) [18]

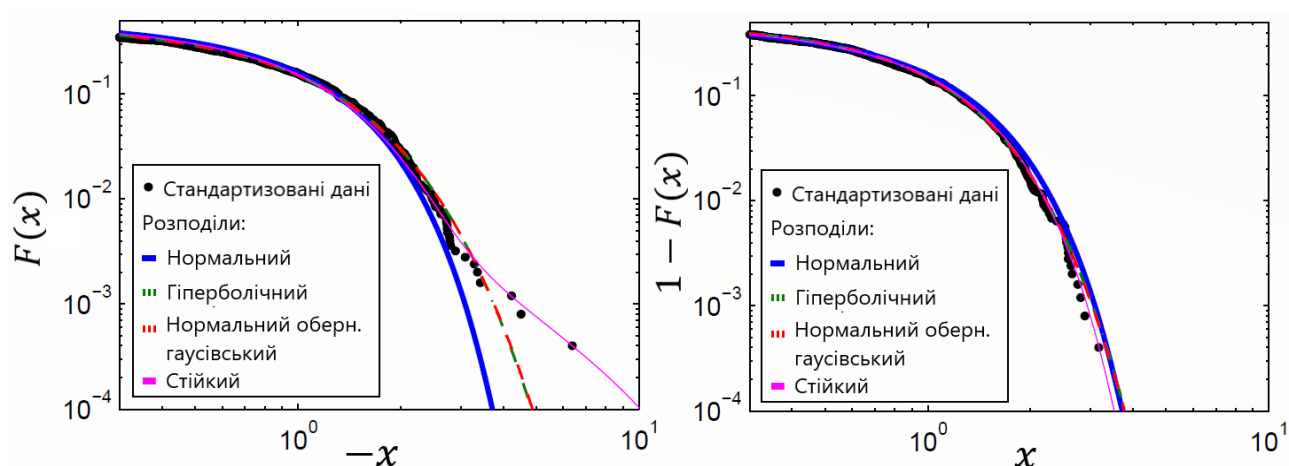


Рисунок 1.8 – Порівняння в логарифмічній шкалі лівого хвоста (зліва) та правого хвоста (справа) емпіричної функції розподілу стандартизованих лог-дохідностей індексу DAX із теоретичними функціями розподілу [18]

1.4 Міра ризику інвестування

Для оцінки ризику вкладів використовується показник, який має назву капітал під ризиком. Вперше поняття з'явились у 1996 р. і згодом стало

загальноприйнятим [30]. Дамо строге математичне означення. Нехай маємо деякий інвестиційний портфель, ціна $X(t)$ якого змінюється в дискретні моменти часу $t, t + 1, t + 2 \dots$

Означення 9. Капітал під ризиком (англ. Value-at-Risk, скор. VaR) для рівня достовірності $c \in [0,1]$ на проміжку часу $T > 0$ визначається як найбільше дійсне число VaR, що виконується (1.11) [18, С. 1]

$$P(L_T > VaR) \leq 1 - c \quad (1.11)$$

де L_T – максимальні втрати вартості портфеля за час T

Наприклад, нехай $VaR = 1$ млн \$ для рівня достовірності 98% на проміжку часу 10 днів. Це означає, що з ймовірністю 0.98 протягом 10 днів не відбудеться падіння ціни інвестиційного портфеля більше ніж на 1 млн \$. Існує також означення VaR через лог-дохідності.

Означення 10. Капітал під ризиком лог-дохідностей для рівня достовірності $c \in [0,1]$ на одиницю часу вперед визначається як

$$P(\ln S_t - \ln S_{t-1} < -VaR_{log}) \leq 1 - c \quad (1.12)$$

де S_i – ціна портфелю в момент часу i ;

Мінус у формулі (1.12) стоїть щоб значення VaR було додатнім. Рівень достовірності частіше задається у відсотках. На практиці зазвичай обчислюють VaR для лог-дохідностей, потім переходять від них до звичайного VaR на один день вперед наближено за формулою [31, С. 328]

$$VaR = S_t \cdot VaR_{log} \quad (1.13)$$

де VaR – капітал під ризиком для рівня c на день вперед, означення 9

S_t – ціна інвестиційного портфелю в момент часу t

VaR_{log} – капітал під ризиком через лог-дохідності, означення 10

Існують різні способи оцінки VaR. Параметричні припускають певний розподіл лог-дохідностей [32, С. 108]. Є методи, що не покладають ніякого розподілу, а працюють лише з вибіркою (наприклад, оцінюючи квантиль). Метод Монте-Карло симулює багато часових рядів, на основі результатів моделювання оцінює VaR [33]. Фільтровані історичні симуляції наближують часовий ряд деякою моделлю (наприклад, ARMA чи GARCH), щоб знайти залежність від попередніх значень та волатильність. На основі цієї моделі можна з ряду лог-дохідностей $\{r_t\}$ отримати стандартизований часовий ряд $\{\varepsilon_t\}$, де елементи незалежні та однаково розподілені. Далі навмання обираються елементи $\{\varepsilon_t\}$, будуються нові часові ряди аналогічно як це робиться в методі Монте-Карло. Із них вже оцінюється VaR. Методи описано в підрозділі 2.3.

Існують статистичні критерії, які перевіряють, що дані не протирічають гіпотезі про правильність оцінки VaR подібно до критерію узгодженості для розподілу (підрозділ 3.3). На практиці задача знаходження VaR для портфелю є складною навіть якщо точно відомі розподіли усіх акцій у ньому. Наприклад, візьмемо акцію A та знайдемо для неї VaR; потім так само обчислимо VaR для акції B. Але якщо розглянемо портфель із акцій A та B і обчислимо для нього VaR, він не дорівнюватиме сумі VaR для акції A та B [18, С. 1].

Причина в тому, що ціна акцій залежать одна від одної. Наприклад, зростання ціни однієї компанії може призвести до падіння ціни компанії-конкурента. Інвестори можуть використати це явище для створення більш надійного портфелю.

Можна виділити два підходи до оцінки VaR портфелю. Перший – розглядати лог-дохідності за формулою (1.1). Другий підхід – розкласти портфель на кілька складових, моделюючи між компонентами взаємозалежність. Якщо обрано останній варіант, для моделювання використовуються так-звані копули (англ. *copula*) – функції, які описують взаємозв'язок між розподілом випадкового вектора $(\xi_1, \xi_2, \dots, \xi_n)$ та його

маргінальними розподілами $\xi_1, \xi_2, \dots, \xi_n$. Для цього обирається деякий вид копули, що залежить від параметрів, які оцінюються по вибірці. Потім проводиться моделювання методом Монте-Карло і рахується значення VaR.

Розглянемо для прикладу дослідження [18], яке кількома способами рахувалось VaR для портфелю, який об'єднував один індекс DAX (30 німецьких акцій) та один індекс WIG20 (20 польських акцій), аналізувались дані з 2001 по 2009 рік. Взаємозв'язок між індексами встановлювався формулою Клейтона та формулою Франка. Маргінальні розподіли DAX та WIG20 наближувались нормальним оберненим гаусівським (скор. NIG) та гіперболічним (скор. Нур) (підрозділ 2.2). Було 4 варіанти моделі: Clayton(Нур, Нур), Frank(Нур, Нур), Clayton(NIG, NIG), Frank(NIG, NIG).

У якості «еталонного» методу VaR розраховувався ще через фільтровані історичні симуляції. Точність VaR оцінювалась за допомогою тесту Крістофферсена. Модель Clayton(NIG,NIG) була кращою, суттєво точнішою за ту, що була отримана фільтрованими історичними симуляціями [18, С. 30].

Є публікації, де обчислюється VaR індексів індійського ринку SENSEX, NIFTY. Розрахунок проводився методом фільтрованих історичних симуляцій, для якого випробовувались 2 моделі: GARCH із додаванням даних про лог-дохідності з західного ринку (S&P500, ціни на золото і т.д.) та модель ARMA-GARCH без додаткових даних. Перша модель, яка враховувала дані з західного ринку, виявилась більш точною [34].

Висновки до розділу

Найбільш розповсюдженими інструментами фондового ринку є прості акції, в той час як міжнародні індекси є показниками економічного стану країн. Результативна робота на фінансовому ринку вимагає використання сучасних

статистичних моделей. Довгий час використовувався нормальний розподіл для лог-дохідностей акцій. Однак він недооцінює ймовірність реалізації великих числових значень, які у менеджменті ризиків асоціюються зі значними прибутками або збитками, які іноді можуть бути катастрофічними.

У першому розділі проводиться аналіз предметної області: розкривається актуальність проблеми, наводяться базові поняття фінансової математики, які необхідні для роботи з моделями поведінки акцій та міжнародних індексів; описується історія розвитку цих моделей та їх сучасний стан. Детально висвітлено та порівняно результати огляду публікацій, де для моделювання лог-дохідностей замість нормального розподілу застосовуються альтернативні. На їх основі обрано розподіли, які доцільно дослідити та апробувати в цій роботі. Частина з них відносяться до розподілів із важкими хвостами, інша, як буде показано далі (підрозділ 3.2), до розподілів із напівважкими хвостами.

На основі проведеного аналізу предметної області можна впевнено зробити висновок, що розподіли лог-дохідностей акцій та індексів не відповідають нормальному розподілу. Однак із введенням альтернативних розподілів виникає необхідність знайти оцінки їх параметрів; ознаки, які дозволяють досліднику звужити коло «кандидатів» на роль найкращого розподілу.

Тому актуальним питанням є пошук та аналіз характеристик оцінок параметрів різних розподілів із важкими хвостами, визначення властивостей, які можуть бути використані як «підказки» для дослідника, щоб звужити множину кращих моделей.

Наступний розділ буде присвячено формальному математичному опису розподілів, які використовуються в моделюванні цін акцій, їх властивостей, методів оцінки параметрів, а також способам обчислення показника ризику інвестування.

2 РОЗПОДІЛИ З ВАЖКИМИ ХВОСТАМИ, ЇХ ВЛАСТИВОСТІ, ОЦІНКИ ПАРАМЕТРІВ ТА ПОКАЗНИКИ РИЗИКУ ІНВЕСТИВАННЯ

2.1 Розподіли з важкими хвостами

2.1.1 Стійкий розподіл

У підрозділі 1.2 було обґрунтовано вибір класу розподілів із важкими хвостами для моделювання лог-дохідностей акцій та міжнародних індексів. Зокрема, такими є стійкі розподіли. Наведемо строге означення.

Означення 1.1. Стійкий розподіл – такий розподіл, для якого справедлива умова (2.1) [13, С. 162]

$$\forall a_1, a_2 \in \mathbb{R}^+, b_1, b_2 \in \mathbb{R}, \exists a \in \mathbb{R}^+, b \in \mathbb{R}: F(a_1x + b_1) * F(a_2x + b_2) = F(ax + b) \quad (2.1)$$

де $F(x)$ – функція розподілу

Згадаємо, що згортка функцій розподілу – функція розподілу суми незалежних випадкових величин. Звідси стає зрозумілим значення терміну «стійкий» – сума двох стійких випадкових величин знову стійка випадкова величина з точністю до розтягування в a раз та зміщення на b .

Теорема 1. Закон розподілу є стійким тоді і тільки тоді коли для нього виконується співвідношення (2.2) [13, С. 164]

$$\ln \chi(t) = i\delta t - c^\alpha |t|^\alpha \left(1 + i\beta \frac{t}{|t|} \omega(t, \alpha) \right), \quad (2.2)$$

$$\omega(t, \alpha) = \begin{cases} -tg\left(\frac{\pi}{2}\alpha\right), & \alpha \neq 1 \\ \frac{2}{\pi} \ln|t|, & \alpha = 1 \end{cases} \quad (2.3)$$

де $\alpha \in (0,2]$; $\beta \in [-1,1]$; $c \in [0, +\infty)$; $\delta \in \mathbb{R}$ – параметри розподілу $\chi(t)$ – його характеристична функція

Таким чином для опису стійкого розподілу необхідно чотири параметри, будемо позначати його як $S(\alpha, \beta, c, \delta)$. Кожному параметру можна приписати задання деяких властивостей:

а) α (індекс стійкості, англ. index of stability) – задає швидкість спадання щільності розподілу при $x \rightarrow \infty$, «важкість» хвоста. Чим ближче α до нуля, тим повільніше спадає щільність;

б) β (параметр скошеності, англ. skewness parameter) – задає асиметрію: якщо $\beta > 0$, тоді мода зміщена вліво; якщо $\beta < 0$ – вправо. Випадок $\beta = 0$ відповідає симетрії;

в) c (масштабний параметр, англ. scale parameter) – є мірою розсіювання розподілу навколо нуля, аналогічно дисперсії (якої при $\alpha < 2$ не існує);

г) δ (параметр положення, англ. location parameter) – параметр задає математичне сподівання при $\alpha > 1$.

Застосування стійких розподілів може адекватно пояснює дані, які характеризуються асиметрією та важкими хвостами. Як покаже теорема 2, вони дозволяють описати явища коли значення, які спостерігаються, є сумою великої кількості факторів, вклад кожного з яких малий, такими є ціни акцій.

Теорема 2. Закон розподілу є стійким тоді і тільки тоді коли існує деяка послідовність $\{\zeta_n\}$ (2.4), що є слабка збіжність до цього закону [13, С. 162]

$$\zeta_n = \frac{\xi_1 + \xi_2 + \dots + \xi_n}{B_n} - A_n \quad (2.4)$$

де $\xi_1, \dots, \xi_i, \dots, \xi_n$ – незалежні однаково розподілені випадкові величини

$\{A_n\}, \{B_n\}$ – деякі послідовності дійсних чисел

Таким чином, використання стійких розподілів не суперечить уявленню про приріст ціни акції як результату дії великої кількості факторів (підрозділ 1.2).

Проведемо обчислювальні експерименти з параметрами стійкого розподілу. Використано бібліотеку Scipy мови програмування Python, яка дозволяє порахувати наближено щільність розподілу (бібліотеки описано у підрозділі 3.1). Для розподілу $S(\alpha, \beta, c, \delta)$ з фіксованими параметрами $\alpha = 0.9$; $\delta = 0$; $c = 2$, значення β варіювалось в інтервалі $[-1, 1]$. Щільності при $\beta \in \{0; -0.4; -0.9; 0.8\}$ зображено на рисунку 2.1 (зліва). Аналогічно при фіксованих $\beta = 0$; $\delta = 0$; $c = 2$ змінювався параметр α в інтервалі $(0, 2]$. Щільності при $\alpha \in \{2; 1.5; 0.9; 0.7\}$ зображено на рисунку 2.1 (справа).

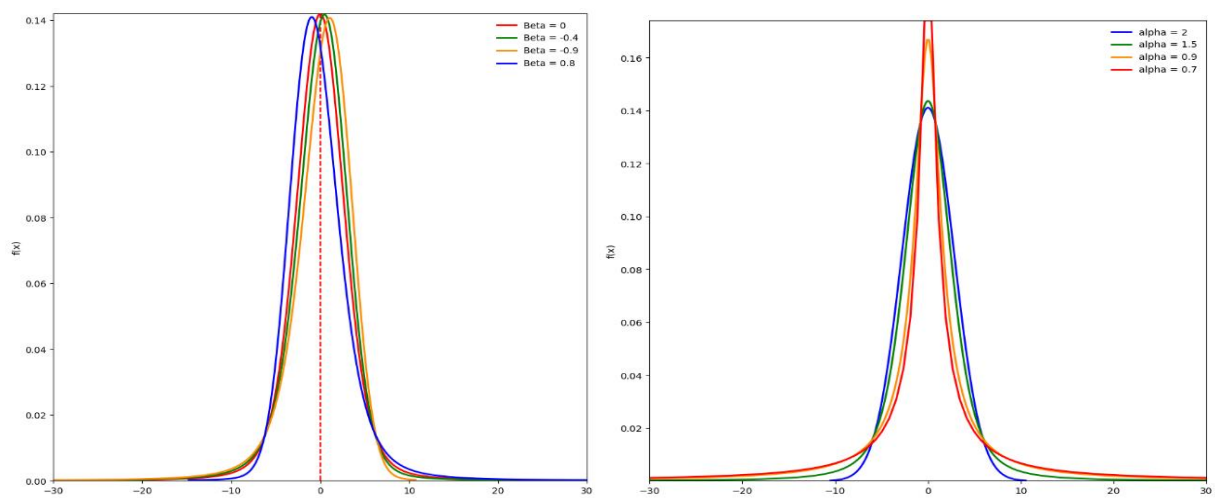


Рисунок 2.1 – Результати обчислювального експерименту для варіювання β (зліва) та для варіювання α (справа) при інших фіксованих параметрах для розподілу $S(\alpha, \beta, c, \delta)$

Аналогічно можна перевірити властивості для параметрів c та δ . Наведемо основні властивості стійких розподілів.

Властивість 1. Дана випадкова величина $\xi \sim S(\alpha, \beta, c, \delta)$. Тоді його лінійна комбінація теж є стійким розподілом.

$$w\xi + b \sim S(\alpha, \beta \cdot \text{sgn}(w), |w|c, w\delta + b) \quad \text{при } \alpha \neq 1, \quad (2.5)$$

$$w\xi + b \sim S\left(1, \beta \cdot \text{sgn}(w), |w|c, w\delta + b - \frac{2\beta wc}{\pi} \ln|w|\right) \quad \text{при } \alpha = 1 \quad (2.6)$$

де w, b – дійсні коефіцієнти

Доведення:

$$\chi_{\xi}(t) = E e^{i\xi t} = e^{i\delta t - c^{\alpha}|t|^{\alpha} \left(1 + i\beta \frac{t}{|t|} \omega(t, \alpha)\right)}$$

$$\chi_{w\xi+b}(t) = E e^{i\xi(wt)} e^{ibt} = \chi_{\xi}(wt) e^{ibt} = e^{i(w\delta+b)t - c^{\alpha}|wt|^{\alpha} \left(1 + i\beta \frac{wt}{|wt|} \omega(wt, \alpha)\right)};$$

$$\chi_{w\xi+b}(t) = e^{i(w\delta+b)t - (|w|c)^{\alpha}|t|^{\alpha} \left(1 + i\beta \cdot \text{sgn}(w) \frac{t}{|t|} \omega(wt, \alpha)\right)};$$

Розглянемо випадок $\alpha \neq 1$. Тоді $\omega(wt, \alpha) = \omega(t, \alpha)$

$$\chi_{w\xi+b}(t) = e^{i(w\delta+b)t - (|w|c)^{\alpha}|t|^{\alpha} \left(1 + i\beta \cdot \text{sgn}(w) \frac{t}{|t|} \omega(t, \alpha)\right)}$$

Отже, у випадку $\alpha \neq 1$ маємо $w\xi + b \sim S(\alpha, \beta \cdot \text{sgn}(w), |w|c, w\delta + b)$

Розглянемо випадок $\alpha = 1$:

$$\begin{aligned} \ln \chi_{w\xi+b}(t) &= i(w\delta + b)t - (|w|c)|t| \left(1 + i\beta \cdot \text{sgn}(w) \frac{t}{|t|} \cdot \frac{2}{\pi} \ln|wt|\right) = \\ &= i(w\delta + b)t - |w|c \left(|t| + i \frac{2\beta}{\pi} \text{sgn}(w) (\ln|w| + \ln|t|)t\right) = \\ &= i \left(w\delta + b - \frac{2\beta|w|c}{\pi} \text{sgn}(w) \ln|w|\right)t - |w|c \left(|t| + i \frac{2\beta}{\pi} \text{sgn}(w) \ln|t|t\right) = \\ &= i \left(w\delta + b - \frac{2\beta|w|c}{\pi} \text{sgn}(w) \ln|w|\right)t - (|w|c)|t| \left(1 + i\beta \text{sgn}(w) \cdot \frac{t}{|t|} \cdot \frac{2}{\pi} \ln|t|\right) \end{aligned}$$

Остаточно, $w\xi + b \sim S\left(1, \beta \cdot \text{sgn}(w), |w|c, w\delta + b - \frac{2\beta wc}{\pi} \ln|w|\right)$.

Властивість 2. Нехай $\xi_1, \xi_2, \xi_3, \dots, \xi_n$ незалежні однаково розподілені випадкові величини $\xi_k \sim S(\alpha, \beta, c, \delta)$. Тоді виконується (2.7) [23, С. 36]

$$\eta = \sum_{k=1}^n \xi_k \sim S\left(\alpha, \beta, c \cdot n^{\frac{1}{\alpha}}, n\delta\right) \quad (2.7)$$

Властивість 3. Нехай $\xi_i \sim S(\alpha, \beta_i, c_i, \delta_i), i = \overline{1,2}$, незалежні випадкові величини мають стійкий розподіл. Тоді сума має стійкий розподіл [23,С. 36]

$$\xi_1 + \xi_2 \sim S\left(\alpha, \frac{\beta_1 c_1^\alpha + \beta_2 c_2^\alpha}{c_1^\alpha + c_2^\alpha}, (c_1^\alpha + c_2^\alpha)^{\frac{1}{\alpha}}, \delta_1 + \delta_2\right) \quad (2.8)$$

Властивість 4. Моменти p -го порядку стійкого розподілу скінчені тоді і тільки тоді коли $p < \alpha$. При $\alpha > 1$ мат. сподівання дорівнює δ [23,С. 36].

Властивість 5. Стійкий розподіл відноситься до розподілів із важкими хвостами при $\alpha < 2$. Це негайно слідує із того, що наявність важких хвостів еквівалентне нескінченості моменту деякого порядку (підрозділ 1.2) та того, що при $\alpha < 2$ не існує дисперсії [16,С. 6].

Властивість 6. Стійкий розподіл $\xi \sim S(2, 0, c, \delta)$ є гаусівською випадковою величиною $\xi \sim N(\delta, 2c^2)$ [22,С. 4].

Властивість 7. Стійкий розподіл $\xi \sim S(1, 0, c, \delta)$ є зміщеним вправо на δ розподілом Коші із параметром c [22,С. 4].

Розглянемо питання симуляції стійких випадкових величин на обчислювальній машині. Через відсутність аналітичного вигляду щільності та функції розподілу більшість загальних алгоритмів застосувати не можна. Однак у 1976 р. було запропоновано зручний метод [22,С. 9]. Опишемо цей алгоритм генерації стійких випадкових величин.

а) генерувати дві випадкові величини $V \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right); W \sim Exp(1)$

б) на основі V, W обчислити X . Якщо $\alpha \neq 1$ тоді застосовується (2.9)

$$X = S_{\alpha, \beta} \cdot \frac{\sin\{\alpha(V + B_{\alpha, \beta})\}}{\{\cos(V)\}^{\frac{1}{\alpha}}} \cdot \left[\frac{\cos\{V - \alpha(V + B_{\alpha, \beta})\}}{W} \right]^{\frac{1-\alpha}{\alpha}} \quad (2.9)$$

де α, β – параметри стійкого розподілу, який бажаємо згенерувати

$$B_{\alpha,\beta} = \frac{\arctg\left(\beta \cdot tg\left(\frac{\pi\alpha}{2}\right)\right)}{\alpha} - \text{коефіцієнт залежний від параметрів } \alpha, \beta$$

$$S_{\alpha,\beta} = \left\{1 + \beta^2 \cdot tg^2\left(\frac{\pi\alpha}{2}\right)\right\}^{\frac{1}{2\alpha}} - \text{коефіцієнт залежний від параметрів } \alpha, \beta$$

Якщо $\alpha = 1$ обчислити X за формулою (2.10)

$$X = \frac{2}{\pi} \cdot \left\{ \left(\frac{\pi}{2} + \beta V \right) tg(V) - \beta \cdot \ln \left(\frac{\frac{\pi}{2} \cdot W \cdot \cos V}{\frac{\pi}{2} + \beta V} \right) \right\} \quad (2.10)$$

Тоді отриманий з (2.9) чи (2.10) X має стійкий розподіл, $X \sim S(\alpha, \beta, 1, 0)$

в) обчислити Y – це зміщена на δ та масштабована на c реалізація стійкої випадкової величини X

$$Y = \begin{cases} cX + \delta, & \alpha \neq 1 \\ cX + \frac{2}{\pi} \beta \ln c + \delta, & \alpha = 1 \end{cases} \quad (2.11)$$

де α, β, c, δ – параметри стійкого розподілу, який бажаємо згенерувати.

Цікаво, що для $\alpha = 2$ процедура переходить в алгоритм Бокса-Мюллера для генерації гаусівських випадкових величин [22, С. 9].

Розглянемо питання оцінки параметрів стійких розподілів. На основі аналізу літератури можна виділити такі: метод МакКаллока, метод Кутрувеліса, метод логарифмічних моментів, метод максимальної правдоподібності.

Опишемо метод МакКаллока, який опубліковано в [35]. Він вважається найпростішим, але найменш точним [22, С. 13]. Спочатку обчислюються v_α, v_β

$$v_\alpha = \frac{\widehat{x_{0.95}} - \widehat{x_{0.05}}}{\widehat{x_{0.75}} - \widehat{x_{0.25}}}; \quad v_\beta = \frac{\widehat{x_{0.95}} + \widehat{x_{0.05}} - 2\widehat{x_{0.50}}}{\widehat{x_{0.95}} - \widehat{x_{0.05}}} \quad (2.12)$$

де $\widehat{x}_p$ – емпіричний квантиль p -го порядку обчислений за формулою

$$\widehat{x_{\frac{2i-1}{2n}}} = X_i, i = \overline{1, n} \quad (2.13)$$

де $\widehat{x_{\frac{2i-1}{2n}}}$ – емпіричний квантиль $\frac{2i-1}{2n}$ -го порядку

X_i – i -ий елемент вибірки $X_1, X_2, \dots, X_n$ впорядкованої за зростанням

Таким чином, щоб знайти квантилі порядку p для цього методу слід спочатку впорядкувати вибірку та покласти кожен елемент квантилем $\frac{2i-1}{2n}$ -го порядку.

Якщо серед отриманих таким способом квантилів нема потрібного, недостаючи можна отримати з сусідніх лінійною інтерполяцією. Знайшовши v_α, v_β рахуються оцінки α, β, c, δ

$$\hat{\alpha} = \psi_1(v_\alpha, v_\beta); \quad \hat{\beta} = \psi_2(v_\alpha, v_\beta); \quad \hat{c} = \frac{\widehat{x_{0.75}} - \widehat{x_{0.25}}}{\phi_3(\hat{\alpha}, \hat{\beta})} \quad (2.14)$$

$$\hat{\delta} = \begin{cases} \left(\widehat{x_{0.05}} + \hat{c} \phi_5(\hat{\alpha}, \hat{\beta}) \right) - \hat{\beta} \left(\widehat{x_{0.05}} + \hat{c} \phi_5(\hat{\alpha}, \hat{\beta}) \right) \operatorname{tg} \left(\frac{\pi \hat{\alpha}}{2} \right), & \alpha \neq 1 \\ \widehat{x_{0.05}} + \hat{c} \phi_5(\hat{\alpha}, \hat{\beta}); & \alpha = 1 \end{cases} \quad (2.15)$$

де $\psi_i(x, y), \phi_i(x, y)$ – функції, табличні значення яких приведено в [35]. Метод МакКаллока дає просту в реалізації конзистентну оцінку [18, С. 10].

Більш точним в літературі вважається метод Кутрувеліса який буде зараз описано [18, С. 12]. Для стійкого розподілу справедливі рівності (2.16), (2.17)

$$\ln(-\ln|\chi(t)|^2) = \ln(2c^\alpha) + \alpha \ln|t| \quad (2.16)$$

$$\operatorname{arctg} \left(\frac{\operatorname{Im} \chi(t)}{\operatorname{Re} \chi(t)} \right) = \delta t + \beta |c|^\alpha \operatorname{tg} \left(\frac{\pi \alpha}{2} \right) \operatorname{sgn}(t) \cdot |t|^\alpha \quad (2.17)$$

де $\chi(t)$ – характеристична функція

α, β, c, δ – параметри розподілу

Запишемо емпіричну оцінку характеристичної функції за формулою (2.18)

$$\hat{\chi}(t) = \frac{1}{n} \sum_{j=1}^n e^{itX_j} \quad (2.18)$$

де X_j – елементи вибірки

n – розмір вибірки

Якщо замінити $\chi(t)$ у формулі (2.16) на оцінку $\hat{\chi}(t)$ з формули (2.18) аналогічно як в методі моментів, тоді отримаємо рівняння регресії (2.19)

$$\ln(-\ln|\hat{\chi}(t_k)|^2) = \ln 2 + \alpha(\ln(c) + \ln|t_k|), n = \overline{1, N} \quad (2.19)$$

Маючи $\hat{\alpha}, \hat{c}$ використаємо (2.17) для оцінки β, δ , вийде рівняння регресії (2.20)

$$\arctg\left(\frac{\operatorname{Im} \widehat{\chi(t_k)}}{\operatorname{Re} \widehat{\chi(t_k)}}\right) = \delta t_k + \beta |\hat{c}|^{\hat{\alpha}} t g\left(\frac{\pi \hat{\alpha}}{2}\right) \operatorname{sgn}(t_k) |t_k|^{\hat{\alpha}} + \varepsilon_k^*, k = \overline{1, L} \quad (2.20)$$

де $\widehat{\chi(t_k)}$ – значення емпіричної характеристичної функції в точці t_k

$\hat{\alpha}, \hat{c}$ – оцінки параметрів α, c , які отримані з рівняння регресії (2.19)

ε_k^* – деякі похибки

Опишемо ітераційну процедуру метода Кутрувеліса:

а) покласти $i = 0, K = K_0$, де K_0 – довільне;

б) розв'язати рівняння регресії (2.19), де t_k обирати на основі (2.21);

$$t_k = \frac{\pi k}{25}, k = 1, 2, \dots, K \quad (2.21)$$

в) на основі оцінок $\hat{\alpha}_i, \hat{c}_i$ оцінити $\hat{\beta}_i, \hat{\delta}_i$ рівнянням регресії (2.20);

г) якщо для кожної послідовності оцінок $\{\hat{\alpha}_i\}, \{\hat{\beta}_i\}, \{\hat{c}_i\}, \{\hat{\delta}_i\}$ виконується критерій збіжності – зупинити процес і покласти $\hat{\alpha}_i, \hat{\beta}_i, \hat{c}_i, \hat{\delta}_i$ оцінками. Якщо

не виконується – оновити K за табличними значеннями, які є в оригінальній роботі [36, С. 921]. Вибір K залежить від оцінки $\hat{\alpha}_i$ та розміру вибірки.

Розглянемо метод максимальної правдоподібності. Задача запису функції правдоподібності ускладнюється відсутністю аналітичного вигляду щільності стійких розподілів. Однак її можна записати через інтеграл (так звана формула Золотарьова). Через дуже складний вигляд формулу наводити не будемо, обмежившись посиланням [37, С. 760]. Отже, можна записати функцію правдоподібності та знайти параметри максимізуючи функцію

$$L(\alpha, \beta, c, \delta) = \sum_{i=1}^n \ln f_{\alpha, \beta, c, \delta}(z_i) \rightarrow \max \quad (2.22)$$

Була доведена конзистентність, асимптотична нормальність цієї оцінки [38, С. 2]. Для максимізації функції правдоподібності застосовуються чисельні методи. На жаль, математичного обґрунтування збіжності максимізації певним алгоритмом немає. В літературі його описують як найбільш довгий з точки зору обчислювальної складності алгоритм, однак найбільш точний.

Розглянемо метод логарифмічних моментів [38, С. 5]. Він може бути застосований лише коли розподіл має вид $\xi \sim S(\alpha, 0, c, 0)$, нижче буде пояснено як його використати в загальному вигляді.

Теорема 3. Нехай дана вибірка незалежних однаково розподілених випадкових величин $X_1, X_2, X_3, \dots, X_n$, де $X_i \sim S(\alpha, 0, c, 0)$. Тоді для параметрів α, c справедлива оцінка (2.23) та (2.24) відповідно

$$\hat{\alpha}_{LOG}^{(n)} = \left(\max \left(\frac{6}{\pi^2 n} \sum_{i=1}^n \left[\ln |X_i| - \frac{1}{n} \sum_{k=1}^n \ln |X_k| \right]^2 - \frac{1}{2}, \frac{1}{4} \right) \right)^{-\frac{1}{2}} \quad (2.23)$$

$$\hat{c}_{LOG}^{(n)} = \exp \left(\frac{1}{n} \sum_{i=1}^n \ln |X_i| - \left(\frac{1}{\hat{\alpha}_{LOG}^{(n)}} - 1 \right) \gamma^* \right) \quad (2.24)$$

де γ^* – стала Ейлера-Маскероні

X_i – вибірка обсягом n , на основі якої оцінюються параметри

Тоді $(\hat{\alpha}_{LOG}^{(n)}, \hat{c}_{LOG}^{(n)}) \xrightarrow{м.н.} (\alpha, c)$, коли $n \rightarrow \infty$. Оцінка є конзистентною, асимптотично нормальною [38, С. 5]. Метод можна застосувати й в загальному випадку. Якщо $\xi_1, \xi_2 \sim S(\alpha, \beta, c, \delta)$, тоді

$$-\xi_2 \sim S(\alpha, -\beta, c, -\delta) \quad \text{за властивістю 1}$$

$$\xi_1 - \xi_2 \sim S\left(\alpha, 0, 2^{\frac{1}{\alpha}}c, 0\right) \quad \text{за властивістю 3}$$

Таким чином, можемо попарним відніманням від вибірки з n несиметричних стійких розподілів перейти до $\frac{n}{2}$ симетричних, оцінити для них параметри $\hat{\alpha}, \hat{c}$. Можна перейти до початкових параметрів розподілу α, c перетвореннями (2.25), (2.26). Маємо оцінки початкових параметрів α^*, c^* (2.25), (2.26).

$$\alpha^* = \hat{\alpha} \quad (2.25)$$

$$c^* = \hat{c} \cdot 2^{-\frac{1}{\hat{\alpha}}} \quad (2.26)$$

де $\hat{\alpha}, \hat{c}$ – оцінка α та c , відповідно, для вибірки розміру $\frac{n}{2}$, яка отримана попарним відніманням. Для моделювання застосовуються й інші розподіли.

2.1.2 Узагальнений скошений розподіл Стюдента

Означення 12. Скошений узагальнений розподіл Стюдента – розподіл, що має щільність, яка визначається формулою (2.27) [39] (використаємо параметризацію, яка була запропонована в мові програмування R [40, С. 6])

$$f(x) = \frac{p}{2\nu\sigma q^{\frac{1}{p}} B\left(\frac{1}{p}, q\right) \left(\frac{|x-\mu+m|^p}{q(\nu\sigma)^p (\lambda \cdot \text{sgn}(x-\mu+m)+1)^p + 1}\right)^{\frac{1}{p}+q}} \quad (2.27)$$

$$m = \frac{2v\sigma\lambda q^{\frac{1}{p}} \cdot B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)} \quad (2.28)$$

$$v = q^{-\frac{1}{p}} \left((3\lambda^2 + 1) \frac{B\left(\frac{3}{p}, q - \frac{2}{p}\right)}{B\left(\frac{1}{p}, q\right)} - 4\lambda^2 \left(\frac{B\left(\frac{2}{p}, q - \frac{1}{p}\right)}{B\left(\frac{1}{p}, q\right)} \right)^2 \right)^{-\frac{1}{2}} \quad (2.29)$$

де $f(x)$ – щільність розподілу

$\mu, \sigma, \lambda, p, q$ – параметри розподілу ($p > 0, q > 0, \sigma > 0, -1 < \lambda < 1$)

m, v – константи для спрощення вигляду першого та другого моментів

$B(t, s)$ – Бета функція

Параметри задають наступні властивості [41, С. 2]:

- а) μ – мода розподілу (при $pq > 1$ математичне сподівання дорівнює μ);
- б) σ – задає міру розсіювання (якщо $pq > 2$ тоді дисперсія дорівнює σ^2);
- в) λ – задає асиметрію, при $-1 < \lambda < 0$ мода зміщена вліво, якщо показник $0 < \lambda < 1$ мода зміщена вправо, при $\lambda = 0$ розподіл симетричним;
- г) p, q – задають існування моментів. Якщо $pq > k$ тоді існує скінчений момент k -го порядку.

Із властивості г) стає зрозумілим, що скошений узагальнений розподіл Стюдента є розподілом із важкими хвостами (підрозділ 1.2, означення 8).

Якщо покласти $\lambda = \mu = 0$; $p = 2$; $\sigma = \frac{1}{\sqrt{w-2}}$; $q = \frac{w}{2}$ тоді маємо звичайний розподіл Стюдента з щільністю (2.30)

$$f(x) = \frac{1}{\sqrt{w} B\left(\frac{1}{2}, \frac{w}{2}\right)} \left(\frac{x^2}{w} + 1 \right)^{-\frac{w+1}{2}} \quad (2.30)$$

Оскільки щільність відома, для генерації випадкових чисел можна застосувати узагальнену обернену функцію. Для оцінки параметрів застосовується метод максимальної правдоподібності [41, С. 1].

2.2 Розподіли з напівважкими хвостами

У фінансовій математиці знайшли застосування також так-звані розподіли з напівважкими хвостами. Строгого означення немає, в літературі так називають розподіли в яких швидкість спадання щільності при $x \rightarrow \infty$ швидша ніж у стійких, але повільніша ніж у нормального розподілу [42, С. 19].

До таких відноситься, зокрема, узагальнений гіперболічний розподіл. Для моделювання лог-дохідностей акцій та індексів він часто використовуються, в тому числі, його часткові випадки: гіперболічний та нормальний обернений гаусівський.

Означення 13. Узагальнений гіперболічний розподіл – розподіл, який залежить від параметрів $\lambda, \alpha, \beta, \delta, \mu$ із щільністю (2.31) [43, С. 16]

$$f_{GH}(x) = s(\delta^2 + (x - \mu)^2)^{\frac{1}{2}(\lambda - \frac{1}{2})} K_{\lambda - \frac{1}{2}}(\alpha \sqrt{\delta^2 + (x - \mu)^2}) e^{\beta(x - \mu)} \quad (2.31)$$

де $s = \frac{(\alpha^2 - \beta^2)^{\frac{\lambda}{2}}}{\sqrt{2\pi} \alpha^{\lambda - \frac{1}{2}} \delta^{\lambda} K_{\lambda}(\delta \sqrt{\alpha^2 - \beta^2})}$ – нормуюча константа

$K_{\lambda}(x) = \frac{1}{2} \int_0^{+\infty} y^{\lambda-1} e^{-\frac{x}{2}(y+y^{-1})} dy$ – модифікована функція Бесселя 3-го

роду з індексом $\lambda \in \mathbb{R}$. На параметри накладені обмеження (2.32), (2.33), (2.34)

$$\delta \geq 0, |\beta| < \alpha, \text{ якщо } \lambda > 0 \quad (2.32)$$

$$\delta > 0, |\beta| < \alpha, \text{ якщо } \lambda = 0 \quad (2.33)$$

$$\delta > 0, |\beta| \leq \alpha, \text{ якщо } \lambda < 0 \quad (2.34)$$

Розподіл позначається $\xi \sim GH(\lambda, \alpha, \beta, \delta, \mu)$. Деяким його параметрам можна приписати такі властивості [43, С. 17]:

а) λ – регулює тип розподілу. Наприклад, $\lambda = -0.5$ та $\lambda = 1$ виокремлені в самостійні розподіли;

б) α – задає «важкість» хвосту (чим менше α , тим важче хвіст) [43, С. 20];

в) β – задає асиметрію ($\beta < 0$ відповідає зміщенню моди вліво, $\beta > 0$ – вправо, $\beta = 0$ задає симетричний розподіл) [43, С. 19];

г) δ – задає «важкість хвосту» аналогічно α (але тут важчий хвіст відповідає більшому значенню δ) [43, С. 18];

д) μ – задає математичне сподівання за формулою (2.35).

Однак вплив цих параметрів не обмежується згаданими вище властивостями: наприклад, на математичне сподівання та дисперсію впливає не лише μ , як це видно з формул (2.35) та (2.36). Наведемо результати обчислювальних експериментів, які проведено в дослідженні [43] (Рис 2.2, Рис 2.3).

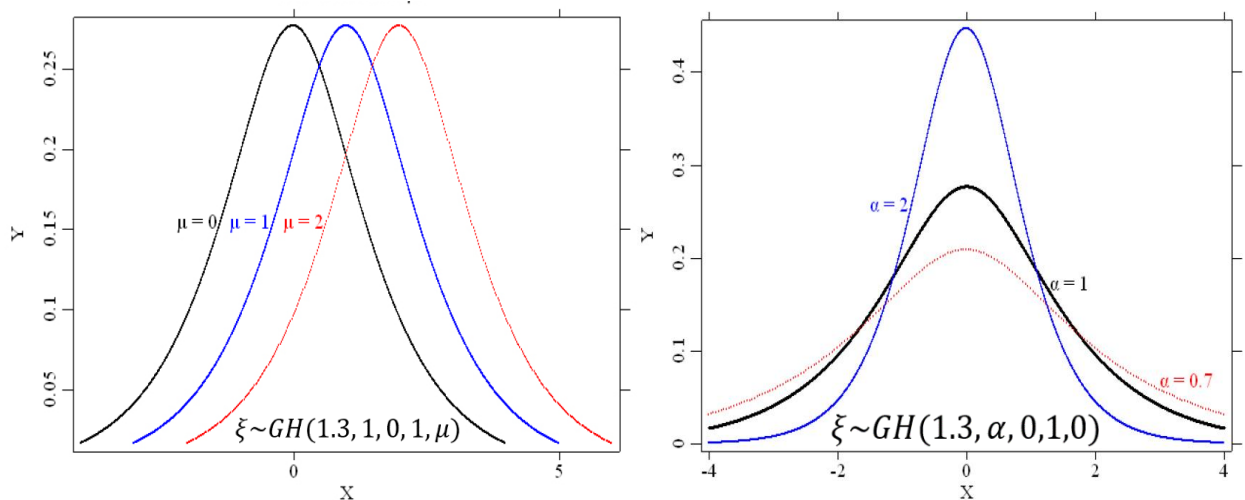


Рисунок 2.2 – Щільність узагальненого гіперболічного розподілу $\xi \sim GH(\lambda, \alpha, \beta, \delta, \mu)$ для змінних параметрів μ (зліва) та α (справа), інші параметри було зафіксовано [43]

З рисунку 2.2 видно, що при зміні параметра μ відбувається зміщення середнього на μ , як це видно з формули (2.35). Зменшення параметра α робить його хвіст більш «важким» — щільність починає спадати більш повільно. На рисунку 2.3 зображено результати аналогічних дослідів, коли змінюються показники δ та β при фіксованих інших параметрах.

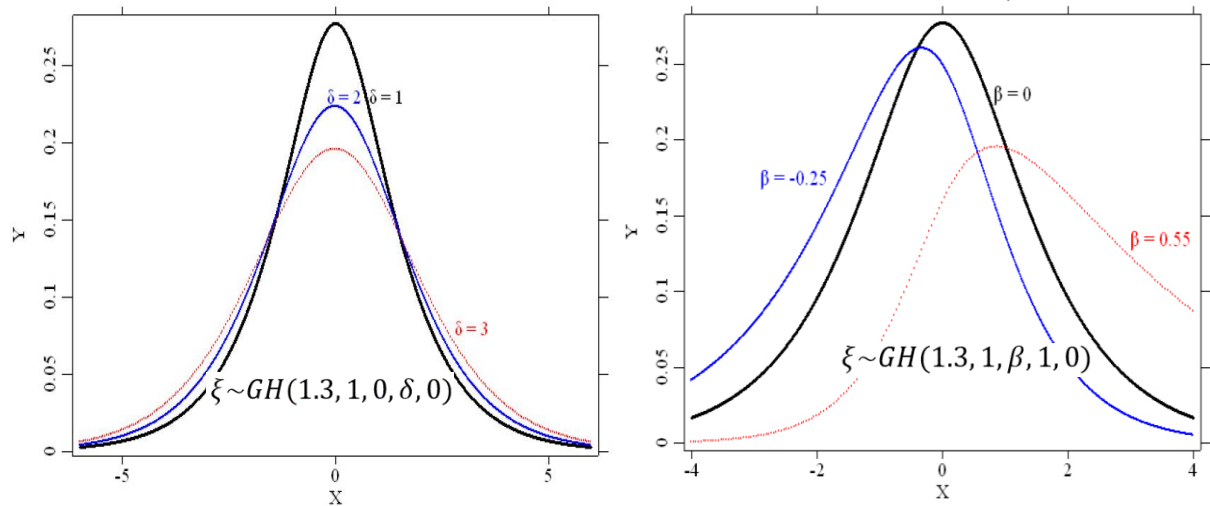


Рисунок 2.3 – Щільність узагальненого гіперболічного розподілу $\xi \sim GH(\lambda, \alpha, \beta, \delta, \mu)$ для змінних параметрів δ (зліва) та β (справа), інші параметри було зафіксовано [43]

Математичне сподівання $E\xi$ та дисперсія $D\xi$ обчислюється за формулами (2.35) та (2.36) [40, С. 24]:

$$E\xi = \mu + \beta \frac{\delta K_{\lambda+1}(\bar{\gamma})}{\gamma K_{\lambda}(\bar{\gamma})}, \quad (2.35)$$

$$D\xi = \delta^2 \left[\frac{K_{\lambda+1}(\bar{\gamma})}{\gamma K_{\lambda}(\bar{\gamma})} + \frac{\beta^2}{\gamma^2} \left(\frac{K_{\lambda+2}(\bar{\gamma})}{K_{\lambda}(\bar{\gamma})} - \left(\frac{K_{\lambda+1}(\bar{\gamma})}{K_{\lambda}(\bar{\gamma})} \right)^2 \right) \right] \quad (2.36)$$

де $\bar{\gamma} = \delta \sqrt{\alpha^2 - \beta^2}$
 $\gamma = \sqrt{\alpha^2 - \beta^2}$

Узагальнений гіперболічний розподіл є нескінченно подільним [47, С. 309], отже, він є границею для суми незалежних випадкових величин [13, С. 116]. Це узгоджується з уявленням приросту ціни акції як суми багатьох випадкових факторів. Для $\alpha = \beta = \mu = 0; \lambda = -\frac{n}{2}; \delta = \sqrt{n}$ розподіл спроститься до розподілу Стюдента зі ступенем свободи n [40, С. 31].

Принаймні для деяких комбінацій параметрів моменти не всіх порядків існують (тому частину представників узагальненого гіперболічного розподілу можна віднести до розподілів із важкими хвостами). Характеристична функція має вид (2.37)

$$\chi_{\lambda,\alpha,\beta,\delta,\mu}(t) = e^{i\mu t} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + it)^2} \right)^{\frac{\lambda}{2}} \frac{K_{\lambda}(\delta \sqrt{\alpha^2 - (\beta + it)^2})}{K_{\lambda}(\delta \sqrt{\alpha^2 - \beta^2})} \quad (2.37)$$

Для оцінки параметрів застосовується метод максимальної правдоподібності, який максимізує функцію чисельними методами. Такий алгоритм, наприклад, реалізовано в мові програмування R [41, С. 4].

Означення 14. Нормальний-обернений гаусівський розподіл (англ. normal-inverse Gaussian distribution) – частковий випадок узагальненого гіперболічного розподілу при параметрі $\lambda = -0.5$. Позначається $\xi \sim NIG(\alpha, \beta, \delta, \mu)$ [43, С. 28]. Щільність тоді спрощується до формули (2.38)

$$f_{NIG}(x) = \frac{\alpha \delta}{\pi} \frac{K_1(\alpha \sqrt{\delta^2 + (x - \mu)^2})}{\sqrt{\delta^2 + (x - \mu)^2}} e^{\delta \sqrt{\alpha^2 - \beta^2} + \beta(x - \mu)} \quad (2.38)$$

Властивості параметрів $\alpha, \beta, \delta, \mu$ написані для узагальненого гіперболічного зберігаються, оскільки цей розподіл є його частковим випадком. Модифікована функція Бесселя 3-го роду для $\lambda = -0.5$ має вигляд (2.39)

$$K_{-\frac{1}{2}}(t) = \sqrt{\frac{\pi}{2}} t^{-\frac{1}{2}} e^{-t} \quad (2.39)$$

Характеристична функція тоді теж спрощується

$$\begin{aligned} \chi_{-\frac{1}{2},\alpha,\beta,\delta,\mu}(t) &= e^{i\mu t} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + it)^2} \right)^{-\frac{1}{4}} \cdot \frac{K_{-\frac{1}{2}}(\delta \sqrt{\alpha^2 - (\beta + it)^2})}{K_{-\frac{1}{2}}(\delta \sqrt{\alpha^2 - \beta^2})} = \\ &= e^{i\mu t} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + it)^2} \right)^{-\frac{1}{4}} \frac{(\delta \sqrt{\alpha^2 - (\beta + it)^2})^{-\frac{1}{2}} e^{-\delta \sqrt{\alpha^2 - (\beta + it)^2}}}{(\delta \sqrt{\alpha^2 - \beta^2})^{-\frac{1}{2}} e^{-\delta \sqrt{\alpha^2 - \beta^2}}}; \end{aligned}$$

Після спрощення характеристична функція прийме вигляд (2.40)

$$\chi_{-\frac{1}{2},\alpha,\beta,\delta,\mu}(t) = e^{i\mu t + \delta\sqrt{\alpha^2 - \beta^2} - \delta\sqrt{\alpha^2 - (\beta + it)^2}} \quad (2.40)$$

З формули (2.40) видно, що при $\alpha = \beta = 0$ він стає розподілом Коші.

Параметри нормального-оберненого гаусівського розподілу можна визначати методом максимальної правдоподібності, тоді вигляд формули спрощується порівняно з загальним випадком, але все одно необхідні чисельні методи. Метод моментів, як показали обчислювальні експерименти, є неточним [44,С. 3987]. В роботі [44] було запропоновано також метод, який використовує емпіричну оцінку характеристичної функції $\hat{\chi}(t)$ розподілу за формулою (2.18). Якщо покласти в ній параметр t достатньо великим, що виконується (2.41)

$$t^2 \gg \alpha^2 - \beta^2, t^2 \gg 4\beta^2 \quad (2.41)$$

Тоді можна зробити оцінку параметрів (2.42), (2.43), (2.44), (2.45)

$$\hat{\delta} = -\frac{d}{dt} \operatorname{Re}(\ln \hat{\chi}(t)), \quad (2.42)$$

$$\hat{\mu} = \frac{d}{dt} \operatorname{Im}(\ln \hat{\chi}(t)), \quad (2.43)$$

$$\hat{\beta} = \frac{\operatorname{Re}(\ln \hat{\chi}(t))}{t\hat{\delta}} [\hat{\mu}t - \operatorname{Im}(\ln \hat{\chi}(t))], \quad (2.44)$$

$$\hat{\alpha} = \sqrt{\left(\frac{\operatorname{Re}(\ln \hat{\chi}(t))}{\hat{\delta}} + \frac{1}{\sqrt{2}} \sqrt{t^2 + \sqrt{t^4 + 4\hat{\beta}^2 t^2}} \right)^2 + \hat{\beta}^2} \quad (2.45)$$

Оцінка є незміщеною та конзистентною. У публікації [44,С. 3988], наприклад, описано результати обчислювального експерименту: для вибірки розміром 1000 з параметрами $\xi \sim NIG(\alpha, 0, 1, 0)$ абсолютна похибка $\hat{\delta}$ становила в середньому 0.0709.

Означення 15. Гіперболічний розподіл (англ. Hyperbolic distribution) – узагальнений гіперболічний розподіл з параметром $\lambda = 1$. Позначається $\xi \sim \text{Hyp}(\alpha, \beta, \delta, \mu)$ [43, С. 27]. Тоді щільність спрощується до (2.46)

$$f_{\text{Hyp}}(x) = \frac{\sqrt{\alpha^2 - \beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2 - \beta^2})} e^{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)} \quad (2.46)$$

Властивості параметрів $\alpha, \beta, \delta, \mu$ написані для узагальненого гіперболічного зберігаються, оскільки цей розподіл є його частковим випадком. Оцінка параметрів теж робиться методом максимальної правдоподібності.

2.3 Методи оцінки показника капіталу під ризиком

У роботі буде розглядатись капітал під ризиком (англ. VaR) через лог-дохідності на одну одиницю часу вперед, який описуються формулою (1.12).

Опишемо параметричний метод пошуку VaR, тобто коли розподіл відомий. Очевидно, тоді задача зводиться до пошуку квантиля $(1 - c)$ -порядку відповідного розподілу. Приклад зображено на рис 2.4.

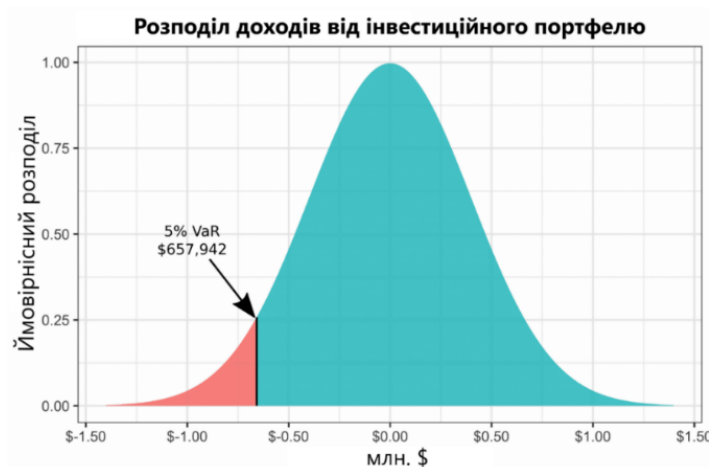


Рисунок 2.4 – Приклад пошуку VaR для рівня достовірності 0.95 на день вперед параметричним методом. Стрілка – квантиль 0.05-го порядку

Якщо розподіл невідомий можна оцінити VaR непараметричними методами. Найпростішим є метод емпіричних квантилів. Тоді будується гістограма, VaR покладається квантилем відповідного порядку. Наведемо приклад який описано в [32,С. 109]. Маємо 254 спостережень, потрібно знайти VaR із рівнем достовірності 0.95 на день вперед. Тоді обчислюємо скільки елементів вибірки приходяться на 5% даних

$$254 \times 0.05 = 12.7 \text{ елементів}$$

За гістограмою зліва від -10 млн \$ маємо 11 спостережень, зліва від -9 млн \$ маємо 15 спостережень. Лінійна інтерполяція дає $VaR = -9.6$ млн \$ (рис 2.5).

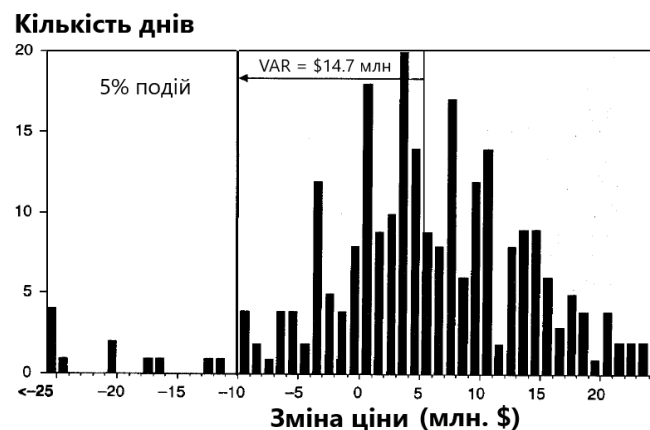


Рисунок 2.5 – Приклад розрахунку VaR за допомогою гістограми [32,С. 109]

Для VaR, обчисленого за допомогою гістограми, існує оцінка точності (2.47) за умови незалежності та однакового розподілу даних [32,С. 126]

$$VaR \in [\hat{q} - SE, \hat{q} + SE]; SE = \sqrt{\frac{c(1-c)}{Tf(q)^2}} \quad (2.47)$$

де $\hat{q}$ – оцінка VaR за допомогою гістограми

c – рівень достовірності VaR

T – розмір вибірки

$f(q)$ – значення щільності розподілу лог-дохідностей в точці q

q – квантиль $(1 - c)$ -го порядку (тобто той, який бажаємо знайти)

Розглянемо метод фільтрованих історичних симуляцій (англ. Filtered Historical Simulation). Він не вимагає наявності моделі розподілу лог-дохідностей, а використовує дані попередніх років щоб дати оцінку VaR. У різних публікаціях деталі можуть відрізнятись. Опишемо один із різновидів цього методу, який описано в [45].

У ній застосовується модель GARCH(1,1), яка показала свою ефективність при аналізі часових рядів акцій та інших фондових активів. Модель здатна виявляти кластери волатильності на ринку. Формули, за якими будувалась ця модель, мають вигляд (2.48), (2.49)

$$r_t = \mu r_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t, \quad (2.48)$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \quad (2.49)$$

де r_s – лог-дохідності в момент часу s

$\mu, \theta, \omega, \alpha, \beta$ – дійсні коефіцієнти, які задають модель

ε_s – випадкова величина, $\varepsilon_s \sim N(0, h_s)$, причому $\{\varepsilon_t\}$ незалежні

h_s – волатильність в момент часу s

Наведемо алгоритм обчислення VaR із для рівня c на період k [45]:

а) наблизити лог-дохідності $\{r_t\}$ активу моделлю GARCH(1,1), яка визначається формулами (2.48), (2.49). Буде отримано коефіцієнти, часовий ряд $\{\varepsilon_t\}$ випадкових збурень та часовий ряд волатильностей $\{h_t\}$;

б) нормувати часовий ряд $\{\varepsilon_t\}$ поелементним діленням на $\{h_t\}$. Буде отримано деякий стандартизований часовий ряд $\{e_t\}$;

в) з множини $\{e_t\}$ навмання обрати k елементів. Рекурентно обчислити $\{z_{t+i}^*\}, \{h_{t+i}^*\}, i = \overline{1, k}$ за формулами (2.50), (2.51)

$$z_{t+i}^* = e_i \sqrt{h_{t+i}^*} \quad (2.50)$$

$$h_{t+i}^* = \hat{\omega} + \hat{\alpha}(z_{t+i-1}^*)^2 + \hat{\beta}h_{t+i-1}^* \quad (2.51)$$

де z_{t+i}^* – прогноз випадкових збурень моделі GARCH(1,1) в момент $t + i$
 e_i – i -ий з навмання обраних k елементів пункту 3
 $\hat{\omega}, \hat{\alpha}, \hat{\beta}$ – числові коефіцієнти моделі
 h_s^* – прогноз волатильності у момент s

Таким чином була змодельована реалізація випадкових збурень $\{z_{t+i}^*\}$. Але для моделювання процесу на проміжку $t + 1, \dots, t + k$ потрібно ще застосувати модель GARCH(1,1), що буде побудовано в наступному пункті г;

г) рекурентно порахувати k майбутніх $r_{t+i+1}, i = \overline{0, k-1}$ за (2.52)

$$r_{t+i+1} = \hat{\mu} r_{t+i} + \hat{\theta} z_{t+i}^* + z_{t+i+1}^* \quad (2.52)$$

де $\hat{\mu}, \hat{\theta}$ – коефіцієнти моделі GARCH(1,1), які було отримано у пункті а
 r_s – прогноз лог-дохідностей у момент часу s
 z_s^* – реалізація збурень $\{z_{t+i}^*\}$ в момент s , отримані в пункті в

Ця дія має аналог у методі Монте-Карло – використання генератора випадкових чисел. Тут розподіл невідомий, тому «генерація» відбувається за допомогою попередніх відомих даних та моделі GARCH(1,1);

д) повернутися в пункт в. Побудову процесів повторювати багато разів;

е) обчислити для побудованих процесів VaR як емпіричний квантиль.

Алгоритм узагальнюється на випадок багатьох активів у портфелі – тоді процеси будуються по координатно для кожного активу, навмання обирається вектор приростів, щоб зберегти кореляцію між активами. Існують більш досконалі методи, де враховується «ефект важеля» (англ. leverage effect) – явища коли новини про спад цін підвищують волатильність [45].

Таким чином, існує багато способів обчислення показника VaR. Виникає питання перевірки узгодженості цього показника із даними, порівняння точності оцінок, отриманих різними способами. Для цього розроблено

статистичні критерії, які працюють аналогічно критеріям узгодженості розподілу. Їх існує не один десяток, наведемо лише два, які перевіряють принципово різні речі. Введемо поняття індикатора, який задається (2.53)

$$I_t = 1\{\text{в день } t \text{ сталися збитки більші за VaR}\} \quad (2.53)$$

У цій роботі розглядається VaR через лог-дохідності. Тому, аналогічно до того як міняється формула (1.11) на (1.12) для VaR при переході до лог-дохідностей, індикатор перетвориться з (2.53) на (2.54)

$$I_t = 1\{\ln S_t - \ln S_{t-1} < -VaR\} \quad (2.54)$$

де S_i – ціна портфелю в момент часу i

Перший критерій має назву біноміальний [46]. Він порівнює відсоток переходів збитками порогу VaR із рівнем достовірності. Висувається гіпотеза H_0 про те, що ймовірність збитків вийти за VaR дійсно дорівнює $1 - c$, відповідно до свого рівня достовірності

$$H_0: P\{I_t = 1\} = 1 - c, t = \overline{1, n} \quad (2.55)$$

де c – рівень достовірності для якого обчислюють VaR

Оскільки лог-дохідності вважаються незалежними, тоді в припущенні H_0 сумарна кількість випадків за n днів коли збитки були більше VaR має біноміальний розподіл $\xi \sim Bin(n, 1 - c)$. Для перевірки будується статистка T

$$T = \frac{H - n(1 - c)}{\sqrt{n(1 - c)(1 - n(1 - c))}} \quad (2.56)$$

де H – кількість випадків, коли лог-дохідності більші за VaR

n – загальний розмір вибірки

c – рівень достовірності для якого обчислюють VaR

При $n \rightarrow \infty$ статистика прямує до нормального розподілу. Була запропонована аналогічна статистика LR , яка прямує до розподілу χ_1^2 для $n \rightarrow \infty$

$$LR = -2 \ln(c^{n-H}(1-c)^H) + 2 \ln\left(\left(1 - \frac{H}{n}\right)^{n-H} \left(\frac{H}{n}\right)^H\right) \quad (2.57)$$

Критерій Крістофферсена перевіряє що ймовірність падіння вартості активу нижче порогу VaR одного дня не залежать від падіння нижче порогу попереднього дня [46]. Нульова гіпотеза строго формулюється так

$$H_0: P\{I_t = 1 | I_{t-1} = 0\} = P\{I_t = 1 | I_{t-1} = 1\} \quad (2.58)$$

де I_t – індикатор, формула якого записана в (2.54). Будується статистика LR^*

$$LR^* = 2 \ln\left(\frac{(1-\pi_0)^{N_{00}}\pi_0^{N_{01}}(1-\pi_1)^{N_{10}}\pi_1^{N_{11}}}{(1-\pi)^{N_{00}+N_{01}}\pi^{N_{01}+N_{11}}}\right) \quad (2.59)$$

$$\pi_0 = \frac{N_{01}}{N_{00}+N_{01}}; \pi_1 = \frac{N_{11}}{N_{10}+N_{11}}; \pi = \pi_0 + \pi_1 \quad (2.60)$$

де N_{ij} – кількість пар днів $(t, t+1)$, $t = \overline{1, n-1}$, коли $I_t = i, I_{t+1} = j$. При виконанні H_0 статистика при $n \rightarrow \infty$ прямує за розподілом до χ_1^2 .

Висновки до розділу

У цьому розділі було розглянуто розподіли, які використовуються як альтернатива нормальному для моделювання поведінки цін акцій та індексів на біржах. Проаналізовано стійкий розподіл, узагальнений скошений розподіл Стюдента, узагальнений гіперболічний розподіл та його часткові випадки.

Залежно від країни, компанії, будь-який з цих розподілів може виявитись кращою моделлю. Для кожного розподілу дано означення, описано основні властивості, оцінки параметрів. Особливо детально було досліджено оцінки параметрів стійкого розподілу, приведено алгоритм генерації випадкових величин. В останньому підрозділі наведено різні методи знаходження VaR та критерії оцінки узгодженості показника з даними. Він має велике значення, оскільки застосування розподілів із важкими та напівважкими хвостами обумовлено бажанням адекватно оцінити ймовірності збитків.

В наступному розділі описані розподіли будуть використані для моделювання лог-дохідностей міжнародного індексу США Standard & Poor 500, результати буде порівняно, обрано найточніший. Буде також обчислено показник VaR різними способами та порівняно їх узгодженість із даними за допомогою описаних критеріїв.

3 РЕАЛІЗАЦІЯ ВЛАСНОГО ЕКСПЕРИМЕНТУ

3.1 Огляд бібліотек Python та R для роботи з розподілами та тестування їх функцій

У власному експерименті розподіли з важкими та напівважкими хвостами будуть використані для моделювання поведінки цін міжнародного фондового індексу США Standard & Poor 500. Проведемо огляд алгоритмів для роботи з такими розподілами, які реалізовано на мовах програмування Python та R. Ці дві мови найчастіше використовуються для аналізу даних.

SciPy – бібліотека високоякісних наукових інструментів Python, яка, зокрема, надає функції для роботи з неперервними розподілами. У наявності є модулі для роботи з нормальним, стійким, узагальненим гіперболічним – для них доступні генератори випадкових чисел, обчислення щільності та функцій розподілу, оцінка параметрів. Оцінка параметрів здійснюється методом максимальної правдоподібності (максимізація функції правдоподібності робиться чисельними методами). В цій бібліотеці також реалізовано критерій Колмогорова-Смірнова. Для стійкого розподілу метод максимальної правдоподібності повільний (підрозділ 2.1), для нього візьмемо іншу бібліотеку.

StableEstim – бібліотека R, в якій вбудовані алгоритми оцінки параметрів стійкого розподілу $S(\alpha, \beta, c, \delta)$. Зокрема, в ній реалізовано методи МакКаллока, Кутрувелліса, комбінація цих методів (параметри c, δ оцінюються методом МакКаллока, а α, β – методом Кутрувелліса), а також метод максимальної правдоподібності.

SGT – бібліотека R для роботи з узагальненим скошеним розподілом Стюдента. У наявності є генератор випадкових чисел, оцінка методом максимальної правдоподібності, розрахунок щільності та функції розподілу.

ARCH – бібліотека Python для аналізу волатильності часових рядів. Зокрема, в ній доступна модель GARCH, якою лог-дохідності зводяться до незалежних однаково розподілених випадкових величин в численних роботах.

Слід пам'ятати, що спочатку необхідно перевірити точність оцінок на згенерованих вибірках, перш ніж їх застосовувати на реальних даних.

Спершу проводився чисельний експеримент з метою перевірки точності оцінок параметрів α, β, c, δ стійкого розподілу $S(\alpha, \beta, c, \delta)$: генерувались вибірки розміру від 100 до 1500 із кроком 100, для кожної було оцінено параметри методами МакКаллока, Куртувелліса, їх комбінацією та методом логарифмічних моментів. Останній алгоритм було реалізовано самостійно, він підходить тільки для оцінки параметрів α, c . Інші алгоритми були доступні в StableEstim. Генерація кожної вибірки та оцінка параметрів повторювалась 60 раз, для кожного параметра знаходилась середня абсолютна похибка. Результати власного експерименту зображено на рис 3.1.

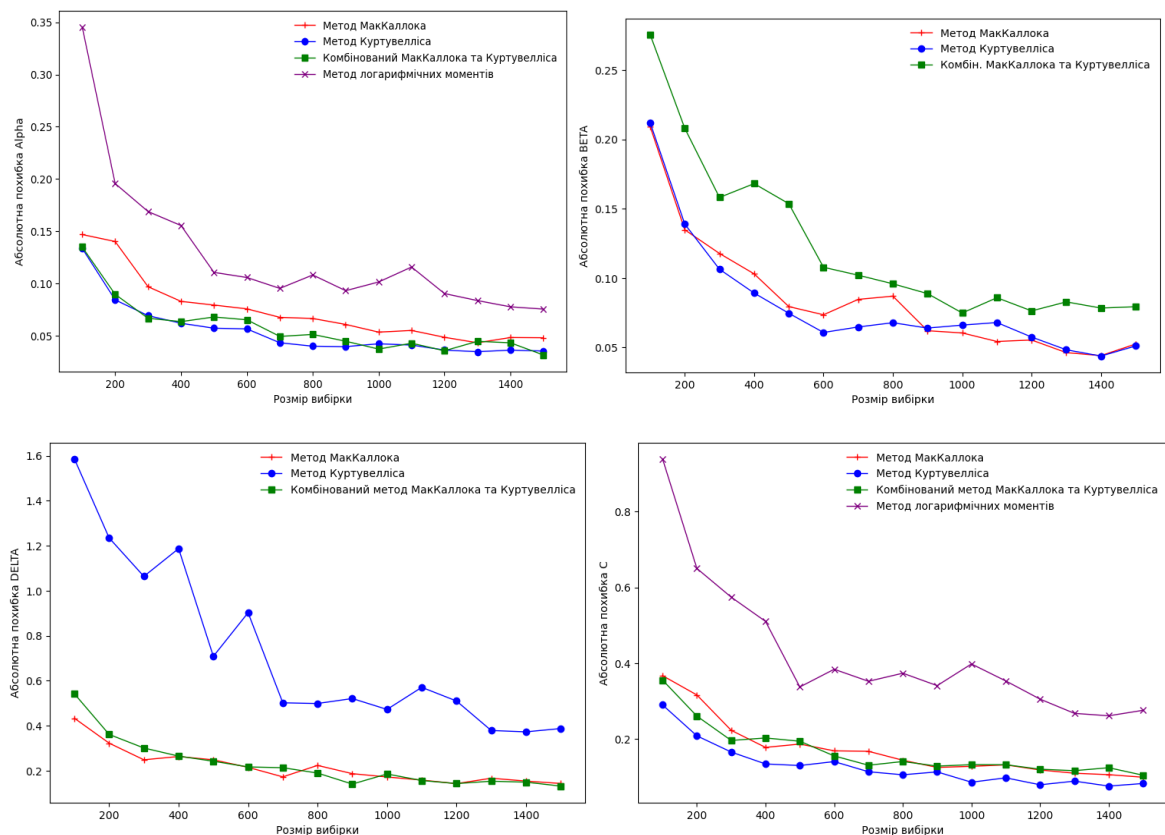


Рисунок 3.1 – Похибка параметрів α (зліва згори), β (справа згори), δ (зліва внизу), c (справа внизу) стійкого розподілу $S(\alpha, \beta, c, \delta)$ від розміру вибірки різними оцінками

Найбільш точним методом оцінки для параметрів α, β, δ виявився метод Кутрувелліса, що співпадає із висновками інших дослідників. Однак для оцінки μ більш точною виявився метод МакКаллока.

Було проведено аналогічний обчислювальний експеримент для перевірки точності оцінки параметрів узагальненого гіперболічного розподілу $\xi \sim GH(\lambda, \alpha, \beta, \delta, \mu)$; для нього застосовувався лише метод максимальної правдоподібності (підрозділ 2.2). Похибка оцінки рахувалась формулою (3.1)

$$\varepsilon = \frac{1}{K} \sum_{i=1}^K \left\| (\lambda_i, \alpha_i, \beta_i, \delta_i, \mu_i) - (\hat{\lambda}_i(N), \hat{\alpha}_i(N), \hat{\beta}_i(N), \hat{\delta}_i(N), \hat{\mu}_i(N)) \right\|_2 \quad (3.1)$$

де $(\lambda_i, \alpha_i, \beta_i, \delta_i, \mu_i)$ – параметри розподілу;

$\hat{\lambda}_i(N), \hat{\alpha}_i(N), \hat{\beta}_i(N), \hat{\delta}_i(N), \hat{\mu}_i(N)$ – оцінки параметрів, які отримано методом максимальної правдоподібності;

$\|\cdot\|_2$ – норма L_2 ;

N – розмір згенерованої вибірки;

K – кількість генерацій вибірки розміром N ;

Для узагальненого гіперболічного розподілу взято N від 100 до 2500 з кроком 300, $K = 20$. Результати обчислювального дослідження зображено на рис 3.2.

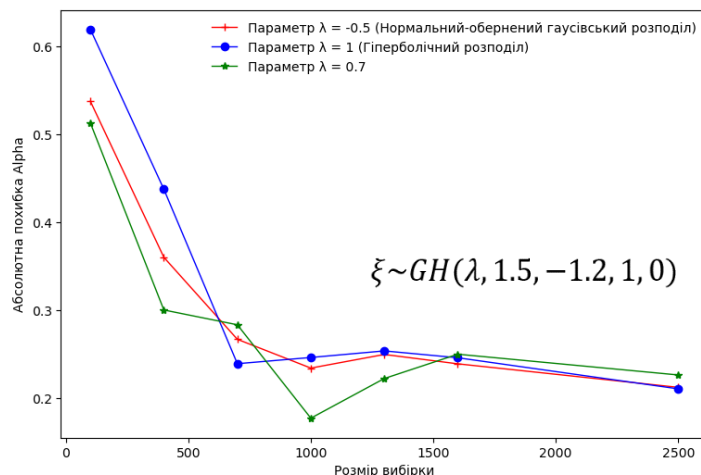


Рисунок 3.2 – Похибка оцінки вектору параметрів $(\lambda, \alpha, \beta, \delta, \mu)$ у нормі L_2 для узагальненого гіперболічного розподілу

Таку саму процедуру було проведено для узагальненого скошеного розподілу Стюдента. Формула приймає вид (3.2) аналогічно (3.1)

$$\varepsilon = \frac{1}{K} \sum_{i=1}^K \left\| (\mu_i, \sigma_i, \lambda_i, p_i, q_i) - (\hat{\mu}_i(N), \hat{\sigma}_i(N), \hat{\lambda}_i(N), \hat{p}_i(N), \hat{q}_i(N)) \right\|_2 \quad (3.2)$$

де $\mu_i, \sigma_i, \lambda_i, p_i, q_i$ – параметри розподілу;

$\hat{\mu}_i(N), \hat{\sigma}_i(N), \hat{\lambda}_i(N), \hat{p}_i(N), \hat{q}_i(N)$ – оцінки параметрів, які отримано методом максимальної правдоподібності;

N – розмір згенерованої вибірки;

K – кількість генерацій вибірки розміром N ;

Для цього розподілу взято N від 100 до 1100 з кроком 100, $K = 20$.

Результати обчислювального експерименту зображено на рис 3.3.

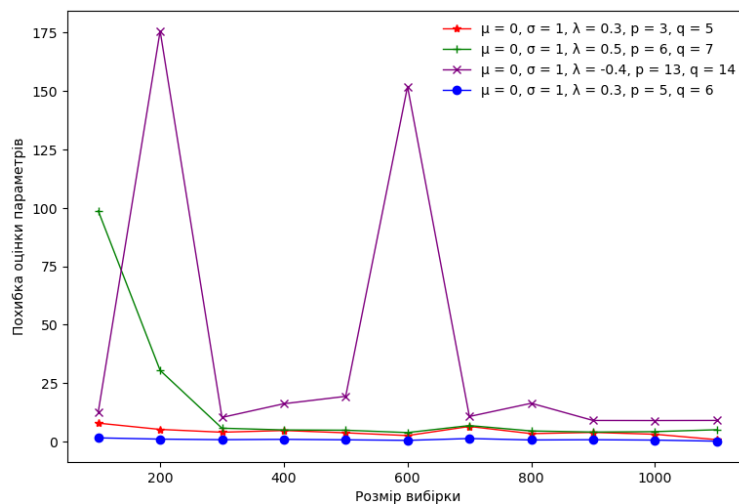


Рисунок 3.3 – Похибка оцінки вектору параметрів $(\mu, \sigma, \lambda, p, q)$ у нормі L_2 для узагальненого скошеного розподілу Стюдента

Із рис 3.3 бачимо, що швидкість спадання похибки залежить від комбінації параметрів. Тим не менш, в усіх випадках похибка прямує до нуля. Отже, власними обчислювальними експериментами було показано працездатність оцінок параметрів розподілів, які використовуються в роботі.

3.2 Моделювання лог-дохідностей індексу S&P 500 розподілами

Перейдемо до застосування розподілів із важкими та напівважкими хвостами на практиці. Дані про ціни акцій та міжнародних індексів надає авторитетний ресурс Yahoo Finance. Зокрема, в ньому доступні дані про ціну закриття за кожен торговий день на великому проміжку часу.

В експерименті буде моделюватись вартість міжнародного фондового індексу США Standard & Poor 500 (далі скорочено S&P 500). Із Yahoo Finance було завантажено дані про ціну закриття індексу з 3 січня 2000 по 30 грудня 2022 року. Дані були оброблені наступним чином:

а) ціни закриття було поділено на тисячу (записано в одиницях виміру тисяч доларів, показано на рис 3.4 зліва);

б) зроблено перехід від ціни закриття до лог-дохідностей, теоретична частина була описана у підрозділі 1.2 (графік показано на рис 3.4 справа);

в) видалено нульові лог-дохідності (дні, коли ціна не змінювалась);

Завантажені та оброблені дані використані для відстеження динаміки ціни індексу S&P 500, показано на рисунку 3.4.

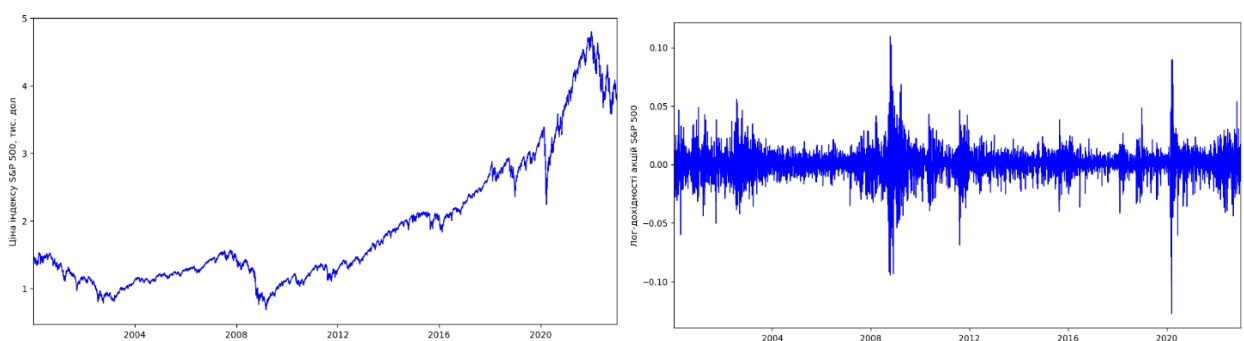


Рисунок 3.4 – Динаміка індексу S&P 500 з 03.01.2000 по 30.12.2022 (зліва) та відповідні лог-дохідності S&P 500 (справа)

Далі лог-дохідності мають бути зведені до реалізації незалежних однаково розподілених випадкових величин перш ніж наближуватись будь-

яким розподілом. На рис 3.4 справа можна помітити кластери волатильності – проміжки часу, коли амплітуда коливань лог-дохідностей була більша, ніж в інші торгові дні (наприклад 2008-2010). Наслідуючи приклад [18], щоб оцінити волатильність кожного торгового дня, побудуємо модель GARCH(1,1) за формулами (3.3), (3.4)

$$r_t = \mu + h_t e_t, \quad e_t \sim N(0,1) \quad (3.3)$$

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \quad (3.4)$$

де r_t – лог-дохідності

$\mu, \omega, \alpha, \beta$ – числові коефіцієнти

h_s – волатильність у момент s

e_t – стандартизовані лог-дохідності у момент s

Для побудови моделі GARCH(1,1) було взято функцію бібліотеки ARCH мови програмування Python. Як повідомила програма, коефіцієнти збіглися доволі швидко (виражаючись термінологією розробників, модель «робастна»).

Зауважимо, що хоча стандартизовані лог-дохідності e_t не є нормально розподіленими (підрозділ 1.3, таблиця 1.1), використання такої моделі тим не менш є корисною. Отримані стандартизовані значення $\{e_t\}$ близькі до реалізації незалежних однаково розподілених випадкових величин. Застосувавши модель GARCH(1,1) можемо перейти від $\{r_t\}$ до стандартизованих значень $\{e_t\}$ за формулою (3.5)

$$e_t = \frac{r_t - \mu}{\sigma_t} \quad (3.5)$$

Порівняння лог-дохідностей до стандартизації та після зображено зліва та справа відповідно (Рис 3.5). Видно що дані після стандартизації стали більш однорідними, їх можна покласти реалізацією незалежних і однаково розподілених випадкових величин і підбирати для них розподіл.

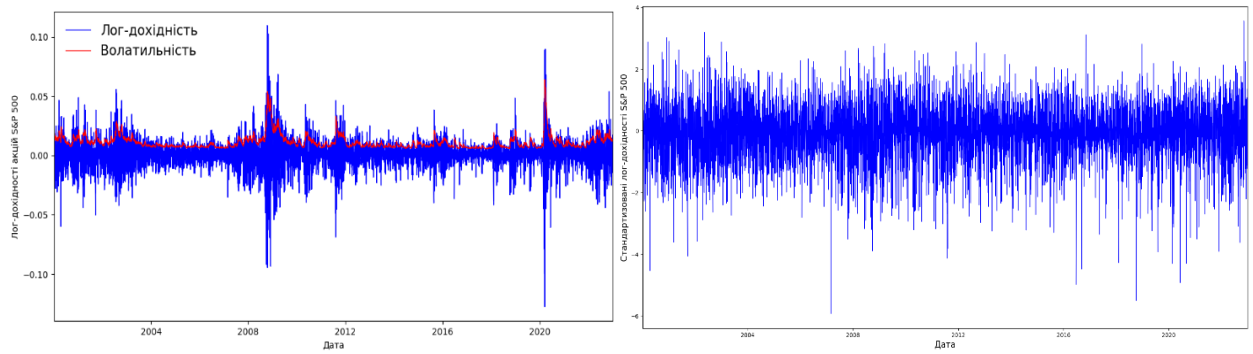


Рисунок 3.5 – Значення лог-дохідностей S&P 500 з оцінкою волатильності (зліва) та після стандартизації за формулою (3.5) (справа)

Для вибірки оцінимо середнє та стандартне відхилення, якщо воно скінчене, тоді виконуватиметься (3.6), (3.7)

$$\bar{x}(n) = \frac{1}{n} \sum_{i=1}^n x_i \rightarrow a, n \rightarrow \infty \quad (3.6)$$

$$\bar{\sigma}(n) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \rightarrow b, n \rightarrow \infty \quad (3.7)$$

де x_i – елементи вибірки $\{x_i\}$
 n – розмір вибірки який береться
 a, b – дійсні числа

Залежність оцінок $\bar{x}(n), \bar{\sigma}(n)$ від розміру вибірки N зображено на рис 3.6. Наявність границі дозволяє припустити скінченність моментів першого і другого порядку розподілу стандартизованих лог-дохідностей S&P 500 [18].

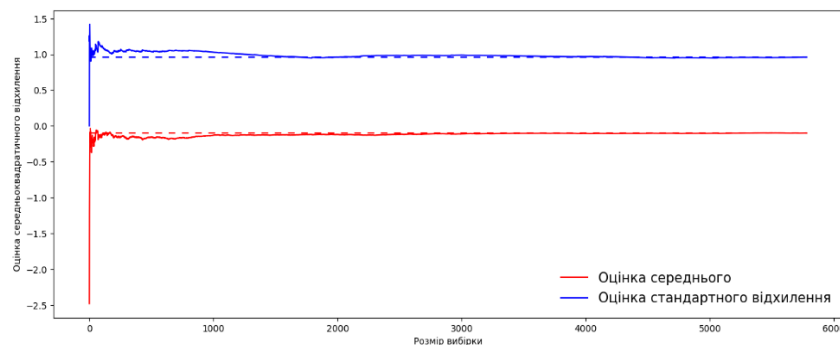


Рисунок 3.6 – Залежність оцінок $\bar{x}(n), \bar{\sigma}(n)$ для стандартизованих лог-дохідностей S&P 500 від розміру вибірки

Наявність середнього та дисперсії характерно для деяких параметрів узагальненого скошеного розподілу Стюдента (підрозділ 2.1) та узагальненого гіперболічного розподілу (підрозділ 2.2). Результати дозволяють припустити що стійкий розподіл не буде точною моделлю, оскільки він в загальному випадку не має скінченої дисперсії (підрозділ 2.1).

За допомогою зробленого програмного продукту наблизимо стандартизовані лог-дохідності різними розподілами та порівняємо їх (Рис 3.7). На рисунку 3.8 зобразимо нормальний розподіл та узагальнений гіперболічний, який, як буде показано далі, дав найкраще наближення статистикою Колмогорова-Смірнова.

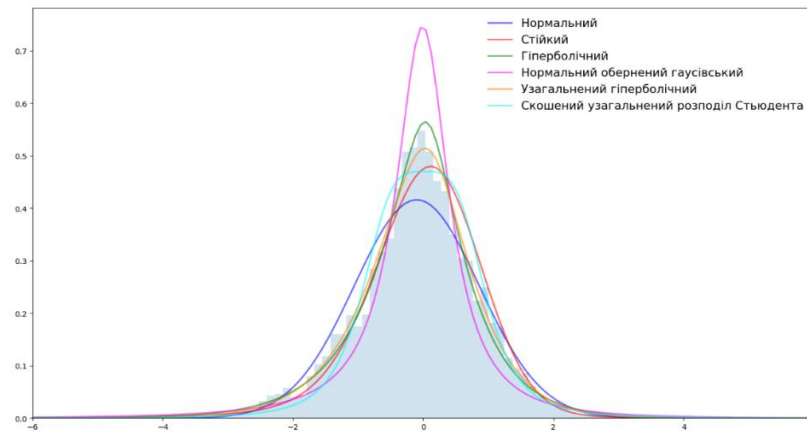


Рисунок 3.7 – Порівняння гистограми стандартизованих лог-дохідностей S&P 500 з щільностями наближених розподілів

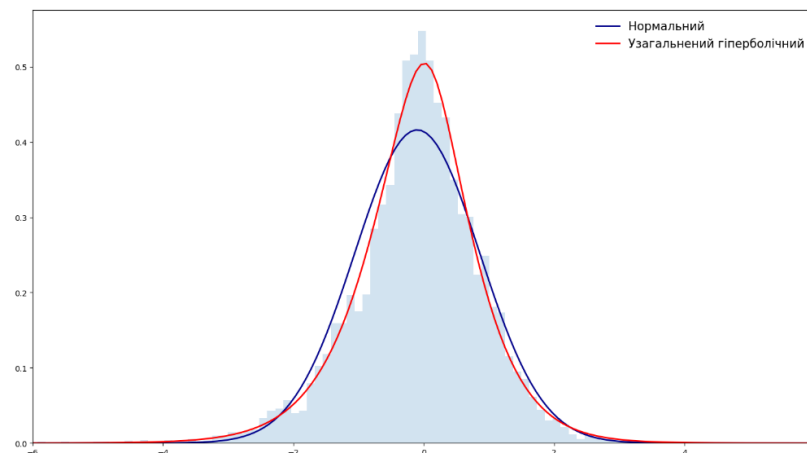


Рисунок 3.8 – Порівняння гистограми стандартизованих лог-дохідностей S&P 500 з щільностями нормального та узагальненого гіперболічного розподілів

Рис 3.7 не відображає поведінку розподілів у хвостах (у квантилях малих і великих порядків). Щоб порівняти емпіричну функцію розподілу $F^*(x)$ та теоретичну $F_i(x)$ для побудованих розподілів, застосовується такий метод:

- а) на двох рисунках зображуються два інтервали – на лівому інтервал $(-\infty, 0)$ (лівий хвіст), на правому – $(0, +\infty)$ (правий хвіст);
- б) на рисунку для лівого хвоста зображуються $F_i(x)$ та $F^*(x)$, на рисунку для правого хвоста зображуються $1 - F_i(x)$ та $1 - F^*(x)$;
- в) осі координат беруться в логарифмічній шкалі.

Графік отриманий описаним методом зображено на рисунку 3.9

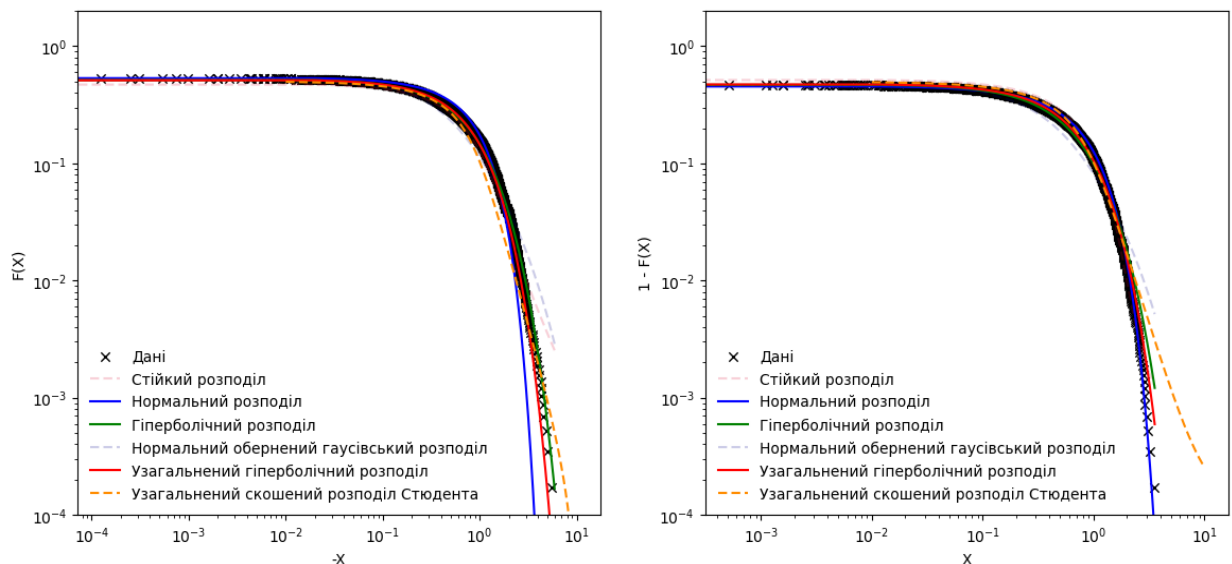


Рисунок 3.9 – Порівняння лівого (зліва) та правого хвосту (справа) емпіричної функції розподілу стандартизованих лог-дохідностей S&P 500 із теоретичними функціями розподілу

З рис 3.9 видно, що для лівого хвоста нормальний розподіл занадто швидко спадає (недооцінює ризики), гіперболічний та узагальнений гіперболічний найкраще відтворюють його поведінку. Для правого хвоста навпаки – нормальний найбільш точно відтворює асимптотичну поведінку. Застосуємо критерій узгодженості Колмогорова-Смірнова та порівняємо значення статистики цього тесту для різних розподілів (чим вона ближче до нуля, тим краще наближення). Результати занесено в таблицю 3.1.

Таблиця 3.1 – Застосування критерію узгодженості Колмогорова-Смірнова для різних розподілів. Зеленим виділено найточніше значення

Розподіл	Статистика Колмогорова-Смірнова	Рівень значущості p
Нормальний	0.0533	$1.013 \cdot 10^{-14}$
Стійкий	0.0590	$5.744 \cdot 10^{-18}$
Гіперболічний	0.0213	0.0102
Нормальний обернений гаусівський	0.0607	$5.4682 \cdot 10^{-19}$
Узагальнений гіперболічний	0.0123	0.3463

Тобто на рівні значущості $p > 0.1$ дані не суперечать гіпотезі про те, що стандартизовані моделлю GARCH(1,1) лог-дохідності індексу S&P 500 розподілені за узагальненим гіперболічним розподілом.

3.3 Обчислення показника VaR для лог-дохідностей різними методами

Розглянемо одну з прикладних задач в моделюванні вартостей акцій та міжнародних індексів – розрахунку показника VaR. Як показано в пункті 1.4, на практиці обчислюють VaR для лог-дохідностей, потім переходять до звичайного VaR за формулою (1.13). Тому сконцентруємо увагу на пошуку VaR для лог-дохідностей. Розрахунок виконувався в Python наступним чином:

а) бралось ковзне вікно розміром 300 нестандартизованих лог-дохідностей, ці дані використовувались для оцінки VaR на рівнях достовірності 90%, 95%, 99% на один день вперед різними методами;

1) параметричним із використанням досліджуваних розподілів;

2) методом емпіричного квантиля;

3) методом фільтрованих історичних симуляцій;

б) обчислені різними методами VaR та значення лог-дохідності наступного дня заносились на графік щоб порівняти;

в) ковзне вікно зміщувалось на 1 вперед, процедура повторювалась.

Результати побудови оцінок VaR різними методами разом із даними показано на рис 3.10 (показано частину обчислених порогів для зручності).

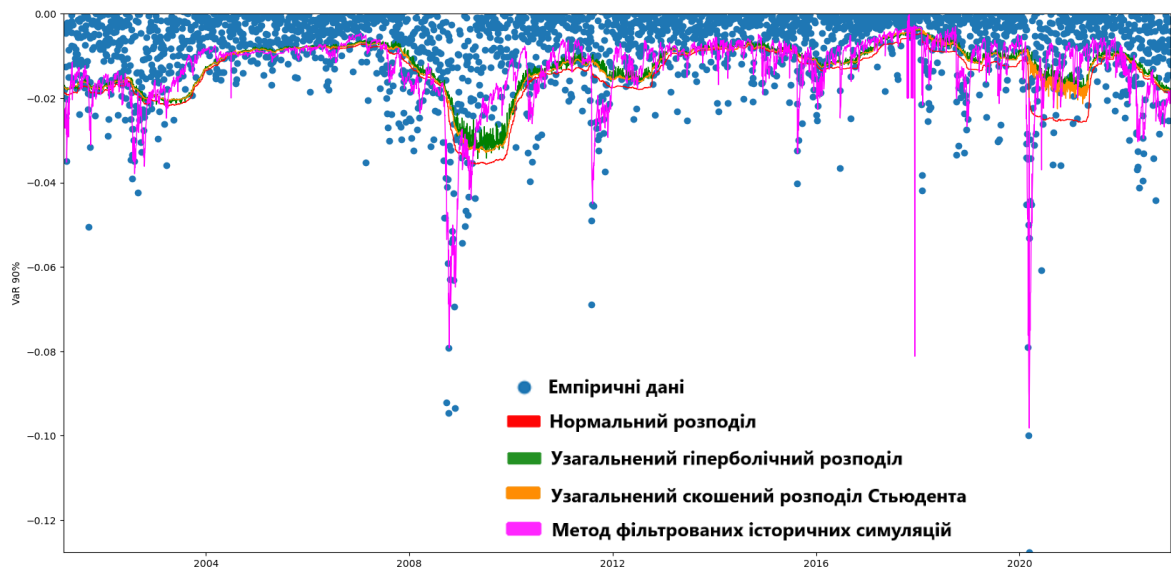


Рисунок 3.10 – VaR на рівні достовірності 90% обчислений різними методами та від’ємні лог-дохідності індексу S&P 500 (тобто збитки)

Для VaR 90% відсоток виходів лог-дохідностей за межі цього порогу має бути приблизно 10%, для VaR 95% – 5%, для VaR 99% – 1%. Їх порівняння занесено в таблицю 3.2.

Таблиця 3.2 – Порівняння емпіричних значень відсотків виходу лог-дохідностей за межі VaR із теоретичними.

Метод пошуку VaR (для параметричного вказано тільки розподіл)	Відсоток виходів лог-дохідностей за межі VaR		
	VaR 90%	VaR 95%	VaR 99%
Нормальний	9.1755%	5.8738%	2.7910%
Стійкий	12.6779%	6.8771%	1.0945%
Гіперболічний	11.1638%	6.0744%	1.4593%
Нормальний обернений гаусівський	11.0544%	5.1988%	0.8938%
Узагальнений гіперболічний	10.6895%	5.6914%	1.6965%
Узагальнений скошений Стюдента	10.6713%	5.7826%	1.4593%
Емпірична оцінка квантилів	10.5801%	5.5637%	1.6782%
Фільтровані історичні симуляції	10.3977%	5.1806%	1.4958%

Примітка. Зеленим виділено теоретичні значення найближчі до емпіричних

Застосуємо біноміальний критерій для перевірки узгодженості даних із порогом VaR, отриманим різними методами. Критерій може використовувати статистику T або LR (підрозділ 2.3), перевіримо для обох. Дані візьмемо до 2008 року, оскільки під час кризи виникають події, які неможливо передбачити.

Таблиця 3.3 – Рівень значущості p , на якому дані не суперечать гіпотезі про адекватність VaR, для різних методів

Метод пошуку VaR	Рівень значущості p для біноміального критерію, на якому гіпотеза про узгодженість даних із VaR відхиляється					
	p для статистики T			p для статистики LR		
	VaR 90%	VaR 95%	VaR 99%	VaR 90%	VaR 95%	VaR 99%
Нормальний	0.9747	0.2844	0.0001	0.0507	0.5850	0.0011
Стійкий	0.0029	0.0173	0.3838	0.0076	0.0433	0.7325
Узагальнений гіперболічний	0.0422	0.4576	0.9314	0.0931	0.9258	0.1369
Нормальний обернений гаусівський	0.1399	0.0768	0.0387	0.2919	0.1359	0.045
Гіперболічний	0.1021	0.9826	0.9314	0.2157	0.0349	0.1369
Узагальнений скошений розподіл Стюдента	0.1625	0.6781	0.8783	0.3371	0.6399	0.245
Емпірична оцінка квантилів (непараметричний метод)	0.5464	0.3364	0.0066	0.9022	0.6875	0.0258
Фільтровані історичні симуляції (непараметричний метод)	0.7751	0.3364	0.0252	0.4485	0.6875	0.0755

Примітка.

У колонці «Метод пошуку VaR»:

для параметричного метода вказано лише відповідний розподіл (підрозділ 1.4)

для непараметричного метода вказано його назву

Зеленим виділено статистично значущий результат ($p > 0.1$)

Бачимо з таблиці 3.3, що навіть відхилені критерієм узгодженості розподіли дають статистично значущий показник VaR, причина цього результату буде пояснена у висновку. Застосуємо ще критерій Крістофферсена (підрозділ 2.3). Дані візьмемо до 2008 року.

Таблиця 3.4 – Рівень значущості p для критерію Крістофферсена, на якому дані не суперечать гіпотезі про незалежність виходів збитків за межі VaR.

Зеленим виділено значення на рівні $p > 0.1$

Метод пошуку VaR (для параметричного методу вказано тільки розподіл)	Рівень значущості p для критерію на якому гіпотеза про незалежність виходів за поріг VaR відхиляється		
	VaR 90%	VaR 95%	VaR 99%
Нормальний	0.0001	0.0001	0.0001
Стійкий	0.0001	0.0001	0.7296
Узагальнений гіперболічний	0.0001	0.0001	0.9176
Нормальний обернений гаусівський	0.0001	0.0001	0.4783
Гіперболічний	0.0001	0.0001	0.9176
Узагальнений скошений розподіл Стюдента	0.0001	0.0001	0.8719
Емпірична оцінка квантилів	0.0001	0.0001	0.0001
Фільтровані історичні симуляції	0.0001	0.0001	0.0001

Результати застосування біноміального критерію та критерію Крістофферсена для VaR дозволяють зробити такі висновки:

а) біноміальний критерій прийняв гіпотезу про узгодженість оцінки із даними одразу для VaR 90%, 95%, 99% на рівні $p > 0.1$ лише для узагальненого скошеного розподілу Стюдента;

б) хоча нормальний розподіл є неадекватною моделлю лог-дохідностей акцій, тим не менш, VaR 95%, отриманий з нього, не був відхилений біноміальним критерієм на рівні $p > 0.1$;

в) для оцінки VaR 99% дуже мало оцінок було прийнято. Можна зробити висновки, що для великих рівнів значущості (VaR 99%) непараметричні методи можуть давати неточні оцінки;

г) критерій Крістофферсена відхилив гіпотезу для всіх VaR 90%, 95% на рівні $p < 0.0001$. Однак для більшості VaR 99% гіпотеза була прийнята. Можна покласти, що принаймні великі збитки (за порогом VaR 99%) незалежні один від одного на рівні $p > 0.1$.

Може виникнути питання: чому якщо найбільш точне наближення дав узагальнений гіперболічний розподіл (таблиця 3.1), біноміальний критерій відхилив гіпотезу про узгодженість даних із VaR 90% за цим розподілом? І навпаки, чому, наприклад, нормальний розподіл (неадекватна модель лог-дохідностей) для VaR 95% не був відхиленим біноміальним критерієм?

Відповідь наступна: VaR вимагає точної оцінки лише квантилів порядку 0.1, 0.05, 0.01. Тому розподіл може неадекватно оцінювати розподіл на всьому проміжку, але його квантилі деяких порядків можуть бути близькими до теоретичних значень.

Висновки до розділу

У власному експерименті було змодельована вартість міжнародного фондового індексу США Standard & Poor 500 з 3 січня 2000 по 30 грудня 2022 року. Дані про вартість було переведено в лог-дохідності, оцінено волатильність, з них отримано стандартизовані значення, які покладено незалежними та однаково розподіленими. Саме ці стандартизовані значення наближувались розподілами із важкими та напівважкими хвостами.

Спочатку була побудована залежність оцінки математичного сподівання та стандартного відхилення від розміру вибірки. Обидві оцінки швидко збіглись, що свідчить про існування моментів першого та другого порядків. Із цього можна було очікувати, що стійкий розподіл не буде найкращою моделлю, оскільки не має скінченого моменту другого порядку.

Для стандартизованих значень оцінено параметри шести розподілів, побудовано гістограму. Також порівняно в логарифмічній шкалі емпіричну функцію розподілу з теоретичними функціями розподілів. Найкращою моделлю виявився узагальнений гіперболічний розподіл. Критерій Колмогорова-Смірнова показав, що на рівні $p > 0.1$ дані не суперечать гіпотезі про те, що стандартизовані лог-дохідності розподілені за цим законом. Зауважимо – лише для цього розподілу рівень p був статистично значущим.

Було також обчислено міру ризику VaR для рівня достовірності 90%, 95%, 99% на день вперед різними методами, їх узгодженість була перевірена

статистичними критеріями. VaR, обчислені деякими методами, давали якісну оцінку ризику інвестування, яка може застосовуватись на практиці.

Може здатись дивним, що для стандартизованих лог-дохідностей S&P 500 найточнішим розподілом виявився узагальнений гіперболічний, а для оцінки VaR на практиці – інший залежно від рівня достовірності. Однак це допустимо, оскільки, по-перше, для VaR необхідна лише точна оцінка квантилів малого порядку, а не всіх квантилів. По-друге, узагальнений гіперболічний розподіл був найкращим саме для лог-дохідностей поділених на волатильність, а VaR рахувався для ковзного вікна розміром 300, де такого нормування не було. Отже у межах того вікна інший розподіл міг давати точніші наближення. З результатів критерію Крістофферсена можна припустити що малі падіння вартості індексу S&P 500 не є незалежними, однак великі падіння вартості незалежні.

Загалом можна зробити висновок, що узагальнений гіперболічний розподіл, який в загальному випадку є представником розподілів із важкими хвостами, є найкращою моделлю поведінки цін індексу S&P 500 на обраному часовому проміжку. Однак для того щоб стверджувати, який метод найкраще підходить для обчислення VaR необхідні подальші дослідження, можливо, застосування інших, більш сучасних підходів.

4 ФУНКЦІОНАЛЬНО-ВАРТІСНИЙ АНАЛІЗ ПРОГРАМНОГО ПРОДУКТУ

У даному розділі проводиться оцінка основних характеристик програмного продукту, розробленого для вирішення задач моделювання цін акцій розподілами з важкими хвостами.

Нижче наведено аналіз різних варіантів реалізації модулю з метою вибору оптимальної, з огляду при цьому як на економічні фактори, так і на 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) – це технологія, яка дозволяє оцінити реальну вартість продукту або послуги незалежно від організаційної структури компанії. ФВА проводиться з метою виявлення резервів зниження витрат за рахунок ефективніших варіантів виробництва, кращого співвідношення між споживчою вартістю виробу та витратами на його виготовлення. Для проведення аналізу використовується економічна, технічна та конструкторська інформація.

Алгоритм функціонально-вартісного аналізу включає в себе визначення послідовності етапів розробки продукту, визначення повних витрат (річних) та кількості робочих часів, визначення джерел витрат та кінцевий розрахунок вартості програмного продукту.

4.1 Постановка задачі проектування

У роботі застосовується ФВА для проведення техніко-економічного аналізу розробки системи моделювання цін акцій розподілами з важкими хвостами. Так як проектування та реалізація компонентів, що розробляються,

впливають на всю систему, кожна окрема підсистема має її задовольняти. Фактичний аналіз є аналізом функцій програмного продукту для моделювання вартості акцій та міжнародних індексів розподілами з важкими хвостами.

Технічні вимоги до програмного продукту є наступні:

- функціонування на персональних комп'ютерах із стандартним набором компонентів;
- зручність та зрозумілість для користувача;
- швидкість обробки даних та доступ до інформації в реальному часі;
- мінімальні витрати на впровадження програмного продукту.

4.2 Обґрунтування функцій програмного продукту

Головна функція F_0 – розробка програмного продукту, який вирішує задачу моделювання вартості акцій та міжнародних індексів розподілами з важкими хвостами. Беручи за основу цю функцію, можна виділити наступні:

F_1 – вибір способу об'єднання двох мов програмування в програмі;

F_2 – вибір бібліотеки для роботи з розподілами з важкими хвостами;

F_3 – вибір середовища розробки;

Кожна з цих функцій має декілька варіантів реалізації:

Функція F_1 :

а) окремі модулі на Python та R – частину програмного модулю буде реалізовано на Python, частину на R, користувач буде запускати їх по черзі;

б) об'єднання Python та R через `gru2` – програма буде реалізована на Python, частина функцій буде викликатись із R за допомогою бібліотеки `gru2`;

Функція F_2 :

а) `SciPy`;

б) `TensorFlow`;

Функція F_3 :

a) Jupyter lab;

б) Visual Studio Code;

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

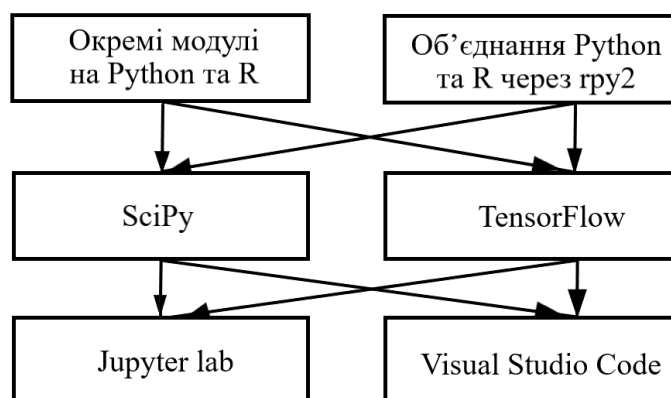


Рисунок 4.1 – Морфологічна карта

Морфологічна карта показує множину можливих варіантів основних функцій.

Таблиця 4.1 - Позитивно-негативна матриця

Функції	Варіанти реалізації	Переваги	Недоліки
F_1	A	Швидкодія за рахунок відсутності посередництва між Python та R, відсутність ризику збою через несумісність версій Python та R	Користувачу для доступу до одних функцій потрібно запустити модуль на Python, для інших – модуль на R
	B	Програма буде в єдиному файлі	Низька швидкість роботи, ризик виникнення несумісності версій Python та R, що встановлені на комп'ютері
F_2	A	Бібліотека містить розподіли з важкими та напівважкими хвостами та майже всі необхідні функції для роботи з ними	Деякі функції повільні
	B	Можливість реалізації паралелізму	Реалізовано значно менше розподілів
F_3	A	Проста в застосуванні, можливість створювати інтерактивні скрипти без модулів інтерфейсу	Немає наявності функцій оптимізації коду
	B	Наявні функції оптимізації коду	Немає можливості створювати інтерактивні скрипти

На основі цієї карти будемо позитивно-негативну матрицю варіантів основних функцій (у таблиці 4.1). 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 F_1 :

Перевагу віддаємо швидкодії та надійності. До того ж неможливо проконтролювати сумісність версій Python та R на комп'ютері користувача. Для спрощення роботи по написанню коду варіант Б має бути відкинтий.

Функція F_2 :

Обидва варіанти можна використовувати в розробці.

Функція F_3 :

Віддаємо перевагу варіанту А оскільки методи оптимізації коду та інші функції доцільно застосовувати у великих проектах, де працюють десятки програмістів та ведеться робота з гігабайтами даних.

Таким чином, будемо розглядати такий варіанти реалізації ПП:

$$F_1 \text{ а} - F_2 \text{ а} - F_3 \text{ а}$$

$$F_1 \text{ а} - F_2 \text{ а} - F_3 \text{ б}$$

Для оцінювання якості розглянутих функцій обрані параметри, описані нижче.

4.3 Обґрунтування системи параметрів ПП

На основі даних, розглянутих вище, визначаються основні параметри вибору, які будуть використані для розрахунку коефіцієнта технічного рівня. Для того, щоб охарактеризувати програмний продукт, будемо використовувати наступні параметри:

- X_1 – швидкодія мови програмування;

- X_2 – об’єм пам’яті для обчислень та збереження даних;
- X_3 – час навчання моделей, оцінки параметрів розподілів;
- X_4 – потенційний об’єм програмного коду;

Гірші, середні і кращі значення параметрів вибираються на основі вимог замовника й умов, що характеризують експлуатацію ПП (у таблиці 4.2).

Таблиця 4.2 – Основні параметри ПП

Назва параметра	Умовні позначення	Одиниці виміру	Значення параметра		
			гірші	середні	кращі
Швидкодія мови програмування	X_1	оп/мс	10000	14000	19000
Об’єм пам’яті	X_2	Мб	1100	450	200
Час навчання моделей, оцінки параметрів розподілів	X_3	с	20	12	3
Потенційний об’єм програмного коду	X_4	Рядки коду	3500	2000	700

За таблицею 4.2 будуються графічні характеристики параметрів (рис 4.2–4.3).

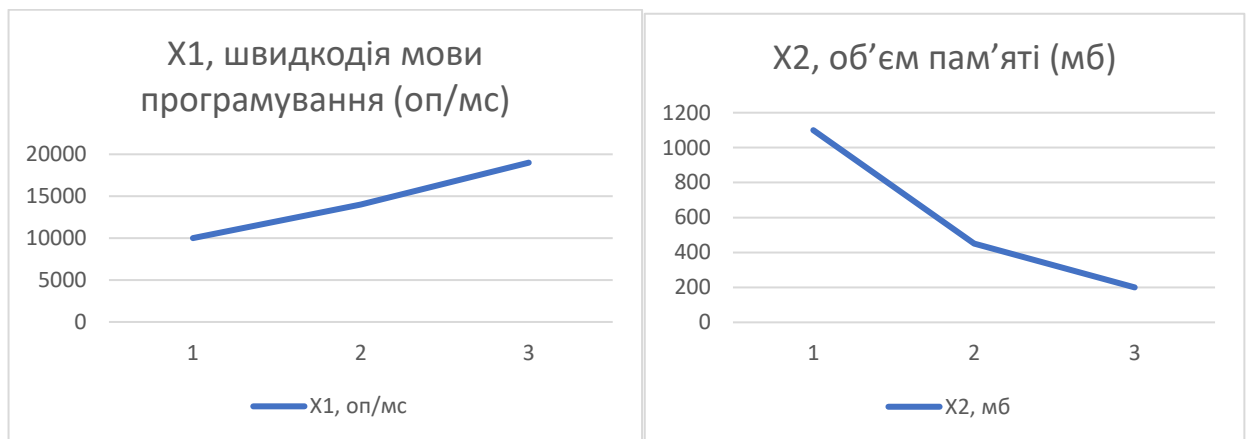


Рисунок 4.2 – Параметри X_1 (зліва), X_2 (справа)

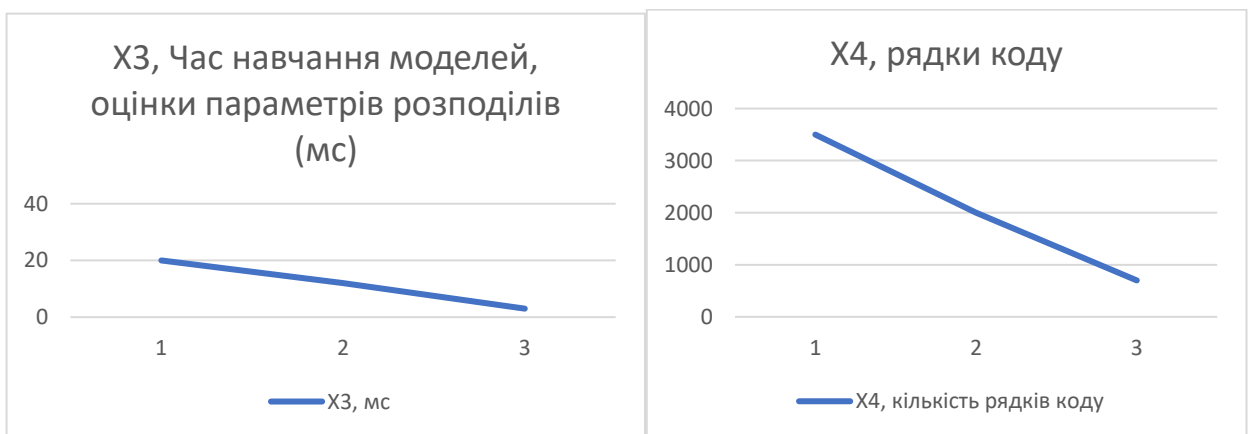


Рисунок 4.3 – Параметри X_3 (зліва), X_4 (справа)

[illegible]

Для перевірки степені достовірності експертних оцінок, визначимо наступні параметри:

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

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

де N – число експертів,
 n – кількість параметрів;
 б) середня сума рангів:

$$T = \frac{1}{n} R_{ij} = 17,5 \quad (4.2)$$

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

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

Сума відхилень по всіх параметрам повинна дорівнювати 0;

г) загальна сума квадратів відхилення:

$$S = \sum_{i=1}^N \Delta_i^2 = 205. \quad (4.4)$$

Порахуємо коефіцієнт узгодженості:

$$W = \frac{12S}{N^2(n^3 - n)} = \frac{12 \cdot 205}{7^2(4^3 - 4)} = 0,83 > W_k = 0,67. \quad (4.5)$$

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

Скориставшись результатами ранжирування, 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	<	<	>	>	<	<	<	<	0,5
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 i -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:

$$K_{bi} = \frac{b_i}{\sum_{i=1}^n b_i} \quad (4.7)$$

$$b_i = \sum_{j=1}^N a_{ij} \quad (4.8)$$

Відносні оцінки розраховуються декілька разів доти, поки наступні значення не будуть незначно відрізнятися від попередніх (менше 2%). На дру-

тому і наступних кроках відносні оцінки розраховуються за наступними формулами:

$$K_{Bi} = \frac{b'_i}{\sum_{i=1}^n b'_i}, \quad (4.9)$$

$$b'_i = \sum_{j=1}^N a_{ij} b_j \quad (4.10)$$

Як видно з таблиці 4.5, різниця значень коефіцієнтів вагомості не перевищує 2%, тому більшої кількості ітерацій не потрібно.

Таблиця 4.5 – Розрахунок вагомості параметрів

Параметри x_i	Параметри x_j				Перша ітер.		Друга ітер.		Третя ітер.	
	X1	X2	X3	X4	b_i	K_{Bi}	b_i^1	K_{Bi}^1	b_i^2	K_{Bi}^2
X1	1	1,5	0,5	1,5	4,5	0,2813	16,25	0,2754	59,125	0,2737
X2	0,5	1	0,5	1,5	3,5	0,2188	12,25	0,2076	44,875	0,2077
X3	1,5	1,5	1	1,5	5,5	0,3438	21,25	0,3602	77,875	0,3605
X4	0,5	0,5	0,5	1	2,5	0,1563	9,25	0,1568	34,125	0,1580
Всього:					16	1	59	1	216	1

4.5 Аналіз рівня якості варіантів реалізації функцій

Визначаємо рівень якості кожного варіанту виконання основних функцій окремо.

Абсолютні значення параметрів X2 (Об'єм пам'яті), X3 (Час навчання моделей, оцінки параметрів розподілів) та X4 (Потенційний об'єм програмного коду) відповідають технічним вимогам умов функціонування даного ПП.

Абсолютне значення параметра X1 (швидкодія мови програмування) обрано не найгіршим.

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

$$K_K(j) = \sum_{i=1}^n K_{\epsilon i,j} B_i, \quad (4.11)$$

де n – кількість параметрів;

$K_{\epsilon i}$ – коефіцієнт вагомості i -го параметра;

B_i – оцінка i -го параметра в балах.

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	A	X1	14000	8	0,2737	2,1896
F2	A	X2	400	4	0,2077	0,8308
F3	A	X4	1200	5	0,3605	1,8025
	Б	X4	1600	6	0,1580	0,948

За даними з таблиці 4.6 за формулою:

$$K_K = K_{Ty}[F_{1k}] + K_{Ty}[F_{2k}] + \dots + K_{Ty}[F_{zk}], \quad (4.12)$$

визначаємо рівень якості кожного з варіантів:

$$K_{K1} = 2,1896 + 0,8308 + 1,8025 = 4,8235$$

$$K_{K2} = 2,1896 + 0,8308 + 0,948 = 3,9688$$

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

4.6 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.

Загальна трудомісткість обчислюється як

$$T_0 = T_P \cdot K_P \cdot K_{СК} \cdot K_M \cdot K_{СТ} \cdot K_{СТ.М},$$

де T_P – трудомісткість розробки ПП;

K_P – 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 степеню 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 $K_{СТ} = 0.75$. Тоді загальна трудомісткість програмування першого завдання дорівнює:

$$T_1 = 90 \cdot 1.6 \cdot 0.75 = 108 \text{ людино-днів.}$$

Проведемо аналогічні розрахунки для подальших завдань.

Для другого завдання (використовується алгоритм третьої групи складності, степінь новизни Б), тобто $T_P = 27$ людино-днів, $K_{\Pi} = 0.9$, $K_{СК} = 1$, $K_{СТ} = 0.75$:

$$T_2 = 27 \cdot 0.9 \cdot 0.75 = 18.225 \text{ людино-днів.}$$

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

$$T_I = (108 + 18.225 + 4.8 + 18.225) \cdot 8 = 1194 \text{ людино-годин.}$$

$$T_{II} = (108 + 18.225 + 6.91 + 18.225) \cdot 8 = 1210,88 \text{ людино-годин.}$$

Найбільш високу трудомісткість має варіант II.

В розробці беруть участь два програмісти з окладом 12 100 грн., один аналітик в області даних з окладом 14 500 грн. Визначимо середню зарплату за годину за формулою:

$$C_{\text{ч}} = \frac{M}{T_m \cdot t} \text{ грн.,} \quad (4.13)$$

де M – місячний оклад працівників;

T_m – кількість робочих днів тиждень;

t – кількість робочих годин в день.

$$C_{\text{ч}} = \frac{12\,100 + 14\,500 + 12\,100}{3 \cdot 21 \cdot 8} = 76,79 \text{ грн.} \quad (4.14)$$

Тоді, розрахуємо заробітну плату за формулою:

$$C_{\text{зп}} = C_{\text{ч}} \cdot T_i \cdot K_{\text{д}}, \quad (4.15)$$

де $C_{\text{ч}}$ – величина погодинної оплати праці програміста;
 T_i – трудомісткість відповідного завдання;
 $K_{\text{д}}$ – норматив, який враховує додаткову заробітну плату.
 Зарплата розробників за варіантами становить:

- I. $C_{\text{зп}} = 76,79 \cdot 1194 \cdot 1.2 = 110018,57 \text{ грн.}$
- II. $C_{\text{зп}} = 76,79 \cdot 1210,88 \cdot 1.2 = 111573,94 \text{ грн.}$

Відрахування на єдиний соціальний внесок становить 22%:

- I. $C_{\text{вйд}} = C_{\text{зп}} \cdot 0.22 = 110018,6 \cdot 0.22 = 24204,09 \text{ грн.}$
- II. $C_{\text{вйд}} = C_{\text{зп}} \cdot 0.22 = 111573,9 \cdot 0.22 = 24546,27 \text{ грн.}$

Тепер визначимо витрати на оплату однієї машино-години. ($C_{\text{м}}$)

Так як одна ЕОМ обслуговує одного програміста з окладом 12 100 грн., з коефіцієнтом зайнятості 0,2 то для однієї машини отримаємо:

$$C_{\text{г}} = 12 \cdot M \cdot K_3 = 12 \cdot 12100 \cdot 0,2 = 29040 \text{ грн.}$$

З урахуванням додаткової заробітної плати:

$$C_{3П} = C_{Г} \cdot (1 + K_3) = 29040 \cdot (1 + 0.2) = 34848 \text{ грн.}$$

Відрахування на соціальний внесок:

$$C_{ВІД} = C_{3П} \cdot 0.22 = 34848 \cdot 0.22 = 7666,56 \text{ грн}$$

Амортизаційні відрахування розраховуємо при амортизації 25% та вартості ЕОМ – 30 000 грн.

$$C_A = K_{ТМ} \cdot K_A \cdot Ц_{ПР} = 1.15 \cdot 0.25 \cdot 30000 = 8625 \text{ грн.}$$

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

K_A – річна норма амортизації;

$Ц_{ПР}$ – договірна ціна приладу.

Витрати на ремонт та профілактику розраховуємо як:

$$C_P = K_{ТМ} \cdot Ц_{ПР} \cdot K_P = 1.15 \cdot 24000 \cdot 0.05 = 1725 \text{ грн.,}$$

де K_P – відсоток витрат на поточні ремонти.

Ефективний годинний фонд часу ПК за рік розраховуємо за формулою:

$$T_{ЕФ} = (D_K - D_B - D_C - D_P) \cdot t_3 \cdot K_B = (365 - 104 - 12 - 14) \cdot 8 \cdot 0.9 = 1692 \text{ годин,}$$

де D_K – 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:

$$C_{\text{ЕЛ}} = T_{\text{ЕФ}} \cdot N_{\text{С}} \cdot K_3 \cdot \text{Ц}_{\text{ЕН}} = 1692 \cdot 0,3 \cdot 4,87 \cdot 0,2 = 494,40 \text{ грн}$$

де $N_{\text{С}}$ – середньо-споживча потужність приладу;

K_3 – коефіцієнтом зайнятості приладу;

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

Накладні витрати розраховуємо за формулою:

$$C_{\text{Н}} = \text{Ц}_{\text{ПР}} \cdot 0,67 = 24000 \cdot 0,67 = 20\,100 \text{ грн.}$$

Тоді, річні експлуатаційні витрати будуть:

$$C_{\text{ЕКС}} = C_{\text{ЗП}} + C_{\text{ВІД}} + C_{\text{А}} + C_{\text{Р}} + C_{\text{ЕЛ}} + C_{\text{Н}}, \quad (4.16)$$

$$C_{\text{ЕКС}} = 34848 + 7666,56 + 8625 + 1725 + 494,40 + 20100 = 73458,96 \text{ грн.}$$

Собівартість однієї машино-години ЕОМ дорівнюватиме:

$$C_{\text{М-Г}} = C_{\text{ЕКС}} / T_{\text{ЕФ}} = 73458,96 / 1692 = 43,42 \text{ грн/год.}$$

Оскільки в даному випадку всі роботи, які пов'язані з розробкою програмного продукту ведуться на ЕОМ, витрати на оплату машинного часу, в залежності від обраного варіанта реалізації, складає:

$$C_{\text{М}} = C_{\text{М-Г}} \cdot T, \quad (4.17)$$

I. $C_{\text{М}} = 43,42 \cdot 1194 = 51838,06 \text{ грн.}$

II. $C_{\text{М}} = 43,42 \cdot 1210,88 = 52570,92 \text{ грн.}$

Накладні витрати складають 67% від заробітної плати:

$$C_H = C_{ЗП} \cdot 0,67, \quad (4.18)$$

I. $C_H = 51838,06 \cdot 0,67 = 34731,50$ грн.

II. $C_H = 52570,92 \cdot 0,67 = 35222,51$ грн.

Отже, вартість розробки ПП за варіантами становить:

$$C_{ПП} = C_{ЗП} + C_{ВІД} + C_M + C_H, \quad (4.19)$$

I. $C_{ПП} = 110018,57 + 24204,09 + 51838,06 + 34731,50 = 220792,22$ грн.

II. $C_{ПП} = 111573,94 + 24546,27 + 52570,92 + 35222,51 = 223913,64$ грн.

4.7 Вибір кращого варіанту ПП техніко-економічного рівня

Розрахуємо коефіцієнт техніко-економічного рівня за формулою:

$$K_{ТЕРj} = K_{Кj} / C_{Фj}, \quad (4.20)$$

$$K_{ТЕР1} = 4,8235 / 220792,22 = 2,18 \cdot 10^{-5},$$

$$K_{ТЕР2} = 3,9688 / 223913,64 = 1,77 \cdot 10^{-5}.$$

Як бачимо, найбільш ефективним є перший варіант реалізації програми з коефіцієнтом техніко-економічного рівня $K_{ТЕР1} = 2,18 \cdot 10^{-5}$.

Після виконання функціонально-вартісного аналізу програмного комплексу що розроблюється, можна зробити висновок, що з альтернатив, що залишились після першого відбору двох варіантів виконання програмного комплексу оптимальним є перший варіант реалізації програмного продукту. У

нього виявився найкращий показник техніко-економічного рівня якості $K_{\text{ТЕР}} = 2,18 \cdot 10^{-5}$.

Цей варіант реалізації програмного продукту має такі параметри:

F_1 – вибір способу об'єднання двох мов програмування в програмі;

F_2 – вибір бібліотеки для роботи з розподілами з важкими хвостами;

F_3 – вибір середовища розробки;

- Об'єднання двох мов програмування буде зроблено так: окремі модулі будуть реалізовані на Python, окремі на R;
- використання бібліотеки SciPy для роботи з розподілами з важкими хвостами;
- використання для розробки інтерактивного середовища Jupyter Lab;

Даний варіант виконання програмного комплексу дає користувачу зручний інтерфейс, непоганий функціонал і швидкодію.

Висновки до розділу

У даній частині роботи було проведено повний функціонально-вартісний аналіз програмного продукту. Визначено та оцінено основні функції програмного продукту. Наведено параметри, які характеризують програмний продукт. Проведено експертне оцінювання параметрів та аналіз якості варіантів реалізації функцій.

Проведено економічний аналіз варіантів розробки – трудомісткість, витрати на заробітну плату та інші витрати.

Варіанти реалізації програмного продукту було порівняно, обрано найкращий.

ВИСНОВКИ

Моделювання поведінки цін акцій має величезне значення для сучасної економіки, зокрема, в останні десятиліття зростає популярність моделі випадкового блукання. Її ймовірнісні прогнози мають бути дуже точними, бо недооцінка ризиків падіння цін призводить до збитків, які можуть скластися в кризу. Питання, який розподіл має використовуватись, є актуальним.

В даній роботі для моделі випадкового блукання застосовувались розподіли з важкими та напівважкими хвостами як альтернатива нормальному розподілу, який довгий час був домінуючим. Було проведено ретельний аналіз історії розвитку моделей ціноутворення акцій. Огляд публікацій показав, що саме розподіли з важкими та напівважкими хвостами доцільно застосовувати в моделях поведінки цін акцій та індексів. Вони об'єктивно оцінюють ризики, не нехтують малоімовірними катастрофічними подіями.

Для досліджуваних розподілів наведено означення, описано основні властивості, оцінки параметрів. Проаналізовано стійкий розподіл, узагальнений скошений розподіл Стюдента, узагальнений гіперболічний.

Для даних про ціну закриття фондового індексу США Standard & Poor 500 з 3 січня 2000 по 30 грудня 2022 року було проведено наближення моделлю випадкового блукання, де використовувались розподіли з важкими та напівважкими хвостами, враховувалась волатильність. Статистичний аналіз емпіричного закону розподілу лог-дохідностей довів неадекватність нормального розподілу. Із п'яти негаусівських розподілів лише узагальнений гіперболічний розподіл дав статистично значущий результат. Для інших фондових індексів підійшов би інший розподіл. Також було розраховано різними способами капітал під ризиком –показник загрози втрат інвестування. Програма для досліду була написана на мовах програмування Python та R.

Отримані результати мають велике практичне значення, оскільки можуть підвищити показники доходності та зменшити ризику інвестування.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Whitehouse K. One 'Quant' sees shakeout for the ages — '10,000 Years'. Wall Street Journal. 2007. August 11.
2. Larsen P. T. Goldman pays the price of being big. Financial Times. 2007. August 13.
3. Fama E. Mandelbrot and the Stable Paretian Hypothesis. The Journal of Business. 1963. Vol 36, No. 4. P. 420-429. DOI: 10.1086/294633
4. B. B. Mandelbrot, Handbook of Heavy Tailed Distributions in Finance / Ed. S.T. Rachev. Elsevier Science B.V.: 2003.
5. Fama E. and Roll R. Parameter Estimates for Symmetric Stable Distributions. Journal of the American Statistical Association. 1971. Vol. 66, P. 331-338. DOI: 10.1080/01621459.1971.10482264
6. Closing price. State investment website. URL: <https://www.investor.gov/introduction-investing/investing-basics/glossary/closing-price>
7. Fama E. Random Walks in Stock Market Prices. Financial Analysts Journal. 1995. Vol. 51, No. 1. P. 75-80. DOI: 10.2469/faj.v51.n1.1861
8. Unro L. Do stock prices follow random walk: Some international evidence. International Review of Economics & Finance. 1992. Vol. 1, No. 4. P. 315-327. DOI: 10.1016/1059-0560(92)90020-D
9. Bachelier, L. Théorie de la speculation. Annales scientifiques de l'École Normale Supérieure. 1990. Vol. 17, No. 3. P. 21-86. DOI: 10.24033/asens.476
10. Roberts H. V. Stock-Market "Patterns" and Financial Analysis: Methodological Suggestions. The Journal of Finance. 1959. Vol. 14, No. 1. P. 1-10. DOI: 10.2307/2976094
11. Merton R., Samuelson P. Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods. Journal of Financial Economics. 1974. Vol. 1, No. 1. P. 67-94. DOI: 10.1016/0304-405X(74)90009-9

12. Mandelbrot B. The Variation of Certain Speculative Prices. *The Journal of Business*. 1963. Vol 36, No. 4. P. 394-419. DOI: 10.1086/294632
13. Gnedenko B., Kolmogorov A. Limit distributions for sums of independent random variables. Addison-Wesley Publishing Company, Inc. 1954. 264 p.
14. Akgiray V., Booth G. The Stable-Law Model of Stock Returns. *Journal of Business & Economic Statistics*. 1988. Vol 6, No. 1. P. 51-57. DOI: 10.2307/1391417
15. Belov I., Kabasinskas A., Sakalauskas L. A study of stable models of stock markets. *Information Technology and Control*. 2006. Vol 35, No. 1. P. 34-56. DOI: 10.5755/J01.ITC.35.1.12033
16. Jayakrishnan N., Wierman A., Zwart B. The fundamentals of Heavy Tails. Properties, Emergence and Estimation, Cambridge University Press, 2022. 262 p.
17. Bulmer M. G. Principles of Statistics. New York; Dover Publications, Inc. New York, 1979. 252 p.
18. Misiorek A., Weron R. Heavy-tailed distributions in VaR calculations. *Statistical Tools for Finance and Insurance*, Second Edition / Ed. P. Cížek. Berlin, Springer, 2011. P. 9-420. DOI: 10.1007/978-3-642-18062-0
19. Basegmez H., Çekici E. Financial applications of stable distributions: implications on turkish stock market. *Pressacademia*. Vol. 6, No. 4. P. 364-374. DOI: 10.17261/Pressacademia.2017.769
20. Fukunaga T., Umeno K. Universal Levy's stable law of stock market and its characterization, February, 2018. DOI: <https://doi.org/10.48550/arXiv.1709.06279>
21. Zi-Yi G. Heavy-tailed Distributions and Risk Management of Equity Market Tail Events. *Journal of Risk & Control, Risk Market Journals*. 2017. Vol. 4, No. 1. P. 31-41.
22. Borak S., Wolfgang H., Weron R. Stable Distributions. SFB 649 Discussion Paper. 2005. P. 1-28. DOI: 10.2139/ssrn.2894444
23. Belov I., Kabasinskas A., Sakalauskas L. A study of stable models of stock markets. *Information Technology and Control*. 2006. Vol 35, No. 1. P. 34-56. DOI: 10.5755/J01.ITC.35.1.12033

24. Necula C. Modeling heavy-tailed stock index returns using the generalized hyperbolic distribution. *Romanian Journal of Economic Forecasting*. 2009. Vol. 6, No. 2, P. 118-131. URL: <https://ideas.repec.org/a/rjr/romjef/v6y2009i2p118-131.html>
25. Virginie K., Wilcox D. Comparison of Generalized Hyperbolic Distribution Models for Equity Returns. *Journal of Applied Mathematics*. 2014. Vol. 15. P. 1-15. DOI: 10.1155/2014/263465
26. Fajardo J., Farias A. Generalized Hyperbolic Distributions and Brazilian Data. *Banco Central do Brasil Working Paper*. 2002. No. 52, P. 3-37. DOI: 10.2139/ssrn.338283
27. Amado P. The distribution of stock returns: international evidence. *Applied Financial Economics*. 1994. Vol. 4, No. 6, P. 431-439. DOI: 10.1080/758518675
28. Kumar D. Long Memory in the Volatility of Indian Financial Market: An Empirical Analysis Based on Indian Data. *Anchor Academic Publishing*, 2014. 104 p.
29. Bielinskyi A., Semerikov S., Solovieva V., Soloviev V. Levy's stable distribution for stock crash detecting. *The 8th International Conference on Monitoring, Modeling & Management of Emergent Economy*. 2019. Vol. 65. P. 1-7. DOI: 10.1051/shsconf/20196506006
30. Morgan J.P. Risk Metrics—Technical Document: 4th Edition. *Journal of Mathematical Finance*. 1996. Vol. 6, No. 1. P. 1-279. URL: <https://www.msci.com/documents/10199/5915b101-4206-4ba0-ae2-3449d5c7e95a>
31. Ruey S. *Analysis of Financial Time Series Third Edition*. John Wiley & Sons, Inc, 2010. 720 p.
32. Jorion P. *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd Edition. The McGraw-Hill Companies Inc, 2007. 594 p.
33. Gurrola-Perez P., Murphy D. Filtered Historical Simulation Value-at-Risk Models and Their Competitors. *Bank of England Working Paper*. 2015. No. 525. P. 1-29. URL: <https://www.bankofengland.co.uk/-/media/boe/files/working->

paper/2015/filtered-historical-simulation-value-at-risk-models-and-their-competitors.pdf

34. Fama E. Estimating Value at Risk (VaR) using Filtered Historical Simulation in the Indian capital market. Reserve Bank of India Occasional Papers. 2011. Vol. 32, No. 2. P. 81-98.

URL: http://filteredhistoricalsimulation.com/downloads/India_Res_Bank.pdf

35. McCulloch J. Simple consistent estimators of stable distribution parameters. Communications in Statistics - Simulation and Computation. 1986. Vol. 15, No. 4. P. 1109-1136. DOI:10.1080/03610918608812563

36. Koutrouvelis I. Regression-Type Estimation of the Parameters of Stable Laws. Journal of the American Statistical Association. 1980. Vol. 75, No. 372. P. 918–928. DOI: 10.2307/2287182

37. Nolan J. Numerical computation of stable densities and distribution functions. Communications in Statistics. Stochastic Models. 1997. Vol. 13, No. 4. P. 759-774, DOI: 10.1080/15326349708807450

38. Véhel J., Philippe A., Robet C. Explicit and combined estimators for stable distributions parameters. Statistics: A Journal of Theoretical and Applied Statistics. 2021. Vol. 55, No 4. P. 1-29, DOI:10.1080/02331888.2021.1976776

39. Hansen C., McDonald J., Whitney N. Instrumental Variables Regression with Flexible Distributions. Journal of Business and Economic Statistics. 2010. Vol. 28, No. 1. P. 13-25, DOI:10.1198/jbes.2009.06161

40. R documentation, package “SGT”, The Skewed Generalized T Distribution Tree Package Vignette, by Carter Davis, published 2015-09-04. URL: <https://cran.r-project.org/web/packages/sgt/vignettes/sgt.pdf>

41. R documentation, package “SGT”, by Carter Davis, published 2015-08-30. URL: <https://cran.r-project.org/web/packages/sgt/sgt.pdf>

42. Borak S., Misiorek A., Weron R. Models for Heavy-tailed Asset Returns. Statistical Tools for Finance and Insurance, Second Edition / Ed. Cizek P. Springer, 2011. P. 21-55

43. Wang C. On the Numerics of Estimating Generalized Hyperbolic Distributions: Doctoral dissertation / Humboldt University of Berlin. Berlin, 2005. 98 p.
44. Hanssen A., Arne T. The normal inverse gaussian distribution: a versatile model for heavy-tailed stochastic processes. Proceedings – ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, May, 2001. P. 3985-3988
45. Barone-Adesi G., Giannopoulos K., Vosper L. VaR Without Correlations for Portfolios of Derivative Securities. Simulating Security Returns: A Filtered Historical Simulation Approach / ed. Barone-Adesi G, New York, NY: Palgrave Pivot, 2014. P. 9-29. DOI: 10.1057/9781137465559_2
46. Zhang, Y., Nadarajah, S. A review of backtesting for value at risk. Communications in Statistics – Theory and Methods. 2017. P. 1-24. DOI: 10.1080/03610926.2017.1361984
47. Barndorff-Nielsen O., Halgreen C. Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. Probability Theory and Related Fields. 1977. P. 309-311. DOI: 10.1007/BF00533162

ДОДАТОК А Лістинг програми на мові R для оцінки параметрів узагальненого скошеного розподілу Стюдента та побудови відповідної щільності

```

library(rgt)
library(readxl)
library(writexl)
library(pracma)

data = read_excel('df_standardized_log_returns.xlsx')
data = data[['0']]

df =
data.frame(var1=c('mean'),var2=c('sigma'),var3=c('lambda'),var4=c('p'),var5=c(
'q'))

print(data)

start = list(mu = 0, sigma = 1, lambda = 0, p = 2, q = 10)
result = rgt.mle(X.f = ~data, start = start, method = "nlminb")

est_par = result['estimate']
est_par_2 = est_par[1]
est_par_3 = est_par_2[['estimate']]

mu = est_par_3[['mu']]
sigma = est_par_3[['sigma']]
lambda = est_par_3[['lambda']]
p = est_par_3[['p']]
q = est_par_3[['q']]

df = rbind(df ,
data.frame(var1=c(mu),var2=c(sigma),var3=c(lambda),var4=c(p),var5=c(q)) )

print(mu)

```

```
print(sigma)
print(lambda)
print(p)
print(q)

t = linspace(-10, 10, n = 1000)
pdf = dsgrt(t, mu=mu, sigma=sigma, p=p, q=q)
df_2 = data.frame(t, pdf)

write_xlsx(df, 'parameters_of_t_distribution.xlsx')
write_xlsx(df_2, 'pdf_of_t_distribution.xlsx')
```

ДОДАТОК Б Лістинг програми на мові Python для оцінки параметрів розподілів та візуалізації результатів

```

import numpy as np
from scipy.stats import kstest
from scipy.stats import levy_stable
from scipy.stats import norm
from scipy.stats import genhyperbolic
from scipy.stats import fit

import matplotlib.pyplot as plt
import yfinance as yf
import pandas as pd
from arch.univariate import arch_model

plot_xleft = -10
plot_xright = 10

def pdf_stable(delta, alpha, beta, c):
    x = np.linspace(plot_xleft, plot_xright, 250)
    return x, levy_stable.pdf(x, alpha, beta, loc = delta, scale = c)

def pdf_norm(mean, std):
    x = np.linspace(plot_xleft, plot_xright, 250)
    return x, norm.pdf(x, loc=mean, scale=std)

def pdf_gen_hyp(lambda_p, a, b, loc, scale):
    x = np.linspace(plot_xleft, plot_xright, 250)
    return x, genhyperbolic.pdf(x, lambda_p, a, b, loc=loc, scale=scale)

def pdf_laplace(loc_par, scale_par):
    x = np.linspace(plot_xleft, plot_xright, 250)
    return x, laplace.pdf(x, loc=loc_par, scale=scale_par)

```

```

def pdf_t(df_par, loc_par, scale_par):
    x = np.linspace(plot_xleft, plot_xright, 250)
    return x, t.pdf(x, df=df_par, loc=loc_par, scale=scale_par)

data = pd.read_excel('SPX_xlsx.xlsx')
data['Adj Close'] = data['Adj Close']/1000

fig, ax = plt.subplots(1, 1, figsize=(12,6))

ax.plot(data['Date'], data['Adj Close'], color = 'blue', lw=1)
ax.set_xlabel('Рік')
ax.set_ylabel('Ціна індексу S&P 500, тис. дол')
ax.margins(x=0)
plt.show()

price_list = np.array(data['Adj Close'])
data_list = list(data.index)

log_returns = []
for i in np.arange(1, len(price_list)):
    log_returns.append( np.log(price_list[i]) - np.log(price_list[i - 1]) )

log_returns = np.array(log_returns)

cleared_log_returns = log_returns[log_returns != 0]
cleared_data = np.array(data['Date'][1:])[log_returns != 0]
log_returns = cleared_log_returns

fig, ax = plt.subplots(1, 1, figsize=(12,6))

ax.plot(cleared_data, log_returns, color = 'blue', lw=1)
ax.set_xlabel('Дата')
ax.set_ylabel('Лог-дохідності акцій S&P 500')
ax.margins(x=0)
plt.show()

```

```

df_log_returns = pd.DataFrame(log_returns)
df_log_returns.to_excel('df_not_standardized_log_returns.xlsx')

am = arch_model(log_returns, mean='Constant', vol = 'GARCH',
p=1,o=0,q=1,power=2.0,dist='normal')
res = am.fit()
print(res.summary())

standardized_residuals = (res.resid -
res.params.mu)/res.conditional_volatility

standardized_log_returns = standardized_residuals

fig, ax = plt.subplots(1, 1, figsize=(20,10))

ax.plot(cleared_data, standardized_log_returns, color = 'blue', lw=1,
label='Фільтровані лог-доходи')
ax.set_xlabel('Дата', fontsize=15)
ax.set_ylabel('Стандартизовані лог-дохідності S&P 500', fontsize=15)
ax.margins(x=0)
plt.show()

fig, ax = plt.subplots(1, 1, figsize=(12,6))

ax.plot(cleared_data, log_returns, color = 'blue', lw=1, label = 'Лог-
дохідність')
ax.plot(cleared_data, res._volatility, color = 'red', lw=1, label =
'Волатильність')
ax.set_xlabel('Дата')
ax.set_ylabel('Лог-дохідності акцій S&P 500')
ax.margins(x=0)
ax.legend(loc='best', frameon=False, fontsize=15)
plt.show()

mean_estimate = []
vol_estimates = []

for i in np.arange(1, len(standardized_residuals)):

```

```

vol_estimates.append( np.std(standardized_residuals[0:i]) )
mean_estimate.append( np.mean(standardized_residuals[0:i]) )

fig, ax = plt.subplots(1, 1, figsize=(16,6))

ax.set_xlabel('Розмір вибірки')
ax.set_ylabel('Оцінка середньоквадратичного відхилення')

ax.plot(np.arange(0, len(mean_estimate)), mean_estimate, color='red', label =
'Oцінка середнього')

ax.plot(np.arange(0, len(vol_estimates)), vol_estimates, color= 'blue', label
= 'Оцінка стандартного відхилення')

ax.plot([0, len(mean_estimate)], [mean_estimate[-1], mean_estimate[-1]],
color= 'red', linestyle = (0, (5, 5)))

ax.plot([0, len(vol_estimates)], [vol_estimates[-1], vol_estimates[-1]],
color= 'blue', linestyle = (0, (5, 5)))

ax.legend(loc='best', frameon=False, fontsize=15)

df_standardized_log_returns = pd.DataFrame(standardized_log_returns)
df_standardized_log_returns.to_excel('df_standardized_log_returns.xlsx')

import rpy2
import os
os.environ['R_HOME'] = '/Users/38050/anaconda3/Lib/R'

from rpy2.robj.robj.robj import importr

StableEstim = importr('StableEstim')
stabledist = importr('stabledist')
utils = importr('utils')
sgt = importr('sgt')
optimx = importr('optimx')
numDeriv = importr('numDeriv')

R_log_returns = rpy2.robj.FloatVector(standardized_log_returns)

```

```

stable_est_par = StableEstim.KoutParametersEstim(R_log_returns)
stable_est_par = stable_est_par.rx2('Estim').rx2('par')
res = np.array(list( stable_est_par ))

hyperbolic_par = fit(genhyperbolic, standardized_log_returns, bounds={'p':
(1,1), 'a': (-0.5, 0.5), 'b': (-0.4,0.4), 'loc': (-0.1, 0.1), 'scale': (-0.5,
0.5)})

nig_par = fit(genhyperbolic, standardized_log_returns, bounds={'p': (-0.5,-
0.5), 'a': (-0.5, 0.5), 'b': (-0.4,0.4), 'loc': (-0.1, 0.1), 'scale': (-0.5,
0.5)})

gh_par = fit(genhyperbolic, standardized_log_returns, bounds={'p': (-3, 3),
'a': (-0.5, 0.5), 'b': (-0.4,0.4), 'loc': (-0.1, 0.1), 'scale': (-0.5, 0.5)})

est_mu = np.mean(standardized_log_returns)
est_std = np.std(standardized_log_returns, ddof=1)

alpha = res[0]
beta = res[1]
gamma = res[2]
delta = res[3]

hyp_p = hyperbolic_par.params.p
hyp_a = hyperbolic_par.params.a
hyp_b = hyperbolic_par.params.b
hyp_loc = hyperbolic_par.params.loc
hyp_scale = hyperbolic_par.params.scale

nig_p = nig_par.params.p
nig_a = nig_par.params.a
nig_b = nig_par.params.b
nig_loc = nig_par.params.loc
nig_scale = nig_par.params.scale

gh_p = gh_par.params.p
gh_a = gh_par.params.a
gh_b = gh_par.params.b
gh_loc = gh_par.params.loc

```

```
gh_scale = gh_par.params.scale
```

```
gh_x, gh_y = pdf_gen_hyp(gh_p, gh_a, gh_b, gh_loc, gh_scale)
```

```
nig_x, nig_y = pdf_gen_hyp(nig_p, nig_a, nig_b, nig_loc, nig_scale)
```

```
hyp_x, hyp_y = pdf_gen_hyp(hyp_p, hyp_a, hyp_b, hyp_loc, hyp_scale)
```

```
stable_x, stable_y = pdf_stable(delta, alpha, beta, gamma)
```

```
norm_x, norm_y = pdf_norm(est_mu, est_std)
```

```
dataframe__skewed_t = pd.read_excel('pdf_of_t_distribution.xlsx')
```

```
skewt_x = np.array(dataframe__skewed_t['t'])
```

```
skewt_y = np.array(dataframe__skewed_t['pdf'])
```

```
fig, ax = plt.subplots(1, 1, figsize=(18,10))
```

```
ax.plot(norm_x, norm_y, color = 'blue', lw=2, alpha=0.6, label='Нормальний')
```

```
ax.plot(stable_x, stable_y, color = 'red', lw=2, alpha=0.6, label='Стійкий')
```

```
ax.plot(hyp_x, hyp_y, color = 'green', lw=2, alpha=0.6,  
label='Гіперболічний')
```

```
ax.plot(nig_x, nig_y, color = 'magenta', lw=2, alpha=0.6, label='Нормальний  
обернений гаусівський')
```

```
ax.plot(gh_x, gh_y, color = 'darkorange', lw=2, alpha=0.6,  
label='Узагальнений гіперболічний')
```

```
ax.plot(skewt_x, skewt_y, color = 'cyan', lw=2, alpha=0.6, label='Скошений  
узагальнений розподіл Стюдента')
```

```
ax.hist(standardized_residuals, density=True, bins='auto',  
histtype='stepfilled', alpha=0.2)
```

```
ax.set_xlim([-6, 6])
```

```
ax.legend(loc='best', frameon=False, fontsize=15)
```

```
plt.show()
```

```
fig, ax = plt.subplots(1, 1, figsize=(18,10))
```

```
ax.plot(norm_x, norm_y, color = 'darkblue', lw=2, label='Нормальний')
```

```
ax.plot(gh_x, gh_y, color = 'red', lw=2, label='Узагальнений гіперболічний')
```

```
ax.hist(standardized_residuals, density=True, bins='auto',  
histtype='stepfilled', alpha=0.2)
```

```
ax.set_xlim([-6, 6])
```

```
ax.legend(loc='best', frameon=False, fontsize=15)
```

```
plt.show()
```

```

skew_x_cdf = skewt_x.copy()
skew_y_cdf = np.cumsum(skewt_y)*(skewt_x[1] - skewt_x[0])

skew_x_left = skew_x_cdf[(skew_x_cdf < 0) & (skew_x_cdf > -9)]
skew_y_left = skew_y_cdf[(skew_x_cdf < 0) & (skew_x_cdf > -9)]

skew_x_right = skew_x_cdf[(skew_x_cdf > 0) & (skew_x_cdf < 9)]
skew_y_right = skew_y_cdf[(skew_x_cdf > 0) & (skew_x_cdf < 9)]

# sort the data in ascending order
x_empirical_cdf = np.sort(standardized_log_returns)
# get the cdf values of y
N = len(standardized_log_returns)
y_empirical_cdf = np.arange(N) / float(N)

x_empirical_cdf_right = (np.copy(x_empirical_cdf))[x_empirical_cdf >= 0]
x_empirical_cdf_left = (np.copy(x_empirical_cdf))[x_empirical_cdf < 0]

y_empirical_cdf_right = (np.copy(y_empirical_cdf))[x_empirical_cdf >= 0]
y_empirical_cdf_left = (np.copy(y_empirical_cdf))[x_empirical_cdf < 0]

#####

x_cdf_left = np.linspace(
min(standardized_log_returns[standardized_log_returns < 0]), 0, 200)
x_cdf_right = np.linspace(0,
max(standardized_log_returns[standardized_log_returns > 0]), 200)

#####

fig, ax = plt.subplots(1, 2, figsize = (18,9))

ax[0].plot(x_cdf_left, levy_stable.cdf(x_cdf_left, alpha=alpha,beta=beta,
loc=delta,scale=gamma), color = 'crimson', label='Стійкий')

```

```

ax[0].plot(x_cdf_left, norm.cdf(x_cdf_left, loc=est_mu, scale=est_std), color
= 'cyan', label='Нормальний')

ax[0].plot(x_cdf_left, genhyperbolic.cdf(x_cdf_left, hyp_p, hyp_a, hyp_b,
hyp_loc, hyp_scale), color = 'orange', label='Гіперболічний')

ax[0].plot(x_cdf_left, genhyperbolic.cdf(x_cdf_left, nig_p, nig_a, nig_b,
nig_loc, nig_scale), color = 'green', label='Нормальний обернений
гаусівський')

ax[0].plot(x_cdf_left, genhyperbolic.cdf(x_cdf_left, gh_p, gh_a, gh_b,
gh_loc, gh_scale), color = 'magenta', label='Узагальнений гіперболічний')


ax[0].scatter(x_empirical_cdf_left, y_empirical_cdf_left, marker='x',
color='black', linewidths=1)

ax[0].legend(loc='best', frameon=False)


ax[1].plot(x_cdf_right, levy_stable.cdf(x_cdf_right, alpha=alpha,beta=beta,
loc=delta,scale=gamma), color = 'crimson', label='Стійкий')

ax[1].plot(x_cdf_right, norm.cdf(x_cdf_right, loc=est_mu, scale=est_std),
color = 'cyan', label='Нормальний')

ax[1].plot(x_cdf_right, genhyperbolic.cdf(x_cdf_right, hyp_p, hyp_a, hyp_b,
hyp_loc, hyp_scale), color = 'orange', label='Гіперболічний')

ax[1].plot(x_cdf_right, genhyperbolic.cdf(x_cdf_right, nig_p, nig_a, nig_b,
nig_loc, nig_scale), color = 'green', label='Нормальний обернений
гаусівський')

ax[1].plot(x_cdf_right, genhyperbolic.cdf(x_cdf_right, gh_p, gh_a, gh_b,
gh_loc, gh_scale), color = 'magenta', label='Узагальнений гіперболічний')


ax[1].scatter(x_empirical_cdf_right, y_empirical_cdf_right, marker='x',
color='black', linewidths=1)

ax[1].legend(loc='best', frameon=False)


#####

fig, ax = plt.subplots(1, 1, figsize = (12,5))


ax.plot(x_empirical_cdf, norm.cdf(x_empirical_cdf, loc=est_mu,
scale=est_std), color = 'cyan', label='Нормальний розподіл')

ax.plot(x_empirical_cdf, genhyperbolic.cdf(x_empirical_cdf, hyp_p, hyp_a,
hyp_b, hyp_loc, hyp_scale), color = 'orange', label='Гіперболічний')

ax.plot(x_empirical_cdf, genhyperbolic.cdf(x_empirical_cdf, nig_p, nig_a,
nig_b, nig_loc, nig_scale), color = 'green', label='Нормальний обернений
гаусівський розподіл')

ax.plot(x_empirical_cdf, levy_stable.cdf(x_empirical_cdf,
alpha=alpha,beta=beta, loc=delta,scale=gamma), color = 'crimson',
label='Стійкий розподіл')

```

```

ax.plot(x_empirical_cdf, genhyperbolic.cdf(x_empirical_cdf, gh_p, gh_a, gh_b,
gh_loc, gh_scale), color = 'magenta', label='Узагальнений гіперболічний')

ax.scatter(x_empirical_cdf, y_empirical_cdf, marker='x', color='black',
linewidths=1, label = 'Емпірична функція розподілу')

ax.margins(x=0)

ax.legend(loc='best', frameon=False)

# sort the data in ascending order
x_empirical_cdf = np.sort(standardized_log_returns)

# get the cdf values of y
N = len(standardized_log_returns)
y_empirical_cdf = np.arange(N) / float(N)

x_empirical_cdf_right = (np.copy(x_empirical_cdf))[x_empirical_cdf >= 0]
x_empirical_cdf_left = (np.copy(x_empirical_cdf))[x_empirical_cdf < 0]

y_empirical_cdf_right = (np.copy(y_empirical_cdf))[x_empirical_cdf >= 0]
y_empirical_cdf_left = (np.copy(y_empirical_cdf))[x_empirical_cdf < 0]

#####

x_cdf_left = np.linspace(
min(standardized_log_returns[standardized_log_returns < 0]), 0, 300)

x_cdf_right = np.linspace(0,
max(standardized_log_returns[standardized_log_returns > 0]), 300)

#####

skew_x_cdf = skewt_x.copy()
skew_y_cdf = np.cumsum(skewt_y)*(skewt_x[1] - skewt_x[0])

skew_x_left = skew_x_cdf[(skew_x_cdf < 0)]
skew_y_left = skew_y_cdf[(skew_x_cdf < 0)]

skew_x_right = skew_x_cdf[(skew_x_cdf > 0)]
skew_y_right = skew_y_cdf[(skew_x_cdf > 0)]

```

```
#####

fig, ax = plt.subplots(1, 2, figsize = (14,6))

ax[0].scatter(-x_empirical_cdf_left, y_empirical_cdf_left, marker='x',
color='black', linewidths=1, label = 'Дані')

ax[0].plot(-x_cdf_left, levy_stable.cdf(x_cdf_left, alpha=alpha,beta=beta,
loc=delta,scale=gamma), alpha = 0.2, color = 'crimson', label='Стійкий
розподіл', linestyle = (0, (4, 2)))

ax[0].plot(-x_cdf_left, norm.cdf(x_cdf_left, loc=est_mu, scale=est_std),
color = 'blue', label='Нормальний розподіл')

ax[0].plot(-x_cdf_left, genhyperbolic.cdf(x_cdf_left, hyp_p, hyp_a, hyp_b,
hyp_loc, hyp_scale), color = 'green', label='Гіперболічний розподіл')

ax[0].plot(-x_cdf_left, genhyperbolic.cdf(x_cdf_left, nig_p, nig_a, nig_b,
nig_loc, nig_scale), alpha = 0.2, color = 'darkblue', label='Нормальний
обернений гаусівський розподіл', linestyle = (0, (4, 2)))

ax[0].plot(-x_cdf_left, genhyperbolic.cdf(x_cdf_left, gh_p, gh_a, gh_b,
gh_loc, gh_scale), color = 'red', label='Узагальнений гіперболічний
розподіл')

ax[0].plot(-skew_x_left, skew_y_left, color = 'darkorange',
label='Узагальнений скошений розподіл Стюдента', linestyle = (0, (4, 2)))

ax[0].legend(loc='best', frameon=False)

ax[0].set_xscale('log')

ax[0].set_yscale('log')

ax[0].set_xlabel('-X')

ax[0].set_ylim([10e-5, 2*10e-1])

ax[0].set_ylabel('F(X)')

ax[1].scatter(x_empirical_cdf_right, 1 - y_empirical_cdf_right, marker='x',
color='black', linewidths=1, label = 'Дані')

ax[1].plot(x_cdf_right, 1 - levy_stable.cdf(x_cdf_right,
alpha=alpha,beta=beta, loc=delta,scale=gamma), alpha = 0.2, color =
'crimson', label='Стійкий розподіл', linestyle = (0, (4, 2)))

ax[1].plot(x_cdf_right, 1 - norm.cdf(x_cdf_right, loc=est_mu, scale=est_std),
color = 'blue', label='Нормальний розподіл')

ax[1].plot(x_cdf_right, 1 - genhyperbolic.cdf(x_cdf_right, hyp_p, hyp_a,
hyp_b, hyp_loc, hyp_scale), color = 'green', label='Гіперболічний розподіл')

ax[1].plot(x_cdf_right, 1 - genhyperbolic.cdf(x_cdf_right, nig_p, nig_a,
nig_b, nig_loc, nig_scale), alpha = 0.2, color = 'darkblue',
label='Нормальний обернений гаусівський розподіл', linestyle = (0, (4, 2)))

ax[1].plot(x_cdf_right, 1 - genhyperbolic.cdf(x_cdf_right, gh_p, gh_a, gh_b,
gh_loc, gh_scale), color = 'red', label='Узагальнений гіперболічний
розподіл')
```

```

ax[1].plot(skew_x_right, 1 - skew_y_right, color = 'darkorange',
label='Узагальнений скошений розподіл Стюдента', linestyle = (0, (4, 2)))

ax[1].legend(loc='best', frameon=False)

ax[1].set_xscale('log')

ax[1].set_yscale('log')

ax[1].set_xlabel('X')

ax[1].set_ylim([10e-5, 2*10e-1])

ax[1].set_ylabel('1 - F(X)')

```

```

# sort the data in ascending order
x_empirical_cdf = np.sort(standardized_log_returns)

# get the cdf values of y
N = len(standardized_log_returns)
y_empirical_cdf = np.arange(N) / float(N)

```

```

x_empirical_cdf_right = (np.copy(x_empirical_cdf))[x_empirical_cdf >= 0]
x_empirical_cdf_left = (np.copy(x_empirical_cdf))[x_empirical_cdf < 0]

```

```

y_empirical_cdf_right = (np.copy(y_empirical_cdf))[x_empirical_cdf >= 0]
y_empirical_cdf_left = (np.copy(y_empirical_cdf))[x_empirical_cdf < 0]

```

```

#####

```

```

x_cdf_left = np.linspace(
min(standardized_log_returns[standardized_log_returns < 0]), 0, 300)

x_cdf_right = np.linspace(0,
max(standardized_log_returns[standardized_log_returns > 0]), 300)

```

```

#####

```

```

skew_x_cdf = skewt_x.copy()

skew_y_cdf = np.cumsum(skewt_y)*(skewt_x[1] - skewt_x[0])

```

```

skew_x_left = skew_x_cdf[(skew_x_cdf < 0)]

skew_y_left = skew_y_cdf[(skew_x_cdf < 0)]

```

```

skew_x_right = skew_x_cdf[(skew_x_cdf > 0)]
skew_y_right = skew_y_cdf[(skew_x_cdf > 0)]

#####

fig, ax = plt.subplots(1, 2, figsize = (14,6))

ax[0].scatter(-x_empirical_cdf_left, y_empirical_cdf_left, marker='x',
color='black', linewidths=1, label = 'Дані')

#ax[0].plot(-x_cdf_left, levy_stable.cdf(x_cdf_left, alpha=alpha,beta=beta,
loc=delta,scale=gamma), color = 'crimson', label='Стійкий розподіл',
linestyle = (0, (4, 2)))

ax[0].plot(-x_cdf_left, norm.cdf(x_cdf_left, loc=est_mu, scale=est_std),
color = 'cyan', label='Нормальний розподіл')

ax[0].plot(-x_cdf_left, genhyperbolic.cdf(x_cdf_left, hyp_p, hyp_a, hyp_b,
hyp_loc, hyp_scale), color = 'green', label='Гіперболічний розподіл')

#ax[0].plot(-x_cdf_left, genhyperbolic.cdf(x_cdf_left, nig_p, nig_a, nig_b,
nig_loc, nig_scale), color = 'darkblue', label='Нормальний обернений
гаусівський розподіл', alpha=0.3, linestyle = (0, (4, 2)))

ax[0].plot(-x_cdf_left, genhyperbolic.cdf(x_cdf_left, gh_p, gh_a, gh_b,
gh_loc, gh_scale), color = 'magenta', label='Узагальнений гіперболічний
розподіл')

#ax[0].plot(-skew_x_left, skew_y_left, color = 'darkorange',
label='Узагальнений скошений розподіл Стюдента', linestyle = (0, (4, 2)))

ax[0].legend(loc='best', frameon=False)

ax[0].set_xscale('log')
ax[0].set_yscale('log')
ax[0].set_xlabel('-X')
ax[0].set_ylim([10e-5, 2*10e-1])
ax[0].set_ylabel('F(X)')

ax[1].scatter(x_empirical_cdf_right, 1 - y_empirical_cdf_right, marker='x',
color='black', linewidths=1, label = 'Дані')

ax[1].plot(x_cdf_right, 1 - norm.cdf(x_cdf_right, loc=est_mu, scale=est_std),
color = 'cyan', label='Нормальний розподіл')

ax[1].plot(x_cdf_right, 1 - genhyperbolic.cdf(x_cdf_right, hyp_p, hyp_a,
hyp_b, hyp_loc, hyp_scale), color = 'green', label='Гіперболічний розподіл')

ax[1].plot(x_cdf_right, 1 - genhyperbolic.cdf(x_cdf_right, gh_p, gh_a, gh_b,
gh_loc, gh_scale), color = 'magenta', label='Узагальнений гіперболічний
розподіл')

ax[1].legend(loc='best', frameon=False)

ax[1].set_xscale('log')
ax[1].set_yscale('log')

```

```

ax[1].set_xlabel('X')
ax[1].set_ylim([10e-5, 2*10e-1])
ax[1].set_ylabel('1 - F(X)')

print(kstest(standardized_log_returns, norm.cdf, args = (est_mu, est_std) ))
print(  kstest(standardized_log_returns, levy_stable.cdf, args = (alpha,
beta, delta, gamma) ) )

print(  kstest(standardized_log_returns, genhyperbolic.cdf, args = (hyp_p,
hyp_a, hyp_b, hyp_loc, hyp_scale) ) )

print(  kstest(standardized_log_returns, genhyperbolic.cdf, args = (nig_p,
nig_a, nig_b, nig_loc, nig_scale) ) )

print(  kstest(standardized_log_returns, genhyperbolic.cdf, args = (gh_p,
gh_a, gh_b, gh_loc, gh_scale) ) )

```

ДОДАТОК В Лістинг програми на мові R для оцінки VAR ковзним вікном для стійкого та узагальненого скошеного розподілу Стюдента

```

library(sgt)
library(readxl)
library(writexl)
library(pracma)

library(StableEstim)

options(warn = 0)

data = read_excel('C:\\Users\\38050\\Desktop\\Диплом\\Прототип диплому
04.04.2023\\Оцінка VAR 10.05.2023\\df_not_standardized_log_returns.xlsx');

data = data[['0']];

window_size = 300;
i = 1;
length__ = length(data);

df1 =
data.frame(var1=c('mean'),var2=c('sigma'),var3=c('lambda'),var4=c('p'),var5=c(
'q'),var6=c('VAR 90%'),var7=c('VAR 95%'),var8=c('VAR 99%'));

df2 =
data.frame(var1=c('par1'),var2=c('par2'),var3=c('par3'),var4=c('par4'));

print(data)

while (window_size + i < length__ ) {

temp = data[i:(window_size + i)];

#####

```

```

start = list(mu = 0, sigma = 1, lambda = 0, p = 2, q = 10);

result = tryCatch( sgt.mle(X.f = ~temp, start = start, method = "nlminb"),
error = function(e) NA);

est_par = result['estimate'];
est_par_2 = est_par[1];
est_par_3 = est_par_2[['estimate']]

#print(est_par_3)

if ( all(is.na(est_par_3)) ) {
  mu = NA;
  sigma = NA;
  lambda = NA;
  p = NA;
  q = NA;

  var90 = NA;
  var95 = NA;
  var99 = NA;
} else {

  mu = est_par_3[['mu']];
  sigma = est_par_3[['sigma']];
  lambda = est_par_3[['lambda']];
  p = est_par_3[['p']];
  q = est_par_3[['q']];

  var90 = qsgt(0.1, mu = mu, sigma = sigma, lambda = lambda, p = p, q =
q);
  var95 = qsgt(0.05, mu = mu, sigma = sigma, lambda = lambda, p = p, q =
q);
  var99 = qsgt(0.01, mu = mu, sigma = sigma, lambda = lambda, p = p, q =
q);
}

```

```
#####

result_2 = tryCatch( KoutParametersEstim(temp), error = function(e) NA);
#print(result_2)

if ( (all(is.na(result_2))) ) {

par1 = 0;
par2 = 0;
par3 = 0;
par4 = 0;

} else {
result_2 = result_2['Estim'];
result_2 = result_2$Estim$par;

par1 = result_2[1];
par2 = result_2[2];
par3 = result_2[3];
par4 = result_2[4];
}

df1 = rbind(df1 ,
data.frame(var1=c(mu),var2=c(sigma),var3=c(lambda),var4=c(p),var5=c(q),var6=c
(var90),var7=c(var95),var8=c(var99)) );

df2 = rbind(df2 ,
data.frame(var1=c(par1),var2=c(par2),var3=c(par3),var4=c(par4)) );

print( i )
i = i + 1;

}

write_xlsx(df1, 'estimate_parameters_skewed_t.xlsx')
write_xlsx(df2, 'estimate_parameters_stable.xlsx')
```

ДОДАТОК Г Лістинг програми на мові Python для оцінки VAR ковзним вікном розподілами та візуалізації результатів

```

import numpy as np
from scipy.stats import kstest
from scipy.stats import levy_stable
from scipy.stats import norm
from scipy.stats import genhyperbolic
from scipy.stats import fit
from scipy.stats import t
from scipy.stats import laplace

import matplotlib.pyplot as plt
import pandas as pd
import random

from arch.univariate import arch_model
data = pd.read_excel('SPX_xlsx.xlsx')
data['Adj Close'] = data['Adj Close']/1000

fig, ax = plt.subplots(1, 1, figsize=(12,6))

ax.plot(data['Date'], data['Adj Close'], color = 'blue', lw=1)
ax.set_xlabel('Дата')
ax.set_ylabel('Ціна індексу S&P 500, тис. дол')
plt.show()

price_list = np.array(data['Adj Close'])
data_list = list(data.index)

log_returns = []
for i in np.arange(1, len(price_list)):
    log_returns.append( np.log(price_list[i]) - np.log(price_list[i - 1]) )

```

```

log_returns = np.array(log_returns)

cleared_log_returns = log_returns[log_returns != 0]
cleared_data = np.array(data['Date'][1:])[log_returns != 0]
log_returns = cleared_log_returns

fig, ax = plt.subplots(1, 1, figsize=(12,6))

ax.plot(cleared_data, log_returns, color = 'blue', lw=1)
ax.set_xlabel('Дата')
ax.set_ylabel('Лог-доходи акцій S&P 500')
plt.show()

df_skewed_t = pd.read_excel('estimate_parameters_skewed_t.xlsx')
df_skewed_t = df_skewed_t.iloc[1:]
df_skewed_t = df_skewed_t.astype(float)

df_stable = pd.read_excel('estimate_parameters_stable.xlsx')
df_stable = df_stable.iloc[1:]
df_stable = df_stable.astype(float)

i = 0
length__ = np.shape(df_skewed_t)[0]
df_skewed_t_cleared = pd.DataFrame(columns=['mean', 'sigma', 'lambda', 'p',
'q', 'VAR90%', 'VAR95%', 'VAR99%'])

temp_prev = 0
while i < length__:
    temp = list(df_skewed_t.iloc[i])
    if any(np.isnan( temp )):
        df_skewed_t_cleared.loc[len(df_skewed_t_cleared)] = temp_prev
    else:

```

```

df_skewed_t_cleared.loc[len(df_skewed_t_cleared)] = temp
temp_prev = temp
i += 1

skewed_t_var90 = np.array(df_skewed_t_cleared['VAR90%'])
skewed_t_var95 = np.array(df_skewed_t_cleared['VAR95%'])
skewed_t_var99 = np.array(df_skewed_t_cleared['VAR99%'])

window_size = 300
i_start = 0
i_end = window_size

Par_VaR_90 = []
Par_VaR_95 = []
Par_VaR_99 = []

while i_end < len(log_returns) - 1:

    #####

    hyperbolic_par = fit(genhyperbolic, log_returns[i_start:i_end],
bounds={'p': (1,1), 'a': (-0.5, 0.5), 'b': (-0.4,0.4), 'loc': (-0.1, 0.1),
'scale': (-0.5, 0.5)})

    nig_par = fit(genhyperbolic, log_returns[i_start:i_end], bounds={'p': (-
0.5,-0.5), 'a': (-0.5, 0.5), 'b': (-0.4,0.4), 'loc': (-0.1, 0.1), 'scale': (-
0.5, 0.5)})

    gh_par = fit(genhyperbolic, log_returns[i_start:i_end], bounds={'p': (-3,
3), 'a': (-0.5, 0.5), 'b': (-0.4,0.4), 'loc': (-0.1, 0.1), 'scale': (-0.5,
0.5)})

    est_mu = np.mean(log_returns[i_start:i_end])
    est_std = np.std(log_returns[i_start:i_end], ddof=1)

    hyp_p = hyperbolic_par.params.p
    hyp_a = hyperbolic_par.params.a
    hyp_b = hyperbolic_par.params.b
    hyp_loc = hyperbolic_par.params.loc

```

```

hyp_scale = hyperbolic_par.params.scale

nig_p = nig_par.params.p
nig_a = nig_par.params.a
nig_b = nig_par.params.b
nig_loc = nig_par.params.loc
nig_scale = nig_par.params.scale

gh_p = gh_par.params.p
gh_a = gh_par.params.a
gh_b = gh_par.params.b
gh_loc = gh_par.params.loc
gh_scale = gh_par.params.scale

#####

stable_par = list(df_stable.iloc[i_start])
temp = []
temp.append(norm.ppf(0.1, loc=est_mu, scale=est_std))
temp.append(levy_stable.ppf(0.1, stable_par[0], stable_par[1],
scale=stable_par[2], loc=stable_par[3]) )
temp.append(genhyperbolic.ppf(0.1, gh_p, gh_a, gh_b, gh_loc, gh_scale))
temp.append(genhyperbolic.ppf(0.1, nig_p, nig_a, nig_b, nig_loc,
nig_scale))
temp.append(genhyperbolic.ppf(0.1, hyp_p, hyp_a, hyp_b, hyp_loc,
hyp_scale))
temp.append( skewed_t_var90[i_start] )
Par_VaR_90.append(temp)

temp = []
temp.append(norm.ppf(0.05, loc=est_mu, scale=est_std))
temp.append(levy_stable.ppf(0.05, stable_par[0], stable_par[1],
scale=stable_par[2], loc=stable_par[3]) )
temp.append(genhyperbolic.ppf(0.05, gh_p, gh_a, gh_b, gh_loc, gh_scale))
temp.append(genhyperbolic.ppf(0.05, nig_p, nig_a, nig_b, nig_loc,
nig_scale))
temp.append(genhyperbolic.ppf(0.05, hyp_p, hyp_a, hyp_b, hyp_loc,
hyp_scale))

```

```

temp.append( skewed_t_var95[i_start] )
Par_VaR_95.append(temp)

temp = []
temp.append(norm.ppf(0.01, loc=est_mu, scale=est_std))
temp.append(levy_stable.ppf(0.01, stable_par[0], stable_par[1],
scale=stable_par[2], loc=stable_par[3]) )
temp.append(genhyperbolic.ppf(0.01, gh_p, gh_a, gh_b, gh_loc, gh_scale))
temp.append(genhyperbolic.ppf(0.01, nig_p, nig_a, nig_b, nig_loc,
nig_scale))
temp.append(genhyperbolic.ppf(0.01, hyp_p, hyp_a, hyp_b, hyp_loc,
hyp_scale))
temp.append( skewed_t_var99[i_start] )
Par_VaR_99.append(temp)

print(i_end)

i_start += 1
i_end += 1

Par_VaR_90_final_lite_ver = Par_VaR_90
Par_VaR_95_final_lite_ver = Par_VaR_95
Par_VaR_99_final_lite_ver = Par_VaR_99

Par_VaR_90_final_full = Par_VaR_90
Par_VaR_95_final_full = Par_VaR_95
Par_VaR_99_final_full = Par_VaR_99

Par_VaR_90_final_full_df = pd.DataFrame(Par_VaR_90_final_full, columns =
['Norm', 'Stable', 'GenHyp', 'NIG', 'Hyp', 'Skewed_t'])
Par_VaR_95_final_full_df = pd.DataFrame(Par_VaR_95_final_full, columns =
['Norm', 'Stable', 'GenHyp', 'NIG', 'Hyp', 'Skewed_t'])
Par_VaR_99_final_full_df = pd.DataFrame(Par_VaR_99_final_full, columns =
['Norm', 'Stable', 'GenHyp', 'NIG', 'Hyp', 'Skewed_t'])

Par_VaR_90_final_full_df.to_excel("Par__Var_90_full.xlsx")
Par_VaR_95_final_full_df.to_excel("Par__Var_95_full.xlsx")
Par_VaR_99_final_full_df.to_excel("Par__Var_99_full.xlsx")

```

```

window_size = 300
i_start = 0
i_end = window_size

Nonpar_VaR_90 = []
Nonpar_VaR_95 = []
Nonpar_VaR_99 = []
real_values = []

while i_end < len(log_returns) - 1:
    sliced_log_returns = log_returns[i_start:i_end]
    am = arch_model(sliced_log_returns, mean='Constant', vol = 'GARCH',
p=1,o=0,q=1, dist='normal')
    res = am.fit()
    forecasts = res.forecast()

    std_rets = np.array( (sliced_log_returns - res.params['mu']) /
res.conditional_volatility )
    std_rets = std_rets[~np.isnan(std_rets)]

    mean_ = float(forecasts.mean.iloc[-1])
    var_ = float(forecasts.residual_variance.iloc[-1])

    q1 = np.quantile(std_rets, 0.1)
    q2 = np.quantile(std_rets, 0.05)
    q3 = np.quantile(std_rets, 0.01)

    temp90 = []
    temp95 = []
    temp99 = []

    temp90.append(mean_ + np.sqrt(var_)*q1)
    temp95.append(mean_ + np.sqrt(var_)*q2)
    temp99.append(mean_ + np.sqrt(var_)*q3)

```

```
#####

a2 = np.sort(sliced_log_returns)
temp90.append(np.quantile(a2, 0.1))
temp95.append(np.quantile(a2, 0.05))
temp99.append(np.quantile(a2, 0.01))

Nonpar_VaR_90.append(temp90)
Nonpar_VaR_95.append(temp95)
Nonpar_VaR_99.append(temp99)
real_values.append(log_returns[i_end])

print(i_end)

i_start += 1
i_end += 1

Nonpar_VaR_90_final_df = pd.DataFrame(Nonpar_VaR_90, columns = ['FHS',
'quantile'])
Nonpar_VaR_95_final_df = pd.DataFrame(Nonpar_VaR_95, columns = ['FHS',
'quantile'])
Nonpar_VaR_99_final_df = pd.DataFrame(Nonpar_VaR_99, columns = ['FHS',
'quantile'])

Nonpar_VaR_90_final_df.to_excel("Nonpar__Var_90.xlsx")
Nonpar_VaR_95_final_df.to_excel("Nonpar__Var_95.xlsx")
Nonpar_VaR_99_final_df.to_excel("Nonpar__Var_99.xlsx")

fig, ax = plt.subplots(1, 1, figsize=(20,10))
length__ = len(Par_VaR_90_final_full_df['Norm'])
i_points = data['Date'][300:-5]

ax.plot(i_points, Par_VaR_90_final_full_df['Norm'], color = 'red', lw=1,
label = 'Нормальний розподіл')
ax.plot(i_points, Par_VaR_90_final_full_df['Stable'], color = 'purple', lw=1,
label = 'Стійкий розподіл')
ax.plot(i_points, Par_VaR_90_final_full_df['GenHyp'], color = 'green', lw=1,
label = 'Узагальнений гіперболічний розподіл')
```

```

ax.plot(i_points, Par_VaR_90_final_full_df['NIG'], color = 'magenta', lw=1,
label = 'Нормальний обернений гаусівський розподіл')

ax.plot(i_points, Par_VaR_90_final_full_df['Hyp'], color = 'darkorange',
lw=1, label = 'Гіперболічний розподіл')

ax.plot(i_points, Par_VaR_90_final_full_df['Skewed_t'], color = 'darkblue',
lw=1, label = 'Узагальнений скошений розподіл Стьюдента')

ax.plot(i_points, Nonpar_VaR_90_final_df['FHS'][:len(i_points)], color =
'mediumvioletred', lw=1, label = 'Фільтровані історичні симуляції')

ax.plot(i_points, Nonpar_VaR_90_final_df['quantile'][:len(i_points)], color =
'peru', lw=1, label = 'Емпіричний квантиль')

ax.scatter(i_points, real_values[:length__])

ax.set_ylim([min(real_values), 0])

ax.legend(loc='lower left', frameon=False)

ax.margins(x=0)

ax.set_ylabel('VaR 90%')

plt.show()

#limit = len(Par_VaR_90_final_df['Norm'])

limit = len(Par_VaR_90_final_full_df['Norm'])

N_values = len(real_values[:limit])

print('VaR 90%')

print('{} - {}'.format('Normal', 100*( (real_values[:limit] <
Par_VaR_90_final_full_df['Norm'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Stable', 100*( (real_values[:limit] <
Par_VaR_90_final_full_df['Stable'][:limit]).sum())/N_values ) )

print('{} - {}'.format('GH', 100*( (real_values[:limit] <
Par_VaR_90_final_full_df['GenHyp'][:limit]).sum())/N_values ) )

print('{} - {}'.format('NIG', 100*( (real_values[:limit] <
Par_VaR_90_final_full_df['NIG'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Hyp', 100*( (real_values[:limit] <
Par_VaR_90_final_full_df['Hyp'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Skewed_t', 100*( (real_values[:limit] <
Par_VaR_90_final_full_df['Skewed_t'][:limit]).sum())/N_values ) )

print('{} - {}'.format('quantile', 100*( (real_values[:limit] <
Nonpar_VaR_90_final_df['quantile'][:limit]).sum())/N_values ) )

print('{} - {}'.format('FHS', 100*( (real_values[:limit] <
Nonpar_VaR_90_final_df['FHS'][:limit]).sum())/N_values ) )

print('-----')

```

```

print(' ')

print(' ')

print('VaR 95%')

print('{} - {}'.format('Normal', 100*( (real_values[:limit] <
Par_VaR_95_final_full_df['Norm'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Stable', 100*( (real_values[:limit] <
Par_VaR_95_final_full_df['Stable'][:limit]).sum())/N_values ) )

print('{} - {}'.format('GH',      100*( (real_values[:limit] <
Par_VaR_95_final_full_df['GenHyp'][:limit]).sum())/N_values ) )

print('{} - {}'.format('NIG',     100*( (real_values[:limit] <
Par_VaR_95_final_full_df['NIG'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Hyp',     100*( (real_values[:limit] <
Par_VaR_95_final_full_df['Hyp'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Skewed_t', 100*( (real_values[:limit] <
Par_VaR_95_final_full_df['Skewed_t'][:limit]).sum())/N_values ) )

print('{} - {}'.format('quantile', 100*( (real_values[:limit] <
Nonpar_VaR_95_final_df['quantile'][:limit]).sum())/N_values ) )

print('{} - {}'.format('FHS',     100*( (real_values[:limit] <
Nonpar_VaR_95_final_df['FHS'][:limit]).sum())/N_values ) )


print(' ')

print(' ')

print('VaR 99%')

print('{} - {}'.format('Normal', 100*( (real_values[:limit] <
Par_VaR_99_final_full_df['Norm'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Stable', 100*( (real_values[:limit] <
Par_VaR_99_final_full_df['Stable'][:limit]).sum())/N_values ) )

print('{} - {}'.format('GH',      100*( (real_values[:limit] <
Par_VaR_99_final_full_df['GenHyp'][:limit]).sum())/N_values ) )

print('{} - {}'.format('NIG',     100*( (real_values[:limit] <
Par_VaR_99_final_full_df['NIG'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Hyp',     100*( (real_values[:limit] <
Par_VaR_99_final_full_df['Hyp'][:limit]).sum())/N_values ) )

print('{} - {}'.format('Skewed_t', 100*( (real_values[:limit] <
Par_VaR_99_final_full_df['Skewed_t'][:limit]).sum())/N_values ) )

print('{} - {}'.format('quantile', 100*( (real_values[:limit] <
Nonpar_VaR_99_final_df['quantile'][:limit]).sum())/N_values ) )

print('{} - {}'.format('FHS',     100*( (real_values[:limit] <
Nonpar_VaR_99_final_df['FHS'][:limit]).sum())/N_values ) )

```

```

from scipy.stats import chi2

def Ind_norm(T, p):
    if T < norm.ppf(0.5*p):
        return False
    elif T > norm.ppf(1 - 0.5*p):
        return False
    else:
        return True

def Ind_chi2(T, p):
    if T < chi2.ppf(0.5*p, 1):
        return False
    elif T > chi2.ppf(1 - 0.5*p, 1):
        return False
    else:
        return True

def VaR_binomial_test(data_list, VAR_list, alpha, p):
    n = len(VAR_list)
    H = (data_list < VAR_list).sum()
    T = (H - n*alpha)/(np.sqrt(n*alpha*(1-alpha)))

    p_0 = 0.0001
    tik = 0
    while tik == 0:
        I_0 = Ind_norm(T, p_0)
        if I_0 == False:
            return p_0
        else:
            p_0 += 0.0001

def VaR_binomial_test_2(data_list, VAR_list, alpha, p):
    n = len(VAR_list)
    H = (data_list < VAR_list).sum()

```

```

    LR = -2*np.log( np.power(1 - alpha, n - H)*np.power(alpha, H)) +
    2*np.log( np.power(1 - H/n, n - H)*np.power(H/n, H) )

    p_0 = 0.0001
    tik = 0
    while tik == 0:
        I_0 = Ind_chi2(LR, p_0)
        if I_0 == False:
            return p_0
        else:
            p_0 += 0.0001

    #limit = len(real_values)
    limit = 1500

    p = 0.05
    print('VaR 90%')

    print( VaR_binomial_test(real_values[:limit],
    Par_VaR_90_final_full_df['Norm'][:limit], 0.1, p) )

    print( VaR_binomial_test(real_values[:limit],
    Par_VaR_90_final_full_df['Stable'][:limit], 0.1, p) )

    print( VaR_binomial_test(real_values[:limit],
    Par_VaR_90_final_full_df['GenHyp'][:limit], 0.1, p) )

    print( VaR_binomial_test(real_values[:limit],
    Par_VaR_90_final_full_df['NIG'][:limit], 0.1, p) )

    print( VaR_binomial_test(real_values[:limit],
    Par_VaR_90_final_full_df['Hyp'][:limit], 0.1, p) )

    print( VaR_binomial_test(real_values[:limit],
    Par_VaR_90_final_full_df['Skewed_t'][:limit], 0.1, p) )

    print('-----')

    print( VaR_binomial_test(real_values[:limit],
    Nonpar_VaR_90_final_df['quantile'][:limit], 0.1, p) )

    print( VaR_binomial_test(real_values[:limit],
    Nonpar_VaR_90_final_df['FHS'][:limit], 0.1, p) )

    print(' ')

    print('VaR 95%')

    print( VaR_binomial_test(real_values[:limit],
    Par_VaR_95_final_full_df['Norm'][:limit], 0.05, p) )

```

```

print( VaR_binomial_test(real_values[:limit],
Par_VaR_95_final_full_df['Stable'][:limit], 0.05, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_95_final_full_df['GenHyp'][:limit], 0.05, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_95_final_full_df['NIG'][:limit], 0.05, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_95_final_full_df['Hyp'][:limit], 0.05, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_95_final_full_df['Skewed_t'][:limit], 0.05, p) )

print('-----')

print( VaR_binomial_test(real_values[:limit],
Nonpar_VaR_95_final_df['quantile'][:limit], 0.05, p) )

print( VaR_binomial_test(real_values[:limit],
Nonpar_VaR_95_final_df['FHS'][:limit], 0.05, p) )

print(' ')

print('VaR 99%')

print( VaR_binomial_test(real_values[:limit],
Par_VaR_99_final_full_df['Norm'][:limit], 0.01, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_99_final_full_df['Stable'][:limit], 0.01, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_99_final_full_df['GenHyp'][:limit], 0.01, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_99_final_full_df['NIG'][:limit], 0.01, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_99_final_full_df['Hyp'][:limit], 0.01, p) )

print( VaR_binomial_test(real_values[:limit],
Par_VaR_99_final_full_df['Skewed_t'][:limit], 0.01, p) )

print('-----')

print( VaR_binomial_test(real_values[:limit],
Nonpar_VaR_99_final_df['quantile'][:limit], 0.01, p) )

print( VaR_binomial_test(real_values[:limit],
Nonpar_VaR_99_final_df['FHS'][:limit], 0.01, p) )

def VaR_christoffersen_test(data_list, VAR_list):
    list_zero_one = data_list < VAR_list
    N_00 = 0
    N_01 = 0
    N_10 = 0
    N_11 = 0

```

```

i = 0
N = len(list_zero_one)
while i + 1 != N:
    t1 = list_zero_one[i]
    t2 = list_zero_one[i+1]

    if t1 == 0 and t2 == 0:
        N_00 += 1
    elif t1 == 0 and t2 == 1:
        N_01 += 1
    elif t1 == 1 and t2 == 0:
        N_10 += 1
    elif t1 == 1 and t2 == 1:
        N_11 += 1
    i += 1

pi_0 = N_01/(N_00 + N_01)
pi_1 = N_11/(N_10 + N_11)
pi_ = pi_0 + pi_1

LR = -2*np.log( np.power(1 - pi_, N_00 + N_01)*np.power(pi_, N_01 + N_11)
) + 2*np.log( np.power(1 - pi_0, N_00)*np.power(pi_0, N_01)*np.power(1 -
pi_1, N_10)*np.power(pi_1, N_11) )

p_0 = 0.0001
tik = 0
while tik == 0:
    I_0 = Ind_chi2(LR, p_0)
    if I_0 == False:
        return p_0
    else:
        p_0 += 0.0001

limit = 700

p = 0.1
print('VaR 90%')

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_90_final_full_df['Norm'][:limit]) )

```

```

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_90_final_full_df['Stable'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_90_final_full_df['GenHyp'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_90_final_full_df['NIG'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_90_final_full_df['Hyp'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_90_final_full_df['Skewed_t'][:limit]) )

print('-----')

print( VaR_christoffersen_test(real_values[:limit],
Nonpar_VaR_90_final_df['quantile'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Nonpar_VaR_90_final_df['FHS'][:limit]) )

print(' ')

print('VaR 95%')

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_95_final_full_df['Norm'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_95_final_full_df['Stable'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_95_final_full_df['GenHyp'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_95_final_full_df['NIG'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_95_final_full_df['Hyp'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_95_final_full_df['Skewed_t'][:limit]) )

print('-----')

print( VaR_christoffersen_test(real_values[:limit],
Nonpar_VaR_95_final_df['quantile'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Nonpar_VaR_95_final_df['FHS'][:limit]) )

print(' ')

print('VaR 99%')

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_99_final_full_df['Norm'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_99_final_full_df['Stable'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_99_final_full_df['GenHyp'][:limit]) )

```

```

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_99_final_full_df['NIG'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_99_final_full_df['Hyp'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Par_VaR_99_final_full_df['Skewed_t'][:limit]) )

print('-----')

print( VaR_christoffersen_test(real_values[:limit],
Nonpar_VaR_99_final_df['quantile'][:limit]) )

print( VaR_christoffersen_test(real_values[:limit],
Nonpar_VaR_99_final_df['FHS'][:limit]) )

```

ДОДАТОК Д Презентація

1

Моделювання вартості акцій та міжнародних індексів розподілами з важкими хвостами



Виконав: студент 4 курсу КА-91
Грішин Костянтин Дмитрович

Керівник: доц., канд. фіз.-мат. наук
Канівська Ірина Юріївна

2

Актуальність дослідження

- Інвестування нерозривно пов'язано з аналізом і мінімізацією ризику, вимагає знань сучасних математичних моделей поведінки акцій та міжнародних індексів
- З другої половини минулого сторіччя впроваджуються стохастичні моделі оцінки дохідності й ризику, однак питання, які розподіли використовувати, залишається відкритим
- Розподіли з важкими хвостами не нехтують малоймовірними катастрофічними подіями
- Недооцінка ризиків у моделях може мати серйозні наслідки (банкрутство Lehman Brothers та глобальна криза 2008).

3

- **Об'єкт дослідження:** моделі вартості акцій та міжнародних фондових індексів
- **Предмет дослідження:** розподіли з важкими хвостами в моделях вартості акцій та міжнародних фондових індексів
- **Мета дослідження:** проаналізувати переваги застосування розподілів із важкими хвостами в моделях цін акцій та міжнародних фондових індексів

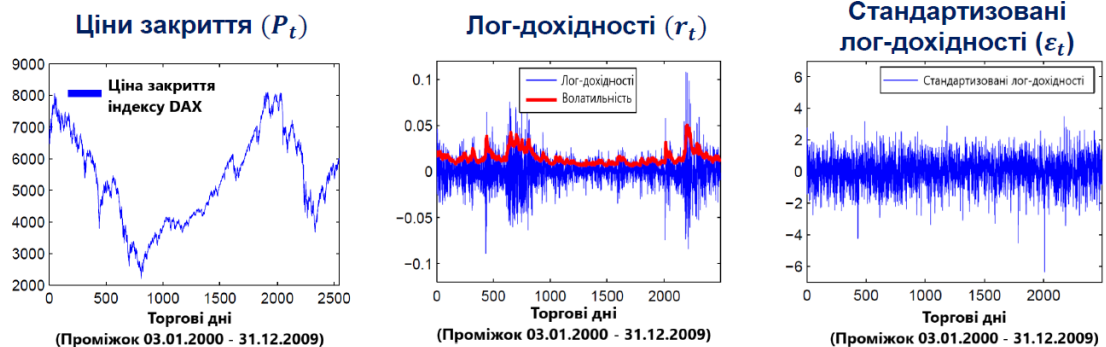
Постановка задачі

4

- Дослідити існуючі моделі поведінки цін акцій та індексів, в яких застосовуються розподіли з важкими хвостами
- Формально описати розподіли з важкими хвостами, їх властивості, методи оцінки параметрів
- Обрати міжнародний індекс та змоделювати поведінку його вартості за допомогою програмного продукту
- Проаналізувати отримані результати, порівняти їх з результатами інших дослідників, зробити висновки

Розподіл лог-дохідностей акцій, індексів

5



Лог-дохідності

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

Оцінка волатильності

$$r_t = \sigma_t \varepsilon_t$$

σ_t – волатильність

Стандартизація лог-дохідностей

$$\varepsilon_t = \frac{r_t}{\sigma_t}$$

σ_t оцінюється зазвичай моделлю GARCH

DAX – міжнародний фондовий індекс, показник стану економіки Німеччини (середньозважене значення цін акцій 30 найбільших німецьких компаній)

Основні етапи розвитку застосування розподілів

6



Розподіли із важкими хвостами

7

Означення:

$$\forall \mu > 0, \lim_{x \rightarrow +\infty} \sup_x (1 - F(x)) e^{\mu x} = \infty$$

де $F(x)$ – функція розподілу



Існування генератриси $E(e^{t\xi})$



Скінченність $E\xi^k, k \in \mathbb{N}$

Переваги застосування розподілів із важкими хвостами:

- не нехтують ймовірністю великих відхилень від середніх значень
- узгоджуються з уявленням про ціну акції, як суми багатьох незалежних випадкових факторів

Аналіз публікацій

8

Результати огляду дозволяють зробити наступний висновок:

- Нормальний розподіл (N) не є адекватною моделлю
- Виділено п'ять альтернативних розподілів:
 - стійкий (S)
 - скошений розподіл Стюдента (SGT)
 - нормальний обернений гаусівський (NIG)
 - гіперболічний (Нур)
 - узагальнений гіперболічний (GH)
- Залежно від компанії, країни різні розподіли можуть виявитись кращою моделлю

Джерело			Значення статистики Колмогорова-Смірнова, яка використовувалась в критерії узгодженості					
Джерело	Акции/індекси	Час	N	S	SGT	NIG	Нур	GH
Misiorzek et al. 2011	DAX	2000-2009	1.858	1.389		0.796	0.770	
	WIG20	2000-2009	1.584	0.881		0.682	0.681	
Basogmez et al. 2017	XTRZM	2007-2017	Відхилено	Краще				
	Nikkei225	1978-2017	Відхилено	Краще				
Fukunaga et al. 2018	S&P 500	1975-2017	Відхилено	Краще				
	Dow 30	1986-2017	Відхилено	Краще				
	SSEC	1998-2017	Відхилено	Краще				
Zi-Yi G. 2017	S&P 500	1957-2017	0.018		0.006			0.007
	DIA	1987-1994	4.6383	0.5583				
Borak et al. 2005	Boeing	1997-2003	2.1361	0.4522				
	ISPIX	1998-2005	Відхилено	0.02584				
Belov et al. 2006	AMEX	1983-2005	Відхилено	0.01824				
	FCHI	1990-2005	Відхилено	0.01198				
	NASDAQ	1999-2005	Відхилено	0.02928				
	S&P 500	1998-2008	0.0541					0.0183
Necula C. 2009	Nikkei225	1998-2008	0.0417					0.0191
	HangSeng	1998-2008	0.0691					0.0254
	DAX	1998-2008	0.0617					0.0335
	AGL	2006-2011			0.0467	0.0414	0.024	0.0214
Virginie et al. 2014	AMS	2006-2011			0.03	0.0267	0.032	0.036
	ANG	2006-2011			0.0274	0.018	0.4746	0.028
	Vale5	1994-2001	0.0751			0.0121	0.0099	0.0108
Fajardo et al. 2002	Ibvsp	1994-2001	0.0661			0.0166	0.0208	0.0093
	Petr4	1994-2001	0.0640			0.0133	0.0142	0.0126
	Elet6	1994-2001	0.0598			0.0123	0.0150	0.0103
	DIA	1987-1992	3996.41*			4092.31*		
Amado P. 1994	Nikkei225	1987-1992	3323.08*			3468.71*		
	FT30	1987-1992	3564.29*			3727.05*		

* – використовувалась статистика likelihood ratio

Складено автором на основі публікацій

Стійкий розподіл $S(\alpha, \beta, c, \delta)$

9

- Параметри

$\alpha \in (0, 2]$ — швидкість спадання щільності

$\beta \in [-1, 1]$ — асиметрію

$c \in [0, +\infty)$ — міру розсіювання навколо середнього

$\delta \in \mathbb{R}$ — положення середнього

- Характеристична функція

$$\ln \chi(t) = \begin{cases} i\delta t - c^\alpha |t|^\alpha \left(1 - i\beta \frac{t}{|t|} \operatorname{tg} \left(\frac{\pi}{2} \alpha \right) \right), & \alpha \neq 1 \\ i\delta t - c^\alpha |t|^\alpha \left(1 + i \frac{2\beta}{\pi} \frac{t}{|t|} \ln |t| \right), & \alpha = 1 \end{cases}$$

- Аналітичного запису щільності нема

- Властивості:

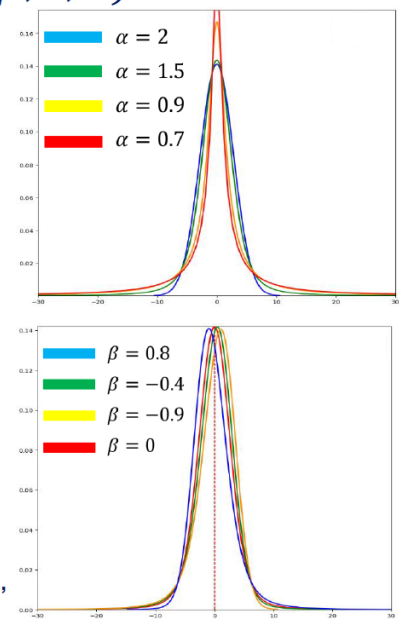
- Розподіл зберігає вигляд після афінного перетворення

- $\xi_k \sim S(\alpha, \beta, c, \delta)$ н. о. р. $\Rightarrow \eta = \sum_{k=1}^n \xi_k \sim S\left(\alpha, \beta, c \cdot n^{\frac{1}{\alpha}}, n\delta\right)$

- $\xi_i \sim S(\alpha, \beta, c, \delta)$ н. о. р. $\Rightarrow \xi_1 + \xi_2 \sim S\left(\alpha, \frac{\beta_1 c_1^\alpha + \beta_2 c_2^\alpha}{c_1^\alpha + c_2^\alpha}, (c_1^\alpha + c_2^\alpha)^{\frac{1}{\alpha}}, \delta_1 + \delta_2\right)$

- $E\xi^p < \infty \Leftrightarrow p < \alpha$. При $\alpha > 1$, $E\xi = \delta$

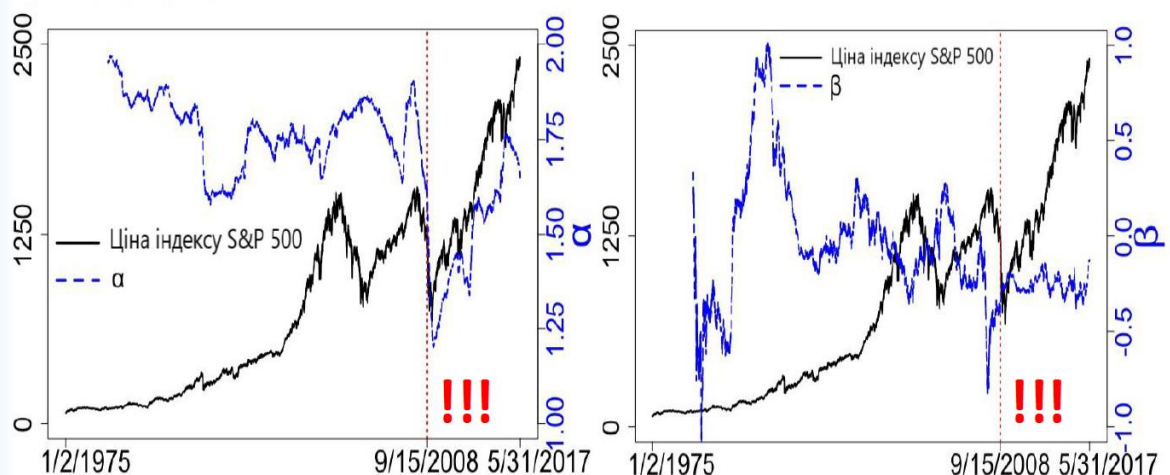
- Проаналізовано алгоритм генерації випадкових величин, чотири оцінки параметрів



Порівняння індексу S&P 500 та параметрів α та β

10

Зменшення параметрів α, β стійкого розподілу $S(\alpha, \beta, c, \delta)$ можуть бути передвісниками кризи, краху біржі.



Джерело: Fukunaga T., Umeno K. Universal Levy's stable law of stock market and its characterization, February, 2018.

Узагальнений скошений розподіл Стюдента

$$f(x) = \frac{p}{2v\sigma q^{\frac{1}{p}} B\left(\frac{1}{p}, q\right) \left(\frac{|x-\mu+m|^p}{q(v\sigma)^p(\lambda \cdot \text{sgn}(x-\mu+m)+1)^p+1}\right)^{\frac{1}{p+q}}}$$

$$\xi \sim SGT(\mu, \sigma, \lambda, p, q)$$

де m, v – константи

p, q – порядок скінчених моментів ($p, q > 0$)

σ – міра розсіювання навколо середнього ($\sigma > 0$)

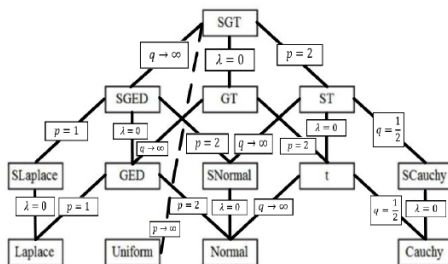
λ – асиметрія ($-1 < \lambda < 1$)

5 параметрів

4 параметри

3 параметри

2 параметри

**Узагальнений гіперболічний розподіл** 11

$$f(x) = s(\delta^2 + (x-\mu)^2)^{\frac{1}{2}(\lambda-\frac{1}{2})} K_{\lambda-\frac{1}{2}}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) e^{\beta(x-\mu)}$$

де $K_{\lambda}(x) = \frac{1}{2} \int_0^{+\infty} y^{\lambda-1} e^{-\frac{x}{2}(y+y^{-1})} dy$

s – нормуюча константа

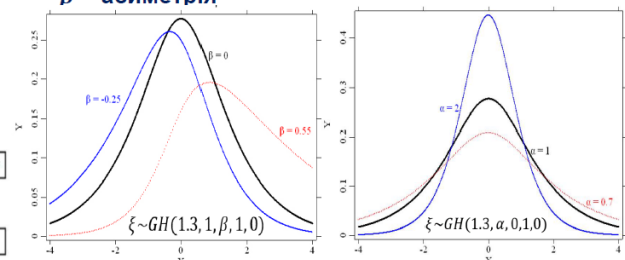
$$\xi \sim GH(\lambda, \alpha, \beta, \delta, \mu)$$

λ – тип розподілу

α, δ – швидкість спадання щільності

μ – математичне сподівання

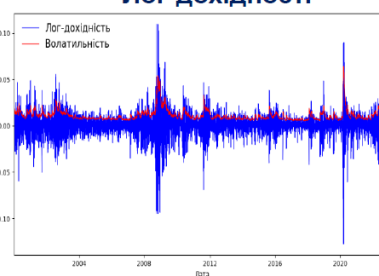
β – асиметрія

**Результати власного експерименту з індексом США Standard & Poor 500** 12

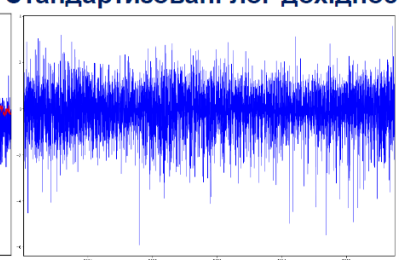
Ціни індексу



Лог-дохідності



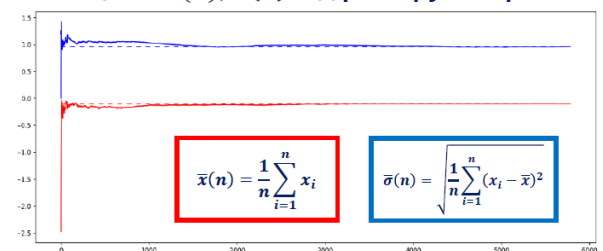
Стандартизовані лог-дохідності



- Дані переведено в лог-дохідності
- Модель GARCH

$$r_t = \mu + h_t e_t, \quad e_t \sim N(0, 1)$$

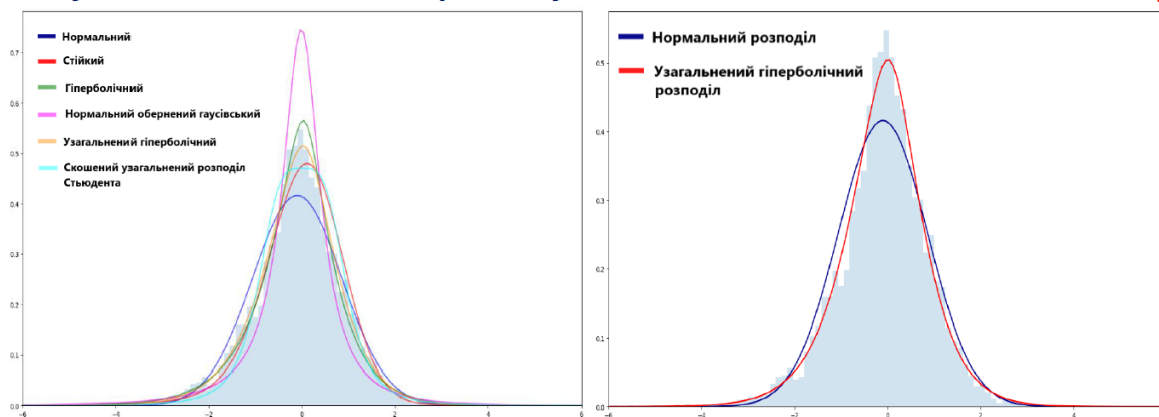
$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}$$

Оцінки $\bar{x}(n)$, $\bar{\sigma}(n)$ від розміру вибірки

$$\bar{x}(n) = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\bar{\sigma}(n) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

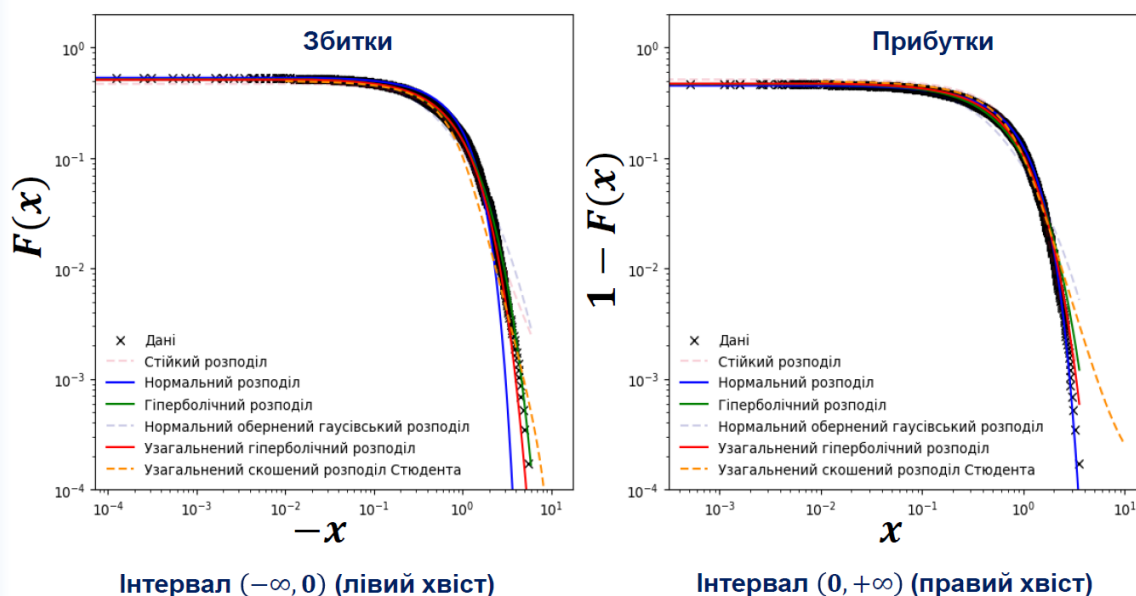
Результати власного експерименту з індексом США Standard & Poor 500 13



Порівняння гістограми стандартизованих лог-дохідностей S&P 500 з щільностями наближених розподілів

- Нормальний розподіл не є адекватною моделлю стандартизованих лог-дохідностей
- Узагальнений гіперболічний розподіл дав найкраще наближення
- На рівні значущості $p > 0.1$ дані не суперечать гіпотезі про те, що стандартизовані лог-дохідності індексу США S&P 500 розподілені за узагальненим гіперболічним розподілом

Результати власного експерименту з індексом США Standard & Poor 500 14



Інтервал $(-\infty, 0)$ (лівий хвіст)

Інтервал $(0, +\infty)$ (правий хвіст)

Value-at-Risk (VaR)

15

Означення:

VaR для рівня достовірності $c \in [0, 1]$ на проміжку $T > 0$:

$$VaR = \sup_{X \in R} \{P(L_T > X) \leq 1 - c\}$$

де L_T – функція втрат вартості за час T

Наближення VaR на день вперед:

$$P(\ln(P_t) - \ln(P_{t-1}) \leq -VaR_{log}) = 1 - c$$

$$VaR = S_t \cdot VaR_{log}$$

де S_t – вартість портфелю в момент t

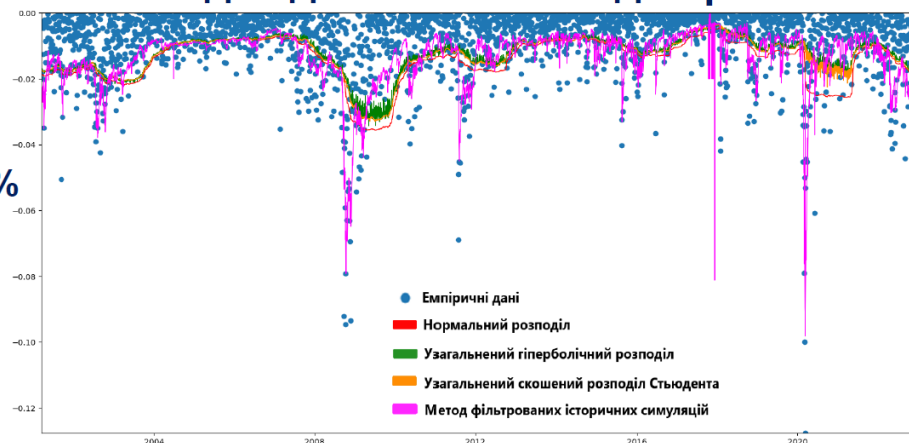
Методи оцінки VaR

- Параметричний (квантиль розподілу)
- Емпіричний квантиль
- Метод «Фільтрованих історичних симуляцій»

Порівняння лог-дохідностей та VaR для різних методів

16

VaR 90%



- VaR рахувався через ковзне вікно. Брались лог-дохідності з 2000 по 2007 рік, не стандартизовані
- Статистичні критерії: біноміальний, критерій Крістоферсена
- Одразу для VaR 90%, 95%, 99% на рівні $p > 0.1$ біноміальний критерій прийняв гіпотезу про узгодженість оцінки із даними лише для узагальненого скошеного розподілу Стюдента
- Можна покласти, що принаймні великі збитки (за порогом VaR 99%) незалежні на рівні $p > 0.1$

Висновки

17

- Створено систему моделювання вартостей акцій та міжнародних індексів розподілами з важкими хвостами, яка здатна:
 - зводити вартості до реалізації незалежних однаково розподілених величин;
 - оцінювати параметри п'яти розподілів, альтернативних нормальному;
 - візуалізувати результати;
 - рахувати різними методами Value-at-Risk
- Систему успішно застосовано для моделювання вартості індексу США Standard & Poor 500 з 2000 по 2022 р., одного з головних показників економіки країни.
- Показано неадекватність нормального розподілу для моделювання вартості індексу США Standard & Poor 500
- Модель із узагальненим гіперболічним розподілом дала найточніше наближення, гіпотеза була прийнята статистичним критерієм

Подальші дослідження

18

- Дослідити моделювання вартості портфельів, які включають акції, валюту, облігації, нерухомість, об'єкти мистецтва...
- Підвищити точність моделі шляхом удосконалення попередньої обробки даних, розширення множини розподілів
- Застосувати інші моделі оцінки ризиків інвестиційного портфеля

Дякую за увагу!

