

Міністерство освіти і науки України  
ДВНЗ «Криворізький державний педагогічний університет»  
Черкаський національний університет  
імені Богдана Хмельницького

**В. М. СОЛОВЙОВ, О. А. СЕРДЮК, Г. Б. ДАНИЛЬЧУК**

# **МОДЕЛЮВАННЯ СКЛАДНИХ СИСТЕМ**

Навчально-методичний посібник  
для самостійного вивчення дисципліни

Черкаси – 2016

**УДК 330.4(075.8)**  
**ББК 65в631я-1**  
**С 60**

*Рецензенти:*

**С. К. Рамазанов**, д.т.н., д.е.н., проф. (ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана»)

**Ю. В. Триус**, д.пед.н., проф. (Черкаський державний технологічний університет)

**Соловйов В.М., Сердюк О.А., Данильчук Г.Б.**

С60      **Моделювання складних систем : Навчально-методичний посібник для самостійного вивчення дисципліни. – Черкаси : Видавець О. Ю. Вовчок, 2016. – 204 с.**

У навчальному посібнику висвітлюються основні теоретичні та інструментальні аспекти моделювання складних систем для студентів спеціальностей 111 «Математика», 113 «Прикладна математика», 014 «Середня освіта», 051 «Економіка», спеціалізації «Економічна кібернетика». Пропонуються сучасні математичні моделі, в яких застосовуються такі методи як вейвлет-аналіз, теорія випадкових матриць, мультифрактальний аналіз, ентропійний та мережний аналіз складності та ін.

Посібником зможуть скористатися як студенти вищих навчальних закладів, так і наукові працівники й особи, які цікавляться методами моделювання складних систем.

УДК 330.4(075.8)  
ББК 65в631я-1

Затверджено Вченою радою ННІ економіки і права Черкаського національного університету імені Богдана Хмельницького (протокол № 9 від 18 травня 2016 р.)

Наукове видання  
**Моделювання складних систем**

Навчально-методичний посібник для самостійного вивчення дисципліни  
Автори: Соловйов В.М., Сердюк О.А., Данильчук Г.Б.

Підп. до друку 01.10.2016. Формат 60x84/16. Папір офсетний. Гарнітура Times.  
Умовн. друк. арк. 11,86. Вид. № 5-16. Наклад 300 прим.  
Видавець Вовчок О.Ю. Свідоцтво про внесення до Державного реєстру видавців.  
Серія ДК № 650 від 30.10.2001 р.  
Україна, 18001, м. Черкаси, вул. Гоголя, 509, к. 21.  
Тел.: 0472 503586. E-mail: book\_brama@ukr.net

ISBN 978-966-8645-88-4

© В.М.Соловйов, О.А.Сердюк,  
Г.Б.Данильчук, 2016  
© ЧНУ ім. Б.Хмельницького, 2016

## Передмова

Задекларований перехід економіки України на інноваційний шлях розвитку в умовах обмеженості ресурсів, у першу чергу фінансових, вимагає зосередження їх на найбільш перспективних напрямках, де імовірність досягнення конкурентного успіху є найбільшою. Як свідчить практика, такого роду оцінки доцільно виконувати із застосуванням економіко-математичних моделей. Множинність шляхів вибору моделей і методів актуалізує необхідність розуміння основних тенденцій розвитку наукового знання і ключових досягнень.

Говорячи мовою синергетики, необхідно виділити такі параметри порядку – провідні змінні, які з плином часу починають визначати динаміку і розвиток складної системи, і підпорядковувати собі її інші параметри, які призведуть до ефективного спрощення складного об'єкта.

Все більш очевидною є революція, що почалася в природничих та гуманітарних науках і пов'язана з вивченням феномена самоорганізації і дослідженням мережних структур. Мережна парадигма складності обумовлена великим значенням, яке придбали такі об'єкти, і тим, що на початку ХХІ століття очевидною стала різьча аналогія в топології мережних структур, що виникають завдяки активному використанню інформаційно-телекомунікаційних, гуманітарних, управлінських, військових технологій.

Мережі стали одним з двигунів економіки. У своїй історії людство долало різні бар'єри, створюючи нові матеріали, технології, об'єкти. Однак «мережна економіка» зіткнулася з обмеженнями самої людини – так званім «когнітивним бар'єром». Як показали психологи, людина в змозі активно, творчо взаємодіяти з 5-7 людьми (з рештою опосередковано або стандартно, незалежно від того, скільки у нього друзів у соціальній мережі). Вона може одночасно стежити за 5-7 змінними (незалежно від того, наскільки великий обсяг інформації їй доступний). Приймаючи рішення, вона може зважити 5-7 факторів (скільки б даних у неї не було).

Зліт нової економіки в США в 1990-х роках, пов'язаний багато в чому з інтернет-компаніями, породив ілюзії, що капіталізація мережних структур пропорційна числу зв'язків між вузлами, тобто квадрату числа вузлів  $N^2$ . Однак, коли на початку 2000-х років міхур «нової економіки» лопнув (криза «доткомів - .com»), то виявилось, що реальна капіталізація мала зовсім іншу залежність від числа об'єктів, пов'язаних мережею –  $N \ln N$ . Інакше кажучи, не «всі зв'язуються з усіма», а майже всі взаємодіють з декількома дуже великими вузлами-хабами, які вже тісно пов'язуються між собою. Подібним же чином виявляється влаштована інфраструктура більшості складних систем, незалежно від їх природи.

Черговим викликом для нас стала нова індустриальна революція, актуалізована давоським форумом 2016 року. Згідно з опитуванням 800 лідерів технологічних компаній, проведеним спеціально для форуму в Давосі, ключовими драйверами змін стануть хмарні технології, розвиток способів збору і аналізу Big Data, краудсорсінг, шерінгова економіка і біотехнології.

Очевидно, що революційні вимоги Індустрії 4.0 потребують нових парадигм моделювання соціально-економічних систем. На наш погляд, такою парадигмою може стати мережна парадигма складних систем.

Мережні технології змінили обчислювальну математику, системний аналіз, інформаційні технології. В останні роки було реалізовано кілька грандіозних мережних проектів, в яких поставлена задача вирішувалася завдяки спільним діям сотень тисяч або навіть мільйонів комп'ютерів. Це і криптографічні проблеми, і пошук ліків проти раку, заснований на математичному моделюванні взаємодії різних речовин з клітинами. Це розподілений аналіз даних космічних експериментів та обробка результатів, отриманих на Великому адронному колайдері.

В зв'язку з новою мережною парадигмою складності перед фахівцями з моделювання виникають принципово нові, актуальні задачі. Ось, на наш погляд, тільки деякі з них:

- дослідження надійності, робастності кіберфізичних мереж відносно випадкових помилок та направлених атак;
- аналіз когнітивних можливостей складних систем;
- моделювання мультиплексних мереж;
- вплив нанотехнологій на формування наноекономіки та ін.

Сьогодні вже зрозуміло, що відповідних інновацій вимагає і система освіти. На початку комп'ютерної ери основну цінність і проблеми становили власне комп'ютери (hardware) і акцент робився на підготовку фахівців з обчислювальної техніки. Потім величезне значення набуло програмне забезпечення (software) і була розпочата підготовка дослідників у цій галузі (computer science) та інженерів-програмістів (computer engineering). В даний час на перший план виходять фахівці з мережних технологій (NetWare). Саме таких фахівців треба починати готувати у провідних національних університетах. За оцінками експертів з Індустрії 4.0 основними трендами освітньої сфери на двадцятирічному горизонті стануть: поширення цінностей мережної культури, прагматизація освіти, автоматизація рутинних інтелектуальних операцій, розвиток індустрії поліпшення когнітивних здібностей, боротьба за таланти, зростання значущості глобальних людських цінностей, уваги до природи і дбайливого поводження з ресурсами.

**Мета курсу** моделювання складних систем – формування системи теоретичних знань та практичних навичок щодо моделювання структурних і динамічних властивостей систем різної природи як засобу дослідження та управління складними явищами на макро-, мезо- й мікрорівнях. Протягом останніх десяти-п'ятнадцяти років відбулися відчутні зміни в розумінні фундаментальних закономірностей складних систем. Виявилось, що такі складні системи мають універсальні емерджентні властивості, які не знаходять адекватного розуміння у рамках традиційних парадигм. Тому для аналізу все активніше використовуються сучасні методи та моделі фундаментальних наук, які у поєднанні з новітніми досягненнями в галузі інформаційних технологій та досить ємними базами даних (мільйони записів навіть в базах некомерційного призначення) забезпечили значний прогрес у розумінні та квантифікації природи цих систем. Сюди відносяться методи фрактального і мультифрактального аналізу, дослідження рекурентних властивостей динамічних систем, нелінійної динаміки, теорії хаосу і біфуркацій тощо. З'явилися нові «кількісні» напрямки економіки: математична економіка, фізична економіка, еконофізика та ін. Деякі з таких моделей, не знайшовши поки що відображення у навчальній літературі, включені до цієї роботи.

**Наукову основу курсу** складають теоретичні моделі, математичний апарат, сучасні концепції та парадигми, які визначають підходи до вивчення характеристик складних систем.

Курс базується на знаннях, одержаних при вивченні дисциплін математичного циклу, основ теорії систем та системного аналізу, моделювання, фінансового аналізу, макро- і мікроекономіки.

**Завдання курсу** – оволодіння теоретичними знаннями та інструментарієм моделювання складних систем, вивчення підходів до дослідження й аналізу, методів прогнозування їхнього розвитку, управління розвитком та функціонуванням складних систем у різних умовах функціонування.

**Предметом курсу** є математичні моделі і методи дослідження складних систем різної природи.

У результаті вивчення дисципліни студент повинен:

- знати структурні та динамічні характеристики складної системи; моделі прогнозування характеристик системи; основні методи оцінки якості функціонування; методи оцінки структурних змін; методи дослідження та моделювання складних природних та штучних систем;
- вміти здійснювати класифікацію характеристик складної системи, проводити порівняльний аналіз методів прогнозування; оцінити якість функціонування ієрархічної системи; визначити катастрофічні зміни в системі, які описуються рівняннями динаміки,



визначити джерела структурних катастроф;

- дослідити та проаналізувати комплекс моделей складної системи;
- бути ознайомленим з сучасними напрямками розвитку сучасних теорій та парадигм, які використовуються для дослідження якісних характеристик динамічних систем.

Система контролю якості навчання студентів містить такі заходи: 1) модульний контроль; 2) проведення контрольних робіт; 3) контроль теоретичних знань у ході практичних занять; 4) виконання індивідуальних завдань на лабораторних заняттях.

Організація самостійної роботи студентів передбачає підготовку до семінарських занять, проведення індивідуальних консультацій, виконання курсової роботи.

Відсутність альтернативних підручників і методичних посібників з курсу моделювання складних систем спонукала авторів розробити методичні вказівки до виконання практичних робіт з даного курсу. Вони спираються на деяке авторське бачення, певні висновки мають дискусійний характер. Автори будуть вдячні за бачення, спрямоване на покращення змісту та методики викладання дисципліни.

## ЛАБОРАТОРНА РОБОТА № 1

**Тема.** Аналіз флуктуацій часового ряду.

**Мета.** Навчитися використовувати аналіз флуктуацій та його похідні для отримання нелінійних характеристик часового ряду.

1. Аналіз динаміки прибутків, модулів прибутків та волатильностей.
2. Аналіз кореляцій.
3. Розподіл флуктуацій.

### Теоретичні відомості

#### Аналіз динаміки прибутків, модулів прибутків та волатильностей

Останнім часом вчені все більше цікавляться економічними часовими рядами, і відбувається це за кількох причин, зокрема: (1) економічні часові ряди, такі як індекси акцій, курсів валют, залежать від розвитку великої кількості взаємодіючих систем, і є прикладами складних систем, що широко вивчаються у науці; (2) з'явилась велика кількість доступних баз з даними про економічні системи, що містять інформацію з різними часовими шкалами (починаючи з 1 хвилини і закінчуючи 1 роком). Внаслідок цього вже на даний час існує також велика кількість розроблених методів (зокрема, у статистичній фізиці), спрямованих на отримання характеристик цін акцій чи курсів валют, що еволюціонують у часі.

Дослідження, проведені над часовими рядами, показують, що стохастичний процес, який лежить у основі зміни ціни, характеризується кількома ознаками. Розподіл зміни ціни має виділений хвіст порівняно із Гаусовим розподілом. Функція автокореляції зміни ціни спадає експоненційно з певним характерним часом. Однак, виявляється, що амплітуда зміни ціни, виміряна за абсолютними значеннями чи квадратами цін, показує степеневі кореляції з довго часовою персистентністю аж до кількох місяців, або навіть років. Такі довгочасові залежності краще моделюються з використанням «додаткового процесу», що в економічній літературі часто називається *волатильністю*. Волатильність змін ціни акції є мірою того, як сильно ринок схильний до флуктуацій, тобто відхилень ціни від попередніх значень.

Першим кроком при проведенні аналізу є побудова оцінювача волатильності. Ми будемо отримувати волатильність як локальне середнє модуля зміни ціни.

Розуміння статистичних властивостей волатильності має також важливе практичне застосування. Волатильність є інтересом торговців, оскільки визначає ризик і є ключовим входом практично до всіх моделей цін опціонів (вторинного цінного паперу), включаючи і класичну модель Блека-Шоулза. Без задовільних методів оцінювання волатильності трейдерам було б надзвичайно важко визначати ситуації, в яких опціони попадають в недооцінку чи переоцінку.

#### Визначення волатильності

Термін волатильність представляє узагальнену міру величини ринкових флуктуацій (відхилень). В літературі існує досить багато визначень волатильності, проте ми будемо використовувати наступне: *волатильність є локальним середнім модуля зміни ціни на відповідному часовому інтервалі  $T$ , що є рухомих параметром нашої оцінки*.

Для індексу  $X(t)$  визначимо зміну ціни  $G(t)$  як зміну логарифмів індексів,

$$G(t) = \ln X(t + \Delta t) - \ln X(t) \cong \frac{X(t + \Delta t) - X(t)}{X(t)}, \quad (1)$$

де  $\Delta t$  є часовим інтервалом затримки. Величину (1) називають прибутковістю (return). Якщо

використовувати границі, то малі зміни  $X(t)$  приблизно відповідають змінам, визначеним другою рівністю. Ми лише підраховуємо час роботи ринку, викидаємо ночі, вихідні та свята із набору даних, тобто, вважається, що ринок працює без перерв.

Модуль  $G(t)$  описує амплітуду флуктуацій. В порівнянні із значеннями  $G(t)$  їх модуль не показує глобальних трендів, але великі значення  $|G(t)|$  відповідають крахам та великим миттєвим змінам на ринках.

Визначимо волатильність як середнє від  $|G(t)|$  для часових вікон  $T = n \cdot \Delta t$ , тобто

$$V_T(t) = \frac{1}{n} \sum_{t'=t}^{t+n-1} |G(t')|, \quad (2)$$

де  $n$  є цілим числом. Таке визначення може бути ще узагальнене заміною  $|G(t)|$  на  $|G(t)|^\gamma$ , де  $\gamma > 1$  дає більш виражені великі значення  $|G(t)|$ , в той час як  $0 < \gamma < 1$  виділяє малі значення  $|G(t)|$ .

У цьому визначенні волатильності використовується два параметри,  $\Delta t$  та  $n$ . Параметр  $\Delta t$  є шаблонним (чи модельним) часовим інтервалом для даних, а параметр  $n$  є кроком переміщення часового вікна. Зауважимо, що вказане визначення волатильності має внутрішню помилку, а саме: вибір більшого часового інтервалу  $T$  веде до збільшення точності визначення волатильності. Однак, велике значення  $T$  також включає погане розбиття часу на інтервали, що веде, у свою чергу, до врахування не всієї прихованої у ряді інформації.

При порівнянні між собою кількох акцій часто використовують нормалізовану волатильність, що визначається для кожної акції наступним чином:

$$v_T^i = v_T^i(t) = \frac{V_T^i}{\sqrt{\langle (V_T^i)^2 \rangle - \langle V_T^i \rangle^2}}, \quad (3)$$

де  $\langle \dots \rangle$  означає середній час, отриманий для вікон, що не перекриваються, для різних часових масштабів  $T$ .

### Визначення кореляцій

Для визначення кореляцій часового ряду використовується функція автокореляції. Саме поняття *автокореляції* означає кореляцію часового ряду самого з собою (між попередніми та наступними значеннями). Автокореляцію іноді називають *послідовною кореляцією*, що означає кореляцію між членами ряду чисел, розташованих у певному порядку. Також синонімами цього терміну є *лагова кореляція* та *персистентність*. Наприклад, часто зустрічається автокореляція геофізичних процесів, що означає перенесення залишкового процесу на наступні часові проміжки.

Позитивно автокорельований часовий ряд часто називають персистентним, що значить існування тенденції слідування великих значень за великими та малих за малими, інакше позитивно корельований часовий ряд можна назвати інертним.

Візьмемо  $N$  пар спостережень двох змінних  $x$  та  $y$ . Кореляційний коефіцієнт між  $x$  та  $y$  визначається як

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}, \quad (3)$$

де сума знаходиться по всім  $N$  спостереженням.

Таким же чином можна визначити й автокореляцію, або ж кореляцію всередині одного часового ряду. Для автокореляції першого порядку береться лаг (часова затримка), рівний 1 часовій одиниці. Таким чином, автокореляція першого порядку використовує перші  $N-1$  спостережень  $x_t$ ,  $t = 1, \dots, N-1$ , та наступні  $N-1$  спостережень  $x_t$ ,  $t = 2, \dots, N$ . Кореляція між  $x_t$  та  $x_{t+1}$  визначається наступним чином:

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2} \sqrt{\sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2}}, \quad (4)$$

де  $\bar{x}_{(1)}$  є середнє для першого проміжку спостережень (перші  $N-1$  значень), а  $\bar{x}_{(2)}$  є середнім для другого проміжку (другі  $N-1$  значень). Коефіцієнт, що визначається у (4), і називається коефіцієнтом автокореляції.

Для суттєво великої кількості спостережень (великого  $N$ ) різниця між середнім на першому інтервалі  $\bar{x}_{(1)}$  та середнім на другому інтервалі  $\bar{x}_{(2)}$  є несуттєвою і може бути проігнорована, таким чином  $r_1$  буде приблизно рівним

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}, \quad (5)$$

де  $\bar{x} = \frac{1}{N} \sum_{t=1}^N x_t$  є середнім для досліджуваного періоду.

Рівняння (5) може бути узагальнене для отримання кореляції між спостереженнями, розділеними  $k$  часовими інтервалами:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}. \quad (5)$$

Значення  $r_k$  називається коефіцієнтом автокореляції з лагом  $k$ . Графік функції автокореляції як залежності  $r_k$  від  $k$  також називають корелограмою.

## Хід роботи

Завантажимо у середовищі MatLab 6.5 (або вищої версії) пакет для дослідження часового ряду, набравши у командному рядку **work2** та натиснувши **Enter**.

Для дослідження прибутків та волатильності будемо користуватись переважно пунктами **Анализ рядов** та **Обработка ряда**.

Завантажимо часовий ряд, використовуючи пункт меню **Базы**→**Открыть базу**. Після завантаження файлу у вікні програми з'явиться інформація про відкритий ряд: кількість значень, основні статистики. У даній роботі використовуватимемо в якості зразка ряди щоденних значень індексу фондового ринку США – індексу Доу-Джонса за періоди, що включають відомі фінансові кризи 1929, 1987 та 2008 рр. Використавши пункт **Отобразить график**, для кризи 1929 р. одержимо рис.1. Видно, що ряд нестационарний, що викликає певні ускладнення для подальшого аналізу. Тому перейдемо до прибутковостей (**Рассчитать возвраты** – return), які вже є стаціонарними, а їх нормалізація стандартним відхиленням (**нормализовать**) дозволяє легко порівнювати їх розподіл з розподілом Гауса.

На рисунку 2 відображено поведінку нормалізованих прибутковостей. При цьому графік побудований з використанням можливостей вбудованого нами редактора PlotFigure і стандартного MatLab. Вказані також межі ( $\pm 3\sigma$ ), характерні для нормального розподілу.

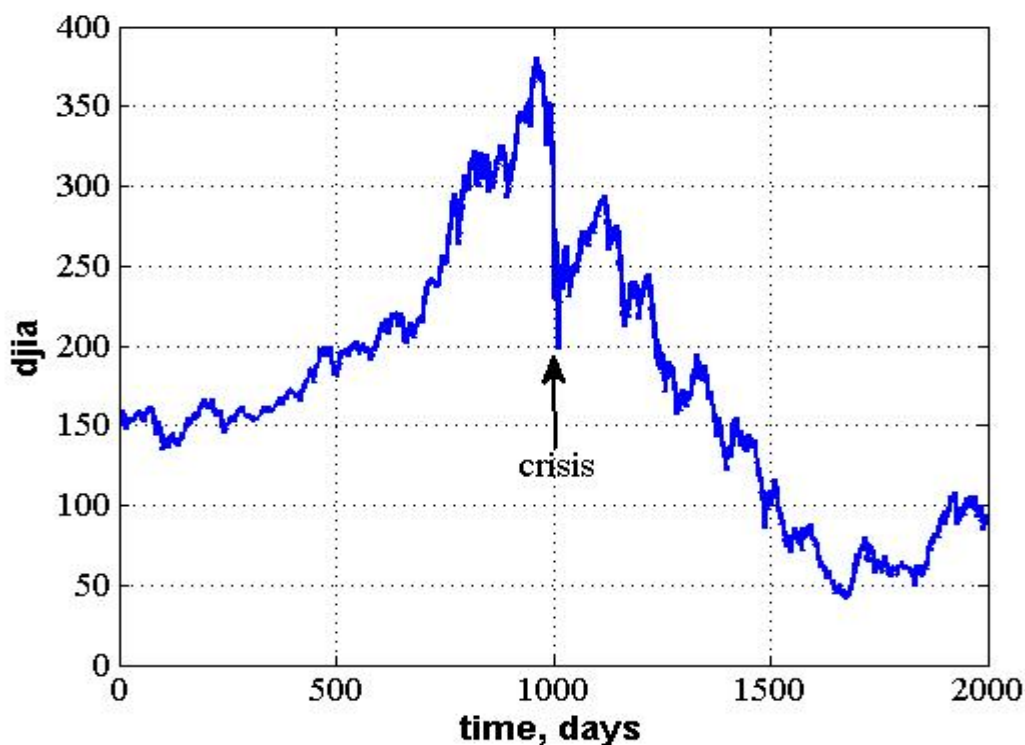


Рис.1. Динаміка щоденних змін індексу Доу-Джонса (djia) у період, що включає кризу 1929 року (відмічена стрілкою)

Зверніть увагу, що флуктуації нормалізованих прибутковостей досить часто перевищують величину  $\pm 3\sigma$ , що, як відомо, надзвичайно рідко спостерігається для незалежних подій. Цей факт можна відобразити шляхом порівняння функції розподілу нормалізованих флуктуацій з розподілом Гауса (рис. 3). Очевидно, що хвости розподілу вихідного ряду містять значні флуктуації, вони досить помітні (часто кажуть «важкі» у порівнянні з самою «головою» розподілу).

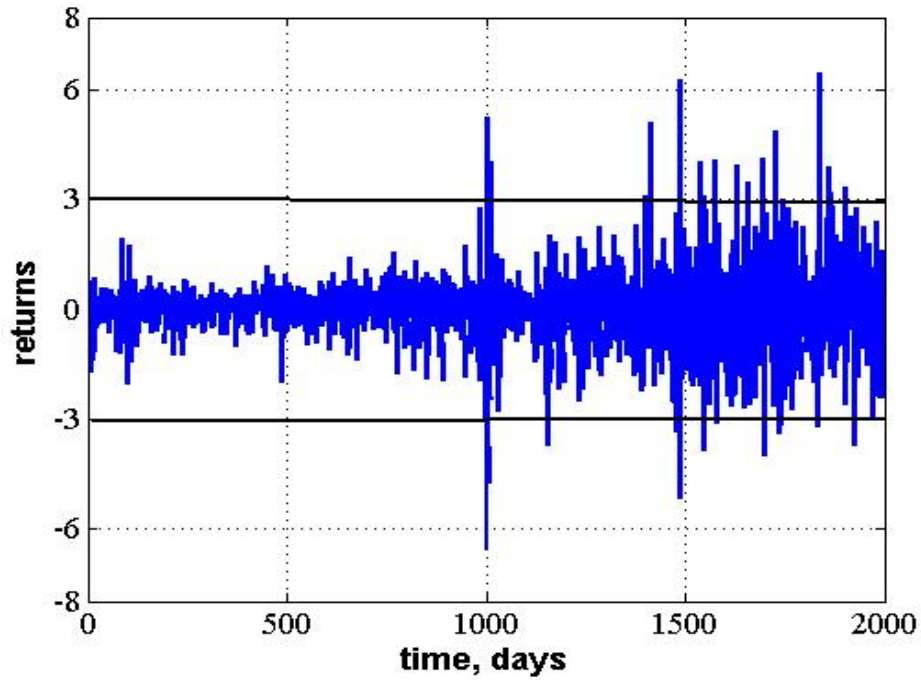


Рис. 2. Нормалізовані прибутковості вибраного часового ряду

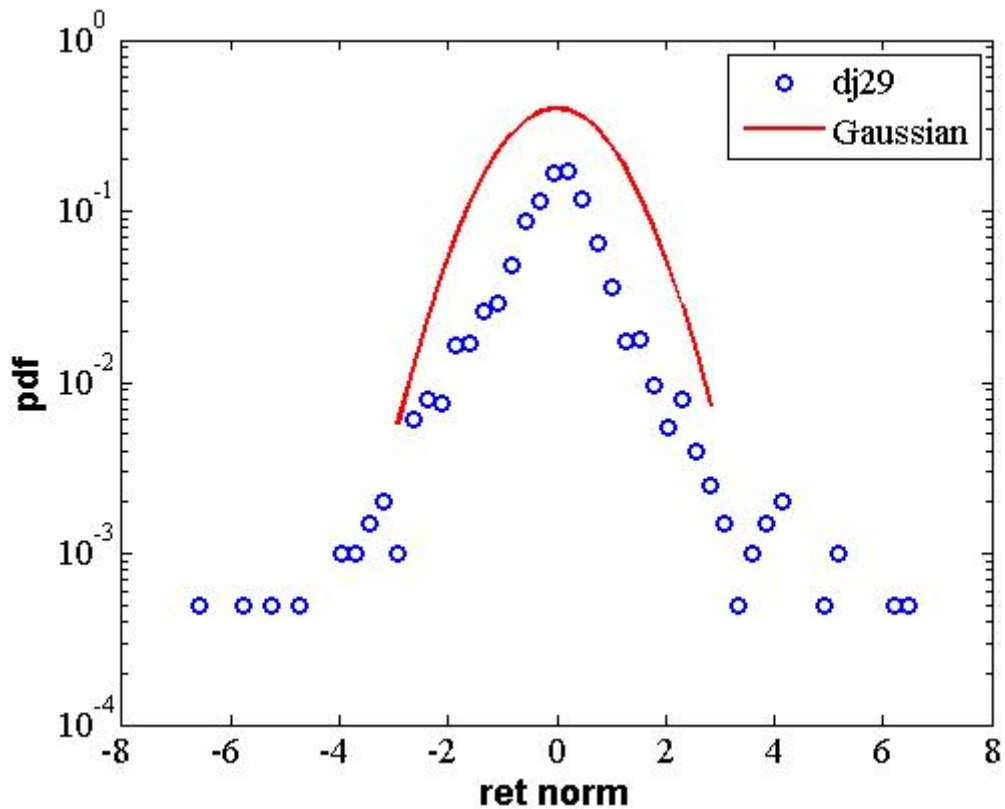


Рис. 3. Порівняння функцій розподілу нормалізованих прибутковостей  $g$  з нормальним розподілом

Отже, ми можемо стверджувати, що наші прибутковості не є незалежними. Підтвердження цьому факту будемо шукати шляхом вивчення кореляційних властивостей нашого часового ряду.

Побудуємо функцію автокореляції. Для цього виберемо *Анализ ряда*→*Функция автокорреляции*, після чого відкриється вікно, подане на рис. 4.

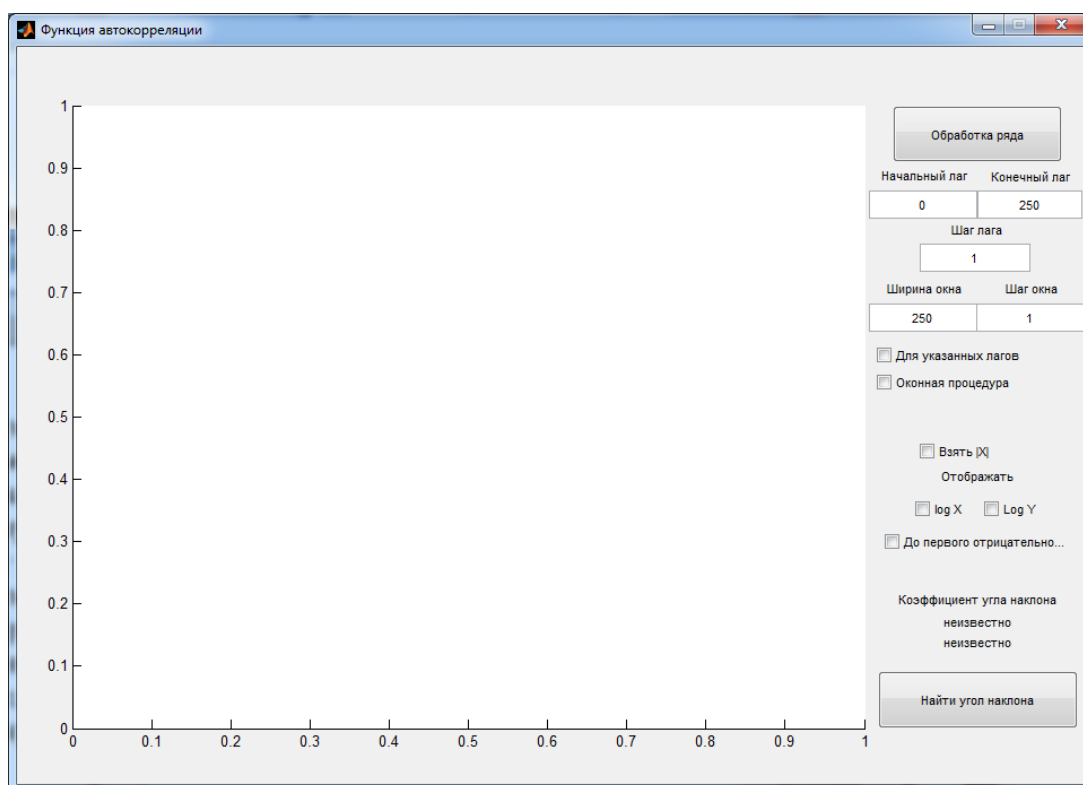


Рис. 4. Вікно для пошуку функції автокореляції

Програмний продукт дає можливість розраховувати функцію автокореляції для усього часового ряду і заданих часових лагів, а також реалізувати віконну процедуру. В цьому випадку автокореляція розраховується для заданого підпроміжку всього часового ряду. Потім він зміщується у бік збільшення часових відліків з вказаним кроком і процедура повторюється до повного вичерпання ряду. У вікні ми можемо вибирати початковий лаг (початкову часову затримку, поле *Начальный лаг*), крок лагу (поле *Шаг лага*) та кінцевий лаг (поле *Конечный лаг*). Здебільшого функція автокореляції будується у звичайному масштабі, тому поля *LogX* та *LogY* виключені. Виберемо початковий, кінцевий лаг та крок лагу відповідно 0, 250, 1 (такі значення встановлено за замовчуванням) і натиснемо кнопку *Обработка*. У полі графіка з'явиться графічне зображення корелограми. Знайдемо корелограми вихідного ряду, нормалізованих прибутковостей та їх модулів і за допомогою вбудованого графічного редактора відобразимо для порівняльного аналізу всі три криві на одному графіку (рис. 5).

З рисунка можна бачити, що вихідні значення ряду дуже сильно корельовано між собою. Таке явище можна пояснити сильним впливом попередніх значень на наступні або досить високою стабільністю ринку. Тривалість довгої пам'яті за допомогою функції автокореляції для вихідного ряду можна вказати, знайшовши, для якого ж лагу значення стають рівними 0 (**Знайдіть цю величину для вашого ряду!**). З рисунка добре видно, що для прибутковостей пам'ять (як довга, так і коротка) відсутня взагалі, оскільки вже для лагу 1 значення автокореляції стає рівним нулю, і всі наступні значення коливаються біля цього значення.

Більш суттєву інформацію дають модулі прибутків, оскільки в даному випадку вимірюється лише наявність відхилень, а не їх характер (у який бік відхилення – в додатній чи від'ємній). На побудованому графіку корелограми можна добре бачити наявність інтервалу з додатними значеннями автокореляції. Для того, щоб точніше визначити зазначений

інтервал, виберіть опцію *До першого отрицательного* і знову натисніть кнопку *Обработка*. Знайдіть і порівняйте з попередніми довжину пам'яті ряду нормалізованих прибутковостей. Зауважимо, що природа довготривалої пам'яті на сьогодні до кінця не досліджена, але очевидно, що для різних цінних активів і їх похідних вона різна. Це треба враховувати в задачах прогнозування цінних коливань.

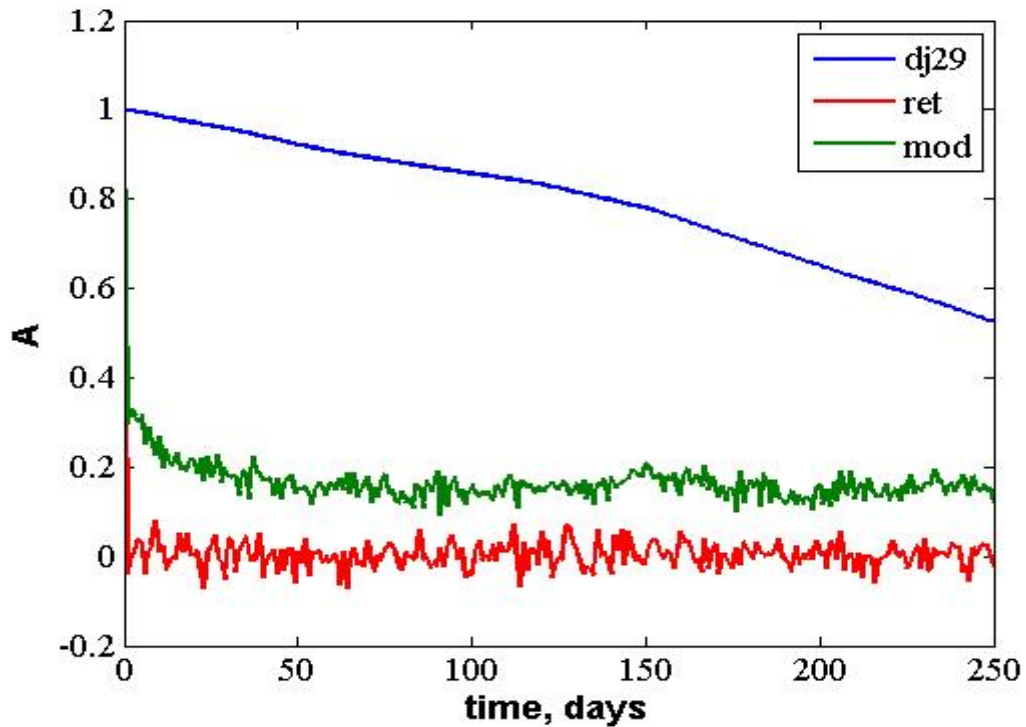


Рис.5. Зміна з часом парних автокореляційних функцій для вихідного ряду  $X$ , нормалізованих прибутковостей  $g$  та їх модулів  $\text{mod}(g)$

У випадку алгоритму рухомого вікна для модулів прибутковостей результати представлені на рис. 6.

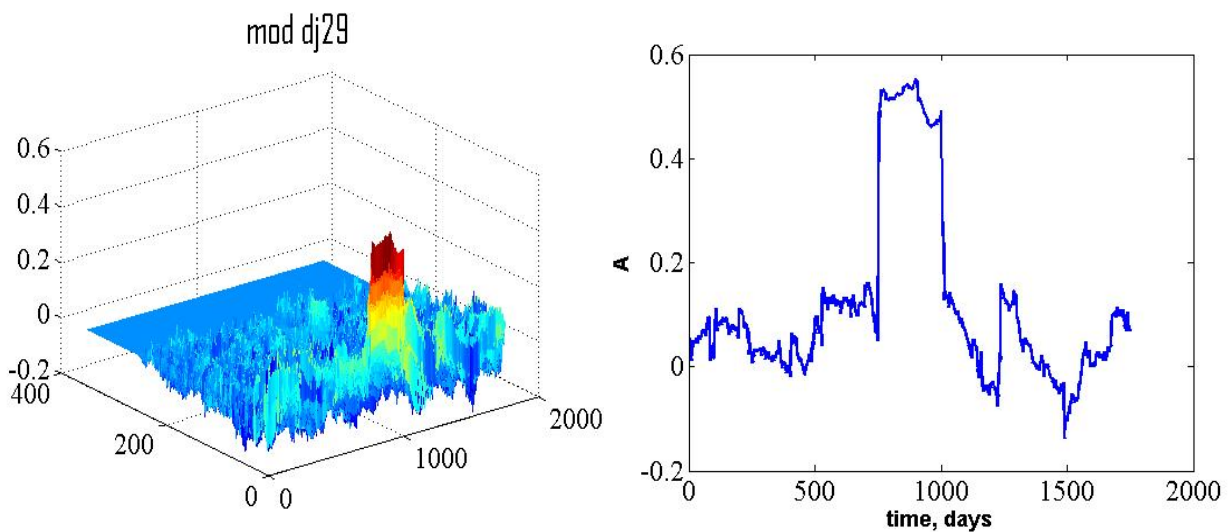


Рис. 6. Динаміка віконної автокореляції, розрахованої з вікном у 250 днів з кроком 1 і лагів від 1 до 350 днів

Наступним кроком у дослідженні ряду є аналіз волатильності.



Згідно з (2) волатильність розраховується для модулів прибутковостей, тому виберемо у основному вікні ряд *Модуль* і розрахуємо волатильність, вказавши *Обработка ряда*→*Расчитать волатильность*. У вікні, що відкриється, можна вибирати параметри *Ширина окна* (ширина вікна для розрахунку середніх значень волатильності) та *Шаг окна* (крок вікна). Для встановлених за замовченням параметрів одержимо результат, відображений на рис. 7.

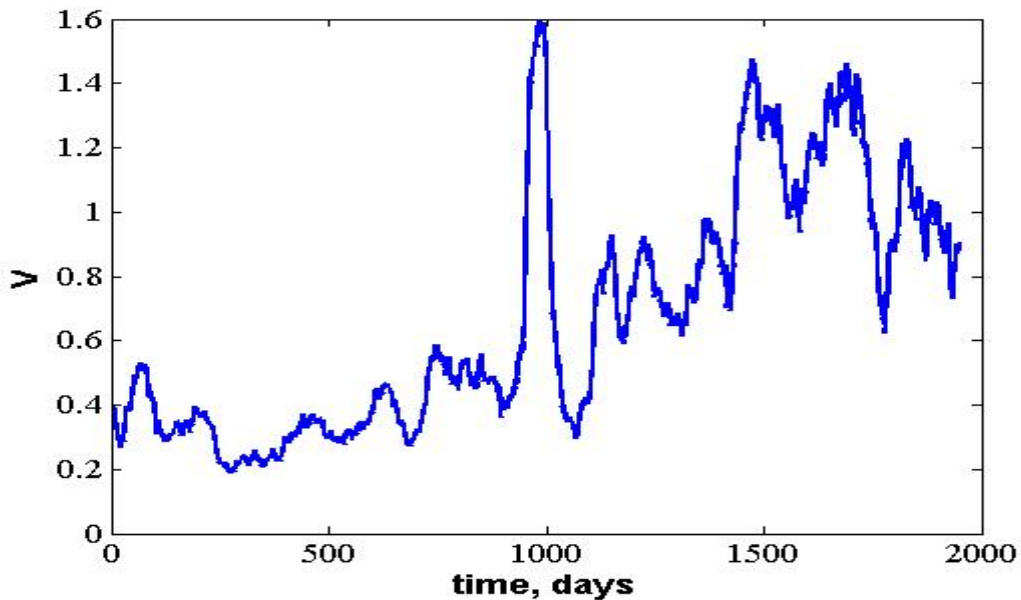


Рис. 7. Розрахована волатильність для модулів прибутковостей

Аналізуючи графік, можна зробити висновок, що у певні моменти спостерігалися стрибки волатильності (як і автокореляції) із поступовим зменшенням її до попереднього рівня, що може бути внаслідок збурень у процесі роботи ринку. Аналіз таких збурень, їх частоти та сили, дозволяє виявляти приховані закономірності роботи ринку.

Нарешті порівняємо динаміку вихідного ряду і розрахованих похідних. Для цього врахуємо, що автокореляцію і волатильність ми рахували для рухомого вікна. Результати представлено на рис. 8.

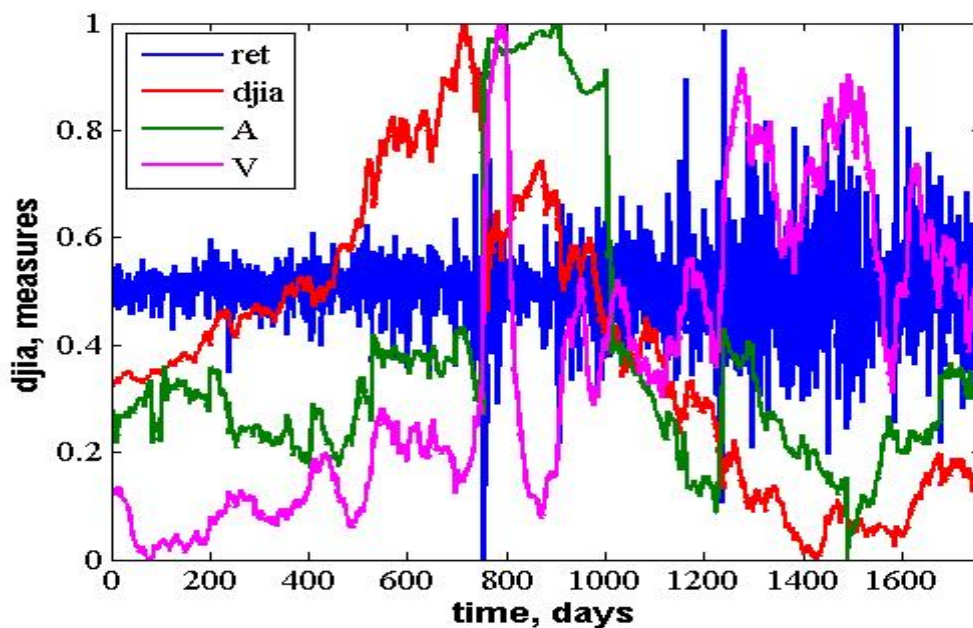


Рис. 8. Динаміка індексу Доу-Джонса, нормалізованих прибутковостей, віконних автокореляції та волатильності

Очевидно, що така процедура порівняння дозволяє виявляти за похідними величинами їх реакцію на критичні і кризові явища. В даному випадку у кризовий період усі похідні стрімко зростають у період кризи і є власне її індикаторами.

### Висновок

Таким чином, аналіз флуктуацій прибутковостей та волатильностей шляхом побудови функції автокореляції та розподілу ймовірності дозволяє отримати певні висновки, що можуть допомогти в роботі із аналізованими часовими рядами і ринком, з якого взято зазначені часові ряди. Зокрема, у даному випадку, можна давати рекомендації аналітикам, що працюють на фінансових ринках.

### Завдання для самостійної роботи

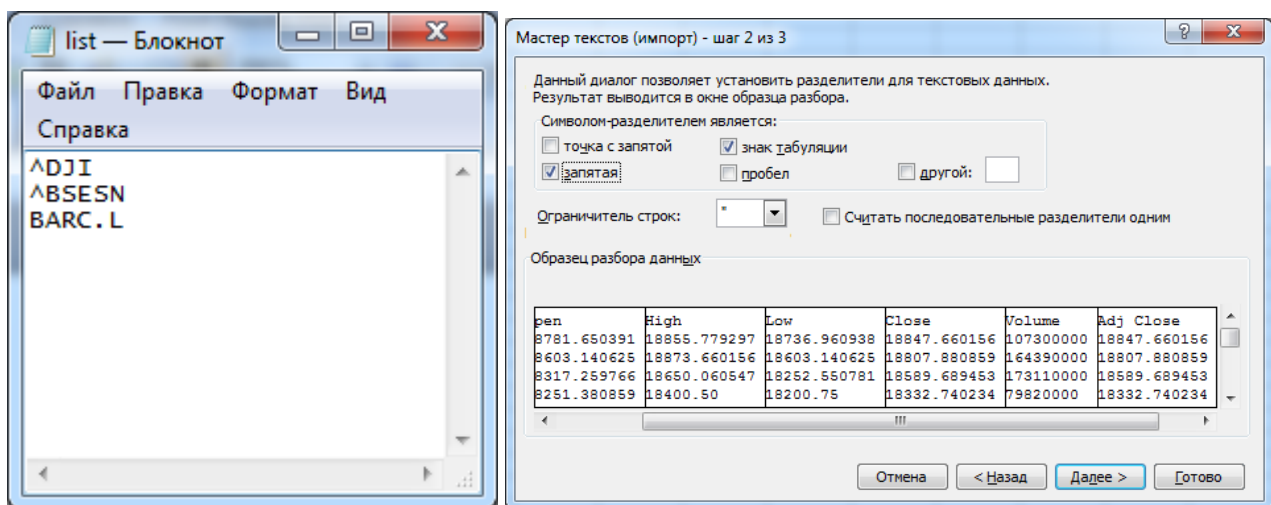
1. Отримати базу даних у викладача
2. Провести дослідження часових флуктуацій, волатильностей та автокореляцій згідно інструкції
3. Проведіть дослідження автокореляцій прибутковостей для вихідних і перемішаних рядів.
4. Зробити висновки

### Контрольні питання

1. Порівняйте вид залежностей флуктуацій цін і прибутковостей. Чому при розрахунках користуються не цінами, а прибутковостями ?
2. Яку характеристику ряду визначає волатильність ?
3. В чому причина різних залежностей прибутковостей і їх модулів ?

### Додаток

Для роботи з поширеними фондовими індексами, курсами акцій окремих компаній, наприклад, банківських установ, архіви яких оновлюються на сайті <https://finance.yahoo.com>, було створено спеціальний скрипт **Yahoo\_get\_task.m**, який за списком тікерів **list.txt** зберігає у визначеній папці історію перелічених активів. Крім вказаних файлів у папці повинні міститись також додатково ще два скрипти: **yahoo\_get.m** та **get\_quote\_close\_alldates\_adjclose.m**. Наприклад, якщо файл **list.txt** має вигляд, представлений на рис (а), то будуть збережені історії індексів Доу-Джонса індустріального (DJI), фондового індексу фондової біржі у Бомбеї (BSESN) та ціни акцій лондонського банку Берклі (BARC).



(а)

(б)

Оскільки архіви зберігаються у форматі **.csv**, для виділення необхідного стовпчика з таблиці, а саме, **Adj Close**, відкриємо необхідні із збережених файлів у MS Excel (рис. б),

впорядкуємо за зростанням (а не за спаданням, як в оригіналі), виділимо за датами необхідну кількість значень часового ряду, скопіюємо та збережемо у форматі **.txt** для подальшого використання.

## ЛАБОРАТОРНА РОБОТА № 2

**Тема.** Фрактальний і мультифрактальний аналіз часового ряду.

**Мета.** Навчитися використовувати аналіз флуктуацій та його похідні для отримання нелінійних характеристик часового ряду.

1. Стандартний аналіз флуктуацій.
2. Аналіз детрендованих флуктуацій (АДФ) другого порядку.
3. Визначення коефіцієнта Херста методами  $R/S$ -аналізу та АДФ.
4. Мультифрактальний АДФ.

### Теоретичні відомості

#### $R/S$ -аналіз

Метод  $R/S$ -аналізу, розроблений Мандельбротом та Уоллесом, базується на попередньо створеному методі гідрологічного аналізу Херста, і дозволяє обчислювати параметр самоподібності  $H$ , який вимірює інтенсивність довготривалих залежностей у часовому ряду. Коефіцієнт  $H$ , який називають коефіцієнтом Херста, містить мінімальні прогнози стосовно природи системи, що вивчається, і може класифікувати часові ряди. За допомогою цього показника розрізняють випадкові (гаусові) та не випадкові ряди; окрім того, він пов'язаний із фрактальною розмірністю, що, у свою чергу, характеризує ступінь згладженості графіка, побудованого на основі часового ряду. Методом  $R/S$ -аналізу можливо також виявити максимальну довжину інтервалу (цикл), на якому значення зберігають інформацію про початкові дані системи (довготривала пам'ять).

Аналіз починається із побудови ряду логарифмічних прибутків,  $G_i(t) \equiv \ln S_i(t + \Delta t) - \ln S_i(t)$ , де  $S_i(t)$  – значення вихідного часового ряду у момент  $t$ ,  $\Delta t$  – часовий крок. Отримана послідовність  $G_i(t)$  розбивається на  $d$  підпослідовностей довжини  $n$ .

Для кожної підпослідовності  $m = 1, \dots, d$ :

1) шукається середнє значення  $E_m$  та стандартне відхилення  $S_m$ ;

2) дані нормалізуються шляхом віднімання середнього значення послідовності  $X_{i,m} = G_{i,m} - E_m$ ,  $i = 1, \dots, n$ ;

3) створюється послідовність накопичень  $Y_{i,m} = \sum_{j=1}^i X_{j,m}$ ,  $i = 1, \dots, n$ ;

4) знаходиться розмах  $R_m = \max\{Y_{1,m}, \dots, Y_{n,m}\} - \min\{Y_{1,m}, \dots, Y_{n,m}\}$ , який нормується середнім квадратичним відхиленням  $R_m/S_m$ ;

5) обчислюється середнє  $(R/S)_n$  нормованих значень розмаху для всіх підпослідовностей довжини  $n$ .

$R/S$ -статистика, обрахована таким чином, відповідає співвідношенню  $(R/S)_n \cong cn^H$ , де значення  $H$  може бути отримане шляхом обчислення  $(R/S)_n$  для послідовності інтервалів зі збільшенням часового горизонту:

$$\log(R/S)_n = \log c + H \log n. \quad (1)$$

Знайти коефіцієнт Херста можна, побудувавши залежність  $(R/S)_n(n)$  у подвійному логарифмічному масштабі і взявши коефіцієнт нахилу прямої, яка інтерполює точки отриманого графіка. Якщо значення  $H = 0.5$ , говорять про послідовність, що представляє собою білий шум;  $0.5 < H \leq 1$  свідчить про персистентний ряд, коли існує тенденція слідування великих значень ряду за великими і навпаки;  $H < 0.5$  вказує на антиперсистентний ряд.

При збільшенні часового горизонту коефіцієнт нахилу інтерполюючої прямої повинен прямувати до значення  $H = 0.5$ ; сам процес переходу свідчить про втрату впливу початкових умов на поточні значення, і, таким чином, можна говорити про горизонт довгої пам'яті – це точка, до якої коефіцієнт нахилу інтерполюючої прямої відмінний від 0.5, а після – близько 0.5.

### Стандартний аналіз флуктуацій

У стандартному аналізі флуктуацій (АФ) розглядається «профіль накопичень» (runoff profile)

$$y_n = \sum_{i=1}^n x_i, \quad n = \overline{1 \dots N} \quad (2)$$

і вивчається швидкість зміни флуктуацій профіля в залежності від зміни ширини вікна  $s$ . Можна розглядати профіль  $y_n$  як позицію точки у ланцюжку випадкових блукань після  $n$  кроків. Випадкові блукання починаються в деякій точці; на  $i$ -му кроці точка переміщується вправо, якщо  $x_i \in$  додатнім, або ж вліво, якщо  $x_i$  – від'ємне.

Для знаходження зміни скейлінгового показника в залежності від  $s$  спочатку часовий ряд довжиною  $N$  елементів розбивається на підпослідовності, що не перекриваються,  $N_s = \text{int}(N/s)$  довжиною  $s$ , починаючи з першого елемента, та  $N_s$  підпослідовностей, що не перекриваються, довжиною  $s$ , починаючи з останнього елемента. Потім визначаються флуктуації у кожному сегменті  $v$ .

У стандартному АФ флуктуації обчислюють безпосередньо від значень профілю на обох кінцях кожної підпослідовності  $v$ ,  $F^2(v, s) = (y_{vs} - y_{(v-1)s})^2$  і знаходиться середнє значення  $F^2(v, s)$  для  $2N_s$  підпослідовностей, щоб досягти реального значення флуктуації  $F_s(s)$ ,

$$F_2(s) \equiv \left( \frac{1}{2N_s} \sum_{v=1}^{2N_s} F^2(v, s) \right)^{1/2}. \quad (3)$$

За означенням,  $F_2(s)$  може розглядатись як середньоквадратичний зсув (переміщення) точки випадкових блукань в ланцюжку після  $s$  кроків. Для некорельованих значень  $x_i$  отримується закон дифузії Фіка (Fick)  $F_2(s) \cong s^{1/2}$ . Для істотних випадків довгочасових кореляцій, де поведінка  $C(s)$  відповідає степеневому закону ( $C(s) = \langle x_i x_{i+s} \rangle \cong s^{-\gamma}$ ,  $0 < \gamma < 1$ ),  $F_2(s)$  збільшується теж згідно зі степеневим законом

$$F_2(s) \cong s^H,$$

де флуктуаційний коефіцієнт  $H$  пов'язаний із кореляційним коефіцієнтом  $\gamma$  і коефіцієнтом спектра потужності  $\beta$  наступним законом:

$$H = 1 - \gamma/2 = (1 + \beta)/2.$$

Для степеневих кореляцій, що спадають швидше за  $1/s$ , отримуємо  $H = 1/2$  для великих значень  $s$ , так як і для некорельованих даних.

Варто зауважити, що стандартний аналіз флуктуацій де в чому подібний до  $R/S$ -аналізу, розробленого Херстом, за виключенням того, що він базується на другому моменті  $F_2(s)$ , в той час як Херст розглядав випадок першого моменту  $F_1(s)$ . Для монофрактальних даних  $H$  ідентичний до коефіцієнта Херста.

### Аналіз детрендованих флуктуацій (АДФ)

АДФ базується на гіпотезі про те, що корельований часовий ряд може бути відображений на самоподібний процес шляхом інтегрування. Таким чином, вимірювання властивостей самоподібності може непрямо свідчити про кореляційні властивості ряду. Переваги АДФ порівняно з іншими методами (спектральний аналіз,  $R/S$ -аналіз) полягають в тому, що він виявляє довгочасові кореляції нестационарних часових рядів, а також дозволяє ігнорувати очевидні випадкові кореляції, що є наслідком нестационарності.

Існують АДФ різних порядків, що відрізняються трендами, які вилучаються з даних.

Розглянемо АДФ найнижчого порядку.

Часовий ряд довжини  $N$  інтегрується,  $Y(k) = \sum_{i=1}^k (X_i - \bar{X})$ , де  $X_i$  –  $i$ -те значення часового ряду,  $\bar{X}$  – його середнє значення,  $k = 1, \dots, N$ . Ця дія повторює процес знаходження накопичень у  $R/S$ -аналізі. Отриманий ряд  $Y(k)$  розбивається на  $m$  підпоследовностей (вікон) однакової ширини  $n$  і для кожної підпоследовності (у кожному вікні) виконується наступне: 1) за допомогою методу найменших квадратів знаходиться локальний лінійний тренд  $Y_t(k)$ ; 2) підпоследовність детрендується шляхом віднімання значення локального тренду  $Y_t(k)$  від значень ряду  $Y(k)$ , що належать підпоследовності  $t$ ; 3) знаходиться середнє  $\bar{Y}_t$  детрендованих значень.

Для отриманих таким чином значень на всіх підпоследовностях знаходиться:

$$F_n = \sqrt{\frac{1}{m} \bar{Y}_t},$$

де  $n$  – кількість точок у підпоследовності (ширина вікна),  $m$  – кількість підпоследовностей,  $\bar{Y}_t$  – середнє детрендованих значень для підпоследовності  $t$ .

Вказана процедура повторюється для вікон різної ширини, внаслідок чого отримується набір пар точок  $(F_n; n)$ . Побудова залежності  $\log F(n)$  від  $\log n$  та інтерполяція отриманих значень прямою дає змогу обчислити показник скейлінга  $\alpha$ , що є коефіцієнтом кута нахилу інтерполяційної прямої і характеризує зміну кореляцій логарифмічних прибутків часового ряду при збільшенні часового інтервалу.

Порівняно із  $R/S$ -аналізом метод АДФ дає більші можливості інтерпретації скейлінгового показника  $\alpha$ : для випадкового ряду (перемішаного чи «сурогатного»)  $\alpha = 0.5$ ; при наявності лише короткочасових кореляцій  $\alpha$  може відрізнятися від 0.5, проте має тенденцію прямувати до 0.5 при збільшенні розміру вікна. Значення  $0.5 < \alpha \leq 1.0$  показує персистентні довгочасові кореляції, що відповідають степеневому закону;  $0 < \alpha < 0.5$  означає антиперсистентний ряд. Спеціальний випадок, коли  $\alpha = 1$ , означає наявність шуму  $1/f$ . Для випадків, коли  $\alpha \geq 1$ , кореляції існують, проте перестають відображувати степеневу

залежність; випадок  $\alpha = 1.5$  свідчить про Броунівський шум, інтегрований білий шум.

У АДФ другого порядку (АДФ2) обчислюються відхилення  $F^2(\nu, s)$  профілю від графіка інтерполяційного многочлена другого порядку. Таким чином, вилучаються впливи можливих лінійних та параболічних трендів для масштабів, більших за розглядувані. Взагалі, у АДФ порядку  $n$  обчислюються відхилення профілю від інтерполяційного многочлена  $n$ -го порядку, що вилучає вплив всіх можливих трендів порядків до  $(n-1)$  для масштабів, більших від розміру вікна.

Потім обчислюється найближчий поліном  $y_\nu(s)$  для профілю на кожному із  $2N_s$  сегментів  $\nu$  і визначається відхилення

$$F^2(\nu, s) \equiv \frac{1}{s} \sum_{i=1}^s (x_{(\nu-1)s+i} - y_\nu(i))^2. \quad (4)$$

Далі застосовується (3) для визначення середньої флуктуації  $F_2(s)$ .

Так як АФ та різні види АДФ мають різні детрендові властивості, їх композиція може використовуватись як для виявлення довгочасових кореляцій, так і для знаходження типів наявних у ряді трендів, що не може бути визначеним з використанням традиційних методів, таких як спектральний аналіз.

### Локальний АДФ

Метод локального АДФ виник з припущення про існування в часовій послідовності як персистентних так і антиперсистентних ділянок.

Для отримання даних на основі локального АДФ використовується вікно постійного розміру, який залежить від максимального значення  $s$  у стандартному АДФ, коли ще було отримано стабільне значення  $\alpha$ . Розмір вікна повинен бути достатнім для використання стандартного методу АДФ, за допомогою якого і визначається локальний коефіцієнт  $\alpha$ . Після цього вікно пересувається по послідовності на певний (невеликий) крок (наприклад, 20 днів) і обчислення повторюються.

Аналіз динаміки локального коефіцієнта Херста може допомогти при інтерпретації суттєвих змін у динаміці системи.

### Мультифрактальний АДФ

Стандартний АДФ використовується для визначення (моно-) фрактальних скейлінгових властивостей і довгочасових кореляцій в зашумлених нестационарних часових рядах. Проте багато економічних (біологічних, медичних тощо) об'єктів не демонструють простої монофрактальної скейлінгової поведінки, що може бути визначена одним коефіцієнтом. В деяких випадках існує кросовер (crossover) на часових шкалах  $s_x$ , що відділяє моделі з різною поведінкою, наприклад, довгочасові кореляції на малих масштабах часу  $s \ll s_x$  та кореляції іншого виду чи некорельовану поведінку на більших масштабах  $s \gg s_x$ . В деяких випадках поведінка скейлінга ще більш складна, і існують різні значення коефіцієнтів скейлінга для різних частин послідовності (наприклад, для першої та другої половини послідовності). Трапляються ще більш складні випадки, коли розглядаються накладені одна на одну множини із різною фрактальністю. Для таких випадків необхідно обчислювати множину коефіцієнтів скейлінга для повного опису поведінки об'єкта. Тоді застосовується мультифрактальний аналіз детрендованих флуктуацій.

У загальному випадку процедура мультифрактального АДФ (МФ-АДФ) складається із п'яти кроків. Перші три кроки, по суті, ідентичні процедурі стандартного АДФ.

Нехай є послідовність  $x_k$  довжини  $N$ , що не має великої кількості нульових значень (compact support).

*Крок 1.* Визначається профіль (накопичення):

$$Y(i) \equiv \sum_{k=1}^i (x_k - \bar{x}), \quad i = \overline{1..N}. \quad (5)$$

Віднімання середнього  $\bar{x}$  є необов'язковим, оскільки може бути виконане пізніше детрендуванням на третьому кроці.

*Крок 2.* Профіль  $Y(i)$  розбивається на  $N_s = \text{int}\left(\frac{N}{s}\right)$  сегментів (підпослідовностей) однакової довжини  $s$ , що не перекриваються. Оскільки загальна довжина послідовності  $N$  часто не ділиться націло на  $s$ , залишок в кінці послідовності, що є меншим за ширину вікна, відкидається. Для врахування відкинутої частини і використання, таким чином, усіх елементів послідовності, вищевказана процедура повторюється також, починаючи з кінця послідовності. Таким чином, разом буде отримано  $2N_s$  підпослідовностей.

*Крок 3.* Для кожної із  $2N_s$  підпослідовностей обчислюється локальний тренд методом найменших квадратів. Потім визначається відхилення

$$F^2(\nu, s) = \frac{1}{s} \sum_{i=1}^s (Y((\nu-1)s+i) - y_\nu(i))^2 \quad (6)$$

для кожного сегмента  $\nu$ ,  $\nu = \overline{1..N}$  і

$$F^2(\nu, s) = \frac{1}{s} \sum_{i=1}^s (Y(N - (\nu - N_s)s + i) - y_\nu(i))^2 \quad (7)$$

для кожного  $\nu = \overline{N_s + 1..2N_s}$ . Тут  $y_\nu(i)$  є інтерполюючий поліном на сегменті  $\nu$ . Для інтерполяції використовуються лінійні, квадратичні, кубічні поліноми чи поліноми вищого порядку (традиційно називаються АДФ1, АДФ2, АДФ3 і т.д.). Оскільки детрендування часового ряду відбувається відніманням значень полінома від реальних значень ряду, АДФ різних порядків відповідно відрізняються у своїх можливостях по вилученню тренду в ряді. У (МФ-)АДФ $m$  ((МФ-)АДФ  $m$ -го порядку) вилучаються тренди профілю порядку  $m$  (або відповідно порядку  $(m-1)$  для вихідного часового ряду). Таким чином порівняння результатів роботи АДФ різних порядків використовується для отримання типу тренду у вихідному часовому ряді.

*Крок 4.* Знаходиться середнє по всіх підпослідовностях для отримання функції флуктуацій  $q$ -го порядку:

$$F_q(s) = \left( \frac{1}{2N_s} \sum_{\nu=1}^{2N_s} (F^2(s, \nu))^{q/2} \right)^{1/q}, \quad (8)$$

де, взагалі кажучи, значення змінної  $q$  може бути довільним за виключенням нульового. Для  $q = 2$  отримується стандартний метод АДФ. При розгляді питання, як впливає часова шкала  $s$  при різних значеннях  $q$  на узагальнену залежність  $F_q(s)$  від  $q$ . Для цього необхідно повторити кроки 2-4 для різних часових масштабів  $s$ . Цілком зрозуміло, що  $F_q(s)$  буде збільшуватись із збільшенням  $q$ . Звичайно,  $F_q(s)$  залежить також від порядку методу АДФ

*m*. Згідно з побудовою,  $F_q(s)$  визначене лише для значень  $s \geq m + 2$ .

*Крок 5.* Визначається скейлінгова поведінка функції флуктуацій шляхом аналізу у подвійному логарифмічному масштабі залежності  $F_q(s)$  від  $q$ . Якщо послідовність  $x_i$  має довгочасові кореляції,  $F_q(s)$  збільшується із збільшенням  $s$  згідно степеневому закону

$$F_q(s) \cong s^{h(q)}. \quad (9)$$

Взагалі, коефіцієнт  $h(q)$  повинен залежати від  $q$ . Для стаціонарних часових рядів,  $h(2)$  ідентичний коефіцієнту Херста. Таким чином, функцію  $h(q)$  можна назвати узагальненим коефіцієнтом Херста.

Для монофрактальної часової послідовності  $h(q)$  залежить від  $q$ , таким чином скейлінгова поведінка відхилень  $F^2(v, s)$  однакова на всіх сегментах  $V$  і процедура усереднення (8) дасть однакові значення скейлінгового коефіцієнта для всіх сегментів послідовності. Лише у випадку, коли масштаби малих і великих флуктуацій відрізняються, буде помітною залежність  $h(q)$  від  $q$ : якщо розглядати додатні значення  $q$ , сегменти  $v$  з великими значеннями  $F_s^2(v)$  (наприклад, великі відхилення від відповідних інтерполяційних поліномів) будую домінувати у середньому значенні  $F_q(s)$ . Таким чином, для додатніх значень  $q$   $h(q)$  описує скейлінгову поведінку сегментів із великими флуктуаціями. Зазвичай великі флуктуації характеризуються меншими скейлінговими коефіцієнтами  $h(q)$  для мультифрактальних рядів. Навпаки, для від'ємних значень  $q$  сегменти  $v$  з малими відхиленнями  $F_s^2(v)$  будуть домінувати у середньому значенні  $F_q(s)$ . Таким чином, для від'ємних значень  $q$   $h(q)$  описує скейлінгову поведінку сегментів з малими флуктуаціями, що зазвичай характеризуються більшим скейлінговим коефіцієнтом.

Однак, метод МФ-АДФ може визначати лише *додатні* узагальнені коефіцієнти Херста  $h(q)$ , і стає неточним для сильно антикорельованих сигналів, коли  $h(q)$  близьке до нуля. В таких випадках використовується модифікований МФ-АДФ аналіз. Більш простим шляхом для аналізу подібних даних може служити інтегрування часового ряду перед виконанням МФ-АДФ процедури (знаходження накопичень). Звідси, просте знаходження суми у (5), що описує профіль початкових даних, замінюється подвійним знаходженням суми,

$$\tilde{Y}(i) = \sum_{k=1}^i (Y(k) - \bar{Y}). \quad (10)$$

Після цього виконується процедура МФ-АДФ, описана вище, внаслідок чого отримується узагальнена функція флуктуацій

$$\tilde{F}_q(s) \cong s^{\tilde{h}(q)} = s^{h(q)+1}. \quad (11)$$

Таким чином, скейлінгова поведінка може бути точно визначена навіть у випадку, коли  $h(q)$  близьке до нуля (проте більше за  $-1$ ) для деяких значень  $q$ . Зауважимо, що  $\tilde{F}_q(s)/s$  відповідає  $F_q(s)$  у формулі (9). Якщо на кожному кроці (10) не віднімається середнє значення, така сума, більш за все, дасть квадратичний тренд у профілі  $\tilde{Y}(i)$ . В цьому випадку необхідно використовувати, щонайменше, МФ-АДФ другого порядку для вилучення такого штучного тренду.



## Хід роботи

У середовищі MatLab завантажимо пакет для дослідження часового ряду, набравши у командному рядку **work2** та натиснувши **Enter**.

У цій роботі будемо користуватись переважно пунктом меню *Анализ рядов*.

Для дослідження часового ряду завантажимо його, використовуючи пункт меню ринку спочатку необхідно відкрити базу з часовими рядами. Заходимо у пункт меню *Базы*→*Открыть базу* та вибираємо потрібну папку і ряд для дослідження. У даній роботі використовуватимемо ті ж самі часові ряди, які містять відомі кризові явища.

Найпростіше дослідження, яке можна провести для даного ряду – дослідити за допомогою *R/S*-аналізу. *R/S*-аналіз проводиться не для самого ряду, а для прибутковостей, тому розрахуємо їх, використовуючи пункт меню *Обработка ряда*→*Рассчитать возвраты*. Нормалізуємо їх. Після цього у вікні програми з'явиться новий ряд *ВозврНор*, з яким і будемо проводити необхідні операції. Виберемо пункт меню *Анализ ряда*→*Коэффициент Херста*, внаслідок чого відкриється вікно, подане на рис. 1.

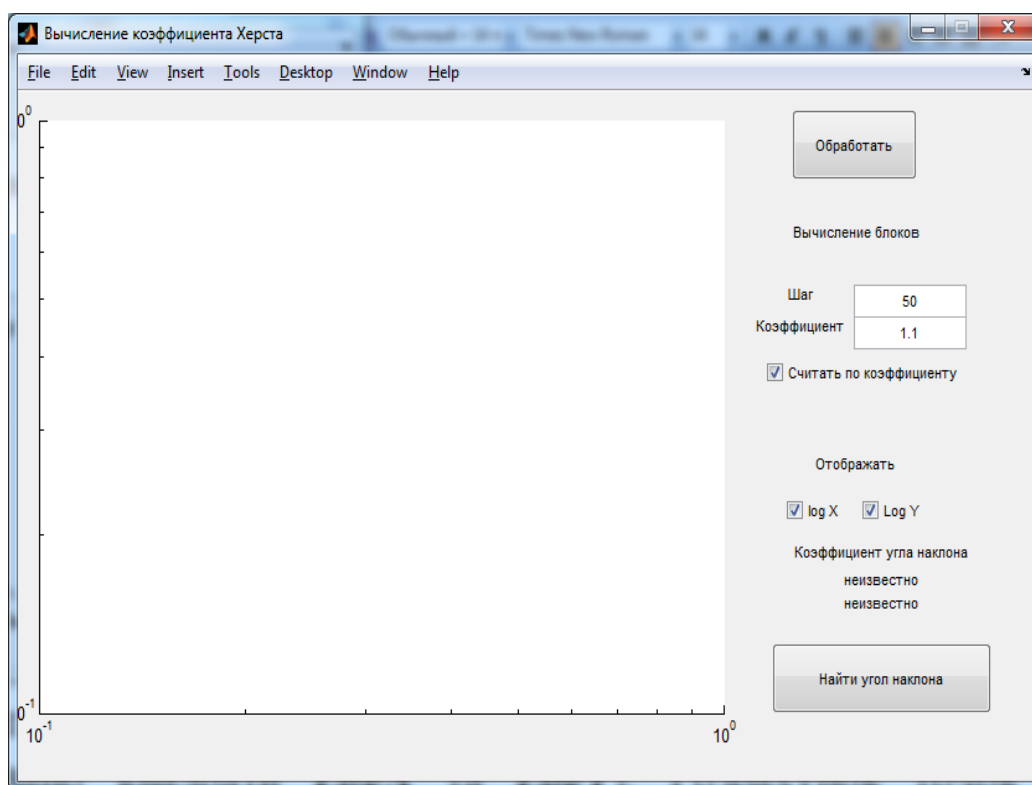


Рис. 1. Вікно програми для обчислення коефіцієнта Херста

У вікні результати аналізу відображаються у подвійному логарифмічному масштабі (активізовано квадрати *LogX* та *LogY*). Розрахунок точок для отримання значень коефіцієнта Херста проведемо з використанням послідовності ширини вікон, де наступне значення більше у кілька разів за попереднє (такий спосіб найкращий для відображення у вибраному нами масштабі), у даному випадку коефіцієнт рівний 1.1, тобто перша ширина оброблюваних вікон може бути, наприклад, 100, наступна 110, далі: 121, 133, 146, 160 і т.д.

Натиснемо кнопку *Обработать* і після проведення розрахунків програма видасть графік, подібний до зображеного на рис. 2.

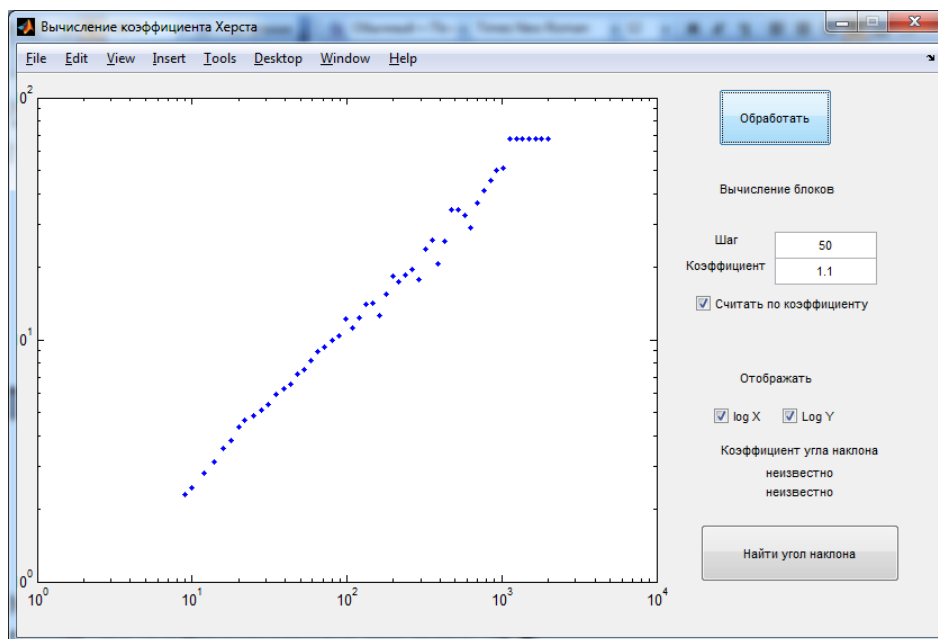


Рис. 2. Ряд, побудований на основі початкового, для визначення коефіцієнта Херста

Для отримання коефіцієнта Херста необхідно натиснути кнопку **Найти угол наклона**, після чого вибрати початкову та кінцеву точку діапазона графіка (мишкою), на якому розраховується коефіцієнт Херста. Внаслідок проведення описаних дій з'явиться пряма лінія, що проходить через графік, кут нахилу якої і є шуканим коефіцієнтом Херста. Кут нахилу відображується під написом **Коэффициент угла наклона** (перше із двох чисел) і рівний для нашого прикладу 0.60, що свідчить про те, що ряд відрізняється від випадкового, або ж, іншими словами, має місце довготривала пам'ять у ряді, обробленому описаним вище чином (для логарифмів прибутків) (рис. 3).

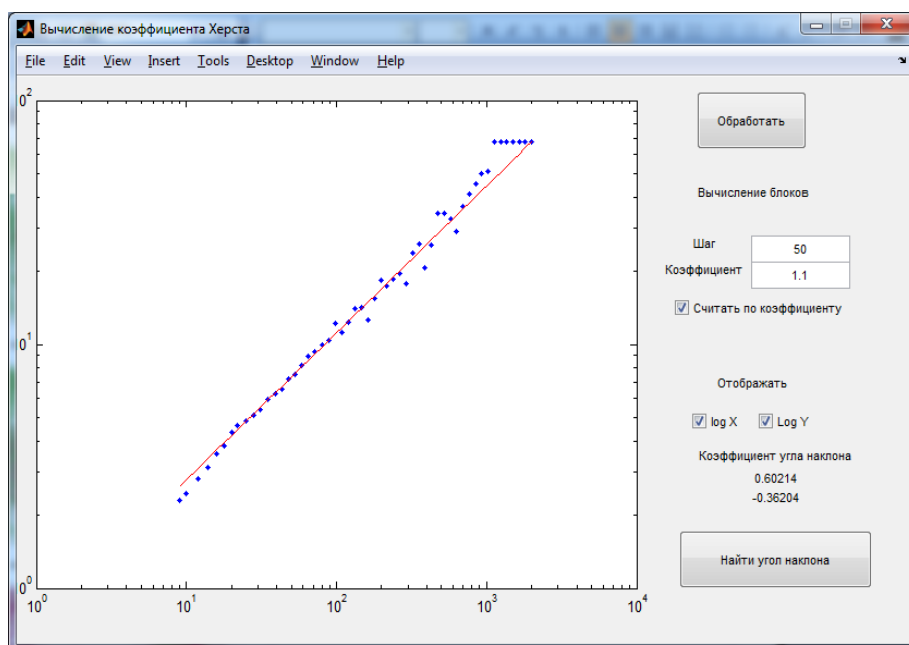


Рис. 3. Ряд із знайденим коефіцієнтом Херста

Якщо перебудувати цей же графік у вбудованому графічному редакторі **PlotFigure** (у подвійному логарифмічному масштабі, провівши пряму, замінивши лінію певним маркером, додавши легенду), отримуємо рис. 4.

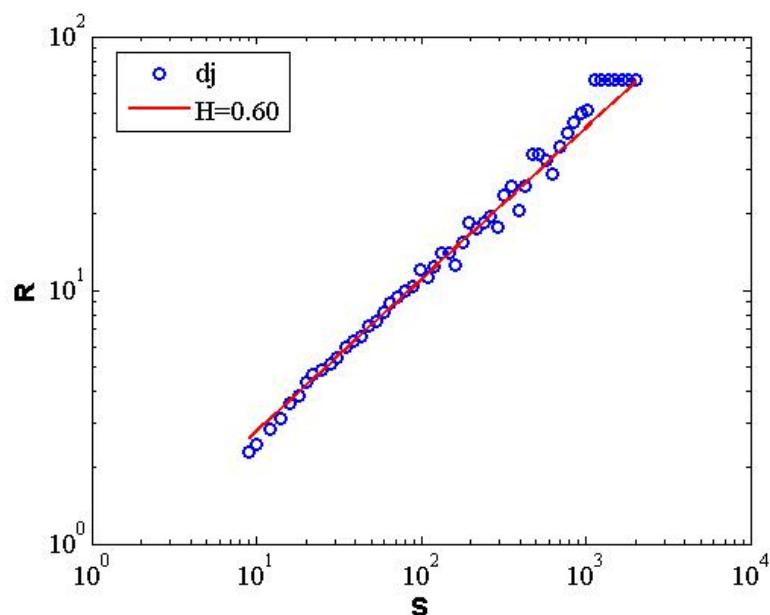


Рис. 4. Коефіцієнт Херста для ряду dj08, знайдений за методом  $R/S$  - аналізу

Після проведення дослідження можемо закрити вікно **Вычисление коэффициента Херста**.

Більш важливу інформацію про ряд дає аналіз методом детрендованих флуктуацій.

Для цього виберемо у основному вікні програми ряд та пункт **Анализ ряда→DFA** (Detrended Fluctuation Analysis). Вікно, що відкриється, дещо відрізняється від вікна пошуку коефіцієнта Херста. Тут необхідно, окрім кроку збільшення вікна, вводити початковий та кінцевий лаг (часову затримку). Початковим значенням виберемо 5, а кінцевим число, яке не менше, ніж половина вихідного ряду. В нашому випадку 1000 точок. Після проведення вказаних дій натиснемо кнопку **Обработка**. Отримаємо результат, поданий на рис. 5. Очевидно, що результати відрізняються. Аналіз методів показує, що останній результат є більш точним, а метод  $R/S$  - аналізу завищує коефіцієнт Херста.

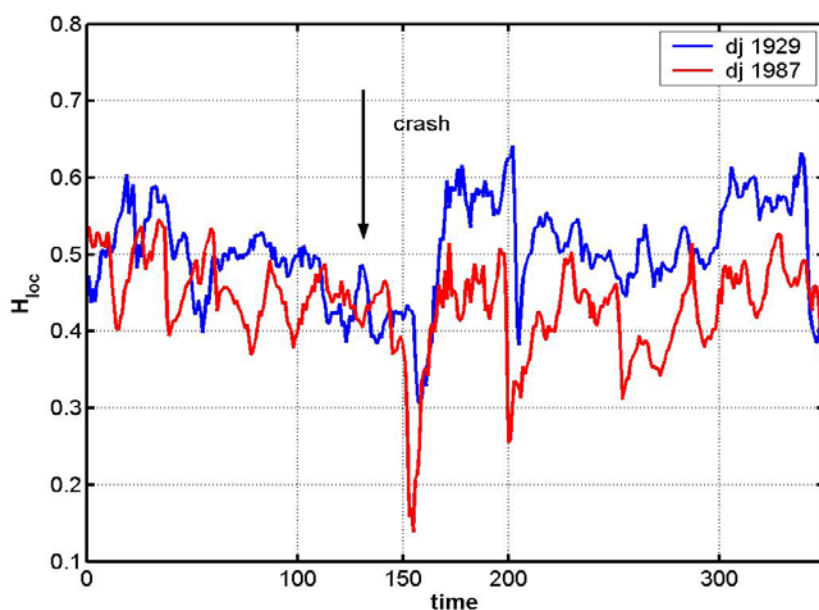


Рис. 5. Динаміка коефіцієнта Херста для часових рядів, які містять фінансову кризу

Зрозуміло, що з плином часу динаміка складної системи може змінюватись, що може

призводити до зміни коефіцієнта Херста. Нами реалізована методика розрахунку  $H(t)$  за допомогою алгоритму рухомого вікна. При цьому для часового вікна фіксованого розміру розраховується коефіцієнт Херста. Потім вікно зміщується з певним кроком і процедура розрахунків продовжується, поки не вичерпується весь часовий ряд.

Вказані розрахунки проводяться за допомогою опції *Анализ ряда* → *Локальный DFA*. У якості часових рядів візьмемо значення індексу Доу Джонса у період криз 1929 і 1987 років. Ряди побудовані так, що час кризи припадає на середину ряду. Якщо їх загальна довжина 2000 точок, то точка кризи відповідає тисячній позначці. При значеннях параметрів **Ширина окна=250, Шаг окна=5** результати розрахунків представлені рисунком 5.

Мультифрактальний АДФ дозволяє визначити наявність мультискейлінга у ряді, що (у випадку існування) свідчить про використання для формування значень ряду композиції різних механізмів утворення цих значень.

Вибравши ряд нормалізованих прибутковостей, завантажимо програму, що дозволяє аналізувати значення ряду, виконавши команду *work3*. У цьому випадку відкриється вікно «*Multifractal DFA research*», у якому вибираються опції для проведення аналізу.

Прочитаємо ряд, активізувавши пункт меню *Файл* → *Читаеть* (рис. 6).

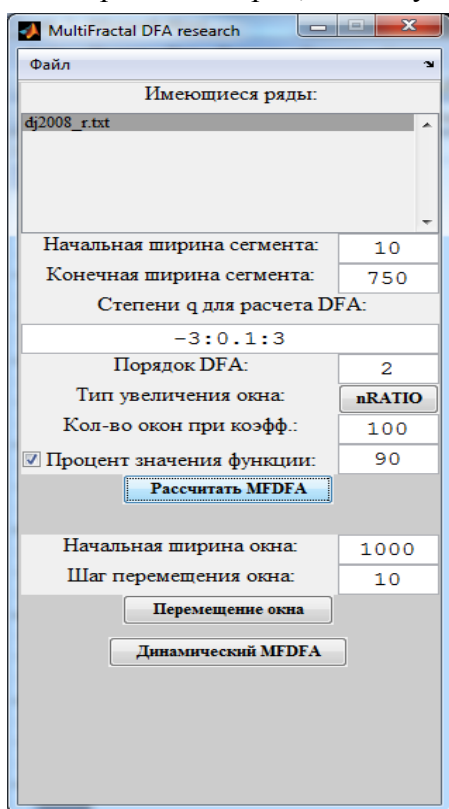


Рис. 6. Вікно налаштування опцій для мультифрактального аналізу ДФ

Для виконання мультифрактального АДФ у відкритому вікні необхідно провести відповідні налаштування. У полі «*Начальная ширина окна*» виставляється найменша ширина сегментів  $s$  (див. крок 2), на які розбивається вихідний ряд для проведення подальшого детрендування, а у полі «*Конечная ширина окна*» – найбільша ширина сегментів. У полі «*Степени q для расчета DFA*» вводиться найменше та найбільше значення степеня  $q$  та крок зміни значень з використанням синтаксису MATLAB для побудови значень ряду. Наприклад, рядок « $-10:1:10$ » означатиме послідовність цілих чисел від  $-10$  до  $10$  із кроком 1. Якщо необхідно вводити дві різні послідовності – їх параметри записуються через пробіл, наприклад: « $-10:0.5:-2$  2:0.5:10». Рекомендована послідовність степенів  $q = -3, -2.9, -2.8, \dots, 3$  (« $-3:0.1:3$ »). Порядок DFA вибирається у відповідному полі (рекомендується аналізувати за допомогою DFA 2-го порядку).

Шляхом натиснення на кнопку справа від напису «*Тип увеличения окна*» вибирається метод, за допомогою якого будуть обиратись можливі значення ширини сегменту  $s$ :

– якщо на кнопці написано «*nRATIO*», то у полі нижче – «*Кол-во окон при коэфф.*» –

встановлюється кількість різних можливих значень сегментів  $s$  (крок 2), які вибиратимуться для проведення аналізу, причому буде розраховано коефіцієнт  $k$ , такий що  $s_{i+1} = ks_i$ , тобто ширина сегментів наступного разу у  $k$  разів більша за ширину сегментів попереднього;

– якщо на кнопці написано «*STEP*», то у полі нижче – «*Шаг увеличения окна*» – встановлюється значення, на яке буде більшою ширина сегментів при наступному видаленні трендів;

– якщо на кнопці написано «*RATIO*», то у полі нижче – «*Коефф. увеличения окна*» – встановлюється значення, на яке необхідно помножити  $s$  для отримання ширини наступних сегментів для подальшого видалення трендів;

– кнопка «*nSTEP*» аналогічна дії кнопки «*nRATIO*» за тією відмінністю, що в даному випадку крок обирається програмою таким чином, щоб між шириною найменшого сегмента та найбільшого можна було утворити ще  $n - 2$  різних значень ширини (якщо це можливо), причому  $n$  береться із поля «*Кол-во окон при шаге*».

Для розрахунку спектру мультифрактальності необхідно натиснути кнопку «*Рассчитать MF DFA*».

По закінченні розрахунку будується кілька вікон (рис. 7):

– у вікні «*Распределение  $Fq(s)$* » (рис. 7а) відображується розподіл залежності  $F_q$  від ширини сегменту  $s$  (крок 4, (8));

– у вікні «*Распределение  $h(q)$* » (рис. 7б) відображується розподіл залежності  $h(q)$  (крок 5, (9));

– у вікні «*Распределение  $\tau(q)$* » (рис. 7в) відображується розподіл залежності  $\tau(q)$  згідно з (12);

– у вікні «*Распределение  $F(\alpha)$ ,  $\alpha$  рассчитана через  $\tau$* » (рис. 7г) відображується спектр мультифрактальності, отриманий за допомогою перетворення Лежандра згідно з (14);

– у вікні «*Распределение  $F(\alpha)$ ,  $\alpha$  рассчитана через  $h$* » відображується спектр мультифрактальності, отриманий за допомогою перетворення згідно з (15); зауважимо, що цей спектр є менш точним, ніж спектр, отриманий у попередньому пункті і на рисунках не відображений.

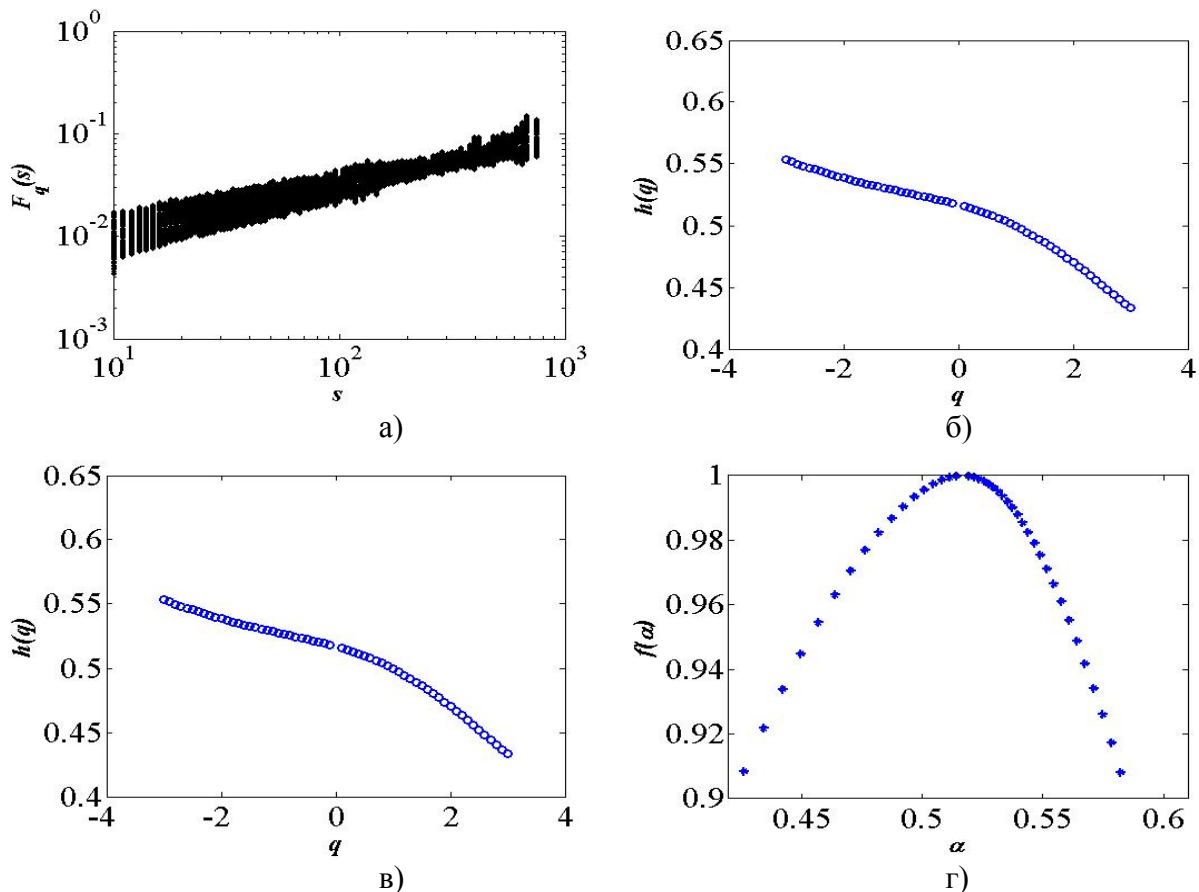


Рис. 7. Результати роботи програми work3

Результати свідчать про мультифрактальність ряду.

Для проведення динамічного MF-DFA необхідно провести додаткові налаштування: вказати ширину вікна, що переміщається, причому можливо два варіанти, які міняються натисненням кнопки «*Перемещение окна/Увеличение окна*»:

— кнопка «*Перемещение окна*»; у полі «*Начальная ширина окна*» виставляється значення ширини вікна, яке буде переміщуватись (воно повинне бути не меншим, ніж значення у полі «*Конечная ширина сегмента*», та не більшим за довжину ряду), а у полі «*Шаг перемещения окна*» – крок, на який буде зміщуватись кожне наступне вікно;

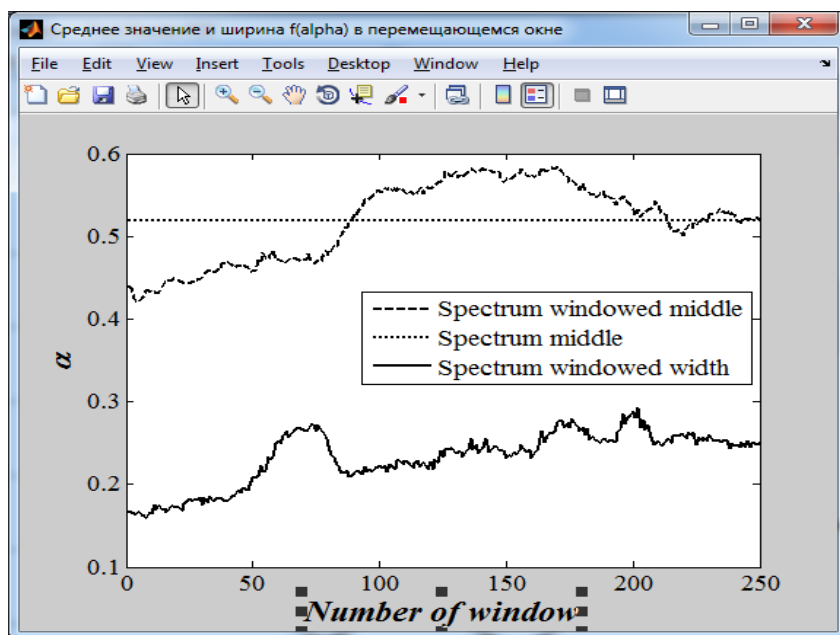


Рис. 8. Зміна ширини спектру

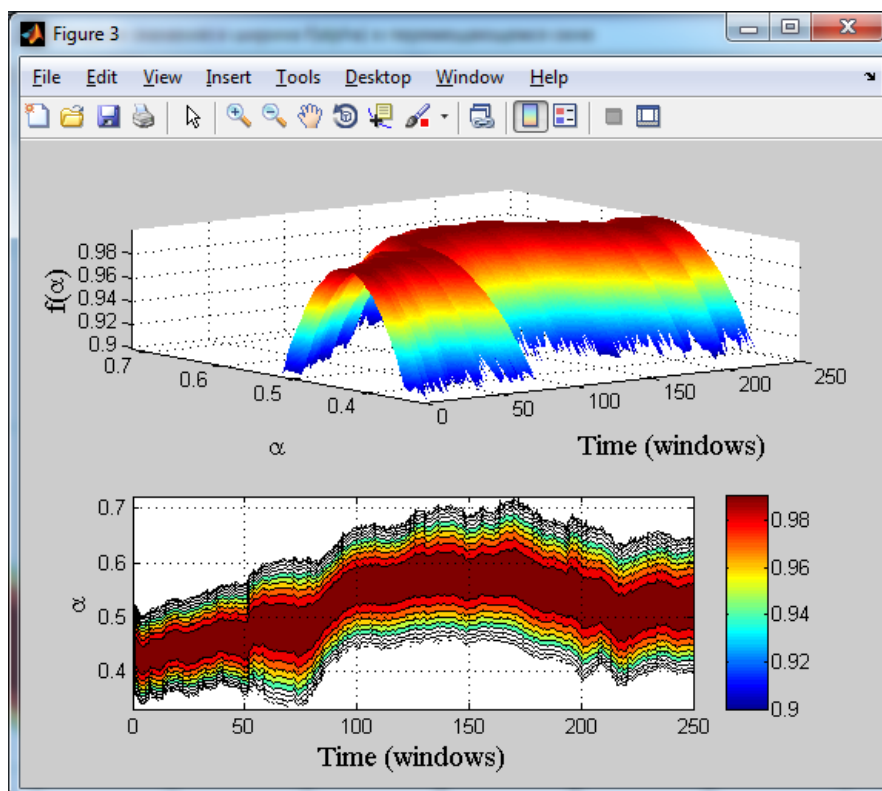


Рис. 9. Зміна спектру мультифрактальності у часі

— кнопка «*Увеличение окна*»; у полі «*Начальная ширина окна*» виставляється значення ширини вікна, яке буде переміщуватись (воно повинне бути не меншим, ніж значення у полі «*Конечная ширина сегмента*», та не більшим за довжину ряду), а у полі «*Шаг перемещения окна*» – крок, на який буде збільшуватись ширина кожного наступного



вікна.

Після натиснення на кнопку «*Динамический MF DFA*» буде проведено розрахунок та з'явиться два вікна:

– вікно, у якому відображується зміна середнього значення ширини спектру (spectrum windowed middle), усереднене значення ширини спектру (spectrum middle) та зміна ширини спектру (spectrum windowed width) у вікнах, що переміщуються (рис. 8);

– вікно із графічним поданням зміни отриманих спектрів у часі (рис. 9).

Програма дає можливість зберегти у вигляді текстових файлів значення екстремуму спектру сингулярності  $\alpha_{\max}$  і його ширину. Співставимо ці величини з вихідним часовим рядом (рис. 10).

Якщо показник  $\alpha_{\max}$  є малоінформативним, то ширина спектру мультифрактальності  $\Delta\alpha$  помітно зростає у період кризи і може слугувати індикатором таких небажаних явищ.

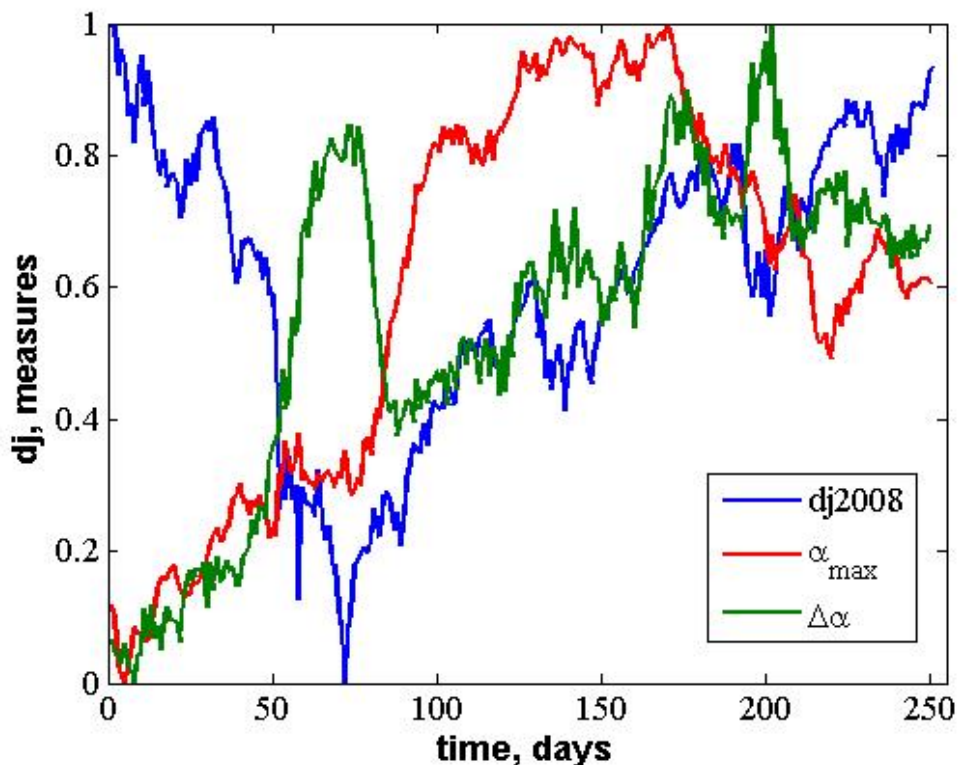


Рис. 10. Порівняння вихідного ряду з мірами мультифрактальності

### Висновок

Таким чином, для одного часового ряду можна отримати, використовуючи різні методи, сукупність показників, що характеризують досліджуваний об'єкт з точки зору впливу зовнішніх подій на систему, яка описується цим рядом.

### Завдання для виконання

1. Вибрати із запропонованої бази даних варіант завдання
2. Виконати дослідження фрактальних характеристик заданих часових рядів. Зберегти результати в окремому файлі
3. Порівняти значення коефіцієнтів Херста, одержаних методом *R/S*-аналізу та АДФ
4. Провести повний аналіз ряду за допомогою методу мультифрактального АДФ.
5. Дати інтерпретацію отриманим результатам

### Контрольні запитання

1. В чому переваги методу детрендованих флуктуацій перед методом *R/S*-аналізу?

2. На що вказує мультифрактальність часового ряду?
3. Яким чином поведуть себе різні характеристики, якщо ряд містить кризу?

### ЛАБОРАТОРНА РОБОТА № 3

**Тема.** Використання рекурентного аналізу для моделювання і прогнозування нелінійних динамічних властивостей складних систем.

**Мета:** Навчитися інструментарію нелінійної динаміки, який відноситься до рекурентних властивостей нестационарних динамічних рядів.

1. Фазовий простір, його реконструкція.
2. Рекурентний аналіз.
3. Рекурентні діаграми.

#### Теоретичні відомості

Дослідження складних систем, як природних, так і штучних, показали, що в їх основі лежать нелінійні процеси, ретельне вивчення яких необхідне для розуміння і моделювання складних систем. У останні десятиліття набір традиційних (лінійних) методик дослідження був істотно розширений нелійними методами, одержаними з теорії нелінійної динаміки і хаосу; багато досліджень були присвячені оцінці нелінійних характеристик і властивостей процесів, що протікають в природі (скейлінг, фрактальна розмірність). Проте більшість методів нелінійного аналізу вимагає або достатньо довгих, або стаціонарних рядів даних, які досить важко одержати з природи. Більш того, було показано, що дані методи дають задовільні результати для моделей реальних систем, що ідеалізуються. Ці чинники вимагали розробки нових методик нелінійного аналізу даних.

Стан природних або штучних систем, як правило, змінюється в часі. Вивчення цих, часто складних, процесів – важлива задача в багатьох дисциплінах, дозволяє зрозуміти і описати їх суть, наприклад, для прогнозування стану на деякий час в майбутнє. Метою таких досліджень є знаходження математичних моделей, які б достатньо відповідали реальним процесам і могли б бути використані для розв'язання поставлених задач.

Розглянемо ідею і коротко теорію рекурентного аналізу, наведемо деякі приклади, розглянемо його можливі області застосування при аналізі і прогнозування складних фінансово-економічних систем.

#### Фазовий простір, його реконструкція

Стан системи описується її змінними стану

$$x^1(t), x^2(t), \dots, x^d(t),$$

де верхній індекс – номер змінної. Набір із  $d$  змінних стану у момент часу  $t$  складає вектор стану  $\vec{x}(t)$  в  $d$ -вимірному фазовому просторі. Даний вектор переміщається в часі в напрямі, визначуваному його вектором швидкості:

$$\dot{\vec{x}}(t) = \partial_i \vec{x}(t) = \vec{F}(x). \quad (1)$$

Послідовність векторів  $\vec{x}(t)$  утворює траєкторію у фазовому просторі, причому поле швидкості  $\vec{F}$  дотичне до цієї траєкторії. Еволюція траєкторії описує динаміку системи і її атрактор. Знаючи  $\vec{F}$ , можна одержати інформацію про стан системи в момент  $t$  шляхом



інтегрування виразу (1). Оскільки форма траєкторії дозволяє судити про характер процесу (періодичні або хаотичні процеси мають характерні фазові портрети), то для визначення стану системи не обов'язково проводити інтегрування, достатньо побудувати графічне відображення траєкторії.

При дослідженні складних систем часто немає інформації про всі змінні стану, або не все з них можливо виміряти. Як правило, є єдине спостереження, проведене через дискретний часовий інтервал  $\Delta t$ . Таким чином, вимірювання записуються у вигляді ряду  $u_i(t)$ , де  $t = i \cdot \Delta t$ . Інтервал  $\Delta t$  може бути постійним, проте це не завжди можливо і створює проблеми для застосування стандартних методів аналізу даних, що вимагають рівномірної шкали спостережень.

Взаємодії і їх кількість в складних системах такі, що навіть по одній змінній стану можна судити про динаміку всієї системи в цілому (даний факт був встановлений групою американських учених при вивченні турбулентності). Таким чином, еквівалентна фазова траєкторія, що зберігає структури оригінальної фазової траєкторії, може бути відновлена з одного спостереження або часового ряду за теоремою Такенса (Takens) методом часових затримок:

$$\hat{x}(t) = (u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau}),$$

де  $m$  – розмірність вкладення,  $\tau$  – часова затримка (реальна часова затримка визначається як  $\tau \cdot \Delta t$ ). Топологічні структури відновленої траєкторії зберігаються, якщо  $m \geq 2 \cdot d + 1$ , де  $d$  – розмірність атратора. На практиці більшості випадків атратор може бути відновлений і при  $m \leq 2d$  (див. рис. 1). Затримка, як правило, вибирається апріорно.

Існує кілька підходів до вибору мінімально достатньої розмірності  $m$ , крім аналітичного. Високу ефективність показали методи, засновані на концепції фальшивих найближчих точок (false nearest neighbours, FNN). Суть її заключається у тому, що при зменшенні розмірності вкладення відбувається збільшення кількості фальшивих точок, що потрапляють в околицю будь-якої точки фазового простору. Звідси витікає простий метод – визначення кількості FNN як функції від розмірності. Існують і інші методи, засновані на цій концепції – наприклад, визначення відносин відстаней між одними і тими ж сусідніми точками при різних  $m$ . Розмірність атратора також може бути визначена за допомогою крос-кореляційних сум.

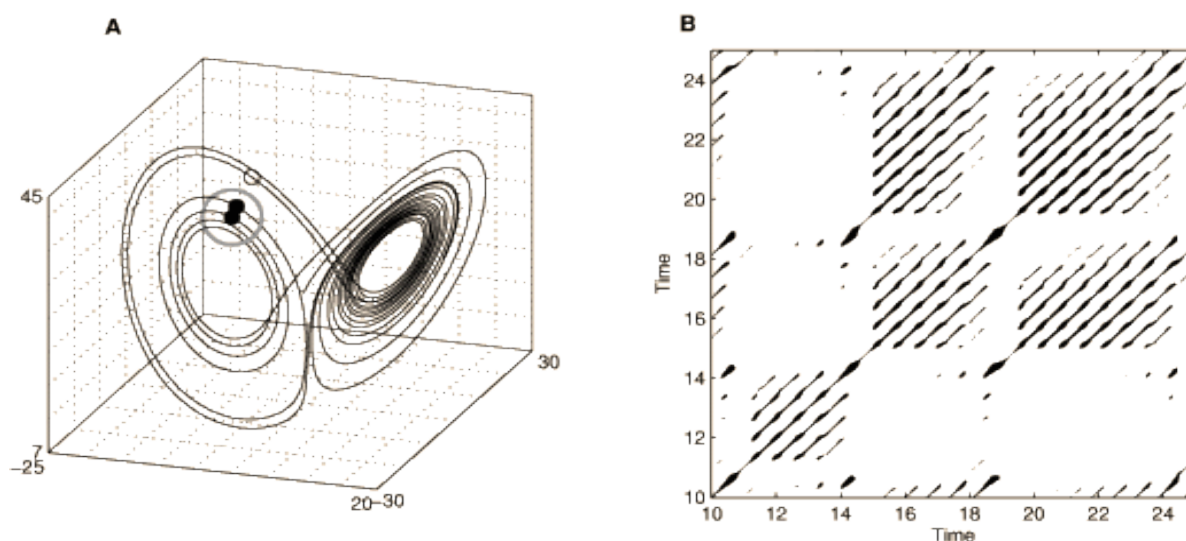


Рис. 1. Реконструйований атратор системи Лоренца у тривимірному фазовому просторі і його рекурентна карта

## Рекурентний аналіз

Процесам в природі властива яскраво виражена рекурентна поведінка, така, як періодичність або іррегулярна циклічність. Більш того, рекурентність (повторюваність) станів в значенні проходження подальшої траєкторії достатньо близько до попередньої є фундаментальною властивістю дисипативних динамічних систем. Ця властивість була відмічена ще в 80-х роках XIX століття французьким математиком Пуанкаре (Poincaré) і згодом сформульовано у вигляді «теореми рекурентності», опублікованої в 1890 р.:

***Якщо система зводить свою динаміку до обмеженої підмножини фазового простору, то система майже напевно, тобто з вірогідністю, практично рівною 1, скільки завгодно близько повертається до якого-небудь спочатку заданого режиму.***

Суть цієї фундаментальної властивості у тому, що, не дивлячись на те, що навіть саме мале збурення в складній динамічній системі може привести систему до експоненціального відхилення від її стану, через деякий час система прагне повернутися до стану, деяким чином близького до попереднього, і проходить при цьому подібні етапи еволюції.

Переконалися в цьому можна за допомогою графічного зображення траєкторії системи у фазовому просторі. Проте можливості такого аналізу сильно обмежені. Як правило, розмірність фазового простору складної динамічної системи більша трьох, що робить практично незручним його розгляд напряду; єдина можливість – проєкції в дво- і тривимірні простори, що часто не дає вірного уявлення про фазовий портрет.

У 1987 р. Екман (Eckmann) і співавтори запропонували спосіб відображення  $m$ -вимірної фазової траєкторії станів системи  $\vec{x}(t)$  завдовжки  $N$  на двовимірну квадратну двійкову матрицю розміром  $N \times N$ , в якій 1 (чорна точка) відповідає повторенню стану при деякому часі  $i$  в деякий інший час  $j$ , а обидві координатні осі є осями часу. Таке представлення було назване рекурентною картою або діаграмою (recurrence plot, RP), оскільки воно фіксує інформацію про рекурентну поведінку системи.

Математично вищесказане описується як

$$R_{i,j}^{m,\varepsilon_i} = \Theta(\varepsilon_i - \|\vec{x}_i - \vec{x}_j\|), \vec{x} \in \mathfrak{R}^m, i, j = 1, \dots, N, \quad (2)$$

де  $N$  – кількість даних станів,  $x_i, \varepsilon_i$  – розмір околиці точки  $\vec{x}$  у момент  $i$ ,  $\|\cdot\|$  – норма і  $\Theta(\cdot)$  – функція Хевісайда.

Непрактично і, як правило, неможливо знайти повну рекурентність у значенні  $\vec{x}_i \equiv \vec{x}_j$  (стан динамічної, а особливо – хаотичної системи не повторюється повністю еквівалентно початковому стану, а підходить до нього скільки завгодно близько). Таким чином, рекурентність визначається як достатня близькість стану  $\vec{x}_j$  до стану  $\vec{x}_i$ . Іншими словами, рекурентними є стани  $\vec{x}_j$ , які потрапляють в  $m$ -вимірну околицю з радіусом  $\varepsilon_i$  і центром в  $\vec{x}_i$ . Ці точки  $\vec{x}_j$  називаються **рекурентними точками** (recurrence points) (див. рис. 2).

Оскільки  $R_{i,i} = 1$  ( $i = 1, \dots, N$ ) за визначенням, то рекурентна діаграма завжди містить чорну діагональну лінію – лінію ідентичності (line of identity, LOI) під кутом  $\pi/4$  до осей координат. Довільно узята рекурентна точка не несе якої-небудь корисної інформації про стани в часи  $i$  і  $j$ . Тільки вся сукупність рекурентних точок дозволяє відновити властивості системи.

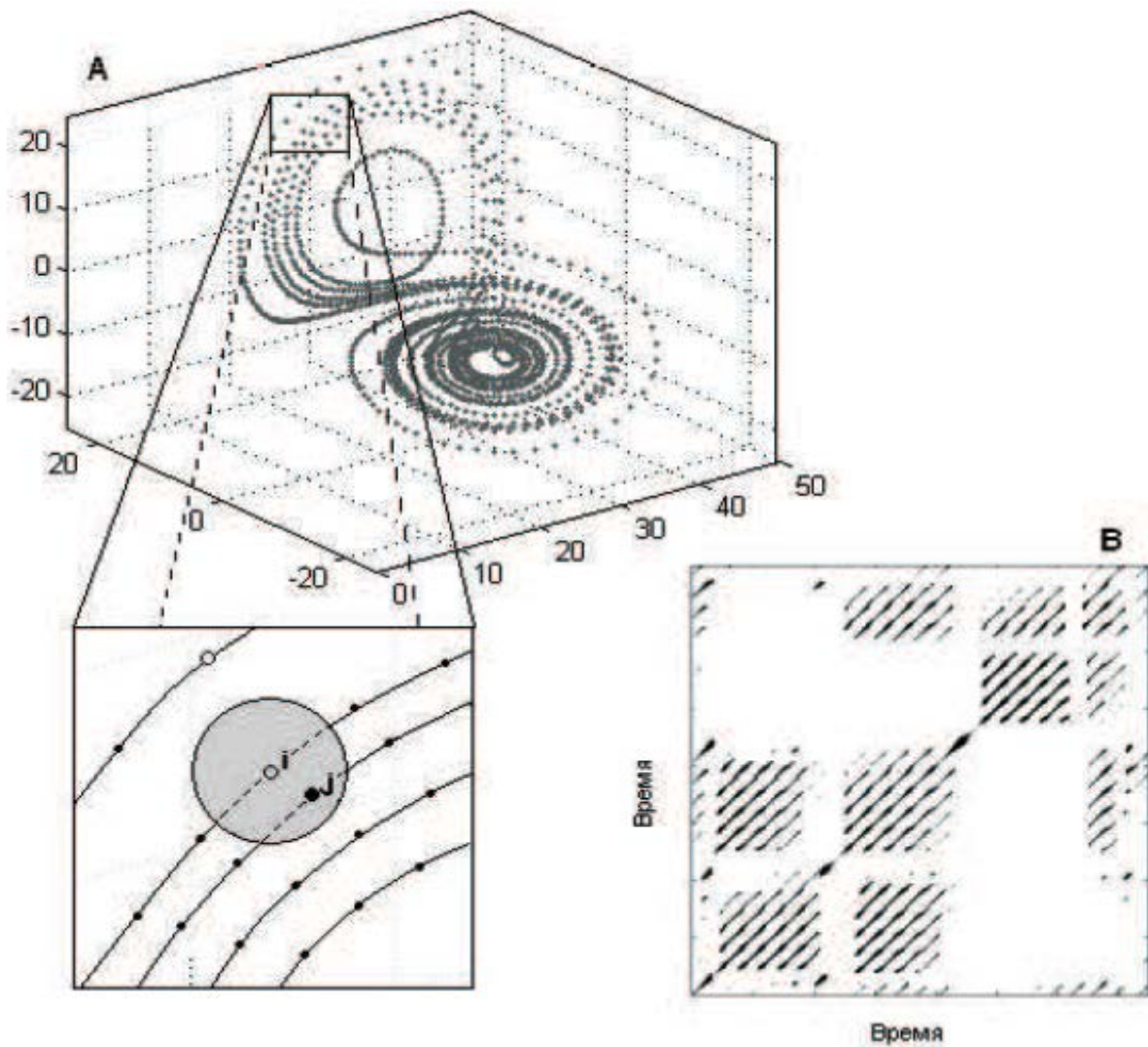
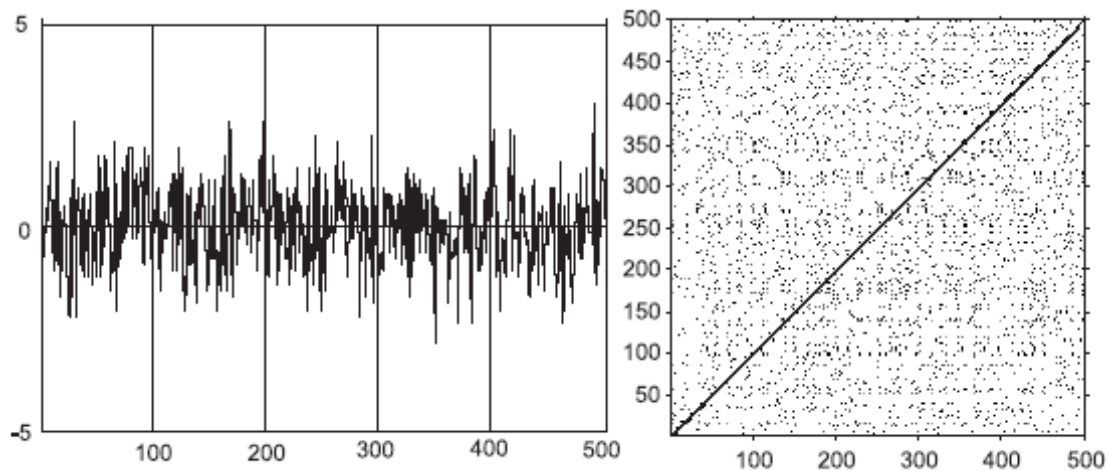
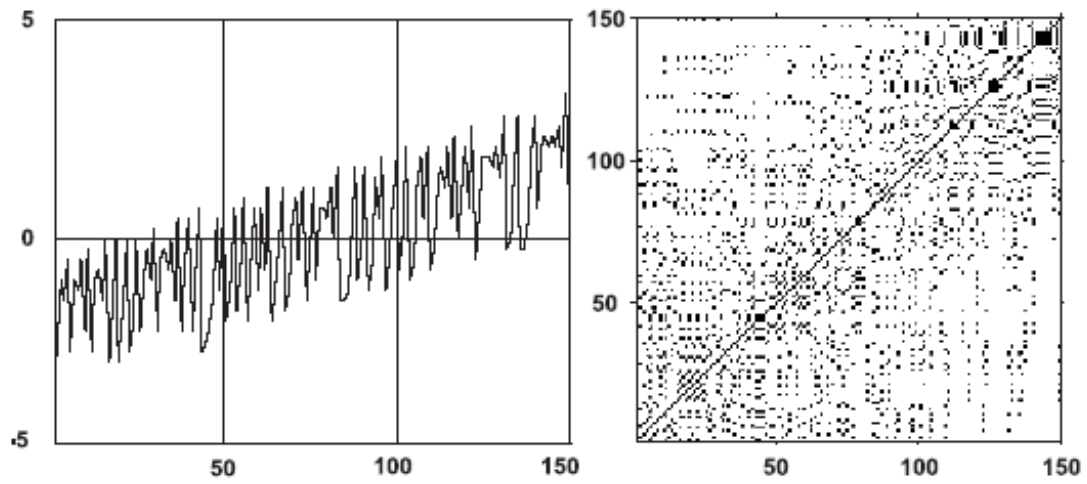


Рис. 2. Атрактор Лоренця, вибір рекурентних точок  $i$  та  $j$  в епсилон-околі (на вставці виділено) та рекурентна карта у координатах «час-час»

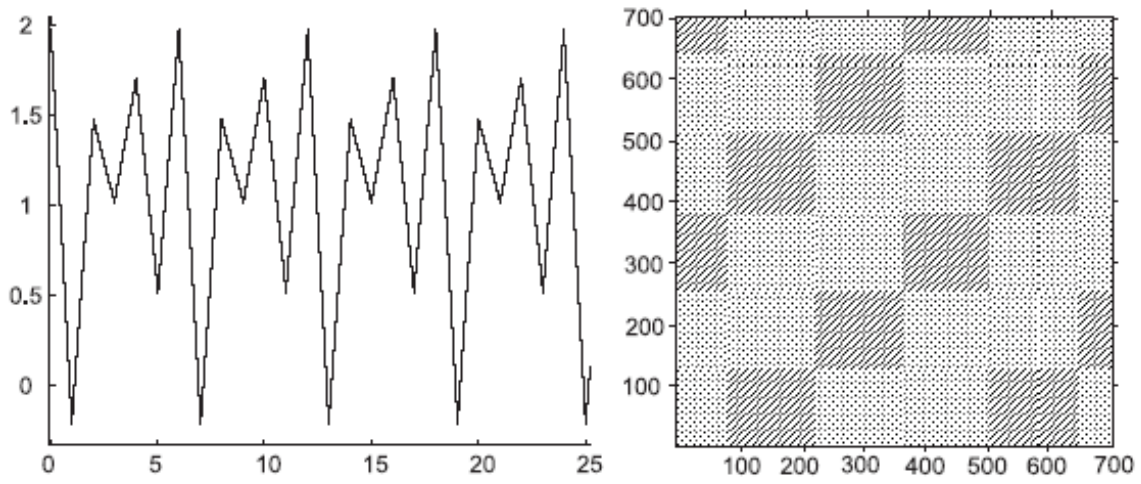
Зовнішній вигляд рекурентної діаграми дозволяє судити про характер процесів, які протікають в системі, наявності і впливі шуму, станів повторення і завмирання (ламінарності), здійсненні в ході еволюції системи різких змін стану (екстремальних подій). На рис. 3 наведені деякі з типових процесів і їх рекурентні карти.



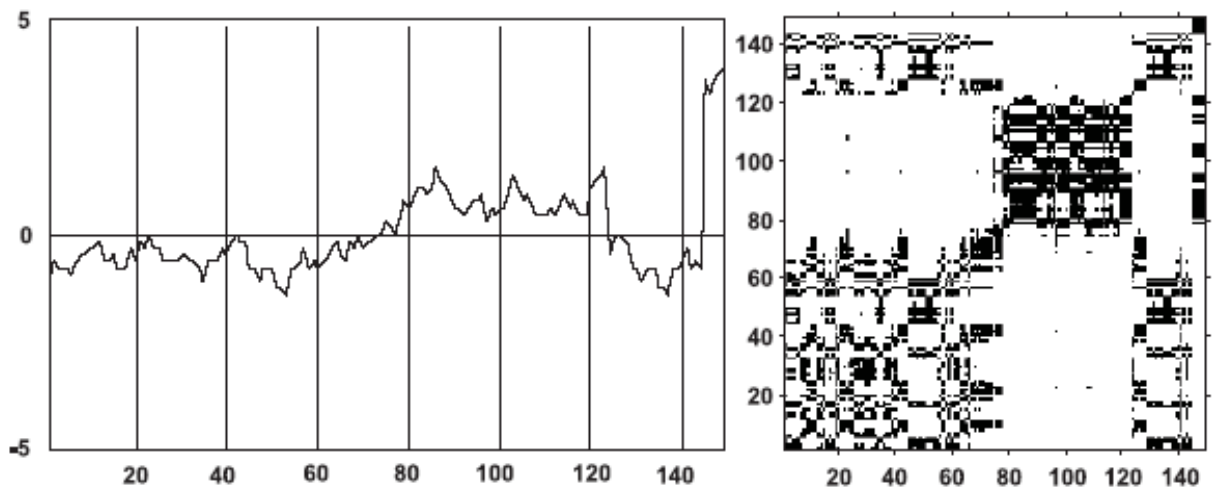
1. Однорідна топологія



2. Дрейф

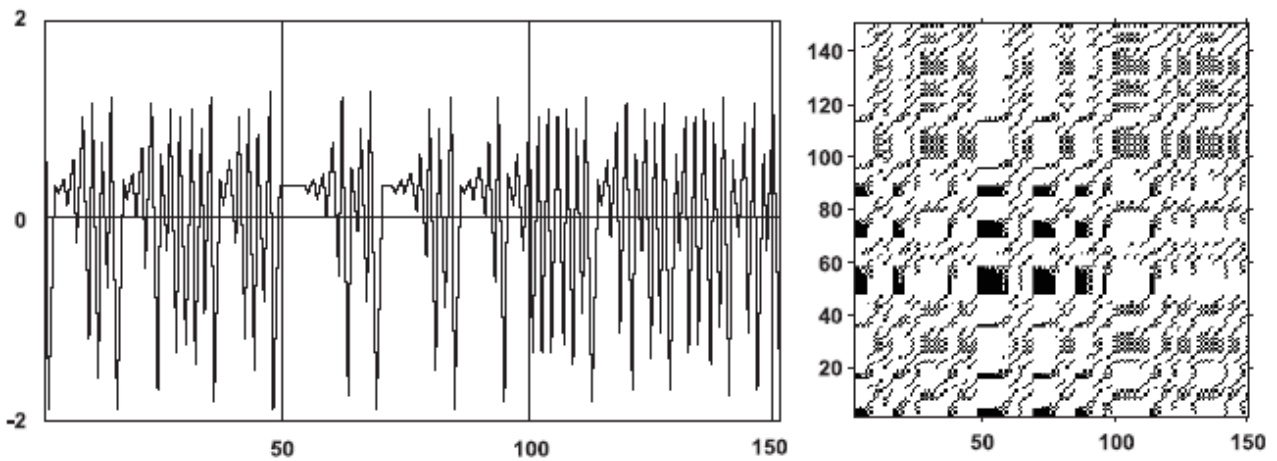


3. Осцилююча поведінка системи



4. Контрастна топологія:





### 5. Ламінарність процесу

Рис. 3. Типові динамічні ряди і їх рекурентні карти

### Аналіз діаграм

З рис. 3 очевидно, що процеси різної поведінки даватимуть рекурентні діаграми з різним рисунком. Таким чином, візуальна оцінка діаграм може дати уявлення про еволюцію досліджуваної траєкторії. Виділяють два основних класи структури зображення: **топология** (*typology*), що представляється крупномасштабними структурами, і **текстура** (*texture*), що формується дрібномасштабними структурами.

Топология дає загальне уявлення про характер процесу. Виділяють чотири основні класи (див. рис. 4):

- **однорідні** рекурентні діаграми типові для стаціонарних і автономних систем, в яких час релаксації малий у порівнянні з довжиною ряду;
- **періодичні** структури, що повторюються (діагональні лінії, узор у шаховому порядку) відповідають різним осцилюючим системам з періодичністю в динаміці;
- **дрейф** відповідає системам з параметрами, що поволі змінюються, що робить білими лівий верхній і правий нижній кути рекурентної діаграми;
- **різкі зміни** в динаміці системи, рівно як і екстремальні ситуації, обумовлюють появу білих областей або смуг.

Рекурентні діаграми **спрощують** виявлення екстремальних і рідкісних подій.

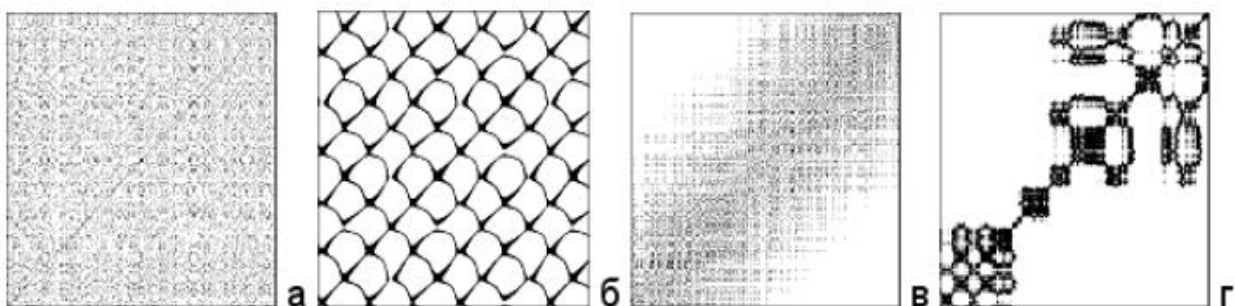


Рис. 4. Характерні топології рекурентних діаграм: а – однорідна (нормально розподілений шум); б – періодична (генератор Ван дер Поля); в – дрейф (відображення Ікеди з накладеною послідовністю, що лінійно росте); г – контрастні області або смуги (узагальнений броунівський рух)

Докладний розгляд рекурентних діаграм дозволяє виявити дрібномасштабні структури – текстуру, яка складається з простих точок, діагональних, горизонтальних і вертикальних ліній. Комбінації вертикальних і горизонтальних ліній формують прямокутні кластери точок.

– **самотні**, окремо розташовані рекурентні точки з'являються в тому разі, коли відповідні стани рідкісні, або нестійкі в часі, або викликані сильною флуктуацією. При цьому вони не є ознаками випадковості або шуму;

– діагональні лінії  $R_{i+k,j+k} = 1$  (при  $k = 1 \dots l$  де  $l$  – довжина діагональної лінії) з'являються у разі, коли сегмент траєкторії у фазовому просторі пролягає паралельно іншому сегменту, тобто траєкторія повторює саму себе, повертаючись в одну і ту ж область фазового простору у різний час.

Довжина таких ліній визначається часом, протягом якого сегменти траєкторії залишаються паралельними; напрям (кут нахилу) ліній характеризує внутрішній час підпроцесів, відповідних даним сегментам траєкторії. Проходження ліній паралельно лінії ідентичності (під кутом  $\pi/4$  до осей координат) свідчить про однаковий напрям сегментів траєкторії, перпендикулярно – про протилежний («відображені» сегменти), що може також бути ознакою реконструкції фазового простору з невідповідною розмірністю вкладення. Нерегулярна поява діагональних ліній є ознакою хаотичного процесу;

– **вертикальні (горизонтальні) лінії**  $R_{i,j+k} = 1$  (при  $k = 1 \dots v$ , де  $v$  – довжина вертикальної або горизонтальної лінії) виділяють проміжки часу, в котрі стан системи не змінюється або змінюється трохи (система як би «заморожена» на цей час), що є ознакою «ламінарих» станів.

## Хід роботи

У даній роботі ми будемо використовувати найбільш розповсюджений і функціональний програмний засіб рекурентного аналізу – **(крос-) рекурентні діаграми** (Cross Recurrence Plots – CRP). З ним можна ознайомитись, прочитавши документ **manual.pdf (crp\_man.pdf)**, який містить основні функції (крос-) рекурентного аналізу.

Спочатку побудуємо дво- та тривимірні фазові портрети як для модельних часових рядів, так і реальних. При цьому вихідний часовий ряд попередньо повинен бути імпортованим у робочу область MatLab: **File/Import Data/**вибраний файл. Файл з тим же іменем з'явиться в робочій області (**Workspace**).

Для побудови фазового портрету необхідно використати функцію Matlab **phasespace**, повний синтаксис якої можна побачити, увівши у командному рядку **edit phasespace**. У загальному випадку функція викликається як **phasespace (n)**, де  $n$  – змінна, у яку записана досліджувана часова послідовність. Результатом роботи функції є траєкторія у фазовому просторі системи.

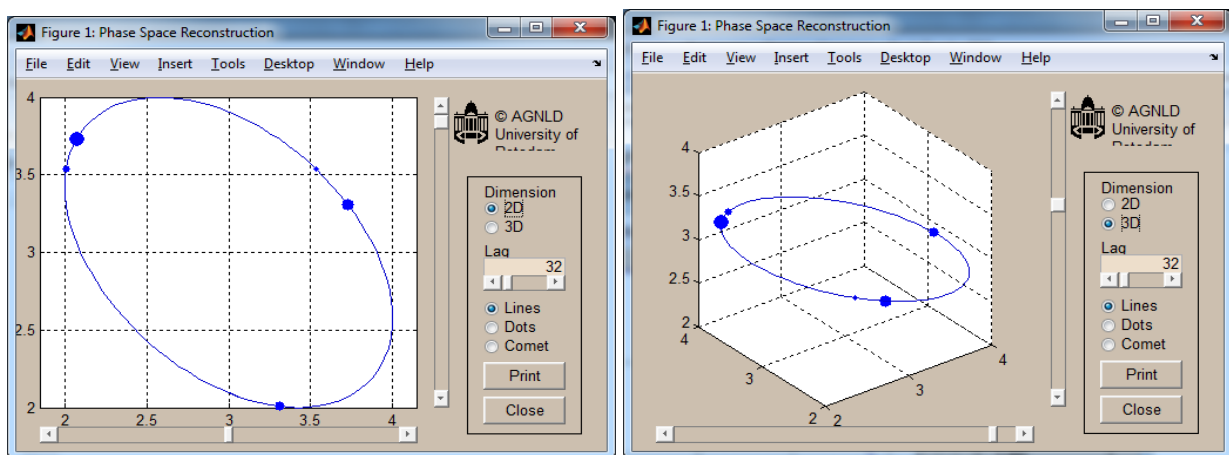


Рис. 5. Дво- та тривимірний фазові портрети періодичної синусоїди

Функція надає можливість користувачу встановити розмір лагу та обрати параметри рисунку (дво- чи тривимірне зображення, тип ліній).

Виберемо, наприклад, із папки стандартних модельних даних числовий ряд, який відповідає періодичній функції синуса і електрокардіограми (ЕКГ). Фазові портрети мають вигляд, зображений на рис. 5-6. Зверніть увагу, що вони очевидно змінюються для різних лагів (Lag).

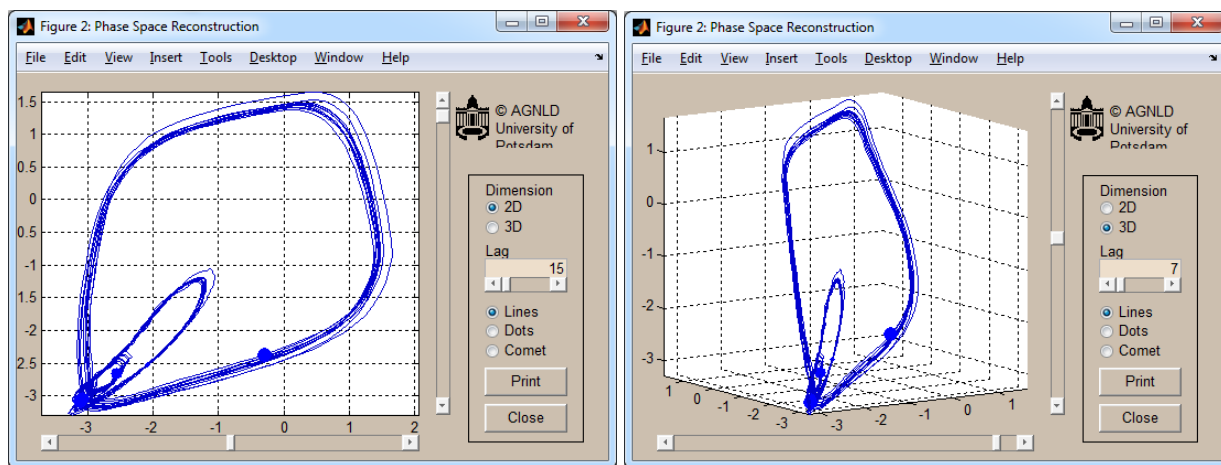


Рис. 6. Дво- та тривимірний фазові портрети сигналу ЕКГ

Для економічного часового ряду, наприклад, фрагменту індексу фондового ринку Німеччини DAX (вихідного (рис. 7) та випадковим чином перемішаного (рис. 8)) маємо відповідні фазові портрети.

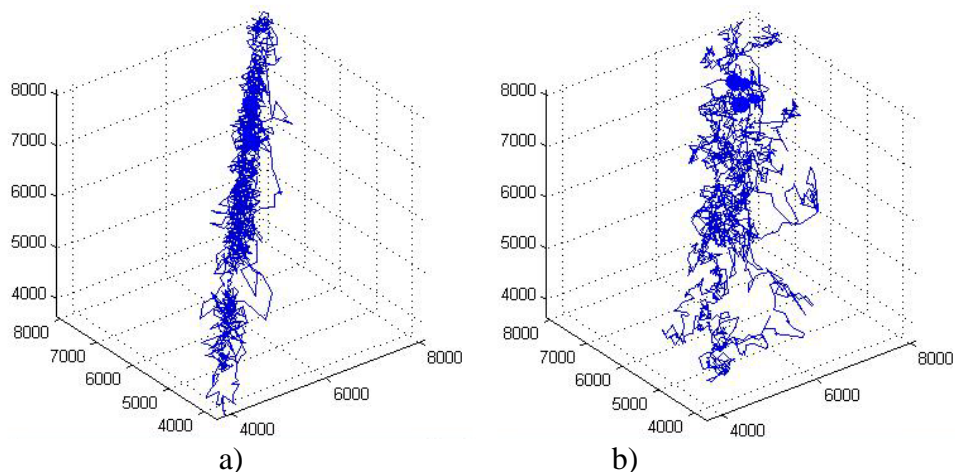


Рис. 7. Фазове відображення фондового індексу DAX для лагу 4 (a) та 40 (b)

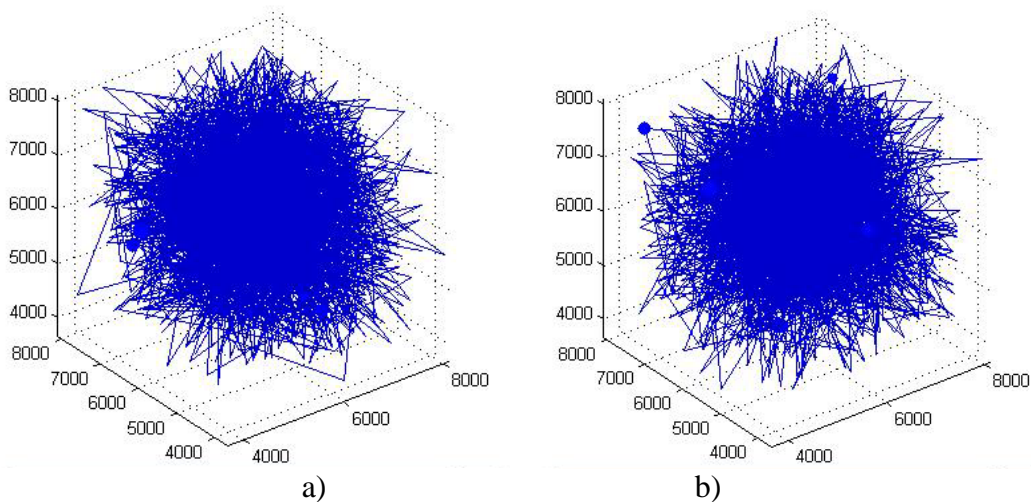


Рис. 8. Фазове відображення перемішаних значень фондового індексу DAX для лагу 4 (a) та 40 (b)

Рекурентну карту(діаграму) можна побудувати за допомогою функції **crp(a)** – **Cross Recurrence Plot**, де **a** – вихідний часовий ряд, який попередньо повинен бути імпортованим у робочу область MatLab. Наприклад, візьмемо часовий ряд індексу Доу-Джонса, який містить кризу фінансового ринку 1987 р. Для зручності довге ім'я файлу можна змінити командою **Rename**, наприклад, на ім'я «a». І тоді, набравши у вікні команд **crp(a)** та натиснувши кнопку **Apply**, одержимо рисунок, подібний зображеному на рис. 9.

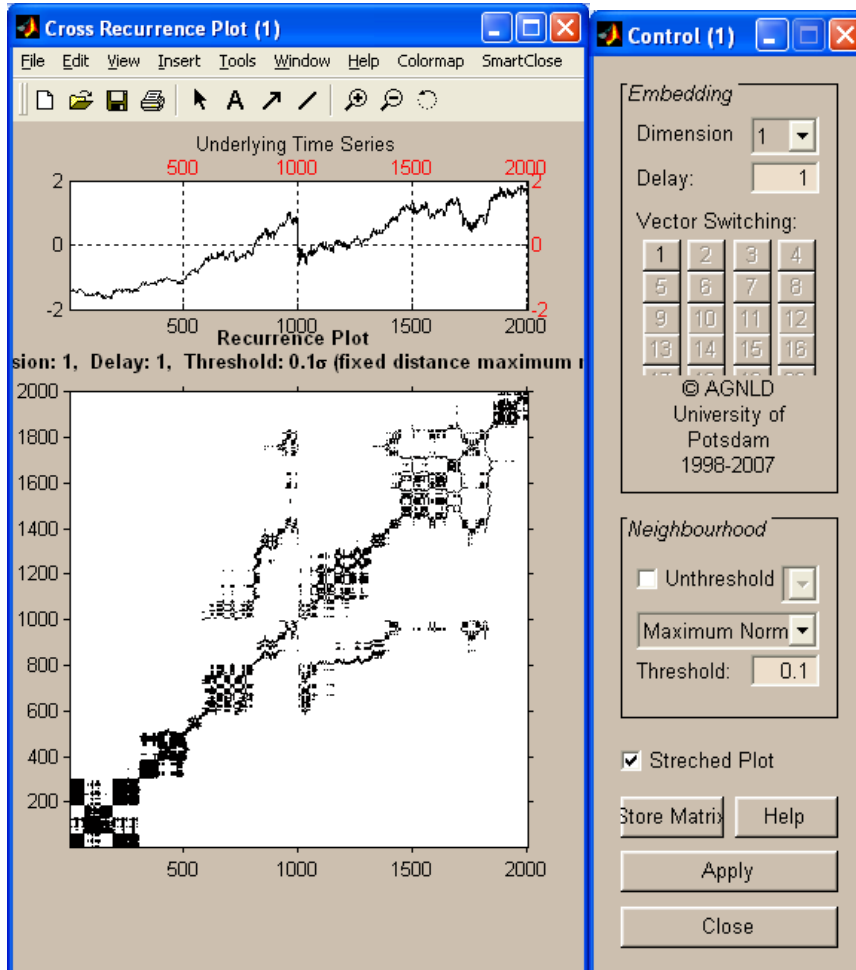


Рис. 9. Результати роботи функції **crp(a)** для часового ряду, що містить кризу

Аналогічно одержаний і рисунок 10 для модельного ряду тригонометричного синуса.

### Завдання для самостійної роботи

1. Виберіть за рекомендацією викладача варіанти часових рядів
2. Проведіть дослідження їх рекурентних властивостей згідно інструкції
3. Порівняйте фазові портрети і рекурентні діаграми сейсмічного, енцефалографічного і ряду прибутковостей. Що спільного між ними і чим вони відрізняються?
4. Зробіть загальні висновки
5. Оформіть звіт у вигляді текстового файлу



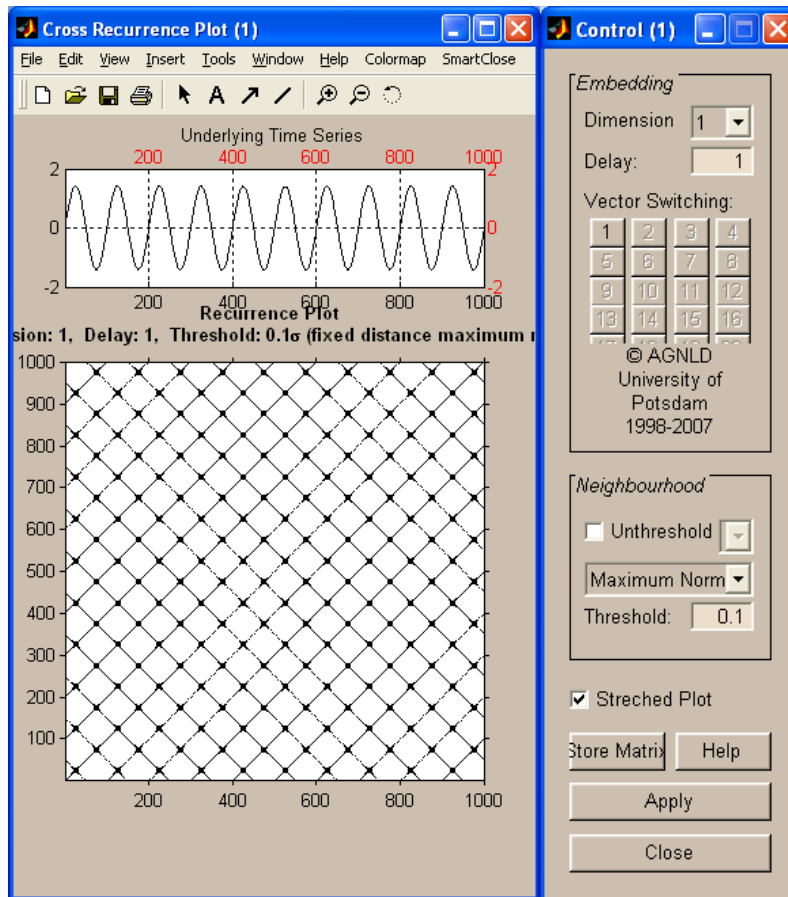


Рис. 10. Рекурентна карта тестового ряду  $\sin(x)$

## ЛАБОРАТОРНА РОБОТА № 4

**Тема.** Кількісний аналіз рекурентних діаграм.

**Мета:** Ознайомитись з кількісними оцінками рекурентних діаграм для аналізу динаміки складних систем.

1. Кількісні методи оцінки рекурентності.
2. Рекурентний аналіз.
3. Аналіз діаграм.

### Теоретичні відомості

У попередній роботі ми дослідили рекурентні властивості нестационарних часових рядів, виявили і проаналізували дрібномасштабні структури рекурентних діаграм (текстуру, яка складається з простих точок, діагональних, горизонтальних і вертикальних ліній). Комбінації вертикальних і горизонтальних ліній формують прямокутні кластери точок (див. рис. 1).

Саме ці комбінації використовуються для обчислення мір *кількісного аналізу* рекурентних діаграм (recurrence quantification analysis, RQA). Збілут (Zbilut) і Веббер (Webber) розробили інструмент обчислення ряду мір на основі підрахунку густини рекурентних точок і побудови частотного розподілу довжин діагональних ліній: **recurrence rate (RR, коефіцієнт самоподібності)**, **determinism (DET, передбачуваність)**, **divergence (DIV, зворотна величина максимальної довжини діагональної лінії)**, **entropy (ENTR, ентропія)** і **trend (TREND, тренд)**. Обчислення цих мір в підматрицях рекурентної діаграми уздовж лінії ідентичності показує поведінку цих мір в часі. Деякі дослідження даних мір

показали, що їх застосування може сприяти виявленню точок біфуркації, переходів «хаос–порядок».

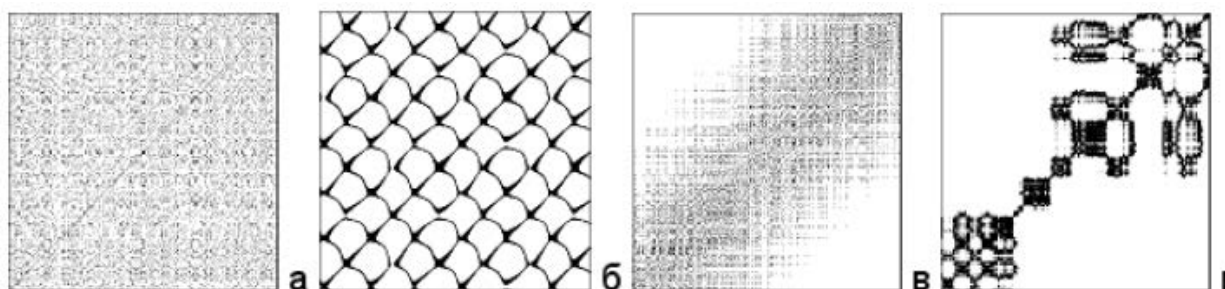


Рис. 1. Характерні топології рекурентних діаграм: а – однорідна (нормально розподілений шум); б – періодична (генератор Ван дер Поля); в – дрейф (відображення Ікеди з накладеною лінійно зростаючою послідовністю); г – контрастні області або смуги (узагальнений броунівський рух)

Пізніше були запропоновані міри, що використовують густину вертикальних (або горизонтальних) структур: **laminarity (LAM, ламінарність)**, **trapping time (TT, показник, або час затримки)**, що дозволяють виявляти переходи «хаос–хаос». Слід зазначити також нечутливість цих мір до вимог стаціонарності і довжини досліджуваних часових рядів.

Введемо кількісні характеристики введених мір.

Визначимо величини

$$R_{ij} = \begin{cases} 1, \text{ якщо } (i, j) \text{ рекурентні,} \\ 0 \text{ інакше} \end{cases}$$

$$D_{ij} = \begin{cases} 1, \text{ якщо } (i, j) \text{ та } (i+1, j+1) \text{ або } (i-1, j-1) \text{ рекурентні} \\ 0 \text{ інакше} \end{cases}$$

Тепер коефіцієнти самоподібності і передбачуваності будуть, відповідно, рівні

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{ij}, \quad (1)$$

$$DET = \frac{\sum_{i,j=1}^N D_{ij}}{\sum_{i,j=1}^N R_{ij}}. \quad (2)$$

Якщо  $N_l$  – число діагональних ліній, а  $l_i$  – довжина  $i$ -ї діагональної лінії, то довжина найдовшої діагональної лінії визначається виразом

$$L = \max(l_i; i = 1, \dots, N_l).$$

Ентропія Шеннона для функції  $f(x)$  має вигляд

$$H(x) = -\sum_x f(x) \log f(x),$$

а для розподілу

$$H(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \log(p_i).$$

Для розподілу  $p(l)$  довжини діагональних ліній ентропія Шеннона дорівнює

$$ENTR = -\sum_{l=1}^{N_l} p(l) \ln p(l). \quad (3)$$

Ознайомимся з функцією **crqa(a)**, яка і реалізує (Cross-)Recurrence Quantification Analysis часового ряду «а». Вона має різні формати (деталі дивись у файлі **crp\_man.pdf**). Для кризового ряду, попередньо імпортованого до робочої області і перейменованого на «а» з початковим значенням ширини вікна Window size=250, Window step=20 одержимо рисунок 1. У верхніх вікнах ми, відповідно, бачимо вихідний ряд (Data) і його варіацію (Variance).

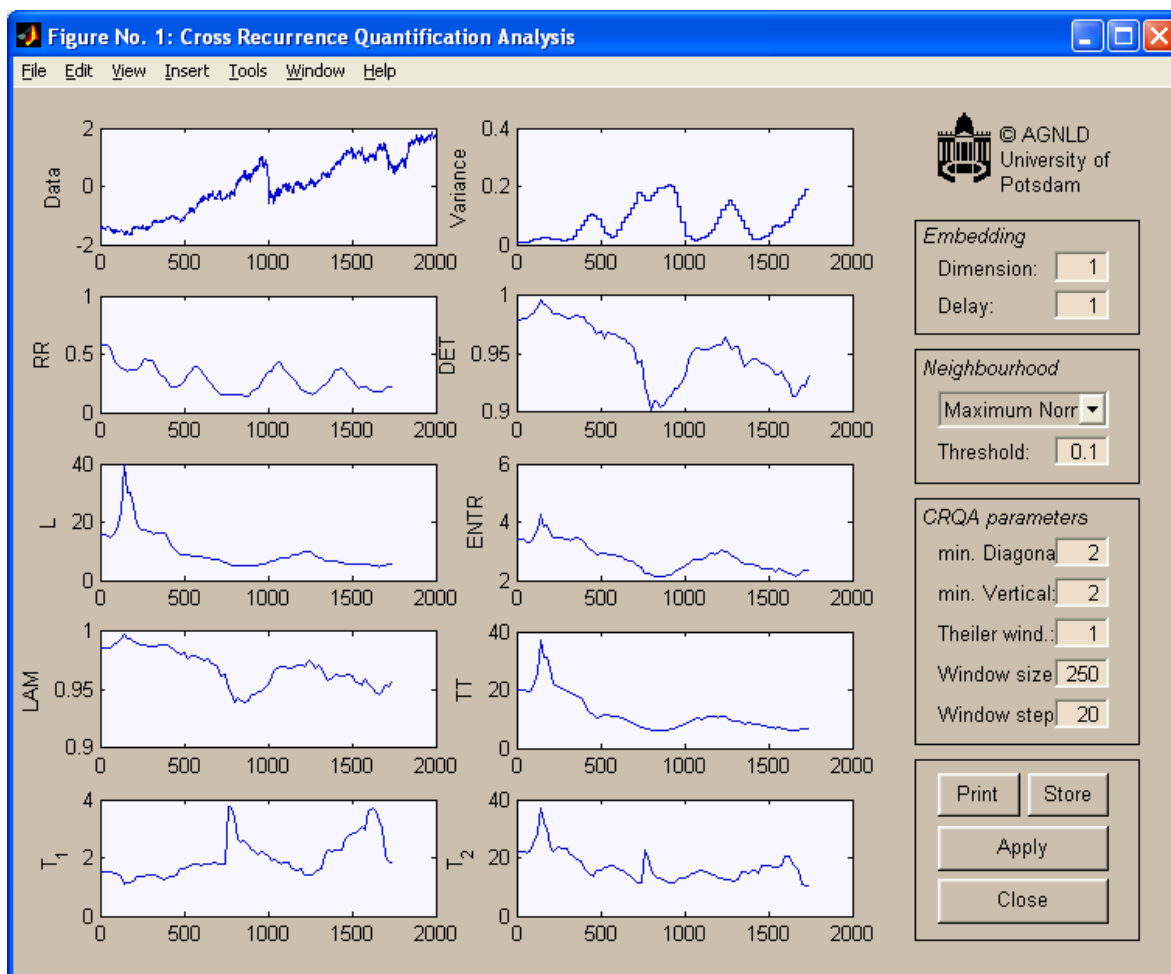


Рис. 1. Результати віконних розрахунків мір кількісного аналізу, які дає функція **crqa** для динамічного ряду з кризою

Для функції  $\sin(x)$  функція **crqa** дає рисунок 2.

На рисунках у вікнах відповідно представлені:

- Data – вихідний часовий ряд
- Variance – варіація ряду
- RR – коефіцієнт самоподібності
- DET – коефіцієнт передбачуваності
- L – середня довжина діагоналей
- ENTR – ентропія довжини діагоналей
- LAM – ламінарність
- TT – час захвату

$T_1$  – рекурентний час 1-го типу  
 $T_2$  - рекурентний час 2-го типу

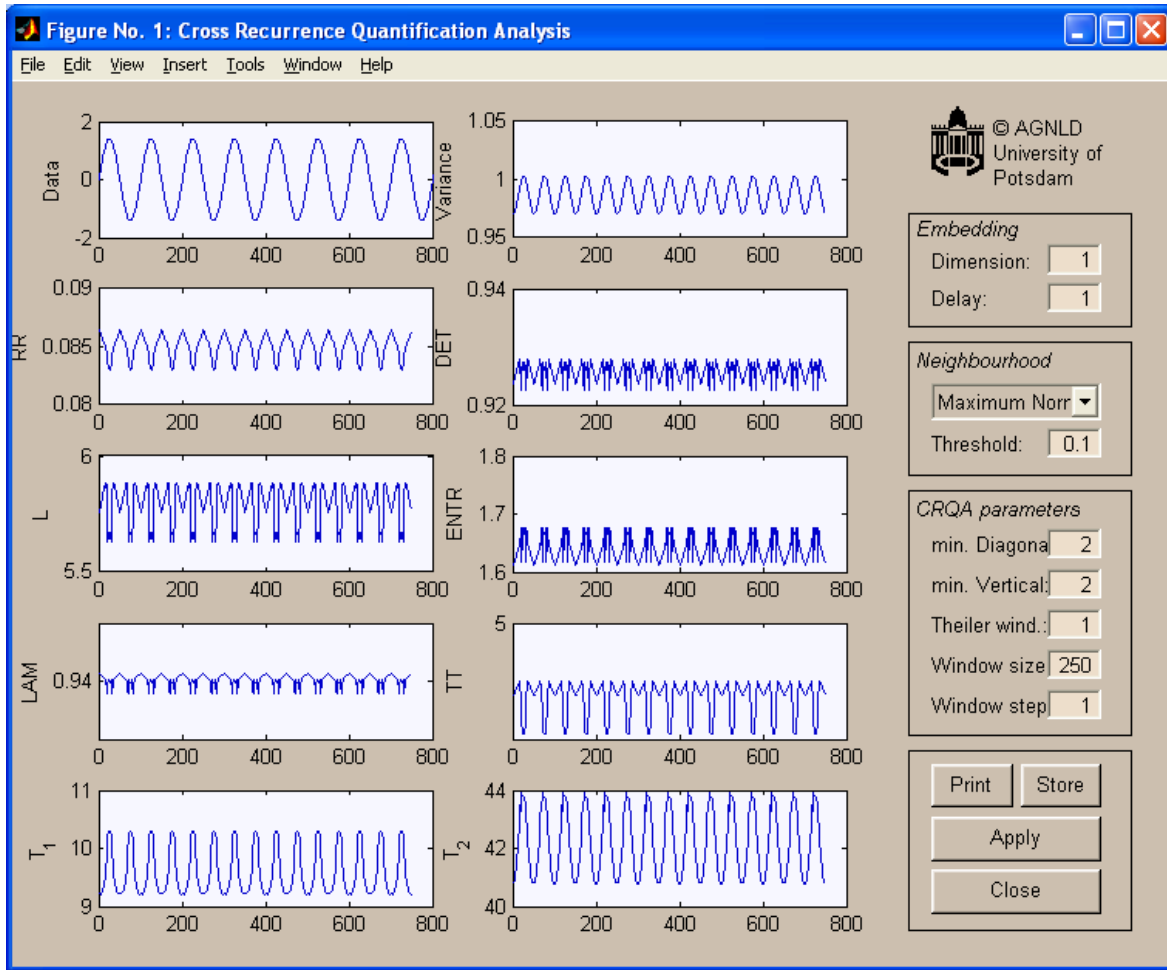


Рис. 2. Результати роботи функції **crqa** для  $\sin(x)$

Діагональна лінія довжини  $l$  означає, що сегмент траєкторії близький протягом  $l$  кроків часу до іншого сегменту траєкторії в інший час; тому ці лінії пов'язані з дивергенцією сегментів траєкторії. Середня довжина діагональної лінії

$$L = \frac{\sum_{l=l_{\min}}^N lP(l)}{\sum_{l=l_{\min}}^N P(l)} \quad (4)$$

є середнім часом, що два сегменти траєкторії близькі один одному, і може бути інтерпретований як середній час прогнозу.

Попередні міри відносяться до діагональних (горизонтальних) ліній на рекурентній карті. В той же час існують і вертикальні лінії з відповідними мірами. Розглянемо ті з них, які використовує функція **crqa(a)**.

Повне число вертикальних ліній довжини  $v$  в RP дається гістограмою

$$P(v) = \sum_{i,j=1}^N (1 - R_{i,j})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k} \quad (5)$$

Аналогічно визначенню детермінізму (рівняння (2)) відношення рекурентних точок, які формують вертикальні структури, до повного набору рекурентних точок може бути обчислено, як

$$LAM = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=1}^N vP(v)}. \quad (6)$$

Ця міра називається ламінарністю (*laminarity*). Обчислення LAM проводяться для тих  $v$ , які перевищують мінімальну довжину  $v_{\min}$ . Для рекурентних карт часто приймають  $v_{\min} = 2$ . Величина LAM зменшується, якщо RP складається з більш поодиноких рекурентних точок, ніж вертикальні структури.

Середня довжина вертикальних структур дається виразом

$$TT = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=v_{\min}}^N P(v)} \quad (7)$$

і називається часом затримки (захвату). Його обчислення вимагає також розгляду мінімальної довжини  $v_{\min}$ , як і у випадку з LAM.  $TT$  оцінює середній час, що система перебуватиме в певному стані, або як довго цей стан буде захоплений.

У контрасті до кількісних мір, заснованих на діагональних лініях, введені щойно міри можуть застосовуватись до переходів «хаос-хаос».

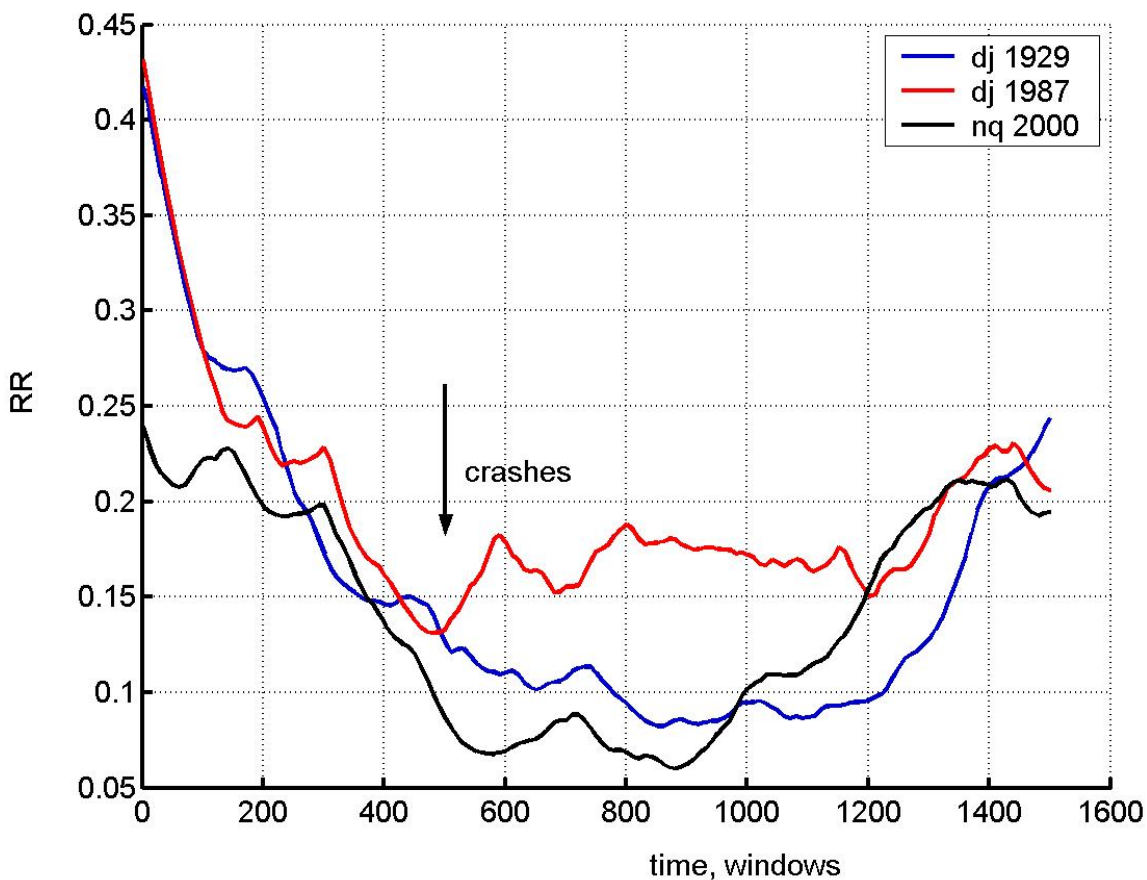


Рис. 3. Порівняльна віконна динаміка коефіцієнта само подібності для різних криз

Останні два параметри характеризують два різні типові проміжки часу, протягом яких траєкторії перебувають в околі  $\varepsilon$ . Їх віконна динаміка дозволяє від слідкувати часову компоненту рекурентних карт.

Зберігати і аналізувати зображення типу рис. 1-2 незручно. Тому є можливість окремо виділяти і працювати з вікнами. Для цього треба вибрати курсором необхідне вікно, натиснути правою кнопкою і активізувати **ShowFigure**.

На наступних рисунках представлено побудовані вказаним чином графіки двох довільно вибраних мір для різних криз, аналіз яких вказує на універсальність прояву кризових явищ.

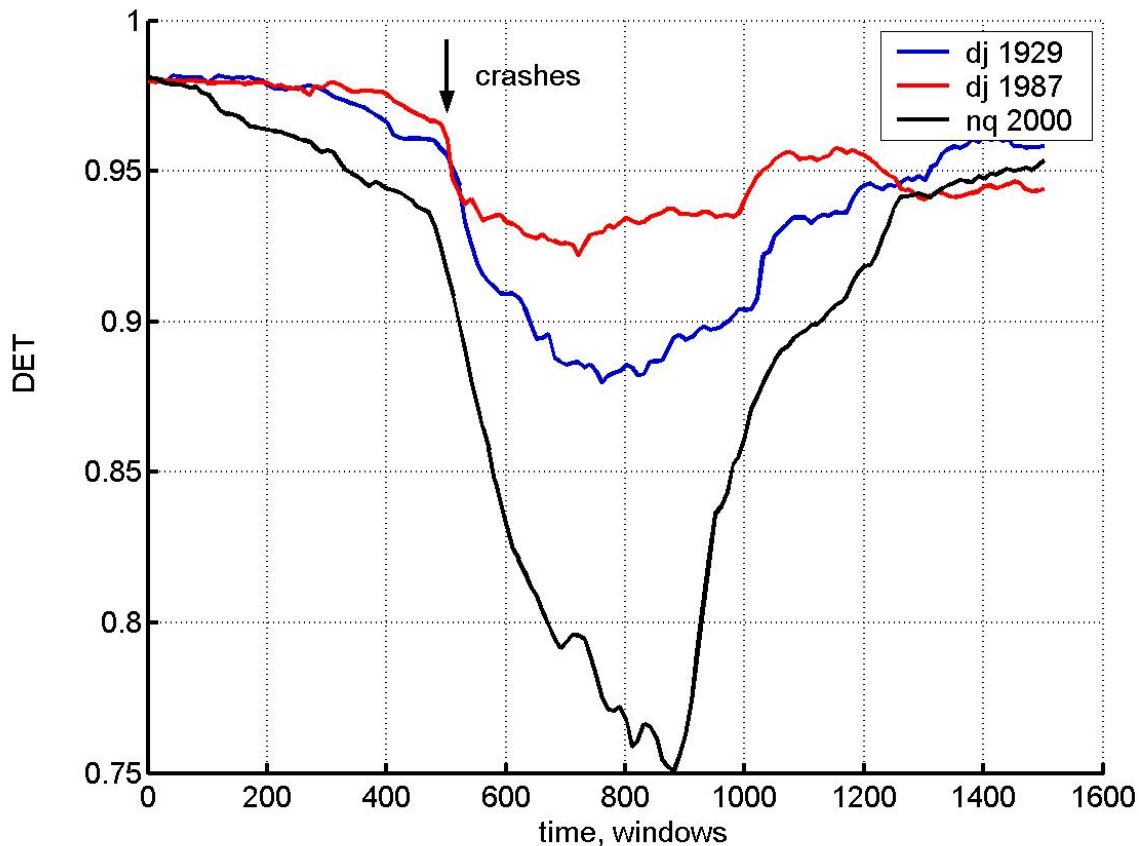


Рис. 4. Динаміка коефіцієнта передбачуваності для кризових динамічних рядів

#### Завдання для самостійної роботи

1. Виберіть за рекомендацією викладача свій варіант часового ряду
2. Проведіть дослідження динаміки кількісних мір рекурентності згідно інструкції
3. Зробити висновки
4. Оформіть звіт у вигляді текстового файлу

#### ЛАБОРАТОРНА РОБОТА № 5

**Тема.** Дослідження процесів самоорганізації в складних системах.

**Мета.** Навчитись використовувати методи теорії випадкових матриць для отримання «прихованої» інформації в складних економічних системах.

1. Знаходження коефіцієнтів матриці крос-кореляцій.
2. Визначення, аналіз та інтерпретація власних значень та власних векторів

симетричної матриці:

2.1. Розподіл власних значень, порівняння його з нормальним розподілом і розподілом для Гаусівського ортогонального ансамблю (ГОО).

2.2. Обернене відношення участі (Inverse participation ratio).

2.3. Аналіз власних значень, що відхиляються від розподілу ГОО.

3. Побудова та інтерпретація кластерної структури економічного ринку.

### Теоретичні відомості

Вивчення статистичних властивостей матриць з незалежними випадковими елементами – *випадкових матриць* – має багату історію, що починається з ядерної фізики, де проблема з'явилася 50 років тому при дослідженні енергетичних рівнів складних ядер, що існуючі на той час моделі були не в змозі пояснити. Теорія випадкової матриці (ТВМ) була розвинена в цьому контексті Вігнером (Wigner), Дайсоном (Dyson), Метою (Mehta) та іншими для пояснення статистики рівнів енергії складних квантових систем. Дослідники постулювали, що функція Гамільтона, яка описує важкі ядра, може бути задана матрицею  $H$  з незалежними випадковими елементами  $H_{ij}$ , отриманими з розподілу імовірності.

Відштовхуючись від цього припущення було зроблено низку вражаючих передбачень, які було підтверджено експериментально. Для складних квантових систем передбачення на основі ТВМ представляють середнє за всіма можливими взаємодіями. Відхилення від універсальних передбачень ТВМ відображують системну специфіку, невідповідності властивості системи, забезпечуючи ключові підходи до розуміння базової взаємодії системи.

Недавні дослідження, що використовували методи аналізу ТВМ до аналізу властивостей матриці взаємних кореляцій  $C$ , показують, що близько 98% власних значень матриці  $C$  співпадають зі значеннями, отримуваними з використанням ТВМ, таким чином пропонуючи задовільний рівень хаотичності у вимірюваних крос-кореляціях. Також було знайдено, що існують відхилення від передбачень за допомогою ТВМ у близько 2% найбільших власних значень. Ці результати викликають наступні питання:

1. Яка можлива інтерпретація для відхилень від ТВМ?
2. Що можна сказати про структуру  $C$  з цих результатів?
3. Яке практичне значення отриманих результатів?

У даній лабораторній роботі ці питання будуть вивчені докладно на прикладі економічних систем. Розглянемо складну систему – фондовий ринок, наприклад, США за індексом Стандарт енд Пурс 500 (Standard & Poor's 500 - SP), який є індексом так званого «широкого ринку», оскільки включає при розрахунку ціни акцій приблизно 500 американських компаній із різних сегментів бізнесу. Взяти певну кількість цих компаній (в принципі, можливо взяти і всі) і дослідивши процеси кореляції між ними, можна отримати досить важливу інформацію щодо роботи фондового ринку в цілому.

Шляхом комп'ютерного моделювання виявлено, що найбільше власне значення матриці  $C$  представляє вплив усього ринку, що є звичайним для всіх акцій. Аналіз змісту власних значень, що відхиляються від ТВМ, показує існування взаємних кореляцій між акціями того ж самого типу промисловості, найбільш капіталізованими акціями, і акціями фірм, що мають бізнес у певному географічному секторі (локалізовані територіально). Обчислюючи скалярний добуток власних векторів від одного періоду часу до наступного, можна побачити, що «власні вектори, що відхиляються», мають різні ступені стабільності в часі, визначеному кількісно величиною скалярного добутку. Найбільші два-три власних вектори стійкі протягом тривалих періодів часу, у той час як для іншої частини власних векторів, що відхиляються, стабільність у часі зменшується як тільки відповідні власні значення наближаються до верхньої межі ТВМ.

## Знаходження коефіцієнтів матриці крос-кореляцій

Визначення кореляцій між різними акціями – тема, цікава не лише з точки зору наукових причин розуміння економіки як складної динамічної системи, але також і з практичних поглядів, зокрема, з точки зору розміщення активів і оцінки портфельного ризику. Ми будемо аналізувати взаємні кореляції між акціями, застосовуючи поняття і методи теорії випадкових матриць, що використовуються в контексті складних квантових систем, де точний характер взаємодій між підодинамиціями невідомий.

Для визначення кількісно кореляцій спочатку обчислюється зміна цін («повернення», прибутковості) акції  $i = 1, \dots, N$  за час  $\Delta t$ ,

$$G_i(t) = \ln S_i(t + \Delta t) - \ln S_i(t), \quad (1)$$

де  $S_i(t)$  позначає ціну акції  $i$ . Оскільки різні акції мають різні рівні змінюваності (стандартні відхилення), визначатимемо нормалізоване повернення (нормалізовану прибутковість)

$$g_i(t) \equiv \frac{G_i(t) - \langle G_i \rangle}{\sigma_i}, \quad (2)$$

де  $\sigma_i \equiv \sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}$  – стандартне відхилення  $G_i$ , а  $\langle \dots \rangle$  позначає середнє значення за період часу, що досліджується. Тоді обчислення матриці взаємних кореляцій  $C$  зводиться до обчислення формули:

$$C_{ij} \equiv \langle g_i(t) g_j(t) \rangle. \quad (3)$$

Згідно з побудовою елементи  $C_{ij}$  обмежені областю  $-1 \leq C_{ij} \leq 1$ , де  $C_{ij} = 1$  відповідає повним кореляціям,  $C_{ij} = -1$  – повним антикореляціям, і  $C_{ij} = 0$  свідчить про некорельованість пар акцій.

Труднощі в аналізі важливості та значення коефіцієнтів крос-кореляції  $C_{ij}$  виникають внаслідок кількох причин, що полягають в наступному:

i) ринкові умови з часом змінюються і взаємна кореляція, що існує між будь-якою парою акцій, може бути не постійною (нестационарною);

ii) скінчена довжина досліджуваного ряду, доступного для оцінювання взаємних кореляцій, додає так званий «шум вимірювання» – чим коротший досліджуваний ряд – тим менш точними будуть отримувані значення.

Якщо буде використано довгий ряд для вилучення проблеми скінченної довжини, на отримані значення буде впливати нестационарність крос-кореляцій. З цих причин, емпірично виміряні крос-кореляції будуть містити «випадкові» складові, і найбільш важливою (і одночасно важкою проблемою) є оцінка в складі матриці  $C$  таких взаємних кореляцій, що не є результатом випадковості.

Яким же чином можна виділяти з  $C_{ij}$  ті акції, що залишилися корельованими на розглядуваному періоді часу? Щоб відповісти на це питання, перевіримо статистику  $C$  у порівнянні із так званою «нульовою гіпотезою» випадкової кореляційної матриці – матриці кореляцій, побудованої із взаємно некорельованих часових рядів. Якщо властивості  $C$  відповідають властивостям для випадкової матриці кореляцій, тоді можна говорити про те, що значення емпірично вимірюваних властивостей  $C$  випадкові. Навпаки, відхилення властивостей  $C$  від таких же властивостей для випадкової кореляційної матриці передає інформацію про «справжні» кореляції. Таким чином, нашою метою є порівняння властивостей  $C$  з такими ж властивостями випадкової матриці кореляцій і розділ



властивостей  $C$  на дві групи: (а) частина  $C$ , що відповідає властивостям випадкової кореляційної матриці («шум») і (б) частина  $C$ , що відхиляється («інформація»).

### Розподіл власних значень

Для отримання інформації про взаємні кореляції  $C$  необхідно порівняти властивості  $C$  з такими ж властивостями випадкової матриці крос-кореляцій. У матричній нотації така матриця може бути виражена як

$$C = \frac{1}{L} GG^T, \quad (4)$$

де  $G$  – матриця розміру  $N \times L$  з елементами  $\{g_{im} = g_i(m\Delta t), i=1, \dots, N; m=0, \dots, L-1\}$  і  $G^T$  позначає транспонування  $G$ . Розглянемо випадкову кореляційну матрицю

$$R = \frac{1}{L} AA^T, \quad (5)$$

де  $A$  – матриця розміру  $N \times L$ , що містить  $N$  часових рядів із  $L$  випадковими елементами  $a_{im}$  з нульовим середнім і одиничним відхиленням, що означають взаємну некорельованість.

Статистичні властивості випадкових матриць типу  $R$  відомі. Зокрема, у наближенні  $N \rightarrow \infty$ ,  $L \rightarrow \infty$ , такому, що  $Q \equiv \frac{L}{N} (> 1)$  фіксоване, показано аналітично, що функція розподілу щільності імовірності  $P_{rm}(\lambda)$  власних значень  $\lambda$  випадкової матриці кореляції  $R$  визначається як

$$P_{rm}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (6)$$

для  $\lambda$  в межах границь  $\lambda_- \leq \lambda_i \leq \lambda_+$ , де  $\lambda_-$  і  $\lambda_+$  – найменше і найбільше власні значення  $R$ , відповідно,

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}}. \quad (7)$$

Звертаємо вашу увагу, що вираз (6) є точним для випадку розподілених за Гаусом матричних елементів  $a_{im}$ .

Порівняємо розподіл власних значень  $P(\lambda)$  для  $C$  з  $P_{rm}(\lambda)$ . Для цього обчислимо власні значення  $\lambda_i$  матриці  $C$ , причому  $\lambda_i$  впорядкуємо за зростанням ( $\lambda_{i+1} > \lambda_i$ ). При дослідженні зверніть увагу на присутність чіткої «великої частини» власних значень, що спадають у межах границь  $[\lambda_-, \lambda_+]$  для  $P_{rm}(\lambda)$ . Також зверніть увагу на відхилення для деяких найбільших і найменших власних значень від отриманих за допомогою ТВМ.

Оскільки рівняння (6) є таким, що строго відповідає лише для  $L \rightarrow \infty$  і  $N \rightarrow \infty$ , необхідно перевірити також відхилення від ідеального випадку, оскільки робота проводиться завжди із **скінченими** рядами. При дослідженні виявляється, що для кількох найбільших (найменших) власних значень ефект впливу скінчених величин  $L$  і  $N$  відсутній.

## Обернене відношення участі

Вивчивши інтерпретацію найбільшого власного значення, що значно відхиляється від результатів ТВМ, зосередимось на власних значеннях, що залишаються. Відхилення розподілу компонентів власного вектора  $u^k$  від ТВМ Гаусового передбачення більш явне, коли відстань від верхньої границі ТВМ  $\lambda_k - \lambda_+$  збільшується. Оскільки близькість до  $\lambda_+$  збільшує ефекти хаотичності, визначаємо кількість компонентів, що беруть значну участь в кожному власному векторі, що, у свою чергу, відображає ступінь відхилення від ТВМ для розподілу компонентів власного вектора. Для цього використовується поняття оберненого відношення участі (ОВУ), що часто застосовується в теорії локалізації. ОВУ власного вектора  $u^k$  визначається як

$$I^k = \sum_{l=1}^N [u_l^k]^4, \quad (8)$$

де  $u_l^k$ ,  $l=1, \dots, N$  – компоненти власного вектора  $u^k$ . Значення  $I^k$  може бути проілюстровано двома граничними випадками:

- i) вектор з ідентичними компонентами  $u_l^k = \frac{1}{\sqrt{N}}$  має  $I^k = \frac{1}{N}$ ;
- ii) вектор з одним компонентом  $u_1^k = 1$  і нульовими іншими має  $I^k = 1$ .

Таким чином, ОВУ визначає кількість даних з числа компонентів власного вектора, що значний впливають на ринок, заданий системою часових рядів.

Наявність векторів з великими значеннями  $I^k$  також виникає в теорії локалізації Андерсона. У контексті теорії локалізації часто знаходять «випадкову смугу матриць», що містять узагальнені стани з маленьким  $I^k$  в більшій частині спектра власних значень, тоді як основні стани локалізовані і мають великі  $I^k$ . Виявлення локалізованих станів для маленьких і великих власних значень матриці крос-кореляцій  $C$  нагадує про локалізацію Андерсона і припускає, що  $C$  може мати випадкову зону матричної структури.

### Інтерпретація власних значень та власних векторів, що відхиляються від ТВМ

Кількість істотних внесків у власний вектор визначається (як уже було описано) за допомогою ОВУ, для отримання ж загальних властивостей досліджуються компоненти  $\frac{1}{I^k}$  власного вектора  $u^k$ . Пряма експертиза таких власних векторів, однак, не приводить до прямої інтерпретації їх економічної доречності. Щоб інтерпретувати їх значення, звернемо увагу, що найбільше власне значення значно більше, ніж інші  $N-1$  власні значення тоді як розмір матриці  $C$  рівний  $N$ . Таким чином, щоб аналізувати власні вектори, що відхиляються, необхідно видалити ефект найбільшого власного значення  $\lambda_{\max}$ .

Щоб позбутися ефекту  $\lambda_{\max}$ , і – відповідно –  $G^{\max}(t)$ , над прибутковостями кожної акції  $G_i(t)$  виконується регресія

$$G_i(t) = \alpha_i + \beta_i M(t) + \varepsilon_i(t), \quad (9)$$

або ж, у нашому випадку просте визначення тренда часового ряду, що відповідає  $\lambda_{\max}$ , та вилучення його з кожної часової послідовності. Тоді обчислюється матриця кореляції  $C$ , використовуючи формули (1)-(3). Далі обчислюються власні вектори  $u^k$   $C$  і проводиться

аналіз їх істотних учасників. Власний вектор  $u^{N-1}$  містить приблизно  $\frac{1}{I^{N-1}}$  істотних учасників, які є всіма акціями з великими значеннями ринкової капіталізації.

Далі проводиться аналіз істотних вкладників іншої частини власних векторів. Результатами аналізу у більшості випадків є те, що кожний з власних векторів, що відхиляються, містить акції, які належать подібним або пов'язаним галузям промисловості як істотним вкладникам. Ці власні вектори поділяють набір всіх акцій у різні групи, що містять акції з великою ринковою капіталізацією ( $u^{N-1}$ ), а також галузеві коаліції ( $u^{N-2} - u^{N-9}$ ).

## Хід роботи

У середовищі MatLab завантажимо пакет для дослідження кореляційних властивостей ринку, заданого часовими рядами його складових, набравши у командному рядку **work2**, натиснувши **Enter**. У цій роботі ми будемо користуватись, в основному, пунктом меню **RMT**.

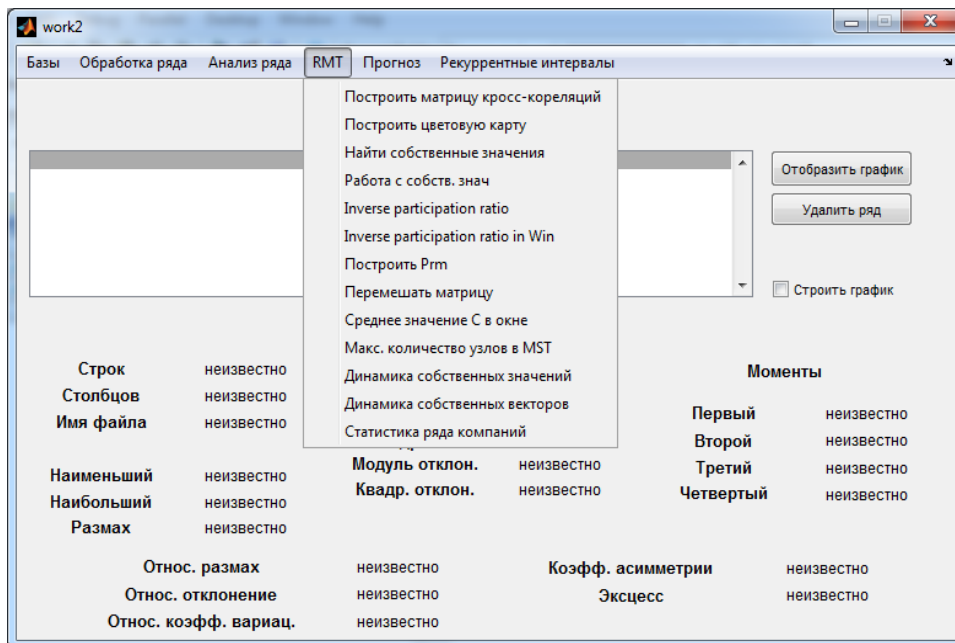


Рис. 1. Вікно програми дослідження ринків, заданих часовими рядами їх складових

База даних для матричних досліджень складається з трьох рядів із розширеннями (.txt), (.nm) та (.dt). Перший з них містить підряд записані часові ряди цін акцій компаній з назвами, що перераховані у наступному файлі і за період часу, вказаний у файлі з розширенням (.dt). Оберемо, наприклад, 119 компаній за період з 30.12.1983 по 21.04.2015pp. Відкриємо базу, завантажуючи файл SP500\_119\_19831230.txt.

Для введеної матриці знайдемо матрицю взаємних кореляцій: **RMT**→**Построить матрицу кросс-корреляций** (спочатку заходимо у меню **RMT**, потім – викликаємо пункт **Построить матрицу кросс-корреляций**). Після побудови бачимо, що утворено матрицю розмірності  $N \times N$ , де  $N$  – кількість рядків (фірм), що знаходяться в матриці. Таким чином, знайдено матрицю  $C$  з використанням формул (1)-(3). Візуалізуємо її та порівняємо з аналогічною матрицею, але перемішаною: **RMT**→**Перемешать матрицу** →**Построить матрицу кросс-корреляций**. Обравши із стовпчика результатів КроссКор і операцію **RMT**→**Построить цветовую карту**, отримаємо результати, що представлені на рис. 2.

Дослідимо статистичні властивості матриці  $C$  шляхом порівняння із властивостями випадкової матриці.

Знайдемо розподіл ймовірності елементів матриці  $C$ . Для цього виконаємо **Анализ ряда**→**Распределение вероятности** і у вікні, що відкриється, виберемо **Количество интервалов** рівне 50 та натиснемо кнопку **Обработка**. У області графіка повинен з'явитись

набір точок, що відповідає щільності ймовірності розподілу значень матриці крос-кореляцій (рис. 3а).

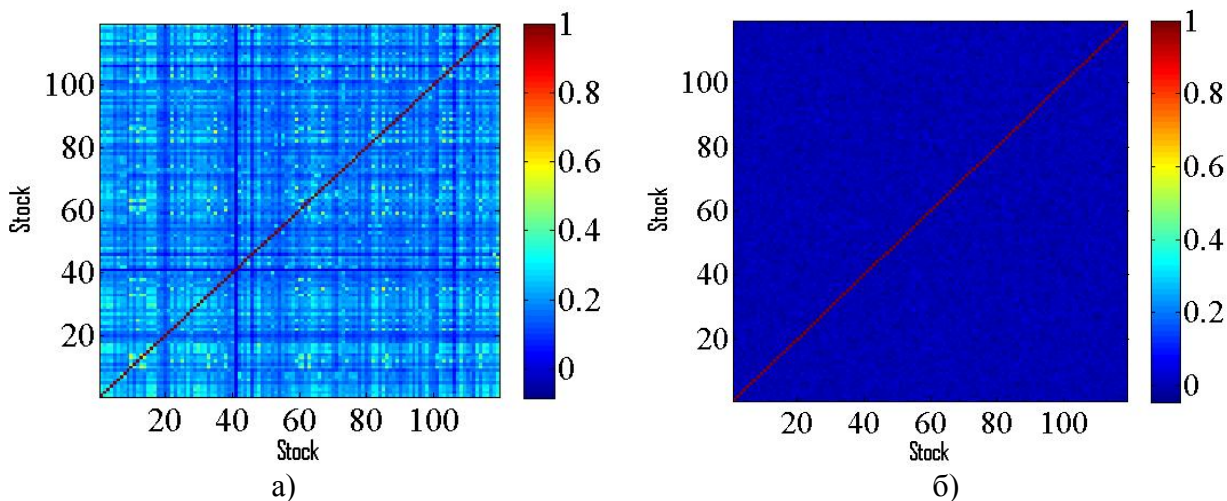


Рис. 2. Карти полів кореляцій вихідної (а) і перемішаної (б) матриць

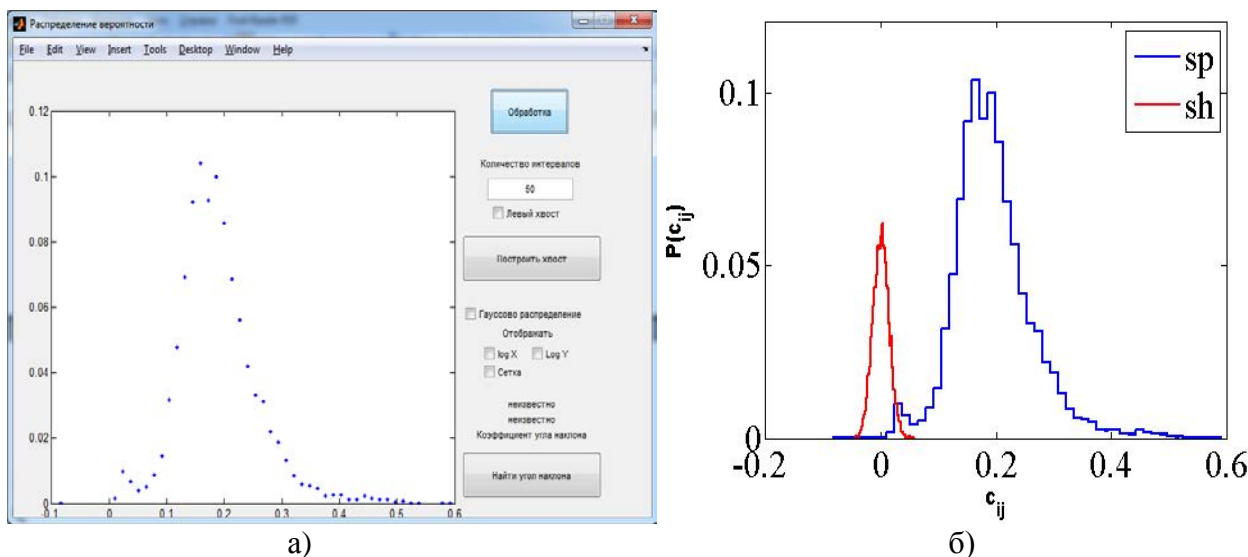


Рис. 3. а) Розподіл ймовірності значень вихідної матриці крос-кореляцій; б) порівняння розподілів значень коефіцієнтів кореляції вихідної та перемішаної (sh) матриць

Закриємо вікно, натиснувши кнопку закриття (на заголовку вікна кнопка з позначкою  $\times$ ). У основному вікні програми у полі *Открытые ряды* з'явилося два ряди: *PDens X* (значення абсцис точок графіка розподілу ймовірності) та *PDens Y* (значення ординат цих точок). Даний графік можна переглянути, вибравши один із рядів *PDens X* чи *PDens Y* та натиснувши кнопку *Отобразить график* (закриття вікна відображення графіка – по натисненні на кнопку *Закреть окно*. Аналогічні дії проробимо для перемішаної матриці.

Внаслідок виконання таких дій отримаємо зображення двох графіків, накладених один на інший (рис. 3б).

Як видно із рисунка, розподіл ймовірності значень матриці крос-кореляцій вихідного ряду зміщений далеко вправо порівняно з розподілом для перемішаної матриці, що свідчить про наявність кореляцій (а отже, і тісних зв'язків) для фірм, що працюють на ринку. Для перемішаної матриці кореляції зникають, їх середнє значення близьке до нуля, відхилення від якого носять випадковий характер. Отже, шляхом досліджень властивостей матриці крос-кореляцій для початкової бази можна отримати значущу інформацію про цей ринок та його складові.

Знайдемо також середні значення коефіцієнтів кореляції для вихідної і перемішаної матриці у режимі рухомого вікна. Нехай вікно буде рівним 250 дням, а крок його зміщення становить 5 днів. Тоді послідовність дій *RMT*→*Среднее значение C в окне* відкриє нове вікно (рис. 4а), активація якого призведе до результату, представленого на рис. 4б.

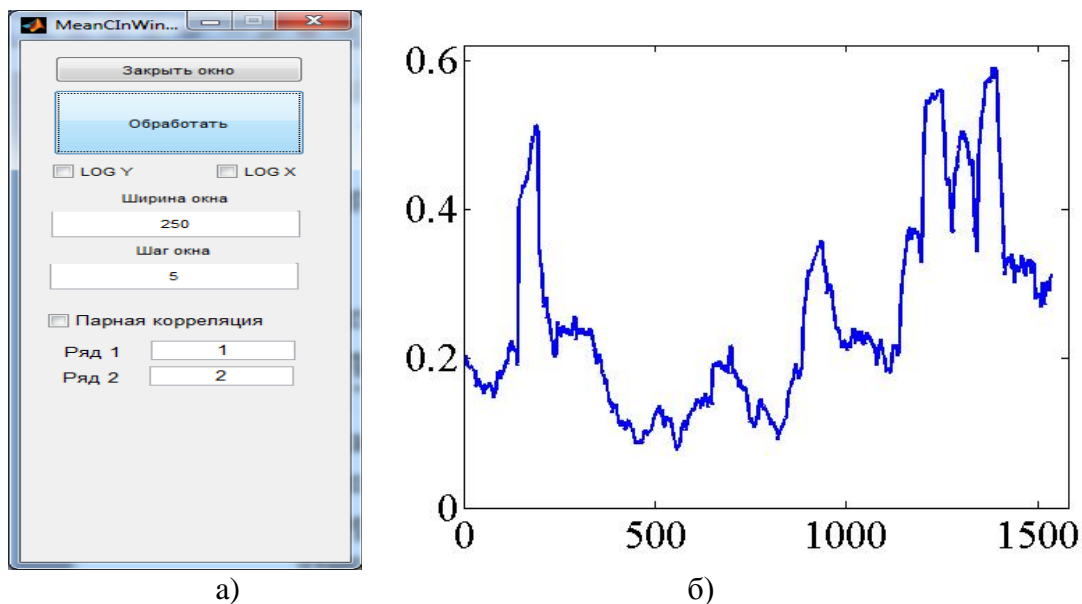


Рис. 4. Розрахунок середнього значення коефіцієнта кореляції: а) скріншот панелі для вибору параметрів і розрахунку; б) результати розрахунку  $\langle C \rangle$

Цікаво порівняти динаміку  $\langle C \rangle$  з динамікою власне індексу S&P 500 за той же період часу. Вкоротивши його на довжину вікна (250) та розрідивши з кроком 5 (опція *Прогноз*→*Взяти частину ряду*), отримаємо залежності, представлені на рис. 5а. Видно, що у періоди відомих криз на фондовому ринку (1987, 2001, 2008, 2011pp.) кореляції помітно зростають, що свідчить про самоорганізацію фінансово-економічної системи. Для перемішаної матриці нічого подібного не спостерігається (рис. 5б).

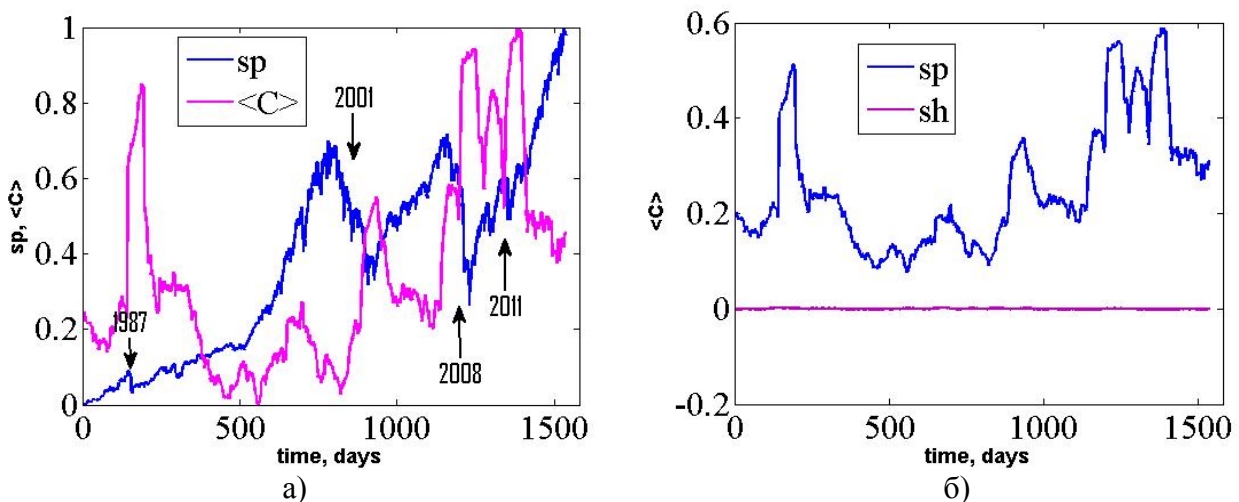


Рис. 5. а) Зростання кореляційних процесів у періоди криз, вказаних стрілками на рисунку; б) порівняння середніх значень коефіцієнта кореляції для вихідної і перемішаної матриць

Програмний засіб **work2** дозволяє знайти й інші «віконні» характеристики корельовано системи. Так, розрахуємо розподіл кореляцій для вихідної та перемішаної матриць (рис. 6). Видно, що, як і на рис. 3, для перемішаної матриці маємо вузький спектр кореляцій, які не змінюються у періоди кризових явищ. Для вихідної матриці крос-кореляцій

помітне зростання кореляцій відбувається у періоди відомих криз. На рис. 6б наведені ті ж результати, що і на рис. 6а, але у напівлогарифмічному масштабі, що дозволяє деталізувати вигляд «хвостів» функцій розподілу.

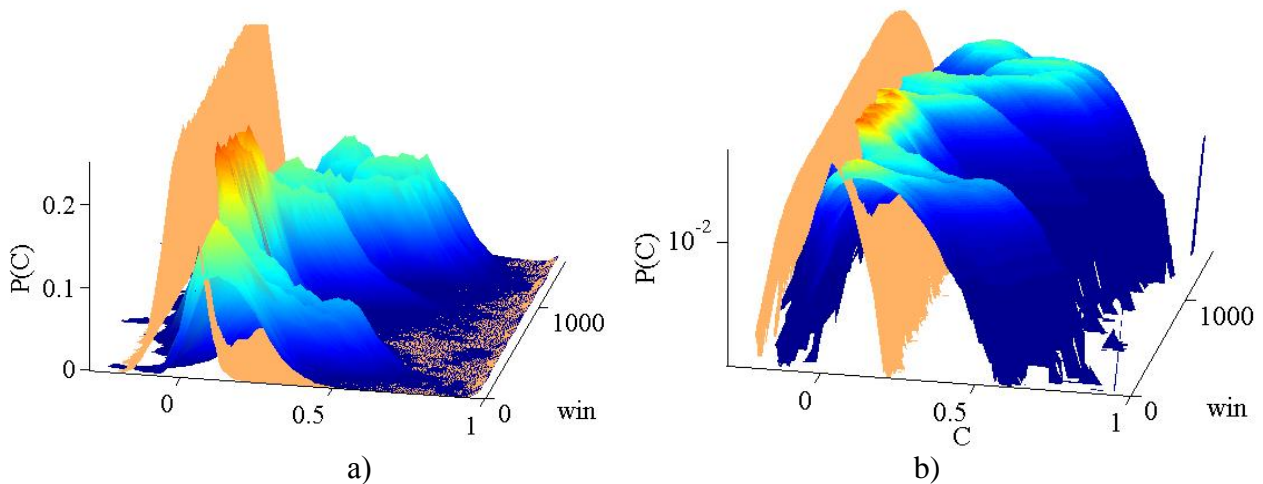


Рис. 6. «Віконні» функції розподілу коефіцієнтів кореляції вихідної і перемішаної матриць (а). Напівлогарифмічний масштаб (б) ідентифікує присутність великих значень («хвоста» розподілу) коефіцієнтів кореляції у вихідній матриці і відсутність таких у перемішаної

Знайдемо далі розподіл власних значень матриці крос-кореляцій та проведемо роботу з ними.

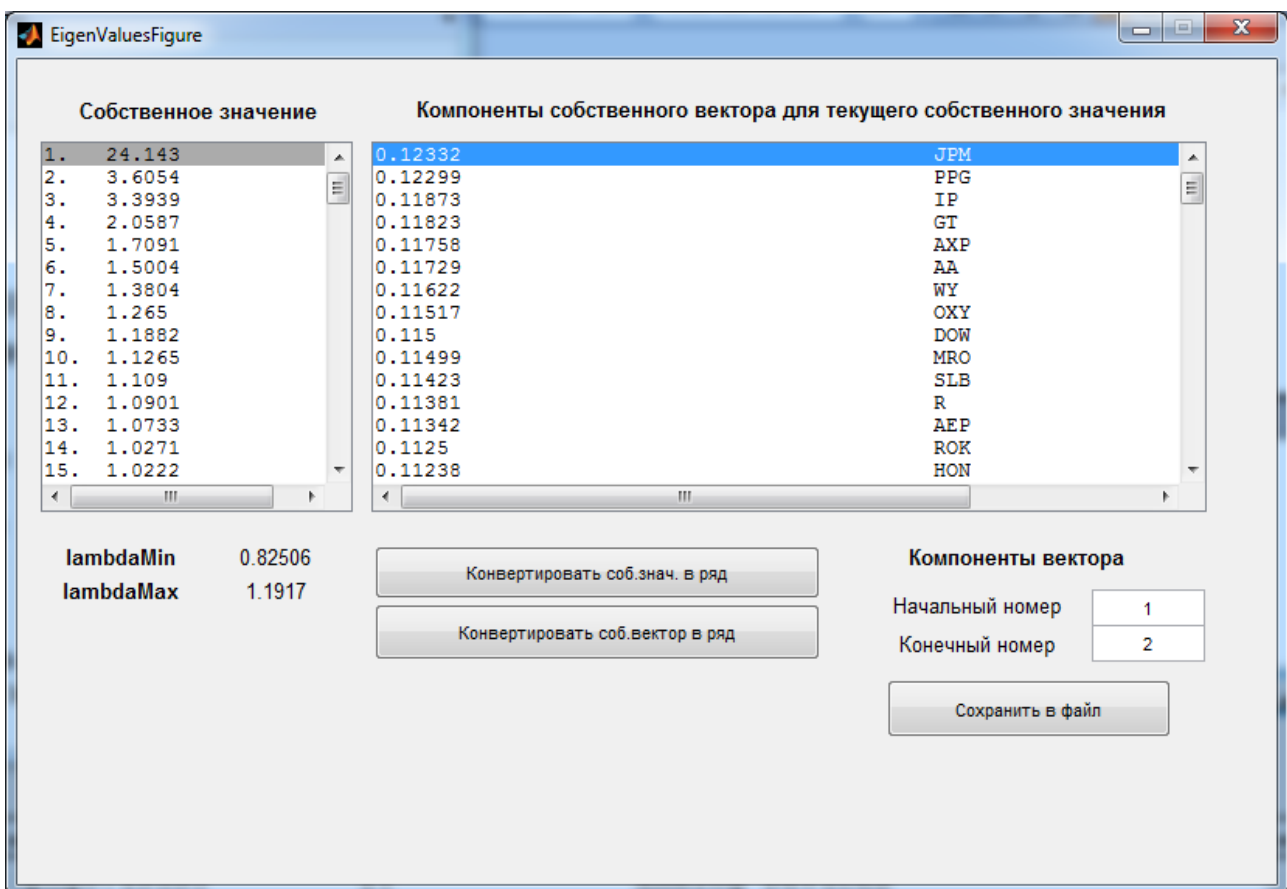


Рис. 7. Вікно для роботи із власними значеннями. При виборі певного значення в полі *Собственное значение* у полі справа відобразатимуться компоненти відповідного власного вектора

Для знаходження найбільших значень виберемо матрицю крос-кореляцій початкової бази (перше входження *КроссКор* у полі *Открытые ряды*) та виконаємо *RMT*→*Найти*



**собственные значения.** У вікні з'явиться ряд **СобстЗнч** (він активний). Для дослідження власних значень виберемо пункт (при виділеному **СобстЗнч**) **RMT**→**Работа с собст. знач.**, після чого відкриється вікно **EigenValuesFigure** (рис. 7).

Як можна побачити, існує найбільше власне значення, що суттєво більше за інші. Компоненти власного вектора показують, на які фірми (компанії) найбільш впливає ринок загалом (для даного прикладу, це, перш за все, фінансові компанії (Financials) та інформаційні технології (IT – Information Technology)). Декілька інших з найбільших власних значень показують вплив компаній окремих секторів економіки на ринок (або найбільшу участь фірм у роботі ринку).

Таким чином, за допомогою статистики власних значень та власних векторів можна робити висновок про взаємну організацію (кластеризацію) фірм на ринку та вплив загальної сукупності на окремі об'єкти ринку і навпаки.

Збережемо власні значення у вигляді окремого ряду, натиснувши **Конвертировать соб. знач. в ряд** та закриємо вікно. У полі **Открытые ряды** з'явився новий ряд **СобстЗнач**. Його відмінність від **СобстЗнч** в тому, що він містить лише один ряд – власних значень, в той час як **СобстЗнч** містить стільки рядів, скільки і фірм або власних значень – у кожному ряді за відповідним власним значенням слідує компоненти власного вектора.

Порівняємо розподіл власних значень із розподілом  $P_m(\lambda)$ .

Знайдемо розподіл власних значень. Виділимо ряд **СобстЗнач** та виберемо **Анализ ряда**→**Распределение вероятности** і побудуємо розподіл. Потім виберемо початкову базу **Матрица** та виконаємо **RMT**→**Построить Prm**.

Після цього побудуємо, наклавши один на інший, останні отримані два графіки. Вийде подібне до зображеного на рис. 8а (без легенди). Реалізувавши віконний алгоритм, отримаємо рис. 8б.

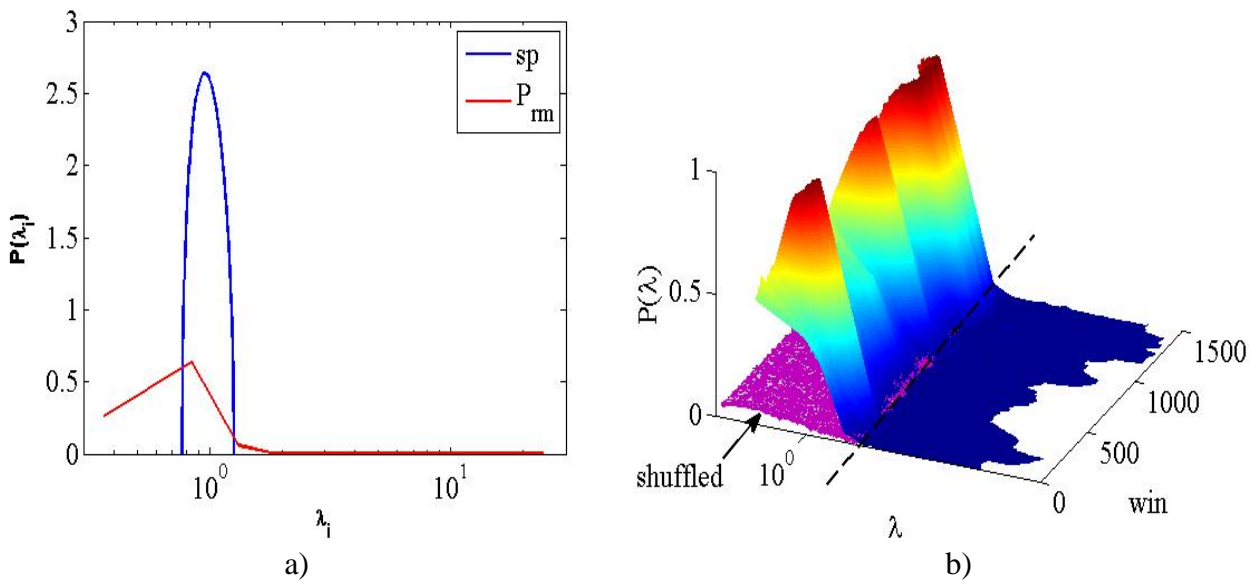


Рис. 8. Порівняння розподілу власних значень матриці крос-кореляцій та графіка функції  $P_m$  для стандартного алгоритму (а) та алгоритму рухомого вікна (б). На графіках добре видно існування значної частини, що виходить справа за межі  $P_m$ , та частини зліва. Це свідчить про наявність у матриці крос-кореляцій певної інформації, присутньої на ринку, про співіснування об'єктів ринку. Найчастіше досліджуються значення, що виходять справа за межі розподілу  $P_m$

Цікаво порівняти максимальне власне значення  $\lambda_{\max}$  з середнім значенням коефіцієнта кореляції (рис. 5а) і так званим коефіцієнтом поглинання (absorption ratio AR), який є кумулятивною мірою ризику, визначається формулою:  $AR_n = \sum_{k=1}^n \lambda_k / \sum_{k=1}^N \lambda_k$  і вказує, яку частину загальної варіації описують  $n$  із загальної кількості  $N$  власних значень (рис. 9).

Додаткові віконні процедури, як для наведених вище розподілів, так і до яких ми ще будемо звертатись, описані і прокоментовані у скрипті **RMTInWindow\_func2.m**.

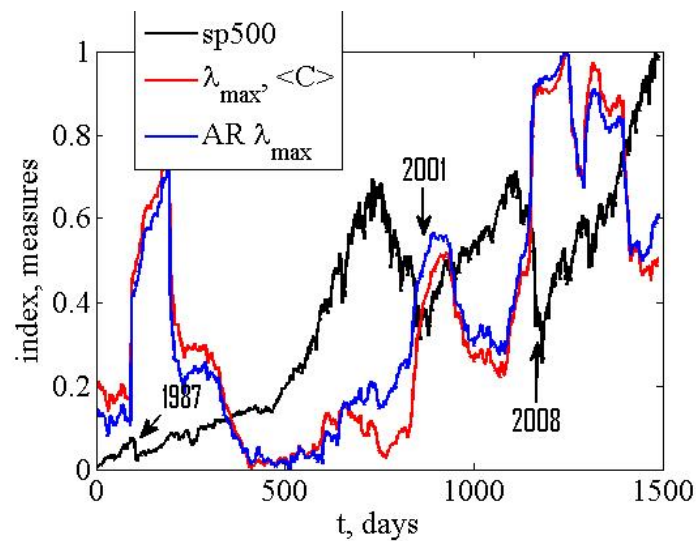
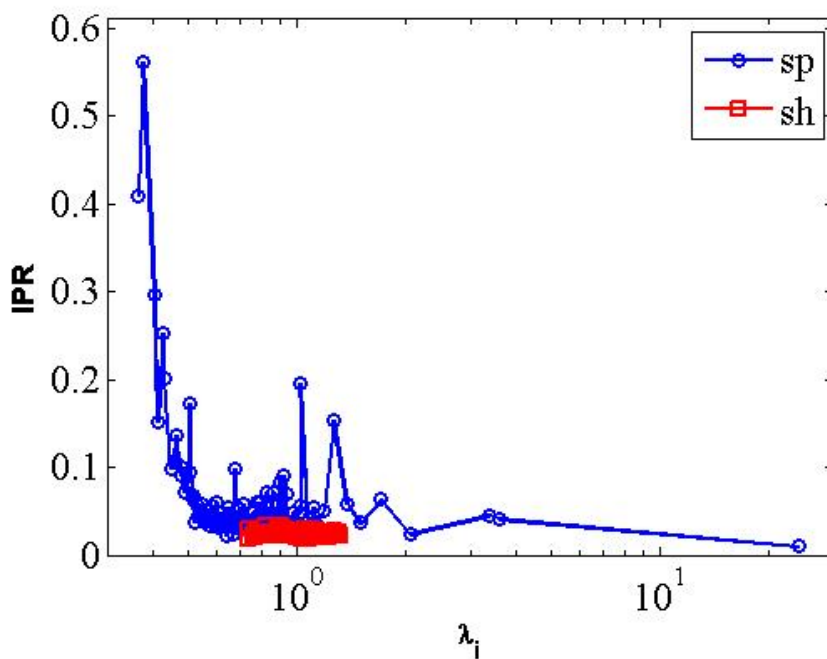


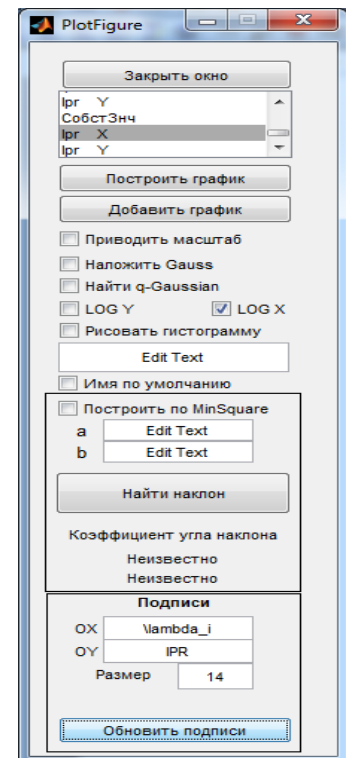
Рис. 9. Порівняльна динаміка віконних значень  $\lambda_{\max}$ ,  $\langle C \rangle$ ,  $AR\lambda_{\max}$  з індексом фондового ринку sp500.

Знайдемо обернене відношення участі для матриці власних значень та векторів матриці крос-кореляцій початкової та матриці крос-кореляцій випадкової матриць.

Для обрахунку ОВУ початкової матриці виберемо об'єкт *СобстЗнч*, розрахований для матриці крос-кореляцій початкової матриці та виконаємо  $RMT \rightarrow$  *Inverse participation ratio*. У кінці списку з'явиться два ряди *Ipr X* та *Ipr Y*. Для знаходження ОВУ перемішаної матриці необхідно для неї розрахувати власні значення та вектори і потім аналогічно знайти значення ОВУ. Відобразивши обидва отримані графіки разом можемо бачити подібне до поданого на рис. 10а.



а)



б)

Рис. 10. а) Графіки ОВУ для вихідного та перемішаного рядів; б) скріншот панелі інструментів для встановлення напівлогарифмічного формату.



Як видно з рисунка, дані відображено у напівлогарифмічному масштабі (що часто використовується в дослідженнях). Для встановлення цього масштабу необхідно на панелі *PlotFigure*, де вибираються параметри відображення графіків та графіки для відображення, активізувати поле *LOG X* (активізація відбувається натисненням кнопки мишки при наведенні курсору мишки на поле), рис. 10б.

З рисунка добре видно, що графік ОВУ для перемішаного ряду добре локалізований, в той час як ОВУ для вихідного ряду дещо розтягнутий вздовж осі абсцис. Це свідчить про наявність інформації стосовно частки вкладень кожної фірми (кожного досліджуваного об'єкта) у ринок в матриці крос-кореляцій. Про розмір участі вкладення можна говорити, аналізуючи саме власні вектори найбільших значень, що виходять за межі графіка  $P_{rm}$ .

Цікаву інформацію містить і аналіз найменших власних значень і відповідних компонентів їх векторів. Можна продемонструвати це, аналізуючи парні коефіцієнти кореляції між найбільшими та відповідно найменшими власними значеннями матриці кореляцій.

**!!! Зробіть це самостійно.**

Віконна динаміка обернених відношень участі та їх розподілу відображена на рис. 11.

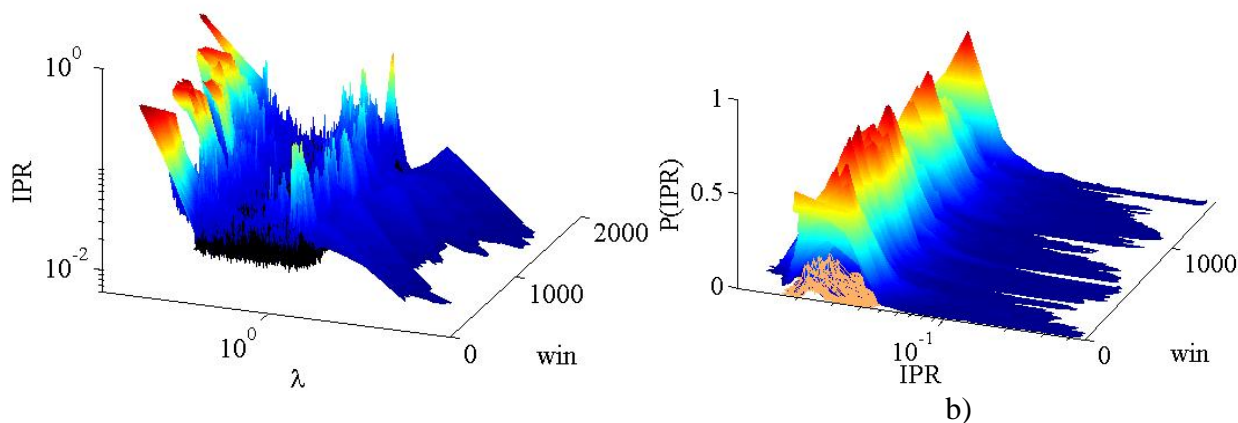


Рис. 11. Віконна динаміка оберненого відношення участі (а) і розподіл значень IPR (б)

### Висновок

Таким чином, при наявності сукупності часових рядів, що є даними діяльності економічних об'єктів однієї області, можна провести дослідження стосовно структури вказаної області та взаємодії об'єктів всередині неї. Дослідження проводяться на основі теорії випадкових матриць, що дозволяє отримувати інформацію шляхом аналізу матриці крос-кореляцій, побудованої для сукупної бази економічних об'єктів.

### Завдання для самостійної роботи

Виберіть відповідну базу даних, проведіть дослідження за допомогою теорії випадкових матриць. Результати-звіт представте у вигляді текстового файлу.

### Контрольні питання

1. Поясніть основну ідею теорії випадкових матриць.
2. Про що свідчить відмінність кореляційних і спектральних властивостей матриці даних і випадкової?
3. Дослідіть, як змінюється розподіл компонентів власного вектора у випадках: а)  $\lambda_- < \lambda < \lambda_+$ ; б)  $\lambda > \lambda_+$ ; в)  $\lambda < \lambda_-$ . Інтерпретуйте результати.
4. Порівняйте кольорову карту поля взаємних кореляцій випадкової матриці і заданої. Зробіть висновки.
5. Відслідкуйте компоненти власних векторів, які відповідають 5-7 найбільшим власним

значенням. Дослідіть, які компанії і з яких секторів входять в кластери, що формуються компонентами власних векторів.

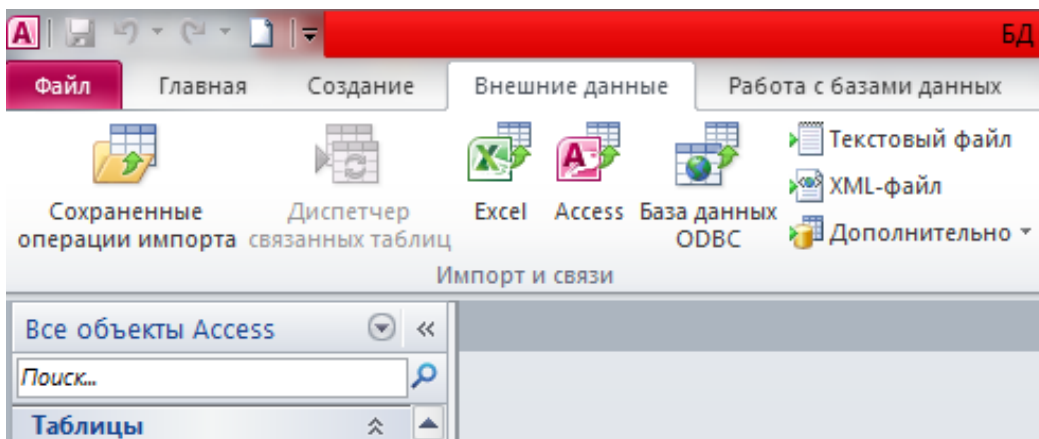
*Додаток*

### Інструкція створення матричної бази даних

Для створення матричної бази, яка використовується в даній та деяких інших роботах, необхідно завантажити відповідні часові ряди з відомих Інтернет ресурсів будь-яким відомим вам способом (наприклад, [finance.yahoo.com](http://finance.yahoo.com)).

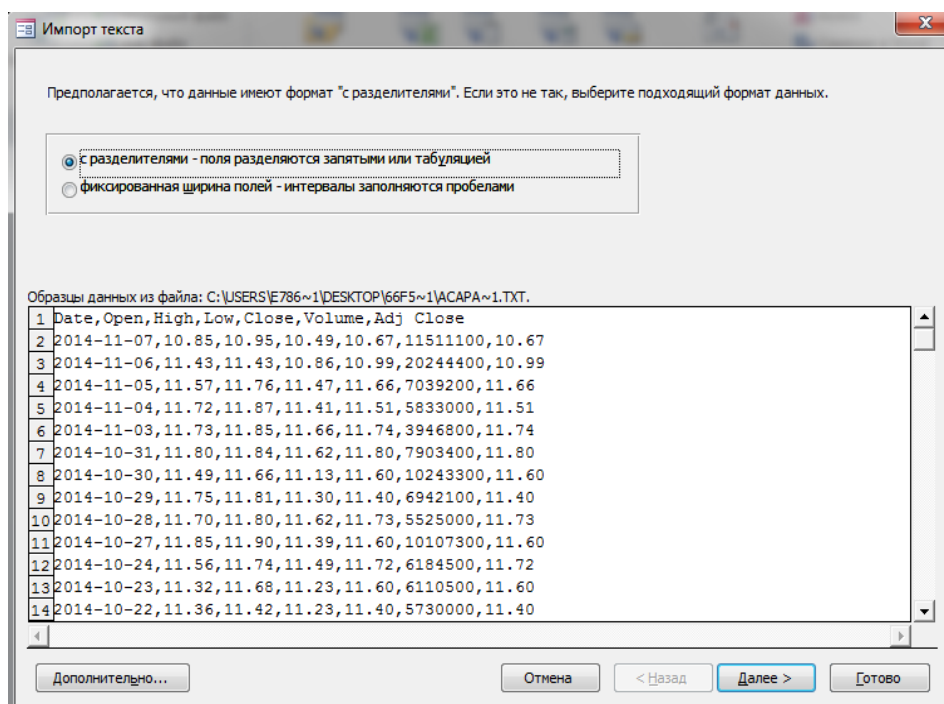
Після завантаження часових рядів на комп'ютер необхідно обробити його за допомогою середовищ Excel або Access. У даному випадку розглянемо варіант обробки через середовище Access.

Після завантаження Access, переходимо на вкладку «внешние данные» і обираємо відповідний формат документу, який необхідно завантажити:

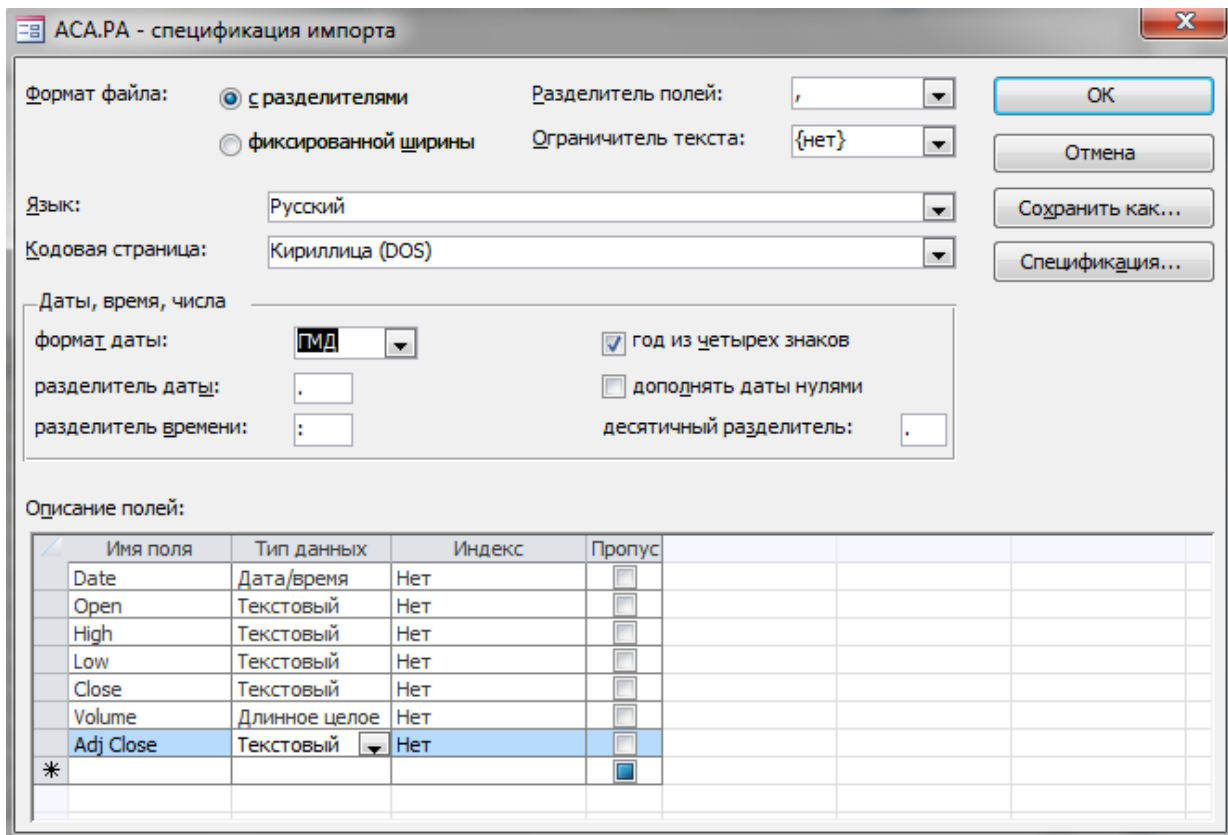


В даному прикладі необроблені часові ряди будемо завантажувати в форматі \*.txt. Після завантаження текстового документу необхідно виконати наступні дії:

1. Встановити галочку в поле «с разделителями – поля разделяются запятыми или табуляцией» і натиснути далі; в наступному вікні встановити галочку на відповідний «разделитель», в основному це «запятая» і натиснути на вкладку «дополнительно»:

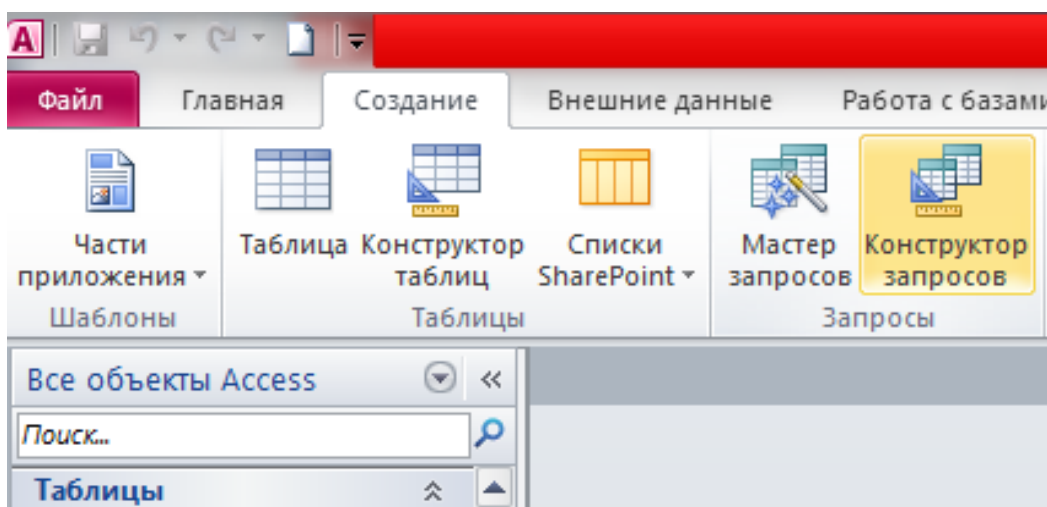


2. Вказуємо формат дати, а саме обрати «ГМД», змінюємо «десятичний роздільник» з коми на крапку; змінюємо «тип даних» так як вказано на малюнку і натиснути ОК; після цього натиснути готово.

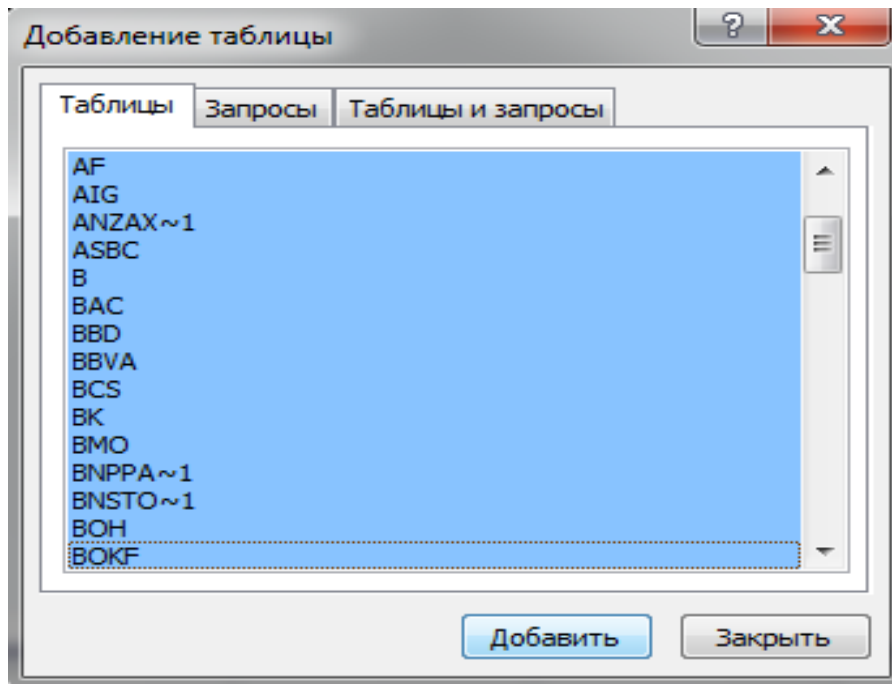


Тепер часовий ряд оброблений і готовий до роботи. Всі дії, які описані вище, необхідно виконати для всіх необроблених часових рядів. Після того як всі часові ряди завантажені в середовище Access, приступаємо безпосередньо до створення самої матричної бази даних та синхронізації рядів.

3. Переходимо на вкладку «создание» і обираємо функцію «конструктор запитів»:

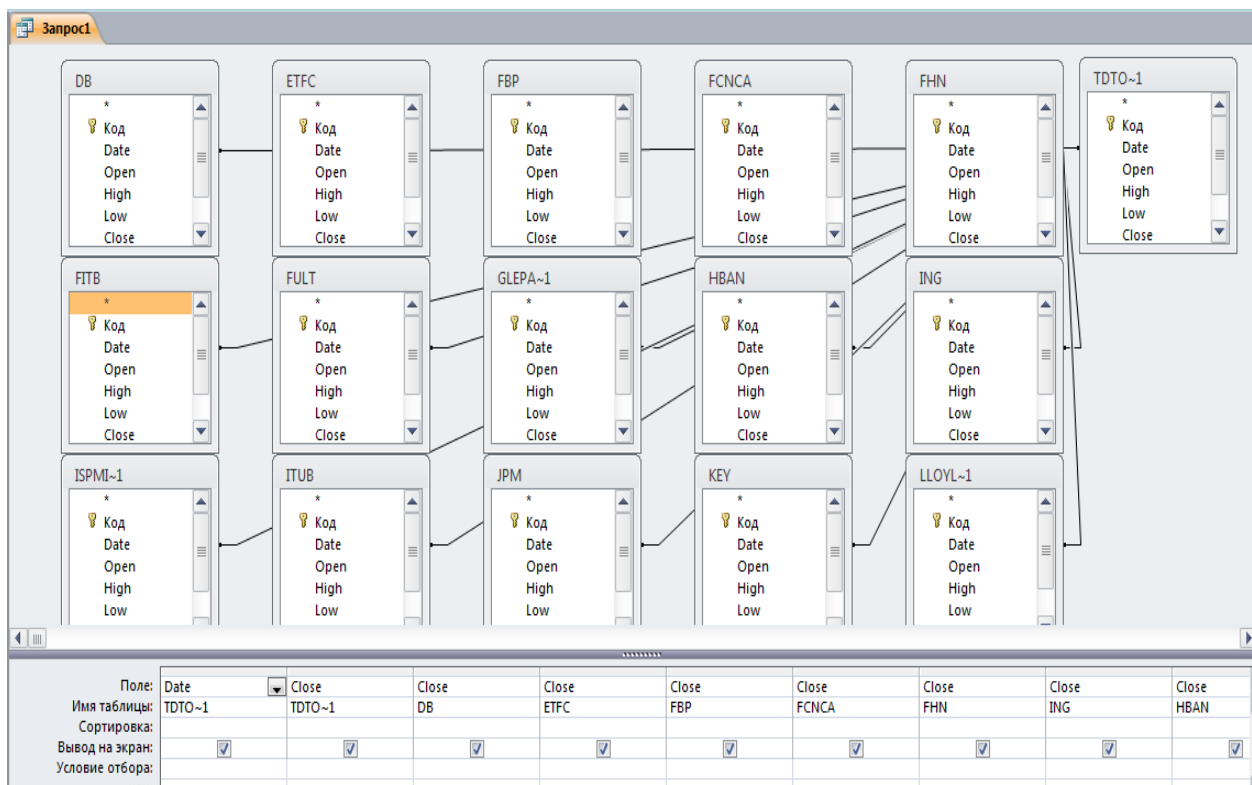


У вікні, що відкрилося, обираємо відповідні часові ряди (рекомендується створювати запити розміром у 15 часових рядів):



Натискаємо опцію «добавить».

4. Синхронізуємо всі ці часові ряди відповідно до однієї дати, і заносимо всі значення adj close у рядок «поле» та «имя таблицы», а також у графі «условие отбора» вказуємо певну дату, до якої необхідно «обрізати» даний запит (у разі необхідності) :



Зберігаємо цей запит.

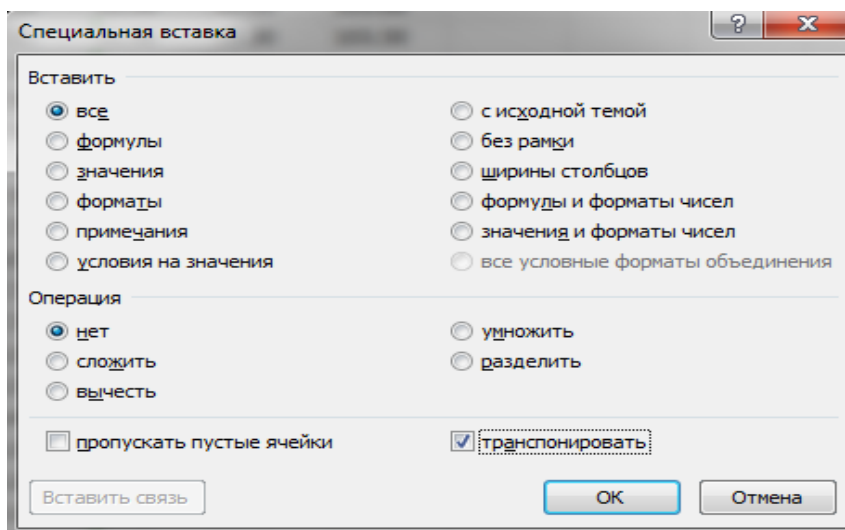
5. Створюємо пусту таблицю і запускаємо її конструктор та виконуємо наступні дії:



8. Експортуємо останній збережений запит, який містить всі необхідні часові ряди, в середовище Excel, в якому необхідно виконати сортування дати від початкової до кінцевої. Далі копіюємо всі значення, але без дати і назви часових рядів, вставляємо в текстовий документ та зберігаємо під відповідною назвою (наприклад bank.txt).

9. Створюємо \*.dt файл. Для створення цього файлу необхідно скопіювати стовбець з датою, вставити його в текстовий документ та зберегти під тією самою назвою, але вказати розширення \*.dt (bank.dt).

10. Створюємо \*.nm файл. Для цього необхідно скопіювати рядок з назвами, на пустій комірці натиснути ПКМ та обрати «спеціальна вставка». В даному вікні встановити галочку на вкладці «транспонировать»:



Натиснути ОК. В результаті всі назви вибудуються у стовпчик. Копіюємо весь стовпчик, вставляємо в текстовий документ та зберігаємо під тією ж самою назвою, що раніше створені документи, але з розширенням \*.nm (bank.nm).

Ці три створенні текстові документи повинні зберігатись разом.

Після виконання всіх вищеописаних дій матрична база даних готова до роботи.

## ЛАБОРАТОРНА РОБОТА № 6

**Тема.** Інформаційні методи оцінки складності економічних систем.

**Мета.** Навчитися використовувати алгоритмічну та ентропійну міри складності за допомогою аналізу часового ряду.

1. Складність. Кількісні міри складності. Інформаційні методи оцінки складності.
2. Складність за Лемпелем-Зівом. Ентропійна складність.
3. Процедура грануляції для мультискейлінгового дослідження часових рядів. Мультимасштабні міри складності.
4. Аналіз складності фондових індексів у періоди, що характеризуються різними рівнями волатильності.

### Теоретичні відомості

**Складність. Кількісні міри складності. Інформаційні методи оцінки складності.**

Дане століття називають століттям складності. Сьогодні питання «що таке складність?» вивчають фізики, біологи, математики і інформатики, хоча при теперішніх

досягненнях у розумінні оточуючого світу, однозначної відповіді на це питання немає.

З цієї причини, відповідно до ідеї І. Пригожина, будемо досліджувати прояви складності системи, застосовуючи при цьому сучасні методи кількісного аналізу складності. Серед таких методів на увагу заслуговують:

- інформаційно-ентропійні;
- засновані на теорії хаосу;
- скейлінгово-мультифрактальні.

Зрозуміло, виходячи з різної природи методів, покладених в основу формування міри складності, вони приділяють певні вимоги до часових рядів, що слугують вхідними даними. Наприклад, перші дві групи методів вимагають стаціонарності вхідних даних. При цьому мають різну чутливість до таких характеристик, як детермінованість, стохастичність, причинність та кореляції. Тому у подальшому, порівнюючи комплексно ефективність різних показників складності, на вказані обставини ми будемо звертати увагу, підкреслюючи спеціально застосовність того чи іншого показника для характеристики різних сторін складності досліджуваних систем.

Розгляд першої групи методів почнемо з добре відомої міри складності, запропонованої А. Колмогоровим.

**Колмогорівська складність.** Поняття колмогорівської складності (або, як ще говорять, алгоритмічної ентропії) з'явилося в 1960-і роки на стику теорії алгоритмів, теорії інформації і теорії ймовірностей.

Ідея А. Колмогорова полягала в тому, щоб вимірювати кількість інформації, що міститься в індивідуальних скінчених об'єктах (а не у випадкових величинах, як у шеннонівській теорії інформації). Виявилось, що це можливо (хоча лише з точністю до обмеженого доданку). А. Колмогоров запропонував вимірювати кількість інформації в скінчених об'єктах за допомогою теорії алгоритмів, визначивши складність об'єкту як мінімальну довжину програми, що породжує цей об'єкт. Дане визначення стало базисом алгоритмічної теорії інформації, а також алгоритмічної теорії ймовірностей: об'єкт вважається випадковим, якщо його складність наближена до максимальної.

Що ж собою являє колмогорівська складність і як її виміряти? На практиці ми часто стикаємося з програмами, які стискають файли (для економії місця в архіві). Найбільш поширені називаються zip, gzip, compress, rar, arj та інші. Застосувавши таку програму до деякого файлу (з текстом, даними, програмою), ми отримуємо його стислу версію (яка, як правило, коротше початкового файлу). За нею можна відновити початковий файл з допомогою парної програми-«декомпресора». Отже, у першому наближенні колмогорівську складність файлу можна описати як довжину його стислої версії. Тим самим файл, що має регулярну структуру і добре стискуваний, має малу колмогорівську складність (порівняно з його довжиною). Навпаки, погано стискуваний файл має складність, близьку до довжини.

Припустимо, що ми маємо фіксований спосіб опису (декомпресор)  $D$ . Для даного слова  $x$  розглянемо всі його описи, тобто всі слова  $y$ , для яких  $D(y)$  визначене і рівне  $x$ . Довжину найкоротшого з них і називають колмогорівською складністю слова  $x$  при даному способі опису  $D$ :

$$KS_D(x) = \min\{l(y) \mid D(y) = x\},$$

де  $l(y)$  позначає довжину слова  $y$ . Індекс  $D$  підкреслює, що визначення залежить від вибору способу  $D$ .

Можна показати, що існують оптимальні способи опису. Спосіб опису тим краще, чим він коротше. Тому природно дати таке визначення: спосіб  $D_1$  не гірше за спосіб  $D_2$ , якщо  $KS_{D_1}(x) \leq KS_{D_2}(x) + c$  при деякому  $c$  і при всіх  $x$ .

Отже, за Колмогоровим, складність об'єкту (наприклад, тексту - послідовності символів)— це довжина мінімальної програми яка виводить даний текст, а ентропія — це складність, що ділиться на довжину тексту. На жаль, це визначення чисто умоглядне.

Надійного способу однозначно визначити цю програму не існує. Але є алгоритми, які фактично якраз і намагаються обчислити колмогорівські складність тексту і ентропію.

### Оцінка складності Колмогорова за схемою Лемпела-Зіва

Універсальна (в сенсі застосовності до різних мовних систем) міра складності кінцевої символічної послідовності була запропонована Лемпелем і Зівом. У рамках їх підходу складність послідовності оцінюється числом кроків процесу, що її породжує. Припустимими (редакційними) операціями при цьому є:

- а) генерація символу (необхідна, як мінімум, для синтезу елементів алфавіту) і
- б) копіювання "готового" фрагмента з передісторії (тобто з уже синтезованої частини тексту).

Нехай  $\Sigma$  - скінчений алфавіт,  $S$  - текст (послідовність символів), складений з елементів  $\Sigma$ ;  $S[i]$  -  $i$ -й символ тексту;  $S[i:j]$  - фрагмент тексту з  $i$ -го по  $j$ -й символ включно ( $i < j$ );  $N = |S|$  - довжина тексту  $S$ . Тоді схему синтезу послідовності можна представити у вигляді конкатенації

$$H(S) = S[1:i_1]S[i_1+1:i_2]...S[i_{k-1}+1:i_k]...S[i_{m-1}+1:N], \quad (1)$$

де  $S[i_{k-1}+1:i_k]$  - фрагмент  $S$ , породжуваний на  $k$ -му кроці, а  $m = m_H(S)$  - число кроків процесу. З усіляких схем породження  $S$  обирається мінімальна за числом кроків. Таким чином, складність послідовності  $S$  за Лемпелем-Зівом

$$c_{LZ}(S) = \min_H \{m_H(S)\}.$$

Мінімальність числа кроків забезпечується вибором для копіювання на кожному кроці максимально довгого прототипу з передісторії. Якщо позначити через  $j(k)$  номер позиції, з якої починається копіювання на  $k$ -му кроці, то довжина фрагмента копіювання

$$l_{j(k)} = i_k - i_{k-1} - 1 = \max_{j \leq i_{k-1}} \{l_j : S[i_{k-1}+1:i_{k-1}+l_j] = S[j:j+l_j-1]\}, \quad (2)$$

а сам  $k$ -й компонент складнісного розкладання (1) можна записати у вигляді

$$S[i_{k-1}+1:i_k] = \begin{cases} S[j(k):j(k)+l_{j(k)}-1] \dots \text{if } j(k) \neq 0, \\ S[i_{k-1}+1] \dots \text{if } j(k) = 0. \end{cases} \quad (3)$$

Випадок  $j(k) = 0$  відповідає ситуації, коли в позиції  $i_{k-1}+1$  стоїть символ, який раніше не зустрічався. При цьому ми застосовуємо операцію генерації символу.

Наприклад, нехай  $\Sigma = \{A, B\}$  і  $S = ABBAABAABBAABBA$ . Схема породження  $S$  має вигляд:

$$\begin{array}{l} H(S) = A \cdot B \cdot B \cdot AB \cdot A \cdot ABBA \cdot ABA \cdot BBA; = 8 \\ j(k): \quad \quad 0 \quad \quad 0 \quad \quad 2 \quad \quad 1 \quad \quad 4 \quad \quad 1 \quad \quad 4 \quad \quad 8 \end{array}$$

Тут компоненти розкладання відокремлені один від одного крапками. Саме розкладання можна трактувати, як подання тексту в термінах повторів, серед яких наявні найбільш значущі (найдовші).

Складнісний аналіз тексту можна проводити у двох режимах: сегментації та



фрагментації. Перший режим розглянуто вище. Він дає інтегральне уявлення про структуру послідовності в цілому і зводиться до розбиття її на непересічні, але взаємопов'язані сегменти (без пробілів). Інший режим зводиться до пошуку окремих фрагментів, що характеризуються аномально низькою складністю, тобто досить високим ступенем структурованості. Такі фрагменти виявляються за допомогою обчислення локальної складності в межах вікон змінної довжини, що ковзають вздовж послідовності. Криві зміни локальної складності вздовж послідовності називаються складнісними профілями. Набір профілів при різних розмірах вікон дозволяє виявити межі аномальних фрагментів і їх взаємозв'язку.

Будемо знаходити складність за Лемпелем-Зівом (LZ) для часового ряду, який являє собою, наприклад, щоденні значення індексу фондового ринку. Для дослідження динаміки LZ та порівняння з іншими фондовими ринками будемо знаходити дану міру складності для підряду фіксованої довжини (вікна). Для цього обчислимо логарифмічні прибутковості та перетворимо їх у послідовність бітів. При цьому можна задавати кількість станів, які диференційовані (система числення). Так, для двох різних станів маємо 0, 1, для трьох – 0,1,2 і т.д. У випадку трьох станів, на відміну від двійкової системи кодування, задається певний поріг  $b$  і стани  $ret$  кодуються так:

$$ret = \begin{cases} 0, ret < -b \\ 1, -b \leq ret \leq b. \\ 2, ret > b \end{cases} \quad (4)$$

Алгоритм виконує дві операції: (1) додає новий біт в уже існуючу послідовність; (2) копіює вже сформовану послідовність. Алгоритмічна складність представляє собою кількість таких операцій, необхідних для формування заданої послідовності.

Для випадкової послідовності довжини  $n$  алгоритмічна складність обчислюється за виразом  $SLZ_r = n/\log(n)$ . Тоді відносна алгоритмічна складність знаходиться як відношення отриманої складності до складності випадкової послідовності:  $SLZ = SLZ_r / SLZ_r$ .

На рис. 1 алгоритмічна складність за Лемпелем-Зівом розрахована для тестових сигналів: періодичної функції  $\sin x$ , білого (white noise) і  $1/f$ -шумів (fnoise) та складного біологічного сигналу – фрагменту електрокардіограми (ECG).

Як і очікувалось, найбільшу складність має випадковий сигнал, а найменшу – періодичний.

Якщо з періодичним сигналом все зрозуміло, то залишається питання із складним біологічним сигналом. Відомо, що такі сигнали являються одними з найскладніших у природі, містять довготривалі кореляції і, принаймні, є складнішими, ніж випадкові послідовності. Окрім того,  $1/f$ -сигнал також є складнішим, аніж білий шум.

Отже, стає зрозумілим, що алгоритмічна складність не в змозі описати складність таких сигналів. Справа в тому, що складні сигнали проявляють притаманну їм складність на різних просторових і часових масштабах, тобто мають масштабно інваріантні властивості. Вони, зокрема проявляються через степеневі закони розподілу. Очевидно, що показники алгоритмічної складності неприйнятні і призводять до помилкових висновків, як у наведеному прикладі з ECG-сигналом та фліккер-шумом.

Можна показати, що використання ентропії Шеннона для оцінки складності принципово нічого нового не дає. Максимальне значення ентропії мають випадкові сигнали.

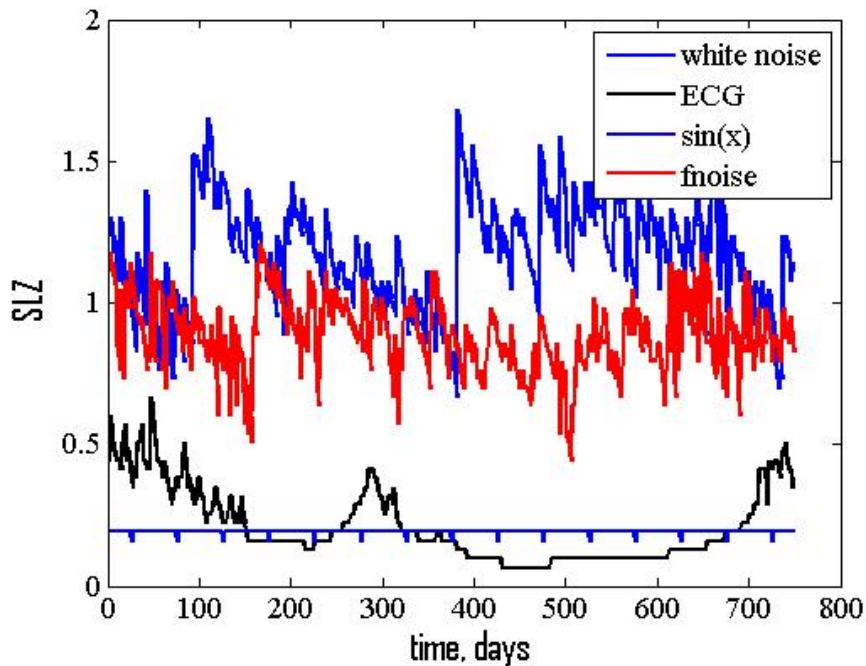


Рис. 1. Динаміка показника алгоритмічної складності у випадку бінарного кодування для фрагментів окремих сигналів довжиною у 1000 точок, шириною вікна 250 точок

Для подолання таких труднощів використовуються мультимасштабні методи, до розгляду яких ми і переходимо.

### Процедура грануляції для мультискейлінгового дослідження часових рядів. Мультимасштабні міри складності

Ідея цієї групи методів включає дві послідовно виконувани процедури: 1) процес «грубого дроблення» (coarse graining – «грануляції») початкового часового ряду – осереднення даних на сегментах, що не перетинаються, розмір яких (вікно осереднення) збільшувався на одиницю при переході на наступний за величиною масштаб; 2) обчислення на кожному з масштабів певного (до сих пір монотмасштабного) показника складності.

Процес «грубого дроблення» («грануляція») полягає в усередненні послідовних відліків ряду в межах вікон, що не перетинаються, а розмір яких  $\tau$  – збільшується при переході від масштабу до масштабу (рис.2). Кожен елемент «гранульованого» часового ряду  $y_j^{(\tau)}$  знаходиться у відповідності до виразу:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq N/\tau,$$

де  $\tau$  характеризує масштабний фактор. Довжина кожного "гранульованого" ряду залежить від розміру вікна і рівна  $N/\tau$ . Для масштабу рівного 1 «гранульований» ряд просто тотожний оригінальному.

Якщо ми застосуємо мультимасштабну версію міри складності Лемпеля-Зіва, то, наприклад, для білого і флікер шумів отримаємо зрозумілі результати (рис. 3).

Бачимо, що для випадкового сигналу (білого шуму - wnoise) складність максимальна на декількох перших масштабах і вже при масштабах більших 5 стає меншою за складність флікер-шуму (1/f-noise).

Введемо також мультимасштабну міру складності, яка є сумою по усіх масштабах міри Лемпеля-Зіва.

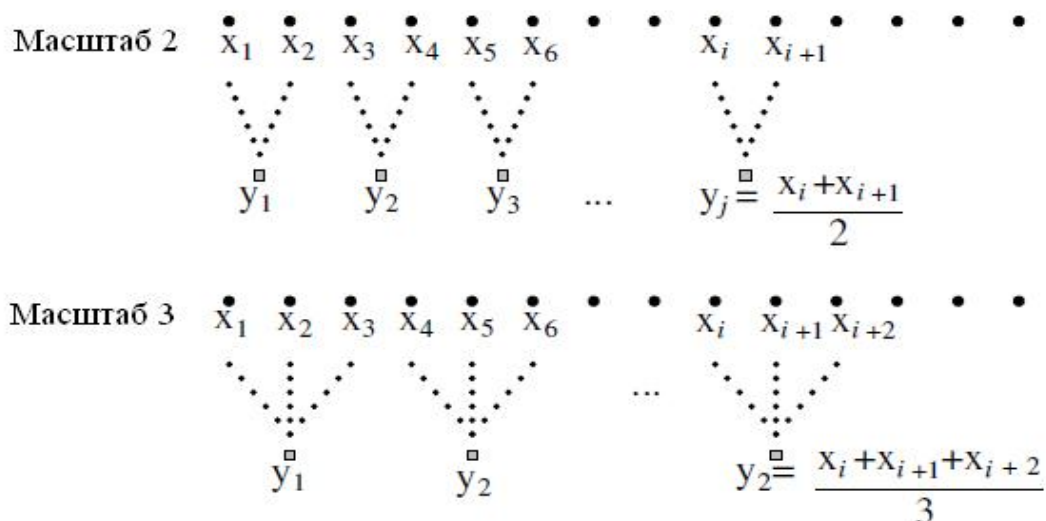


Рис. 2. Схематична ілюстрація процесу грубого дроблення («грануляції») початкового часового ряду для масштабів 2 і 3

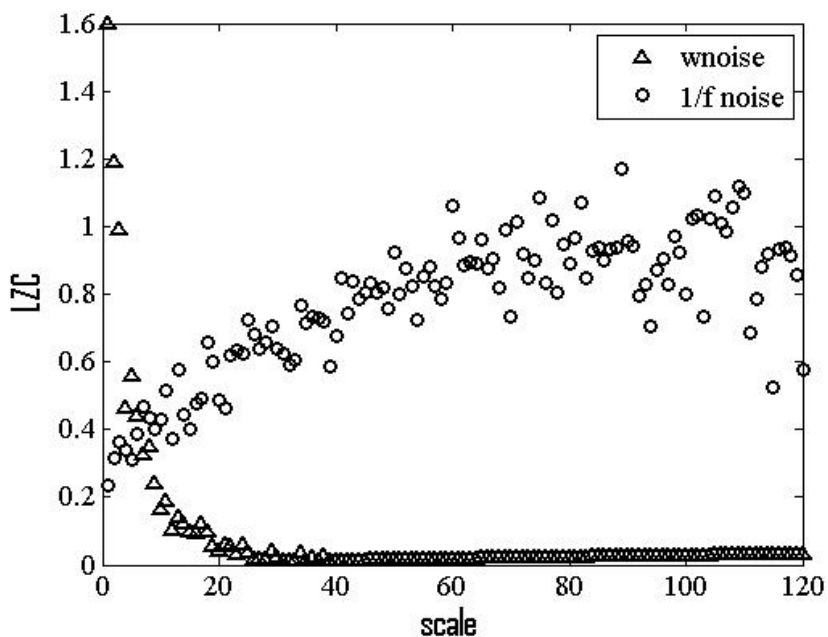


Рис. 3. Залежність міри складності Лемпеля-Зіва (LZC) від масштабу (scale)

Очевидно, що дослідження введених алгоритмічних мір складності дозволить краще зрозуміти динаміку складної системи з урахуванням різних режимів: ламінарного, хаотичного тощо.

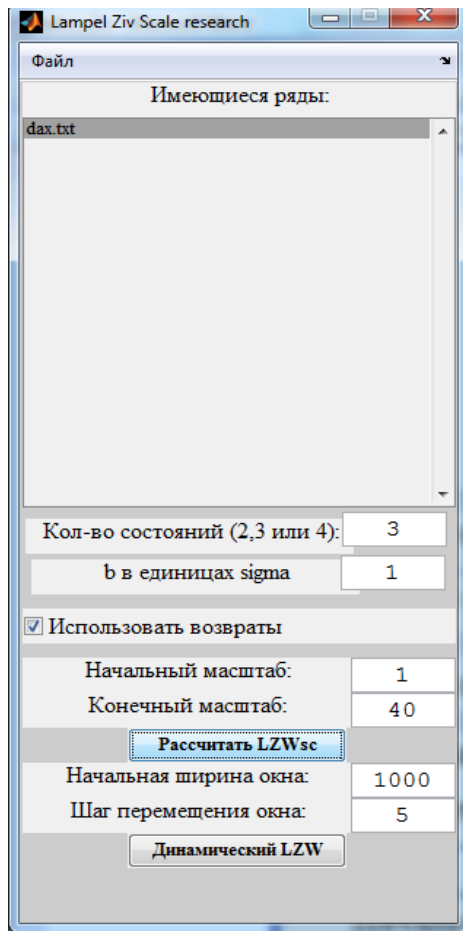
Проведемо такий аналіз на прикладі фондового ринку.

### Хід роботи

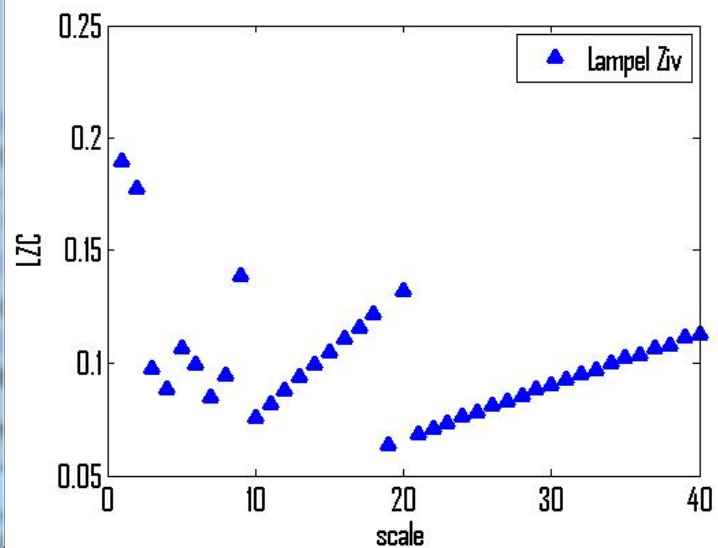
Завантажимо у середовищі Матлаб програмний продукт **work5\_lz**. Завантажимо, наприклад, ряд індексу німецького фондового ринку з 2004 року до поточної дати - **dax.txt** (рис. 4а).

Програмний продукт дає можливість при необхідності працювати з нормалізованими прибутковостями (опція **использовать возвраты**). Можна задавати різну кількість станів (від двох до чотирьох) та критичне значення параметру  $b$  (дивись формулу (4)). Якщо ми

працюємо з прибутковостями, то параметр  $b$  вибирається в одиницях стандартного відхилення.



a)



b)

Рис. 4. Інтерфейс програмного продукту **work5\_lz** (a). Залежність від масштабу міри складності Лемпеля-Зіва (b)

Вибираємо інтервал масштабів, наприклад, від 1 до 40 і розраховуємо поведінку міри складності Лемпеля-Зіва зі зміною масштабу (рис. 5).

Для розрахунку мультимасштабної міри складності у рамках алгоритму рухомого вікна задаємо початкову ширину вікна 500 і крок зміщення 1. Будуємо графіки, які містять порівняння вихідного ряду мірою Лемпеля-Зіва на масштабі 1 (5a), середньою за всіма 40 масштабами (5b) та тримірну картину останньої (рис. 5c). При цьому не обов'язково порівнювати тільки з вихідних рядом. Можна використовувати нормалізовані прибутковості, волатильності тощо.

Порівняння рис. 5 a, b свідчить про подібність основних трендів у періоди вказаних кризових явищ, але помітною є і відмінність між ними. Цей факт буде використано нами у **роботі 8** при визначенні **індексу фрактальності**.

### Шеннонівська складність

Ентропійний аналіз часових рядів за допомогою ентропійних показників різного роду буде проведено у наступній роботі. Зараз же ми розглянемо найпростішу з ентропій - ентропію Шеннона та порівняємо її можливості кількісно оцінювати складність часових послідовностей у порівнянні з мірою Лемпеля-Зіва. Ентропія Шеннона була введена для оцінки невизначеності кодової інформації в каналах зв'язку. Вона обчислюється за формулою:

$$S = -k \sum_{i=1}^n p_i \ln p_i,$$

де  $p_i$  – ймовірність того, що символ  $i$  зустрічається в коді, який містить  $n$  символів,  $k$  — розмірний множник.

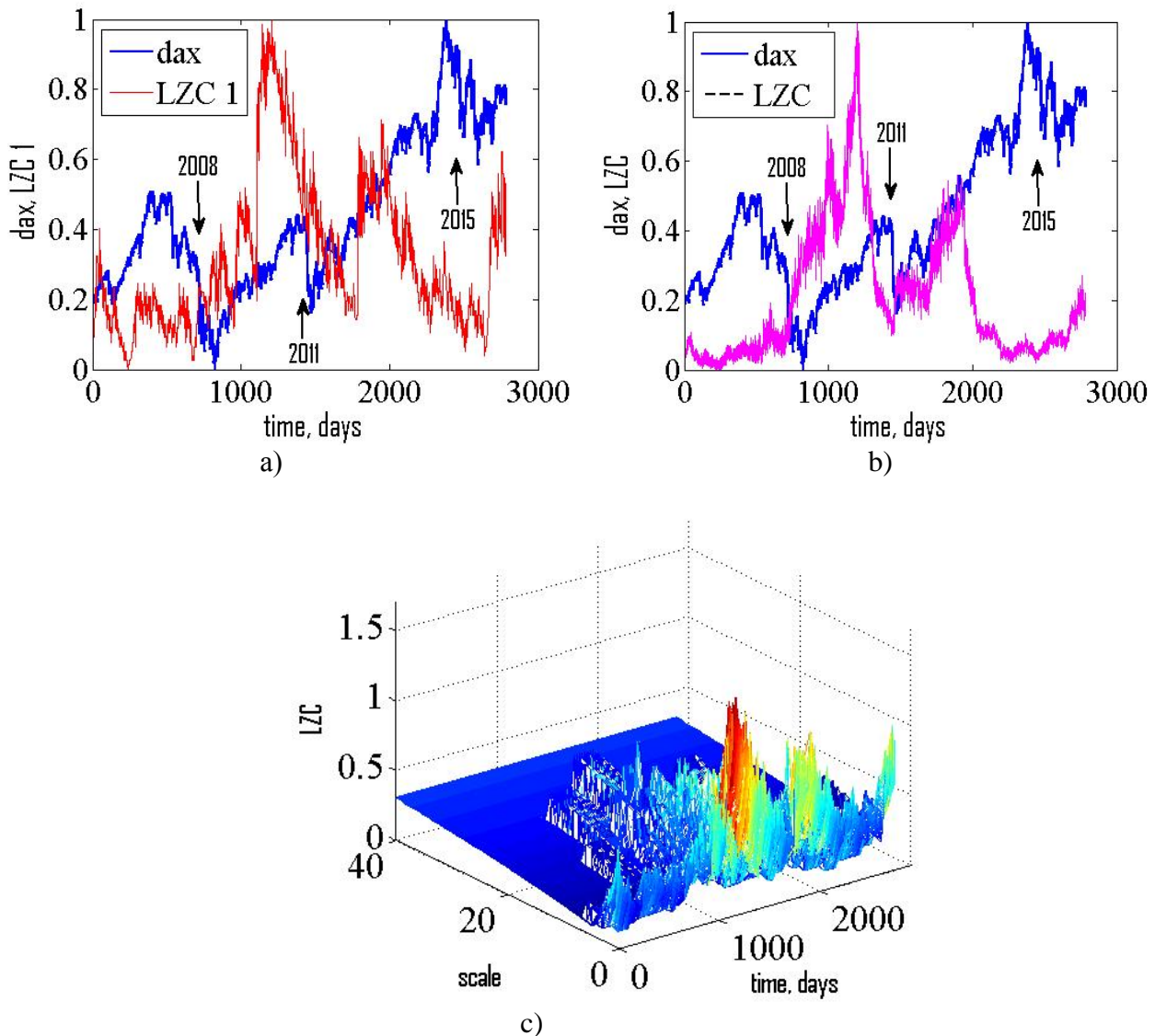
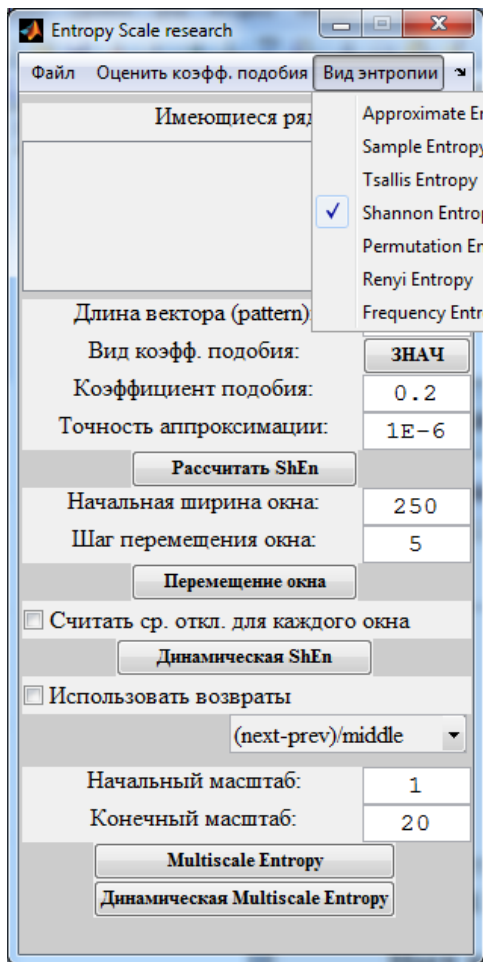


Рис. 5. (а) Мономасштабна та (б) мультимасштабна складності Лемпеля-Зіва. (с) Мультискейлінгова поверхня, яка використовується для розрахунку мультимасштабної міри складності за Лемпелем-Зівом

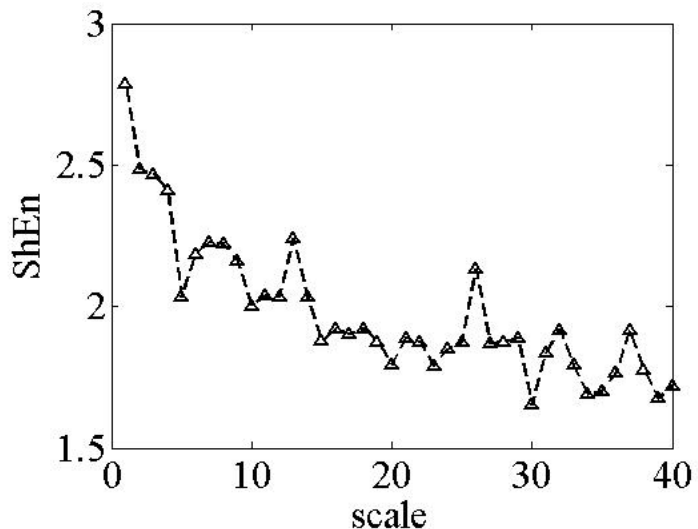
Для часового ряду мова йде про ймовірності мати певні значення вихідного ряду чи прибутковостей.

Дослідимо поведінку шеннонівської ентропії для вихідного ряду і прибутковостей, яка проводиться аналогічним до попереднього показника чином. Для цього застосуємо програмний продукт **work5\_en**. Особливість його застосування в тому, що цей продукт дозволяє будувати й інші ентропійні міри складності. Для вибору типу ентропії скористуємось вікном вибору (**ентропія Шеннона** – рис.6а), завантажимо ряд індексу і повторимо процедури, що ми виконували для міри Лемпеля-Зіва.

Будемо працювати як з вихідним рядом, так і з нормалізованими прибутковостями (**использовать возвраты**).



a)



b)

Рис. 6. Интерфейс программного продукту **work5\_en** (a) та залежність від масштабу міри складності за ентропією Шеннона (b)

На рисунку 7 представлені результати для шеннонівських мір у випадку використання прибутковостей вихідного ряду.

**!!!Розрахунки для значень індексу зробіть самостійно і порівняйте результати.**

### Висновок

Таким чином, розглянуті інформаційні міри складності дозволяють дослідити певні аспекти складності економічних систем. Особливо продуктивним являється мультимасштабна версія введених мір. Ретельний аналіз часових рядів для систем різної природи, різного рівня складності, порівняння їх із тестовими сигналами, вивчення поведінки систем у різних (не обов'язково рівноважних, стаціонарних) умовах дозволить зрозуміти природу складності і спрогнозувати можливу поведінку систем у критичних умовах. Так, порівняння вихідного часового ряду з відповідними мірами складності свідчить про очевидне їх реагування на кризові явища. Однак питання використання їх у якості передвісників вимагає додаткових досліджень.

### Завдання для самостійної роботи

1. Дослідіть і порівняйте результати для тестових рядів різного ступеня складності
2. Проведіть аналіз фондових індексів для країн з розвинутою економікою і таких, що розвиваються. Порівняйте результати. Поясніть, в чому їх схожість та відмінності
3. Яким чином поведуть себе міри складності у період фінансових шоків і криз?
4. Як зміняться результати для міри Лемпеля-Зіва, якщо замість трьох різних станів взяти 2 або 4? Якщо зменшувати (збільшувати) значення порогу  $b$ ?



5. Наскільки чутливими є результати розрахунків до вибору ширини вікна та кроку ?

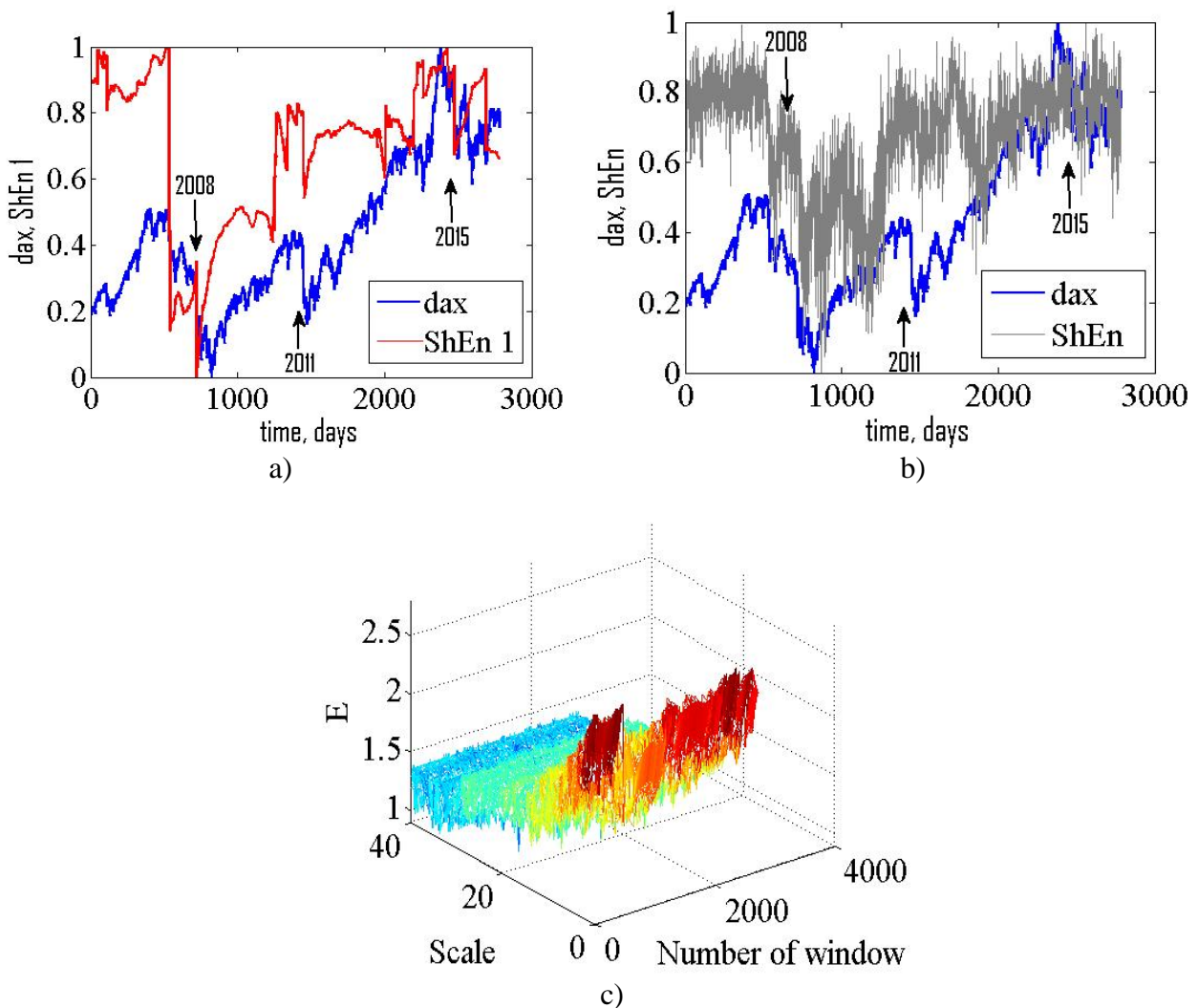


Рис. 7. Мономасштабна (а) та мультимасштабна (б) ентропійні міри складності. (с) Карта мультискейлінгових компонент шеннонівської міри складності

## ЛАБОРАТОРНА РОБОТА № 7

**Тема.** Ентропійний аналіз часового ряду.

**Мета.** Навчитись розраховувати значення різних типів ентропії часового ряду та досліджувати динаміку зміни її значень для оцінки якості прогнозів часових рядів.

1. Нелінійна динаміка і складність.
2. Реконструкція фазового простору методом запізнення.
3. Ентропії подібності, шаблонів, перестановок, мультимасштабна ентропія.
4. Аналіз складності фондових індексів у періоди ентропійними методами.

### Теоретичні відомості

Питання динаміки розвитку і функціонування складних систем може розглядатись у двох варіантах:

- як дослідження шумової активності;



– як детерміністичного випадку з певним ступенем порядку.

Останніми роками було використано кілька підходів для ідентифікації механізмів, що лежать в основі розвитку та функціонування складних систем. Особливо корисні результати було отримано при їх дослідженні методами теорії випадкових матриць, моно- та мультифрактального аналізу, теорії хаосу з реконструкцією траєкторії системи у фазовому просторі та визначення її параметрів, рекурентного аналізу. Ми розглянули ці методи у попередніх роботах. Однак, застосування деяких із методів висуває вимоги до стаціонарності досліджуваних даних, потребує довгих часових рядів та комплексного обчислення кількох параметрів.

Іншим підходом до розгляду питання вивчення особливостей складних систем є обчислення характеристик ентропії. Для практичного застосування у якості міри невизначеності, а значить і складності сигналу, використовують десятки різновидів ентропії.

Концепція термодинамічної ентропії як міри хаосу системи добре відома у фізиці, однак, останніми роками поняття ентропії було застосоване до складних систем інших об'єктів (біологічних, економічних, соціальних тощо). Так, один із найбільш часто використовуваних методів визначення ентропії базується на обчисленні спектру потужності Фур'є та застосовується для вивчення сигналів (часових рядів) різної природи. Проте, використання дискретного перетворення Фур'є для аналізу часових рядів має свої недоліки, зокрема, на результати впливає нестационарність рядів, варіювання їх довжини від сотень до сотень тисяч, та обмеження самого методу (незмінність частотно-часових характеристик протягом всього часу функціонування системи). Тому виникає питання про розрахунок значень ентропії за допомогою інших методів.

Введемо поняття ентропії, скориставшись інформацією, яку можна знайти у Вікіпедії – вільній енциклопедії ([uk.wikipedia.org](http://uk.wikipedia.org)).

**Термодинамічна ентропія**  $S$ , часто просто іменована **ентропія**, в хімії і термодинаміці є мірою кількості енергії у фізичній системі, яка не може бути використана для виконання роботи. Вона також є мірою безладдя, присутнього в системі.

Поняття ентропії була вперше введено у 1865 році Рудольфом Клаузіусом. Він визначив зміну ентропії термодинамічної системи при оборотному процесі як відношення зміни загальної кількості тепла  $\Delta Q$  до величини абсолютної температури  $T$ :

$$\Delta S = \Delta Q / T .$$

Рудольф Клаузіус дав величині  $S$  ім'я «*ентропія*», що походить від грецького слова *τροπή*, «зміна» (зміна, перетворення). Зверніть увагу на те, що рівність відноситься до зміни ентропії.

У 1877 році, Людвіг Больцман зрозумів, що ентропія системи може відноситися до кількості можливих «мікростанів» (мікроскопічних станів) що узгоджуються з їх термодинамічними властивостями. Розглянемо, наприклад, ідеальний газ у посудині. Мікростан визначений як позиція і імпульси кожного атома, що становить систему. Зв'язність пред'являє до нас вимоги розглядати тільки ті мікростани, для яких: (i) місцезрештування всіх частин розташовані в рамках судини, (ii) для отримання загальної енергії газу кінетичні енергії атомів підсумовуються. Больцман постулював що

$$S = k_B \ln \Omega ,$$

де константу  $k_B=1,38 \cdot 10^{-23}$  Дж/К ми знаємо тепер як сталу Больцмана, а  $\Omega$  є числом мікростанів, які можливі в наявному макроскопічному стані. Цей постулат, відомий як принцип Больцмана, може бути оцінений як початок статистичної механіки, яка описує термодинамічні системи використовуючи статистичну поведінку компонентів, із яких вони складаються. Принцип Больцмана зв'язує мікроскопічні властивості системи ( $\Omega$ ) з однією з її термодинамічних властивостей ( $S$ ).

Згідно визначенню Больцмана, ентропія є просто функцією стану. Більш того, оскільки  $(\Omega)$  може бути тільки натуральним числом (1, 2, 3), ентропія повинна бути додатною — виходячи з властивостей логарифма.

У випадку дискретних станів квантової механіки кількість станів підраховується звичайним чином. В рамках класичної механіки мікроскопічний стан системи описується координатами  $q_i$  й імпульсами  $p_i$  окремих частинок, які пробігають неперервні значення. Для підрахунку станів у класичних системах фазовий простір розбивають на невеликі комірки із об'ємом, який відповідає сталій Планка. В такому випадку

$$S = k_B \ln \frac{1}{(2\pi\hbar)^s} \int \prod_{i=1}^s dq_i dp_i ,$$

де  $s$  - число незалежних координат,  $\hbar$  - приведена стала Планка, а інтегрування проводиться по області фазового простору, який відповідає певному макроскопічному стану.

Клод Шеннон (Shannon, 1948) запропонував формулу для оцінки невизначеності кодової інформації в каналах зв'язку, звану ентропією Шеннона:

$$S = -k \sum_{i=1}^n p_i \ln p_i ,$$

де  $p_i$  — вірогідність того, що символ  $i$  зустрічається в коді, який містить  $N$  символів,  $k$  — розмірний множник.

Зв'язок між ентропією і інформацією можна прослідкувати на наступному прикладі. Розглянемо тіло при абсолютному нулі температури, і хай ми маємо повну інформацію про координати і імпульси кожної частинки. Для простоти покладемо, що імпульси всіх частинок рівні нулю. В цьому випадку термодинамічна ймовірність рівна одиниці, а ентропія – нулю. При кінцевих температурах ентропія в рівновазі досягає максимуму. Можна зміряти всі макропараметри, що характеризують даний макростан. Проте ми практично нічого не знаємо про мікростан системи. Точніше кажучи, ми знаємо, що даний макростан можна реалізувати за допомогою дуже великого числа мікростанів. Таким чином, нульовій ентропії відповідає повна інформація (ступінь незнання рівний нулю), а максимальної ентропії – повне незнання мікростанів (ступінь незнання максимальний).

У теорії інформації ентропія (інформаційна ентропія) визначається як кількість інформації. Нехай  $P$  – апіорна вірогідність деякої події (ймовірність до проведення досвіду), а  $P_1$  – ймовірність цієї події після проведення досвіду. Для простоти вважатимемо, що  $P_1 = 1$ . За Шенноном, кількість інформації  $I$ , яка дає точну відповідь (після проведення експерименту)

$$I = K \log P.$$

Наприклад, при грі в орлянку можливі дві відповіді (монета падає тією або іншою стороною на підлогу), тобто  $P = 2$ . При киданні кістки з шести можливих варіантів випадає 1, тобто  $P = 6$ . Якщо за основу логарифма вибрати 2 і покласти  $K = 1$ , то при киданні монети ми одержуємо інформацію

$$I = \log_2 2 = 1.$$

Ця кількість інформації, за визначенням, дорівнює одному біту.

Фізичний сенс  $I$  – це міра нашого незнання. Іншими словами,  $I$  – це та інформація, яку ми можемо одержати, вирішивши завдання. У прикладі (тіло при абсолютному нулі температури), що розглядається вище, міра нашого незнання рівна нулю, оскільки  $P = 1$ .

Після проведення досвіду ми одержуємо нульову інформацію  $I = 0$ , оскільки все було відомо до досвіду. Якщо розглядати тіло при кінцевих температурах, то до проведення досвіду число мікростанів, а отже, і  $P$  дуже велике. Після проведення досвіду ми одержуємо велику інформацію, оскільки нам стають відомими координати і імпульси всіх частинок.

Аналогія між кількістю інформації і ентропією  $S$ , визначуваною з принципу Больцмана, очевидна. Досить покласти множник  $K$  рівним постійній Больцмана  $k_B$  і використовувати натуральний логарифм. Саме з цієї причини величину  $I$  називають інформаційною ентропією. Інформаційна ентропія (кількість інформації) була визначена по аналогії із звичайною ентропією, і вона має властивості, характерні для звичайної ентропії: адитивність, екстремальні властивості і т.д. Проте ототожнювати звичайну ентропію з інформаційною не можна, оскільки неясно, яке відношення має друге начало до інформації. Нагадаємо, що екстенсивна величина – ця така характеристика системи, яка росте із збільшенням розмірів системи, тобто, якщо наша система складається з двох незалежних підсистем  $A$  і  $B$ , то ентропію всієї системи можна одержати складанням ентропій підсистем:

$$S(A+B) = S(A) + S(B).$$

Саме ця властивість і означає екстенсивність, або адитивність, ентропії.

До сьогодення моменту продовжують з'являтися узагальнення ентропійних формул. Так, в 1970 р. Анрі Реньї запропонував міру

$$S_\beta = \frac{1}{1-\beta} \ln \sum_{i=1}^N (n_i)^\beta,$$

що співпадає при значенні параметра  $\beta = 1$  з ентропією Шеннона. Константіно Тсалліс (1987) узагальнив формулу Шеннона на випадок неекстенсивності (не адитивності) ентропії.

У статистичній фізиці ентропія трактується через число мікростанів системи. Прагнучи, проте, визначити ентропію так, щоб вона залишалася екстенсивною величиною, статистична фізика (а точніше, кінетика) вимушена привертати недоказову гіпотезу молекулярного хаосу. Ця гіпотеза полягає у тому, що будь-які молекули, що стикаються, були до зіткнення ніяк не скорельовані, тобто, ніяк не відчували один одного. Для багатьох (але не для всіх!) систем це, звичайно, цілком здорова аксіома. І саме з неї виходить згаданий вище знаменитий вираз для ентропії замкнутої системи (вигравійований, до речі, на надгробній плиті Больцмана). У статфізиці показується, що така певна ентропія насправді є адитивною величиною і приймає максимальне значення в стані рівноваги.

Причина цього в тому, що у більшості систем сили взаємодії між частинами системи короткодіючі: кожна молекула відчуває лише декілька найближчих сусідів. Саме звідси слідує те, що термодинаміка таких макроскопічних систем — екстенсивна. Саме такі системи оточують нас в повсякденному житті. Але в природі є і далеко діючі сили: гравітаційні, інформаційні та ін. Виявляється, що в складних системах еволюція системи в даний момент часу залежить не тільки від параметрів системи в цей конкретний момент часу, але і від її параметрів якийсь час назад. Ефекти пам'яті можуть легко привести до порушення гіпотези молекулярного хаосу. Дійсно, вони можуть означати, що окремі частинки перед зіткненням «пам'ятають» одне одного, їх рух не є повністю некорельованим. А як ми пам'ятаємо, припущення молекулярного хаосу лежить в основі больцмановського виразу для ентропії. Значить, цей вираз не годиться для систем з пам'яттю. Отже головний висновок для нас полягає в наступному твердженні.

***Існують системи, в яких є сильні кореляції, сильна взаємодія між всіма частинами системи. Це приводить до порушення термодинамічної адитивності системи, тому їх не вдається описати больцманівською статистикою і термодинамікою. Значить, необхідний інший підхід, який***

**якимсь чином зумів би справитися з режимом сильного зв'язку між всіма частинками системи.**

Весь цей підхід базується на сміливій ідеї Тсалліса:

**цілком можливо, що сильна взаємодія в термодинамічно аномальних системах настільки міняє картину, що приводить до абсолютно нових ступенів свободи, до абсолютно іншої статистичної фізики не-больцманівського типу: окремі частинки з больцманівською статистикою + сильна взаємодія => нові ступені свободи з не-больцманівською статистикою + відсутність взаємодії**

Ця гіпотеза не доведена. Більш того, досконало незрозуміло, що це за нові ступені свободи. Тому це твердження залишається поки лише гіпотезою. Але з другого боку, інтуїція підказує, що такий перехід в тому або іншому вигляді насправді може мати місце і важливо зрозуміти, до чого це може привести.

Якщо майже нічого невідоме про ці нові ступені свободи, то як же тоді можна одержати щось конструктивне? Тсалліс запропонував наступне. Наша єдина зачіпка полягає у тому, що статистика буде не-больцманівська. Це значить, що вираз для ентропії буде вже іншим. Так давайте придумаємо яку-небудь формулу для ентропії, яка, по-перше, переходила б в стандартну формулу в межі слабкого зв'язку, а по-друге, змогла б описати неекстенсивні системи.

Тсалліс узяв стандартний вираз для ентропії і замість логарифма ввів нову функцію – степеневу:

$$\ln(x) \Rightarrow \ln_q(x) \Rightarrow (x^{1-q} - 1)/(1 - q)$$

з деяким числовим параметром  $q$ . Зауважте, що при  $q$ , що прямує до 1,  $\ln_q(x)$  переходить в справжній логарифм, в чому можна переконатися простим диференціюванням. Нова формула для  $q$ -ентропії виглядає так:

$$S_q = -\sum_i (p_i^q \ln_q(p_i)) = (1 - \sum_i p_i^q)/(q - 1).$$

Якщо  $q \rightarrow 1$ , то  $q$ -ентропія переходить в стандартну больцманівську ентропію.

Головний наслідок такої заміни:  $q$ -ентропія є вже неекстенсивною функцією. Якщо всю систему розбити на дві незалежні підсистеми  $A$  і  $B$  (нагадаємо, що ми вже перейшли до нових *невзаємодіючих*, а значить, і *незалежних* ступенів свободи!), то ми одержимо:

$$S_q(A + B) = S_q(A) + S_q(B) + (1 - q)S_q(A)S_q(B).$$

Отже, параметр  $q$  – це міра неекстенсивності системи.

### **Застосування ентропії**

Не менш різноманітні і приклади застосування поняття ентропії. Ідеологія всіх таких додатків пов'язана з екстремальними принципами природознавства – зростаюче значення ентропії параметризує зміну стану систем самої різної природи в процесі їх «природної» еволюції. Відповідний принцип розвитку одержав назву «принцип максимуму ентропії». Для закритих термодинамічних систем їх ентропія зростає («друге начало» термодинаміки). Розповсюдження «другого начала» на Всесвіт і всі процеси в ньому привело до уявлення про

ентропійну «стрілу часу». Принцип максимуму ентропії застосовується, зокрема, у фізиці, лінгвістиці, екології, математиці, механіці, економіці і інформатиці.

Розглянемо далі три найбільш поширені форми кількісного аналізу складних сигналів: ентропії подібності (Approximate Entropy - *ApEn*) та ентропії шаблонів (Sample Entropy – *SampEn*) та ентропії перестановок (Permutation Entropy - *PermEn*).

На відміну від ентропії Шеннона, яку ми розраховували для вихідного ряду, вказані види ентропій розраховуються з використанням підходів, розроблених у рамках теорії детермінованого хаосу. Справа в тому, що ми не можемо апріорі сказати щодо вихідного ряду, чи є він результатом реалізації динамічної системи, чи являється просто випадковим шумовим сигналом. Якщо це динамічна система, то що можна сказати щодо її властивостей і характеристик – скільки змінних необхідно для опису стану, яка фрактальна розмірність атрактора, що відповідає за спостережуваний режим, чи є він хаотичним, чи можна сконструювати модель у вигляді диференціальних рівнянь, яка дозволяла б адекватно відтворити спостережувану часову залежність і прогнозувати майбутній стан системи?

Якщо такий аналіз можливий, то апарат нелінійної динаміки очевидно стає потужним інструментом дослідження, що дозволяє робити висновки щодо структури об'єкта і визначати шляхи управління його динамікою через зовнішні впливи.

Один із можливих прийомів такого аналізу, який називається **реконструкцією фазового простору методом запізнення** (delay-time reconstruction) ми розглянули раніше при аналізі рекурентних властивостей складних систем.

**Ентропія подібності.** Ентропія подібності (Approximate Entropy, *ApEn*) є «статистикою регулярності», що визначає можливість передбачувати флуктуації в часових рядах. Інтуїтивно вона означає, що наявність повторюваних шаблонів (послідовностей певної довжини, побудованих із чисел ряду, що слідує одне за іншим) флуктуацій у часовому ряді призводить до більшої передбачуваності часового ряду порівняно із рядами, де повторюваності шаблонів немає. Порівняно велике значення *ApEn* показує ймовірність того, що подібні між собою шаблони спостережень не будуть слідувати один за одним. Іншими словами, часовий ряд, що містить велику кількість повторюваних шаблонів, має порівняно мале значення *ApEn*, а значення *ApEn* для менш передбачуваного (більш складного) процесу є більшим.

При розрахунку *ApEn* для даного часового ряду  $S_N$ , що складається із  $N$  значень  $t(1)$ ,  $t(2)$ ,  $t(3)$ , ...,  $t(N)$ , вибираються два параметри,  $m$  та  $r$ . Перший з цих параметрів,  $m$ , вказує довжину шаблону, а другий –  $r$  – визначає критерій подібності. Досліджуються підпоследовності елементів часового ряду  $S_N$ , що складаються з  $m$  чисел, взятих, починаючи з номера  $i$ , і називаються векторами  $p_m(i)$ . Два вектори (шаблони),  $p_m(i)$  та  $p_m(j)$ , будуть подібними, якщо всі різниці пар їх відповідних координат є меншими за значення  $r$ , тобто якщо

$$|t(i+k) - t(j+k)| < r \text{ для } 0 \leq k < m.$$

Для розглядуваної множини  $P_m$  всіх векторів довжини  $m$  часового ряду  $S_N$  можна обраховуються значення

$$C_{im}(r) = \frac{n_{im}(r)}{N - m + 1},$$

де  $n_{im}(r)$  – кількість векторів у  $P_m$ , що подібні вектору  $p_m(i)$  (враховуючи вибраний критерій подібності  $r$ ). Значення  $C_{im}(r)$  є часткою векторів довжини  $m$ , що мають схожість із вектором такої ж довжини, елементи якого починаються з номера  $i$ . Для даного часового ряду обраховуються значення  $C_{im}(r)$  для кожного вектора у  $P_m$ , після чого знаходиться

середнє значення  $C_m(r)$ , яке виражає розповсюдженість подібних векторів довжини  $m$  у ряду  $S_N$ . Безпосередньо ентропія подібності для часового ряду  $S_N$  з використанням векторів довжини  $m$  та критерію подібності  $r$  визначається за формулою:

$$ApEn(S_N, m, r) = \ln \left( \frac{C_m(r)}{C_{m+1}(r)} \right),$$

тобто, як натуральний логарифм відношення повторюваності векторів довжиною  $m$  до повторюваності векторів довжиною  $m+1$ .

Таким чином, якщо знайдуться подібні вектори у часовому ряді,  $ApEn$  оцінить логарифмічну ймовірність того, що наступні інтервали після кожного із векторів будуть відрізнятися. Менші значення  $ApEn$  відповідають більшій ймовірності того, що за векторами слідує подібні їм. Якщо часовий ряд дуже нерегулярний – наявність подібних векторів не може бути передбачуваною і значення  $ApEn$  є порівняно великим.

Зауважимо, що  $ApEn$  є нестійкою до вхідних даних характеристикою, оскільки досить сильно залежить від параметрів  $m$  та  $r$ .

**Ентропія шаблонів.** При розрахунку  $ApEn$ , як було показано вище, враховуються подібності певного вектора  $p_n(i)$  до самого себе, що використовується для звільнення від можливого значення  $\ln 0$ , яке може виникнути при відсутності подібних до даного векторів. Однак, вказана особливість призводить до нівелювання двох важливих характеристик у ентропії подібності:

- $ApEn$  сильно залежить від довжини розглядуваного шаблону (вектора) і є нижчою, ніж очікується, для векторів малої розмірності;
- $ApEn$  не враховує відносну щільність даних.

Це означає, що коли значення  $ApEn$  для одного ряду є більшим, ніж для іншого, то воно повинно залишатись таким (проте не є) для будь-яких можливих початкових умов. Такий висновок тим більш важливий, оскільки  $ApEn$  рекомендується в якості міри порівняння двох наборів даних різними авторами.

Враховуючи вказані обмеження, розроблена для розрахунку інша характеристика, – **ентропія шаблонів** (Sample Entropy,  $SampEn$ ).

При розрахунку  $SampEn$ , на відміну від алгоритму  $ApEn$ , додаються дві умови:

- не враховується подібність вектора самому собі;
- при розрахунку значень умовних ймовірностей  $SampEn$  не використовується довжина векторів.

На основі аналізу вищезазначеного можна зробити висновок про те, що  $SampEn$ :

- більше, ніж  $ApEn$ , відповідає теорії випадкових чисел для ряду із відомою функцією щільності розподілу;
- зберігає відносну щільність, в той час як  $ApEn$  втрачає дану характеристику;
- додає значно меншу помилку до розрахованого значення у випадку використання векторів малої розмірності.

**Ентропія перестановок** (Permutation Entropy) – це метод дослідження, який ефективно використовується для кількісного та якісного визначення динамічних змін у складних системах.

**Алгоритм обрахунку.** Розглянемо знову вибірку  $\{x_k\}$  з досліджуваного часового ряду  $\{x\}$ . Перестановочна ентропія вибірки порядку  $n$  визначається як Шеннонівська ентропія  $n!$  перестановок, що визначаються відношеннями  $n$  поспідовних значень вибірки. Так, наприклад, перестановочна ентропія порядку  $n = 2$ , визначається так. Нехай у вибірці  $\{x_k\}$  міститься  $N$  значень ряду, тобто  $1 \leq k \leq N$ . Порівнюємо два сусідніх значення  $x_k$  і  $x_{k+1}$ . Знаходимо число  $m_1$  випадків у вибірці, коли  $x_k < x_{k+1}$ , і число випадків у вибірці  $m_2$ , коли

$x_{k+1} < x_k$ . Якщо всі  $x_k$  розрізніми, тобто немає рівних значень  $x_k$ , тоді  $m_1 + m_2 = N, p_{2,1} = m_1 / N, p_{2,2} = m_2 / N$ . Перестановочна ентропія порядку  $n=2$ , дорівнює:  $H_2 = -p_{2,1} \log_2 p_{2,1} - p_{2,2} \log_2 p_{2,2}$ .

Ентропія перестановок порядку  $n=3$ , відповідно, визначається наступним чином. Знаходимо число  $m_1$  випадків у вибірці, коли  $x_k < x_{k+1} < x_{k+2}$ ,  $m_2$  - коли  $x_{k+1} < x_k < x_{k+2}$ , і так далі. Число таких перестановок в даному випадку буде  $3!=6$ . Тоді

$$\sum_{j=1}^{3!} m_j = N, \dots p_j = m_j / N, \dots H_3 = -\sum_{j=1}^{3!} p_j \log_2 p_j.$$

Відповідно, перестановочна ентропія вибірки порядку  $n$  визначається так:

$$\sum_{j=1}^{n!} m_j = N, \dots p_j = m_j / N, \dots H_n = -\sum_{j=1}^{n!} p_j \log_2 p_j.$$

Тут, як і у випадку класичної ентропії Шеннона, ентропія перестановок виражається в бітах. Зрозуміло, її можна виражати і в «натах», тобто використовувати натуральні логарифми, і в «датах», тобто користуватися десятковими логарифмами, але потрібно враховувати, що «біти» все ж є загальноприйнятою розмірністю для Шеннонівських ентропій.

Максимально можливе значення ентропії перестановок будь-якого порядку  $MaxH_n$  досягається, як зазвичай, у рівномірнісному випадку, тобто, коли у вибірці всі випадки  $m_j$  рівні. Очевидно, що  $MaxH_n = \log_2 n!$ . Наприклад, такі значення перестановочної ентропії може мати нескінченний ряд випадкових чисел. Мінімальне значення ентропії  $MinH_n = \log_2 n!$  досягається у випадках, коли у всій вибірці реалізується тільки одна з  $n!$  перестановок. Наприклад, це може бути ряд монотонно зростаючих або монотонно убуючих значень.

При практичній роботі з ентропією перестановок виникає питання щодо необхідного і достатнього числа порядків  $n$  при роботі з конкретними рядами і вибірками. Очевидним обмеженням є співвідношення  $n! < N$ , де  $N$  - обсяг вибірки. Однак практика показує, що прийнятні за точністю значення виходять тільки при  $N > 2n!$ . Комп'ютерні програми, побудовані за вищеописаним алгоритмом з урахуванням цих обмежень, дають стійкі і досить точні значення ентропії перестановок різних порядків.

Як і у випадку інформаційних мір, досліджувані ентропійні міри часто виявляються нездатними ідентифікувати дійсно складні сигнали. Тому аналогічно можна використовувати масштабний аналіз ентропії (Multiscale Entropy Analysis – MSE), де у якості міри ентропії на різних масштабах декомпозиції початкового часового ряду використовувався параметр ентропії подібності, шаблонів або перестановок.

**Фур'є-ентропія.** В теорії складних систем достатньо поширеними є Фур'є- та вейвлет-аналіз, оскільки завдяки ним можна проводити кількісний аналіз часових рядів, що базується на перенесенні інформації з часової області до частотної або області, що задається параметрами вейвлет-перетворення.

Основою гармонійного аналізу є інтегральне перетворення Фур'є та ряди Фур'є. Класичний аналіз Фур'є заснований на можливості дослідження функцій у часі і у частотному просторі.

За Фур'є будь-яка  $2\pi$ -періодична квадратично інтегровна функція може бути представлена у вигляді нескінченної суми періодичних комплексних експоненціальних функцій:



$$S(\omega) = \int_{-\infty}^{\infty} s(t) e^{-j\omega t} dt,$$

де  $S(\omega)$  – отримувана спектральна функція.

При роботі з часовими рядами використовується дискретне перетворення Фур'є, що описується наступним виразом:

$$X(n) = \sum_{k=0}^{N-1} x(k) \exp\left(-j \frac{2\pi n k}{N}\right),$$

в якому  $X(n)$  – коефіцієнти розкладу часового ряду у ряд Фур'є.

Для подальшого аналізу спектру коефіцієнти ряду Фур'є перетворюються наступним чином:

$$Y(n) = |X(n)|^2,$$

після чого на основі отриманих значень розраховується ентропія Шеннона для отриманого спектру.

Фур'є-аналіз не завжди є ефективним для опису процесів при аналізі сигналів, оскільки не враховується еволюція частоти сигналу в часі; не розрізняється, наприклад, сигнал, що є сумою двох синусоїд із різними частотами, від сигналу з тими ж синусоїдами, але які включаються послідовно одна за другою; частота сигналу є обернено пропорційною до його довжини, тому необхідно враховувати довжину часового ряду тощо. Ще однією проблемою Фур'є-аналізу є вимога стаціонарності ряду, що в економіці є доволі рідкісним явищем. Віконне перетворення Фур'є (або зважене перетворення Фур'є) покликано подолати описані вище труднощі.

Застосування Фур'є-ентропії є інноваційним методом, що дозволяє провести спектральний розклад зашумлених, хаотичних рядів, а саме такими є більшість економічних рядів.

### **Вейвлет-ентропія**

Необхідність дослідження часових рядів із різними частотними характеристиками призвела до розвитку методів, що базуються на частотно-часовому аналізі. Вейвлетне часово-частотне представлення ряду не робить жодних припущень стосовно його стаціонарності і дозволяє визначати динамічні зміни разом з їх локалізацією.

При проведенні вейвлет-аналізу поданий часовий ряд (сигнал) розкладається на сигнали з кінцевою енергією у ортонормованому базисі  $L^2(R)$  шляхом відображення та масштабування наперед заданої функції  $\psi$  із заданими властивостями, що називається материнським вейвлетом. Внаслідок декомпозиції сигналу отримується набір вейвлет-коефіцієнтів (здебільшого, комплексних),  $C_j(t)$ , що характеризуються двома параметрами: масштабом материнського вейвлета  $j$  та його локалізацією  $t$ . Оскільки енергія сигналу визначається як квадрат функції, що його описує  $E = \int s^2 ds$ , застосування поняття енергії до коефіцієнтів вейвлет-перетворення, що теж інтерпретуються як сигнал у просторі  $L^2(R)$ , дає змогу використовувати методи розрахунку ентропії у частотно-часовому просторі.

При розрахунку коефіцієнтів використовуються наступні типи вейвлет-перетворень:

– неперервне вейвлет-перетворення – розрахунок неперервних вейвлет-коефіцієнтів часового ряду  $T$  на цілих додатніх масштабах з використанням в якості материнського вейвлета похідної 8-го порядку функції Гауса;

– дискретне вейвлет-перетворення – розрахунок дискретних вейвлет-коефіцієнтів часового ряду  $T$  на дійсних додатніх масштабах у просторі з ортонормованим базисом із сімейства материнських функцій Морле з шістьма хвилями та періодами, що є дійсними степенями двійки;

– дискретне вейвлет-перетворення з цілими масштабами – розрахунок дискретних вейвлет-коефіцієнтів часового ряду  $T$  на цілих додатніх масштабах у просторі з базисом із сімейства материнських функцій, що є похідними 2-го порядку функції Гауса.

В якості базової формули розрахунку значення ентропії нами було обрано ентропію Шеннона внаслідок досить простого методу отримання розподілу щільності ймовірності енергії сигналу.

На основі енергії вейвлет-коефіцієнтів було визначено два показники вейвлет-ентропії – масштабової та точкової ентропії. У випадку розрахунку масштабової вейвлет-ентропії формула Шеннона застосовується до оброблених даних по масштабам, в іншому випадку обробка даних проводиться по часовій шкалі.

При розрахунку масштабової ентропії спочатку визначається розподіл щільності ймовірності енергій  $p_{ij} = \frac{C_{ij}^2}{E_{tot}}$ , де  $C_{ij}$  – вейвлет-коефіцієнт на масштабі  $i$  в момент часу  $j$ ,

$E_{tot} = \sum_i \sum_j C_{ij}^2$ . Для розрахунку ентропії знаходяться значення функції

$$E_{WT} = -\sum_i \left( \left( \sum_j p_{ij} \right) \cdot \log_2 \sum_j p_{ij} \right).$$

Для розрахунку точкової ентропії визначаються суми енергій на кожному масштабі  $SM_i = \sum_j C_{ij}^2$  та знаходиться розподіл щільності ймовірності енергій шляхом ділення енергій

вейвлет-коефіцієнтів на сумарну енергію масштабу  $p_{ij} = \frac{C_{ij}^2}{SM_i}$ . Значення ентропії у кожній

точці знаходиться за формулою Шеннона  $E_{WT} = -\sum_j \left( \left( \sum_i p_{ij} \right) \cdot \log_2 \sum_i p_{ij} \right)$ .

При розрахунку енергій вейвлет-коефіцієнтів проводиться нормалізація значень середнім квадратичним відхиленням, а у формулу Шеннона додається нормалізуючий множник  $1/\log_2 N$ , де  $N$  – кількість елементів ряду.

### Хід роботи

Ентропійний аналіз у випадках як номомасштабному, так і мультимасштабному розраховуються засобами програмного продукту **work5\_en** (рис.1).

значення описаних вище видів ентропій, так і досліджувати динаміку їх змін, використовуючи процедуру рухомого вікна.

Для розрахунку значення обраної ентропії виберемо відповідну опцію та завантажимо ряд із текстового файлу, використовуючи меню **Файл** → **Читати**.

Розрахунок значення конкретного ентропійного показника передбачає введення певних параметрів, суть і значення яких описані у теоретичній частині роботи.

Програма дозволяє для обраного часового ряду отримувати як безпосередньо

На прикладі часового ряду, що являє собою щоденні значення індексу фондового ринку США – індексу Доу-Джонса (djia) за період з 02/01/2004 по 11/11/2016рр. розглянемо окремі види ентропійних показників і оцінимо їх чутливість до характерних змін на вказаному ринку акцій.

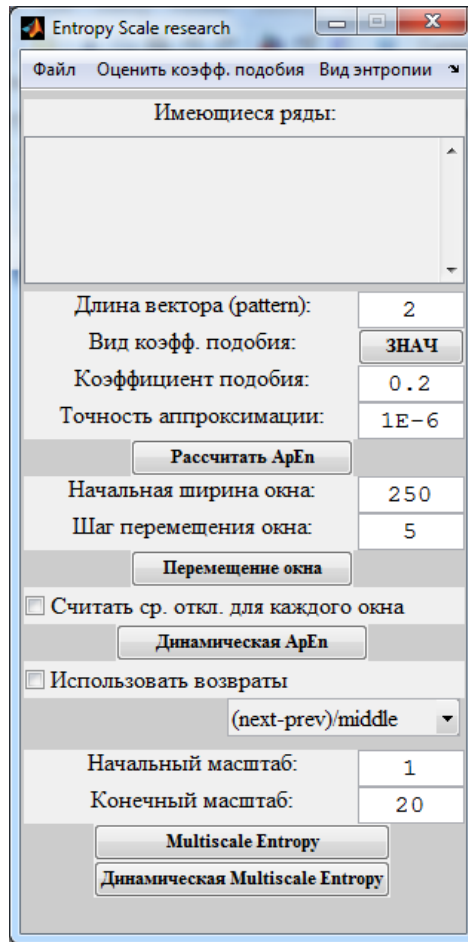


Рис. 1. Видягу інтерфейсу для розрахунку ентропійних показників

На рис. 2 та 2' представлені результати аналізу часового ряду шляхом розрахунку ентропії шаблонів.

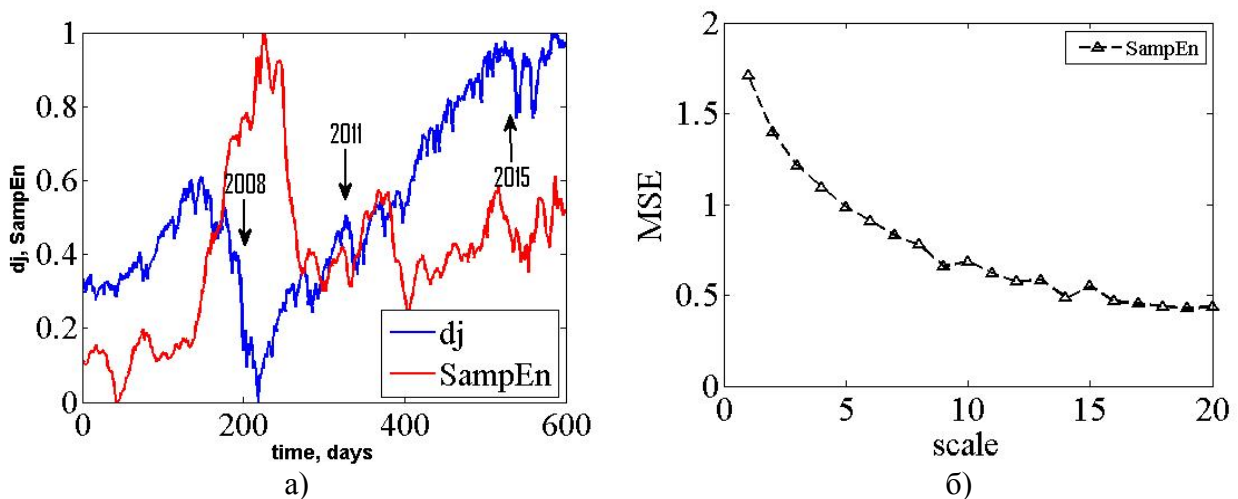


Рис. 2. Ентропійні показники складності системи за даними ентропії шаблонів: а) порівняльна динаміка ентропії та вихідного часового ряду; б) мультимасштабна ентропія шаблонів

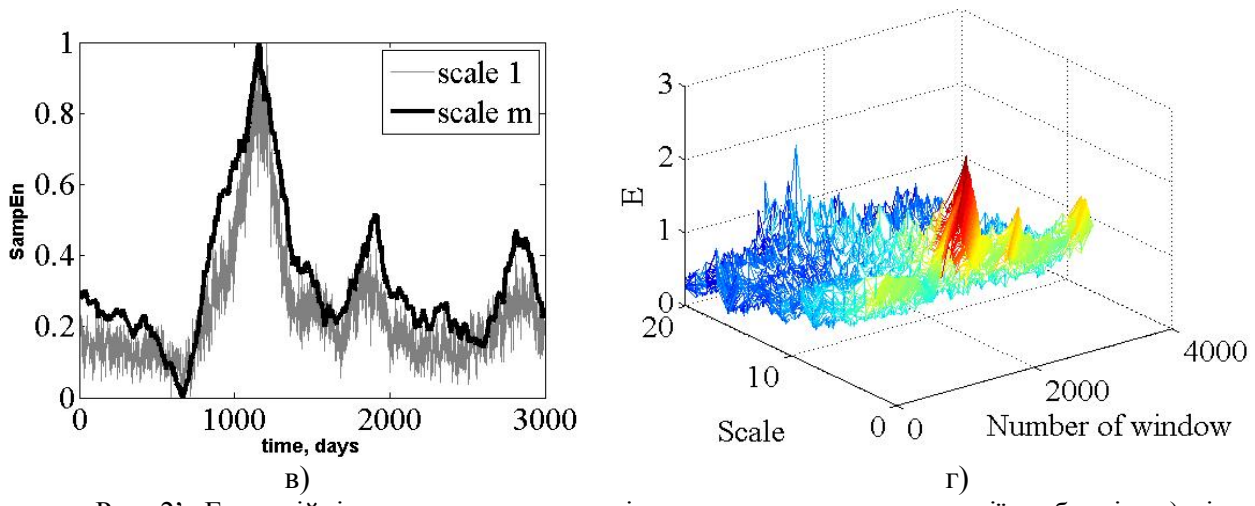


Рис. 2'. Ентропійні показники складності системи за даними ентропії шаблонів: в) віконна ентропія шаблонів для масштабу 1 і середня за усіма масштабами; г) віконна мультимасштабна ентропія

Візьмемо абсолютно інше за природою явище – землетрус, яке кількісно описується часовим рядом амплітуд коливань земної кори. Аналогічні до рис. 2-2' ентропійні показники відповідного часового ряду SEI представлені на рис. 3.

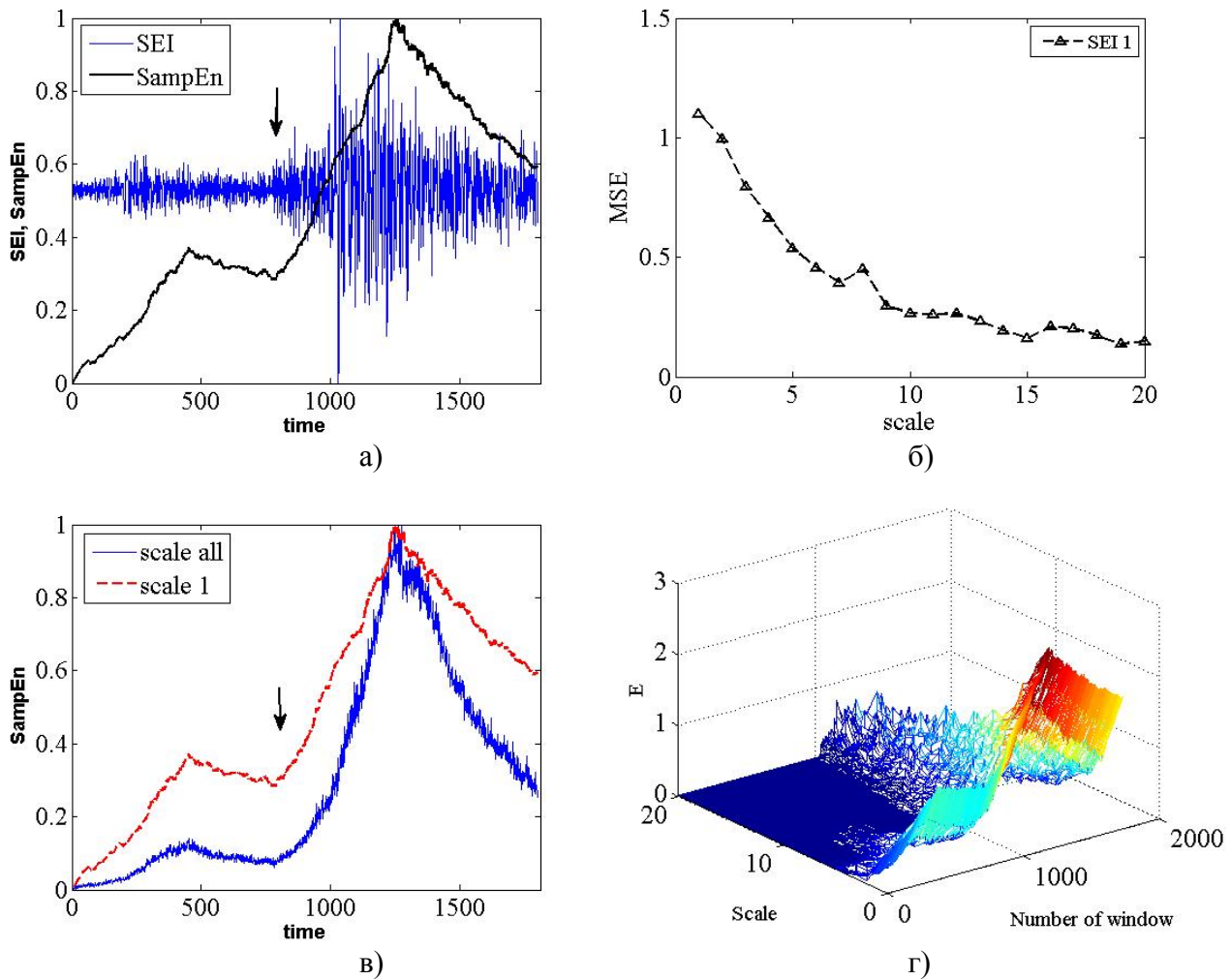


Рис. 3. Ті ж показники, що і на рис. 2-2', але для часового ряду SEI

Рис. 4 демонструє ентропійні показники Тсалліса і Фур'є, розраховані за алгоритмом рухомого вікна.

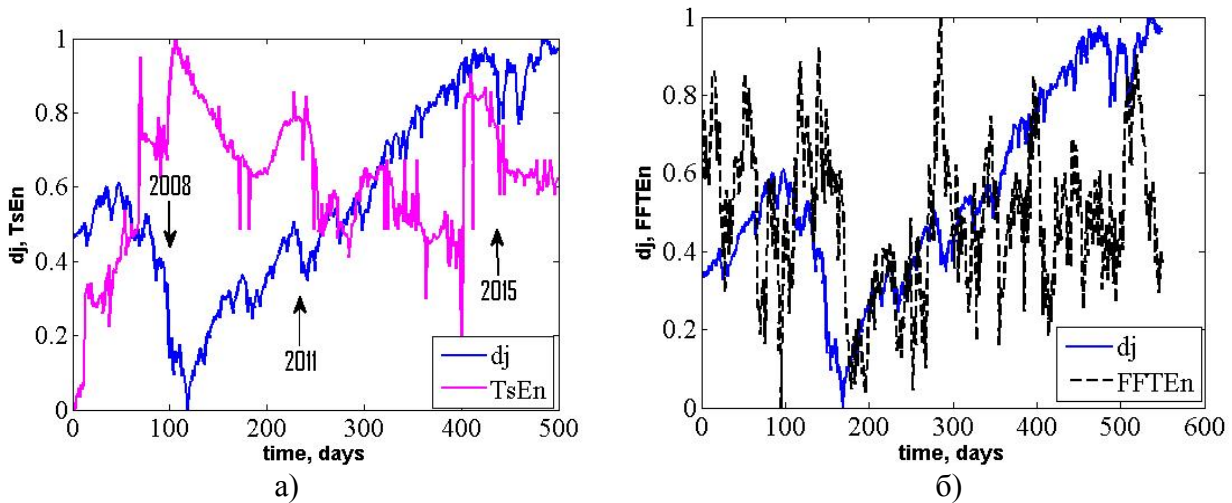


Рис. 4. Ентропія Тсалліса (а) та Фур'є (б) розраховані з вікнами 750 і 500 днів відповідно і кроком розміром у 5 днів

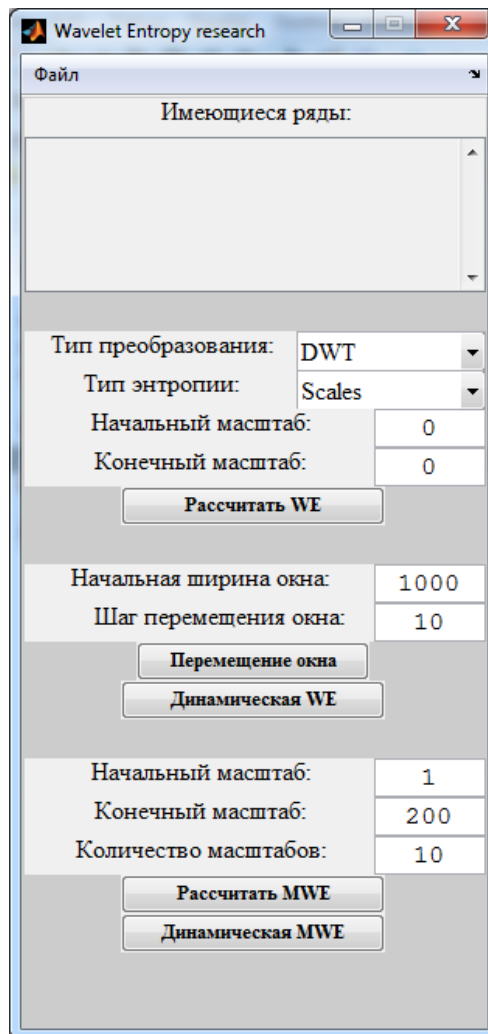


Рис. 5. Інтерфейс розрахунку вейвлет-ентропії

Очевидно, що наведені ентропійні міри містять конкретну інформацію щодо характерних станів складної системи незалежно від її природи. Аналіз, порівняння, співставлення відповідних даних дозволяє виявляти приховані закономірності розвитку

складних систем, будувати необхідні показники та індикатори критичних станів.

Нарешті, розглянемо можливості вейвлет ентропії. Для цього у командній строчці наберемо **work6**. У вікні, яке має вигляд, зображений на рис. 5, завантажимо досліджуваний файл, наприклад, dj29 – фрагмент часового ряду індексу Доу-Джонса, що містить кризу 1929 року. Пік кризи знаходиться в точці 1000, а довжина усього ряду складає 2000 днів. Виберемо далі тип вейвлет-перетворення (наприклад, дискретний – DWT(cs)), тип ентропії (points) одержимо результати, скріншоти яких відображені на рис. 6.

**!!! Пересвідчіться самостійно в тому, що аналогічна картина спостерігається і для інших відомих кризових явищ.**

Аналіз вейвлет-ентропії вихідного або ж перетвореного ряду може вказати на певні характеристики і особливості сигналу, який досліджується.

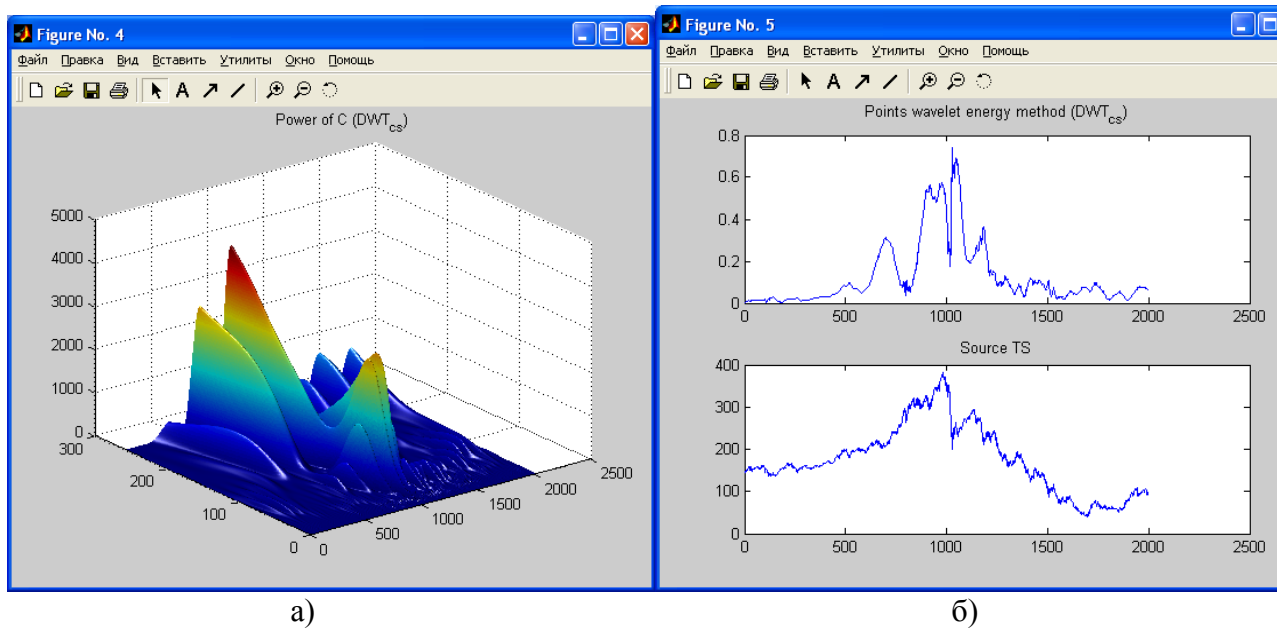


Рис. 6. Енергія вейвлет-коефіцієнтів (а) та динаміка точкової вейвлет-ентропії (б)

Зверніть увагу на той факт, що у передкризовий період (задовго до піку кризи – точки 1000) як на енергетичній поверхні, так і на графіку вейвлет-ентропії з'являються характерні зростаючі коливання, які помітно згасають у після кризовий період. Саме такі коливання можуть слугувати індикатором-передвісником кризових явищ в складних системах.

Таким чином, ентропійні показники є мірами складності системи. Аналіз їх дозволяє будувати індикатори поведінки системи у різних фазах розвитку.

### Завдання

1. Проведіть порівняльний аналіз ентропійних показників для тестових часових рядів. Виявіть, які з них є мірами складності системи.
2. Аналогічний аналіз проведіть для часових рядів реальних об'єктів, наприклад, фондових індексів, курсів валют, цін активів, спотових цін тощо.
3. Побудуйте часові ряди, які у визначених точках характеризуються кризами і покажіть, чи є (і які саме) ентропійні міри індикаторами, або передвісниками кризових явищ.
4. Зробіть загальні висновки і оформіть звіт.

## ЛАБОРАТОРНА РОБОТА № 8

**Тема.** Визначення індексу фрактальності.

**Мета.** Розрахунок індексу фрактальності для оцінки ефективності функціонування складних систем та їх складових.

1. Індекс фрактальності як міра складності системи.
2. Універсальність індексу щодо бази оцінки.
3. Динаміка індексу фрактальності.

### Теоретичні відомості

Одна з ключових особливостей складних систем, зокрема і фінансово-економічних, полягає в їх (мульти-)фрактальній природі. В роботах 6-7 при аналізі властивостей систем на різних масштабах ми виявили мультимасштабність систем. Так, динаміка різних мір складності на масштабі 1 та середня величина, усереднена за всіма масштабами, відрізняються. Звідси виникає питання, а чи не є різниця між даними величинами мірою мультифрактальності. Якщо така величина (назвемо її *індексом фрактальності* – index fractality - IF) прямує до нуля, то ми маємо мультифрактальну, самоподібну на всіх масштабах систему, яка вочевидь є більш складною, аніж система з високою величиною IF, коли складність із зростанням масштабу швидко падає. Основне завдання роботи полягає в тому, щоб вяснити, чи може вказана міра складності виступати мірою ефективності функціонування суб'єктів господарювання.

Для розрахунку індексу фрактальності використовуємо нормалізовані значення міри складності з першого по останній масштаб. Загалом даний показник має універсальний характер, оскільки може бути побудований на основі будь-якої міри.

Алгоритм побудови індексу фрактальності передбачає три етапи:

- розрахунок міри (міри Лемпеля-Зіва, ентропії тощо) на обраних масштабах;
- нормалізацію отриманих результатів;
- калькулювання індексу.

Оберемо для конкретності мультимасштабну ентропію шаблонів, знайдемо для неї віконні міри для масштабу 1 і середнє, наприклад, з 5 масштабів.

На рисунку 1 заштрихована зона схематично відображає індекс фрактальності, який чисельно дорівнює:

$$IF = \left| SC_1 - \frac{\sum_{i=1}^n SC_i}{n} \right|,$$

де  $SC_1$  - значення ентропії шаблонів для першого масштабу,  $SC_i$  - значення ентропії для  $i$ -го масштабу,  $n$  - число масштабів (в даному випадку 5).

Для розрахунку використовуються нормалізовані величини, процедура нормалізації має наступний вигляд:

$$SC_{norm} = \frac{SC_n - SC_{min}}{SC_{max} - SC_{min}},$$

де  $SC_{norm}$  – нормалізоване значення міри для обраного масштабу,  $SC_{max}$  – максимальне значення міри для обраного масштабу,  $SC_{min}$  – мінімальне значення міри для обраного масштабу. В результаті такої процедури отримуємо нормалізовані значення в інтервалі від нуля до одиниці, що дозволяє спростити наступну процедуру розрахунку індексу



фрактальності.

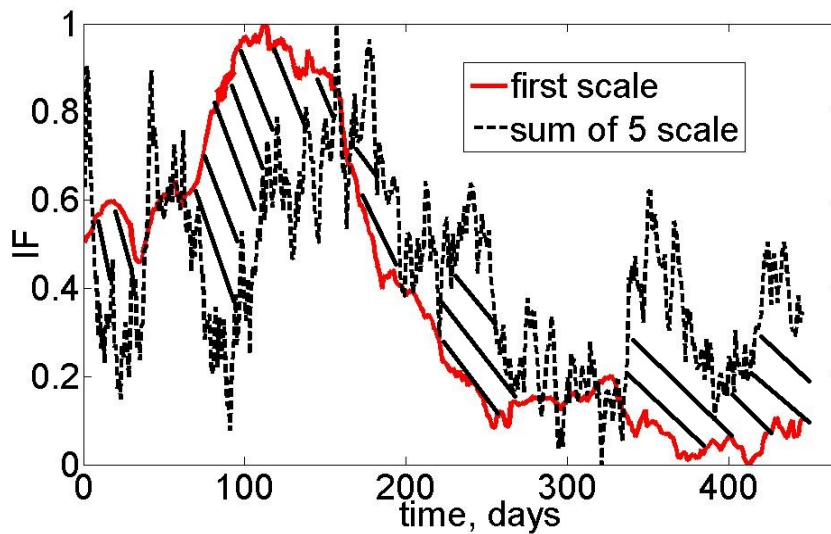


Рис. 1 Схематична ілюстрація ділянок (заштриховані) необхідних для розрахунку індексу фрактальності (штриховою лінією позначене середнє значення суми міри на 5 масштабах, суцільною – значення показника на першому масштабі, приведені після нормалізації)

### Хід роботи

1. Для розрахунку індексу фрактальності потрібно в окрему папку скопіювати два файли: IndexFractality, IndexFractality\_runner. Саме IndexFractality\_runner є програмою для запуску розрахунку індексу.
2. Наступним кроком потрібно розрахувати значення необхідної міри (ентропії шаблонів, подібності, Тсалліса та ін.) на обраних масштабах (запустивши в середовищі matlab програмний засіб work5\_en).
3. Відкрити поверхню розрахунку обраної міри на всіх масштабах та натиснувши на неї мишкою, потім запустити програму IndexFractality\_runner і натиснути F5.
4. У вікні, котре відкриється ввести назву даного індексу та вказати шлях для його збереження.

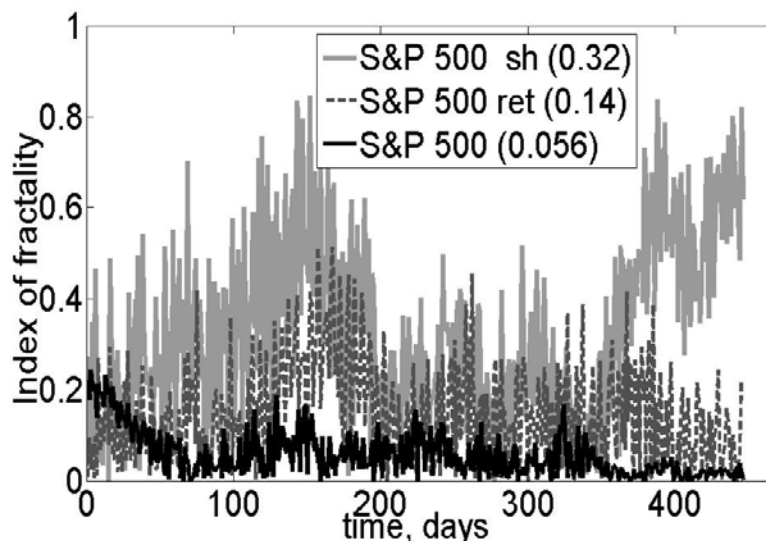


Рис. 2 Індекс фрактальності для вихідного часового ряду фондового індексу США (s&p 500), перемішаного ряду (s&p 500 sh) та прибутковостей, розрахованих для часового ряду (s&p 500 ret), в дужках вказані середні значення індексу фрактальності для всього ряду.

Важливо зазначити, що для розрахунку початкової міри складності можна використовувати стаціонарні та нестаціонарні ряди, результати використання таких рядів представлені на рисунку 2.

Як бачимо з рис. 2, для розрахунку можна використовувати стаціонарні та нестаціонарні ряди. Менше значення індексу фрактальності сигналізує про вищу міру складності, оскільки перемішаний ряд має нижчу складність, ніж фондовий індекс. Мале відхилення динаміки на різних масштабах сигналізує про високий рівень фрактальності, а, отже, високу ефективність функціонування системи.

Приведемо приклад розрахунку індексу фрактальності, побудованого на основі ентропії шаблонів для фондових індексів країн різного ступеня розвитку (рис.3).

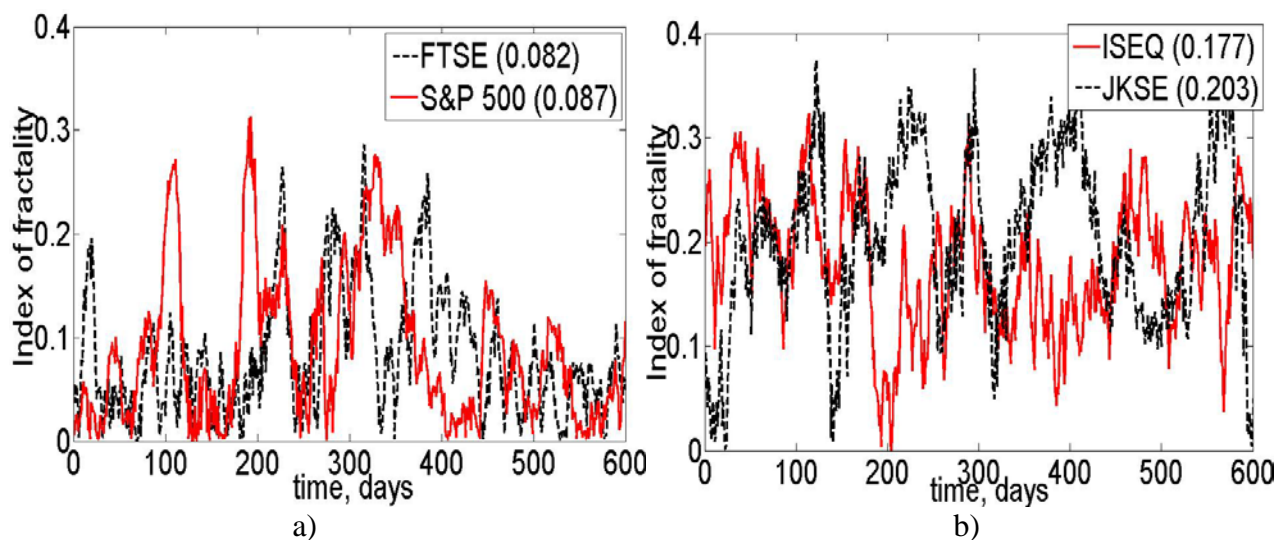


Рис. 3. Індекс фрактальності: а) для розвинених країн (ftse – фондовий індекс Великобританії, s&p 500 – США); б) для країн, які розвиваються (iseq – фондовий індекс Ірландії, jkse – Індонезії), в дужках вказане середнє значення для всього ряду.

Як видно з рисунку 3, країни з більш розвиненими фондовими ринками мають нижчий індекс фрактальності, ніж країни, які розвиваються. Очевидно це вказує на вищу складність їх організації.

Проведемо розрахунки індексу фрактальності для часових рядів, які є цінами акцій відомих світових банківських установ (<https://finance.yahoo.com>). Результати розрахунків представлені у таблиці 1.

Таблиця 1

**Індекс фрактальності для основних світових банків**

№	Символ банку на біржі	Індекс фрактальності
1	BARC	0.086
2	BNP	0.087
3	RoyalBankofCanada (RY)	0.090
4	GS	0.100
5	SocieteGeneraleGroup (SCGLY)	0.106
6	JPM	0.108
7	CS	0.110
8	LLOY	0.112
9	Mitsubishi UFJ FinancialGroup, Inc. (MTU)	0.113
10	UBS	0.122
11	HSBA	0.125

12	DBK	0.126
13	SocieteGeneraleGroup(GLE)	0,126
14	HBOS	0,132
15	DPB	0,138
16	WellsFargo&Company (WFC)	0,142
17	BAC	0,146
18	C	0,152
19	SunTrustBanks, Inc. (STI)	0,152
20	MS	0,154
21	DB	0,16
22	SAN	0,162
23	UniCreditS.p.A. (UCG.MI)	0,165
24	ChinaConstructionBankCorporation	0,179

Як бачимо з таблиці 1, найвища ефективність за індексом фрактальності у BARC та BNP, найнижча у UniCreditS.p.A. та ChinaConstructionBankCorporation, що відповідає макроекономічним показникам з прибутковості на акції цих банків. Такі значення ІФ сигналізують про вищу інвестиційну привабливість першої групи банків.

### Висновки

Отже, в роботі введено новий індикатор (міру) складності системи, методику її оцінки та показано зв'язок з ефективністю функціонування суб'єктів складних фінансово-економічних систем. Алгоритм, заснований на індексі фрактальності, має універсальний характер, оскільки може застосовуватись для різних мір складності та різних баз даних.

### Завдання

1. Розрахуйте, порівняйте і оцініть ефективність індексу фрактальності для інших складних систем, заданих часовими рядами (фондових індексів, крос-курсів, цін на товарних ринках тощо).

## ЛАБОРАТОРНА РОБОТА № 9

**Тема.** Дослідження асиметрії (необоротності) часу складних сигналів

**Мета.** Навчитись розраховувати значення індексу асиметрії часу для складних сигналів та досліджувати їх динаміку у випадку незворотних змін, побудови передвісників критичних і кризових явищ

**«Скільки ж довелося мені прочитати нудних сторінок, на яких автор займався пошуком в економіці чогось такого, що відповідає ентропії або тому чи іншому виду енергії!»**

**П. Самуельсон. 3 Нобелівської лекції  
(«Thesis». Vol. 1. N° 1 Winter 1993. P. 192)**

**«Energy is the queen of the world, and entropy is her shadow»  
W. Ostwald**

**«Енергія є королевою світу, а ентропія – її тінню»  
В. Оствальд**

## Теоретичні відомості

Необоротність часу – фундаментальна властивість нерівноважних систем. Вона має важливе значення для оцінки стану складних динамічних процесів, які відбуваються на різних масштабах.

Введемо деякі основні означення.

Системи поділяють на **ізолювані, закриті і відкриті**.

В *ізолюваних* системах відсутній обмін із зовнішнім середовищем, в *закритих* є обмін тільки енергією, але не масою, у *відкритих* – і масою і енергією

Виділяють також інтенсивні і екстенсивні (адитивні) системи. При цьому екстенсивні параметри характеризують систему як ціле (маса, об'єм), інтенсивні параметри можуть мати окремі значення в кожній точці системи.

Для нерівноважних систем модифікуються і закони термодинаміки. Нагадаємо, що перший з них постулює існування функції стану – енергії, а другий – ентропії.

Якщо рівняння, які описують поведінку у часі деяких процесів інваріантні по відношенню до знаку у змінної «час», то процеси називаються зворотними, інакше – незворотними.

Ентропія походить від грецького *εντροπη* - еволюція. Ентропія є величиною адитивною, хоча останнім часом К. Тсаллісом показано, що у випадку далекодійчих сил ентропія стає екстенсивною (неадитивною) величиною.

Зміна ентропії складається з двох частин:

$d_e S$  - обумовленої взаємодією з зовнішнім середовищем (ендогенна компонента);

$d_i S$  - обумовлена внутрішніми змінами системи (екзогенна компонента)

$$dS = d_e S + d_i S.$$

Якщо зміни зворотні, то  $d_i S = 0$ , для незворотних процесів  $d_i S > 0$ .

Другий закон термодинаміки звучить так. Для ізолюваних систем потік ентропії дорівнює нулю

$$dS = d_i S \geq 0.$$

Отже, зростання ентропії вказує на незворотність процесів.

Складні системи в природі залежать від маси, енергії, ентропії й інформаційних потоків, які проходять через їх межі. При нормальних умовах, дисипативні відкриті системи перебувають у станах, далеких від рівноваги. І навпаки, у певних критичних випадках може мати місце стан, що наближається до максимально рівноважного.

Вказані системи використовують енергію, щоб розвиватися до більш ієрархічно впорядкованих структурних конфігурацій і меншої кількості можливих станів у порівнянні з навколишнім середовищем. Їхня здатність до самоорганізації пов'язана з єдиною спрямованістю потоку енергії через межі системи й необоротністю основних процесів. **Втрата необоротності часу** може бути показником деструктивних процесів с самій системі або зовнішнього руйнівного впливу.

Далі ми покажемо, яким чином можна ввести міру необоротності часового ряду – індексу асиметрії, а також використати її для ідентифікації динамічних змін у складних системах різної природи

## Алгоритм розрахунку

Послідовність розрахунку індексу асиметрії часу, який є мірою наближення системи до рівноважного стану, складається з наступних кроків:

1. Для початкового ряду динаміки функціонування певної складної системи

знаходиться ряд прибутковостей:

$$Y_t = \ln P_t - \ln P_{t-1}, \quad (1)$$

де  $P_t$  – поточне значення ряду динаміки,  $P_{t-1}$  – попереднє значення,  $t \in [1, N]$ ,  $N$  – кількість елементів ряду

2. Щоб дістати інформацію відносно множини часових масштабів для ряду (1), розраховується набір згладжених рядів з різною шириною вікна згладжування:

$$y_\tau(i) = \sum_{j=0}^{\tau-1} y_{i+j} / \tau, \quad (2)$$

де  $\tau$  – параметр масштабу (ширина вікна згладжування).

На основі підходів статистичної фізики, ми робимо припущення, що кожна зміна динаміки процесу являється незалежною і потребує певної кількості енергії,  $E$ . Функція розподілу ймовірностей для систем цього класу має наступний відомий вигляд:

$$p \propto \exp(-\beta E - \gamma Q), \quad (3)$$

де  $Q$  представляє нерівноважне виділення тепла через межі системи,  $\beta$  і  $\gamma$  являються множниками Лагранжа, які отримані з обмежень на середнє значення енергії  $E$ , необхідної для переходу і вкладу кожного переходу у виділення теплоти  $Q$ .

3. В ряду (1) додатні величини відповідають збільшенню початкового значення, а від'ємні – зменшенню. Різниця середньої енергії зростає початкового значення, тобто  $\langle \beta E + \gamma Q \rangle_{y_\tau > 0}$ , і спадів -  $\langle \beta E + \gamma Q \rangle_{y_\tau < 0}$ , може бути використана як міра асиметрії оборотності часу.

Враховуючи, що вираз (3) пов'язує енергію з емпіричним розподілом, а визначимо наступну міру необоротності часу

$$a(\tau) = \frac{\int_0^\infty [\rho(y_\tau) \ln \rho(y_\tau) - \rho(-y_\tau) \ln \rho(-y_\tau)]^2 dy_\tau}{\int_{-\infty}^\infty \rho(y_\tau) \ln \rho(y_\tau) dy_\tau}. \quad (4)$$

Ряд даних є оборотним, якщо  $a(\tau) = 0$ .

Для складних систем важливо не лише визначити ступінь необоротності ряду даних, але також дізнатися, який із рядів показує «прямий» напрямок, а який - зворотній. Вираз (4) не надає такої інформації. Тому замість нього розглядаємо наступний

$$A(\tau) = \frac{\int_0^\infty [\rho(y_\tau) \ln \rho(y_\tau) - \rho(-y_\tau) \ln \rho(-y_\tau)] dy_\tau}{\int_{-\infty}^\infty \rho(y_\tau) \ln \rho(y_\tau) dy_\tau}. \quad (5)$$

Якщо  $A(\tau) > 0$ , тоді для масштабу  $\tau$  ряд даних являється необоротним. Проте, якщо  $A(\tau) = 0$ , то ряд даних може бути, а може і не бути оборотним на масштабі.

Для реальних сигналів кінцева частота являється визначеною, а  $y_\tau$  – дискретна змінна. При аналізі таких сигналів використовується наступний вираз для визначення оцінки  $A(\tau)$ :

$$\hat{A}(\tau) = \frac{\sum_{y_\tau > 0} \Pr(y_\tau) \ln[\Pr(y_\tau)]}{\sum_{y_\tau} \Pr(y_\tau) \ln[\Pr(y_\tau)]} - \frac{\sum_{y_\tau < 0} \Pr(y_\tau) \ln[\Pr(y_\tau)]}{\sum_{y_\tau} \Pr(y_\tau) \ln[\Pr(y_\tau)]}, \quad (6)$$

де  $\Pr(y_\tau)$  означає ймовірність значення  $y_\tau$ .

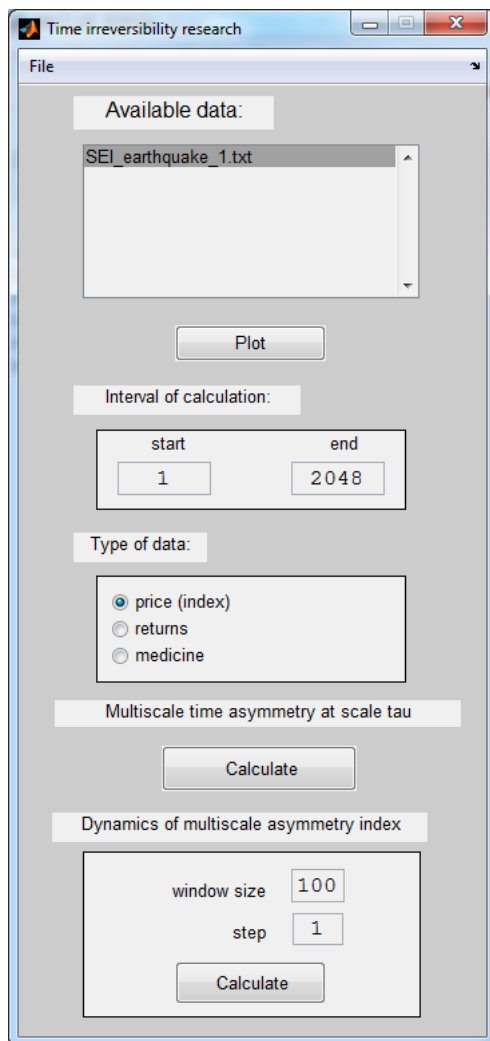
4. Для діапазону масштабів часу ми тоді легко можемо розрахувати мультифрактальний індекс асиметрії ( $A_I$ ) наступним чином:

5.

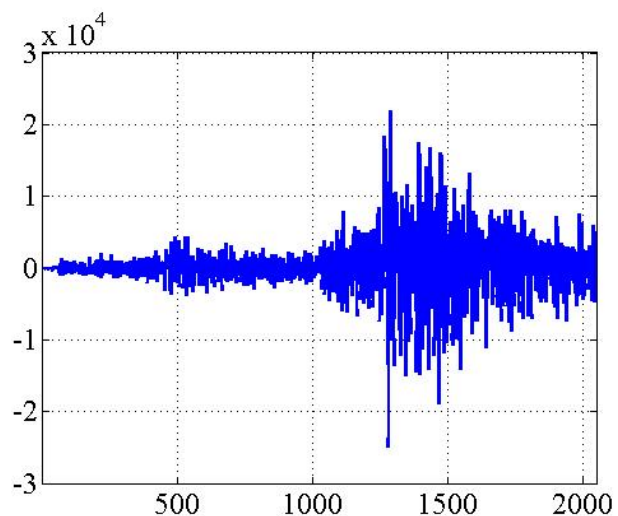
$$A_I = \sum_{\tau=1}^L \hat{A}(\tau) / \tau .$$

### Хід роботи

Для дослідження динамічних рядів за допомогою індексу необоротності часу необхідно завантажити програму **tir**, яка дозволяє як отримувати загальне значення індексу асиметрії для обраного часового ряду, так і досліджувати динаміку зміни значення  $A_I$ , використовуючи процедуру рухомого вікна.



a)



б)

Рис. 1. Вікно програми із завантаженим рядом даних

Програмний продукт дозволяє працювати як з вихідним часовим рядом (**price**

(**index**)), так із прибутковостями (**returns**). Крім того, можна завантажувати часові ряди спеціального типу, які не схожі ні на вихідні фінансові ряди, ні на прибутковості. У переважній більшості вони зустрічаються у біомедичних дослідженнях. Тому у програмі вони виділені окремою кнопкою «**medicine**» (рис. 1а).

Для початку розрахунку необхідно завантажити ряд даних із текстового файлу з розширенням *dat*, *txt*, використовуючи меню **File**→**Read**. Його зразу ж можна відобразити графічно (рис. 1б). Так, на рис. 1 візуалізовано ряд, що відповідає фіксації окремим сенсором прояву землетрусу.

Якщо потрібно розрахувати загальне значення індексу необоротності часу для всього ряду, то необхідно натиснути кнопку **Multiscale time asymmetry at scale tau**. По закінченню розрахунку на екран буде виведене значення  $A_i$ , а також графік  $\sum_{y_\tau > 0} \Pr(y_\tau) \ln[\Pr(y_\tau)]$ ,

$\sum_{y_\tau < 0} \Pr(y_\tau) \ln[\Pr(y_\tau)]$  і  $A(\tau)$  для різних масштабів  $\tau$ . На рис. 2 представлений результат обчислень для індексу Dow Jones за період з 1986 до 1988 року, який містить кризу (жовтень 1987 р.)

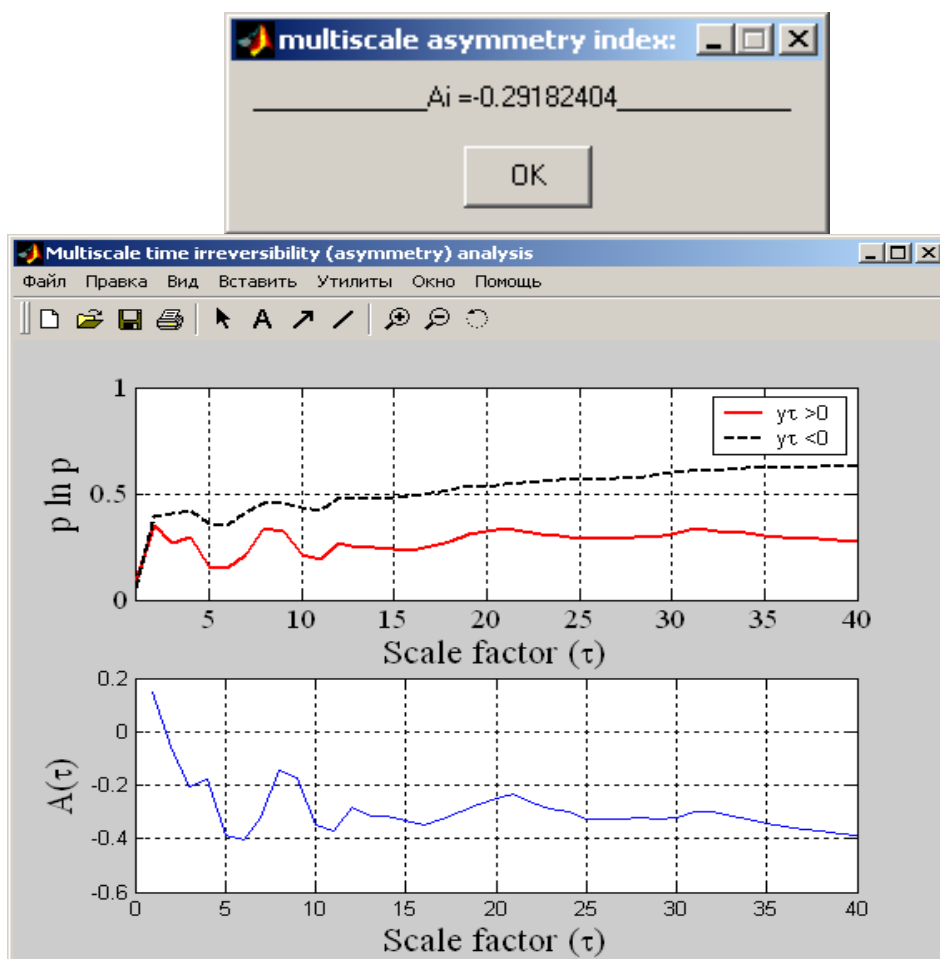


Рис. 2. Результат розрахунку, який виводиться при натисненні на кнопку **Multiscale time asymmetry at scale tau**

Другий блок, позначений назвою **Dinamics of multiscale asymmetry index**, дозволяє відслідковувати зміну індексу необоротності часу в процесі розвитку системи. Ця можливість реалізується завдяки використанню методу рухомого вікна. В блок також включені інструменти для настройки ширини вікна і шагу його зміщення. Приклад розрахунку динаміки зміни  $A_i$  для індексу Dow Jones для трьох найбільших із відомих криз



1929, 1987 і 2008 рр. представлений на рис. 3а. На рис. 3б відображені (зверху вниз) власне один з часових рядів (криза 1929р.), кумулятивний індекс асиметрії та його складові. Стрілками вказані точки перетину цих складових

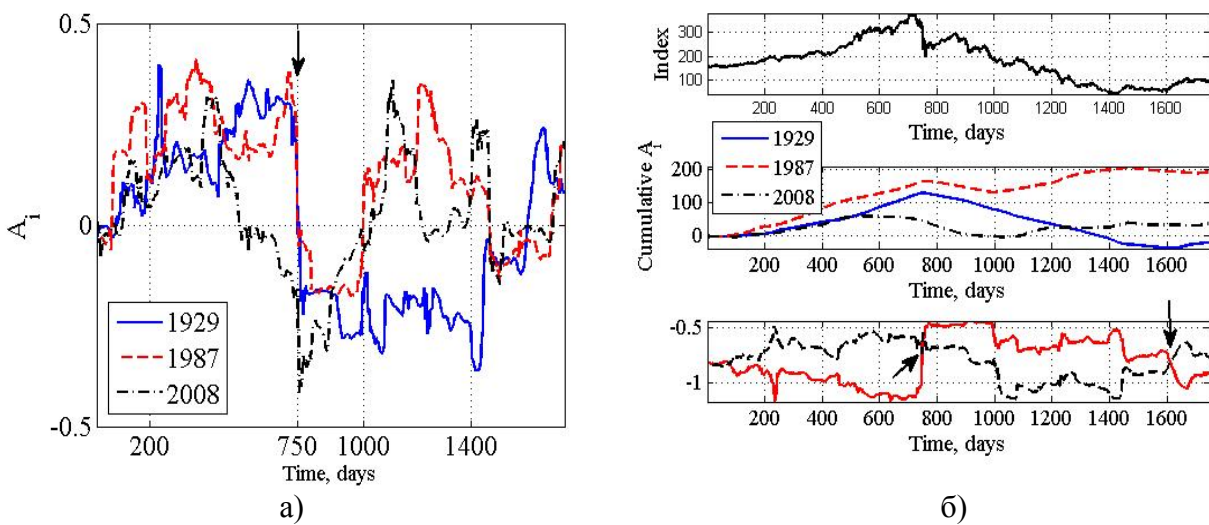


Рис. 3. а) динаміка зміни значення  $A_t$  для відомих криз індексу Dow Jones; б) криза 1929р., кумулятивні індекси асиметрії всіх трьох криз та компоненти кумулятивного індексу асиметрії

Очевидно, що з початком кризи індекс асиметрії стрімко спадає (рис. 3а) і є індикатором кризових явищ. При цьому кумулятивний індекс (рис. 3б) має максимум у точці кризи і мінімум у точці закінчення рецесії – відновлення складної системи. Ці точки на графіку позначені стрілками. Період часу між вказаними стрілками точками дорівнює періоду рецесії.

Картина поведінки кумулятивного індексу асиметрії у випадку дослідження землетрусу представлена на рис. 4.

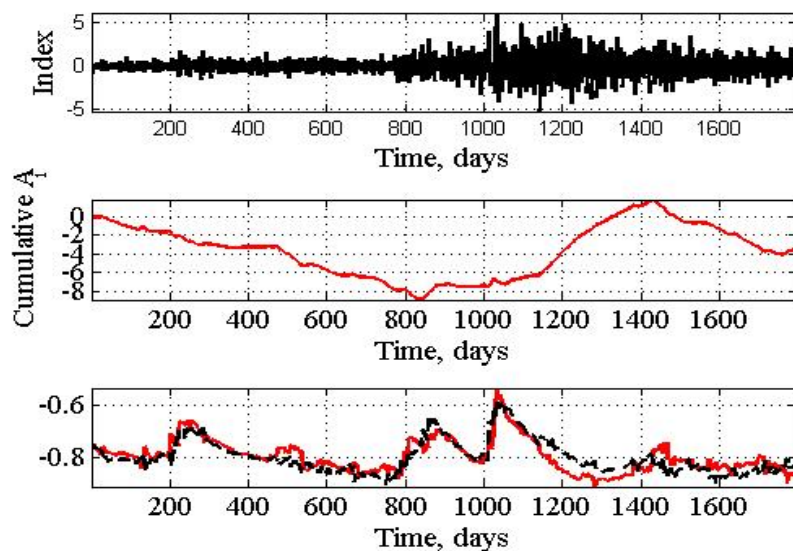


Рис. 4. Кумулятивний індекс асиметрії для часового ряду, який характеризує землетрус

Видно, що індекс асиметрії по-різному поводить себе у періоди власне критичного явища, до та після нього. Це дозволяє використовувати властивість незворотності у часі для моніторингу та попередження кризових явищ.

## Завдання для самостійної роботи

1. Виберіть відповідну базу даних, яка містить кризові явища. Виконайте аналіз поведінки індексу необоротності часу. Виділіть період кризи на графіку, проаналізуйте одержаний результат у звіті. Порівняйте прояви кризи на різних часових рядах, для різних природних та штучних явищ.
2. Проведіть варіацію часового вікна та кроку. Які висновки можна зробити?
3. Для часового ряду Доу Джонса з 1900 по 2016 рр. підберіть розмір вікна та кроку з метою дослідження хвильових процесів на фондовому ринку
4. Для американського фондового ринку дослідіть протяжність усіх відомих прецесійних процесів. Порівняйте результати з експертними оцінками. Зробіть висновки.

## ЛАБОРАТОРНА РОБОТА № 10

**Тема.** Дослідження процесів післякризової релаксації складних систем.

**Мета.** Навчитись аналізувати релаксаційні процеси в складній системі, заданій часовим рядом та прогнозувати період закінчення кризи.

1. Закон Оморі.
2. Визначення довжини релаксації.
3. Віконна процедура для Оморі-релаксації.

### Теоретичні відомості

Відомо, що після землетрусу, кризи на економічних ринках певний час спостерігаються ефекти, подібні до кризової ситуації, проте значно менші за масштабами. Так, дослідження економічних часових рядів показують, що протягом цього періоду відбуваються невеликі падіння, подібні до кризового, частота яких постійно зменшується і, в решті-решт, система стабілізується. Період переходу системи із кризового стану у стабільний (той, в якому система перебувала до кризи або подібний) є періодом релаксації.

Закон Оморі дозволяє визначити швидкість зменшення впливу критичного явища або час протікання періоду релаксації, опираючись на спостережувані аномальні відхилення. У випадку, коли період релаксації не завершено, використовуючи закон Оморі, можна визначити час переходу системи до стабільного стану.

У загальному випадку закон Оморі є степеневим законом залежності кількості аномальних відхилень від часу і подається у вигляді:

$$N(t) = kt^p, \quad (1)$$

де  $N$  – кількість відхилень, що спостерігаються протягом періоду тривалістю  $t$  з моменту кризи,  $p$  – степінь, що показує швидкість зростання кількості відхилень ( $p < 1$ ),  $k$  – деяке дійсне число.

Очевидно, графік поданої залежності зростатиме з від'ємним прискоренням, асимптотично наближаючись до деякої прямої, паралельної осі абсцис. «Вирівнювання» графіка може свідчити про закінчення фази релаксації та стабілізацію економічної системи. Дослідження показника  $p$  дозволяє робити припущення про можливість класифікації криз за силою та тривалістю, що, однак, потребує додаткових досліджень.

Для використання методу Оморі розраховується ряд накопичень «екстремальних» (що виходять за межі коридору  $[\bar{x} - k\sigma, \bar{x} + k\sigma]$ ) флуктуацій. У міру руху від кризи вправо по осі абсцис кількість «екстремальних» флуктуацій зменшується, тому швидкість росту значень по осі ординат падає (рис. 1).

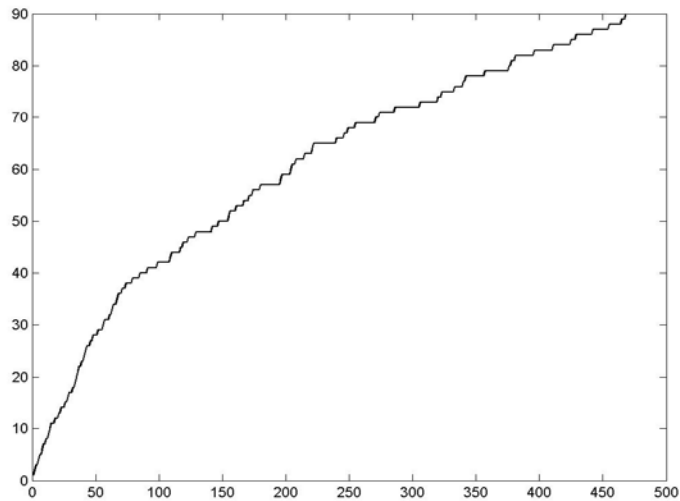


Рис. 1. Ряд накопичень екстремальних флуктуацій

Іншими словами, на початку графіка, наприклад, на інтервалі  $[1,50]$  за 50 днів спостерігалось близько 25 флуктуацій, тоді як на інтервалі  $[400,450]$  – всього близько 5. Таким чином, можна сказати, що вірогідність отримати «екстремальну» флуктуацію спочатку (відразу після кризи) дорівнює  $\frac{25}{50} = 0.5$ , а така ж вірогідність на другому інтервалі –  $\frac{5}{50} = 0.1$  (дані приблизні). Тому закінчення періоду релаксації можна розраховувати, використовуючи таку вірогідність.

Питання виникає в тому, що графік обмежений по осі абсцис і кінець періоду релаксації може знаходитися за межами даного інтервалу. Тому необхідно використовувати апроксимуючу криву і аналізувати її.

Аналіз проводимо на основі похідної цієї кривої. Функція кривої має вигляд

$$f(x) = e^b \cdot x^a,$$

де  $a, b$  – коефіцієнти прямої  $y = ax + b$ , знайденої методом найменших квадратів для графіка накопичень екстремальних флуктуацій, побудованого в подвійному логарифмічному масштабі.

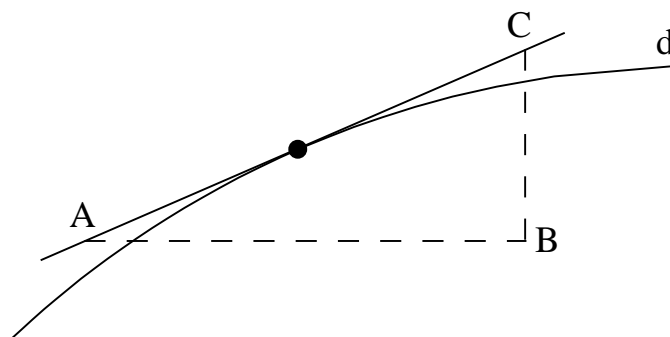


Рис. 2.

Для знаходження вірогідності розглянемо дотичну AC до деякої точки графіка (рис. 2). Як відомо, тангенс кута нахилу дотичної є похідною функції. З іншого боку, тангенс кута

нахилу – відношення протилежного катета прямокутного трикутника до прилеглого. Тому отримаємо для деякої точки наступну рівність:

$$\operatorname{tg}(\angle CAB) = \frac{CB}{AB} = f'(x_0),$$

де  $x_0$  – точка дотику, позначена на рисунку.

У даному виразі відношення  $CB/AB$  якраз і покажуть вірогідність отримати в день, відповідний точці дотику, «екстремальну» флуктуацію (тобто,  $CB$  флуктуацій за  $AB$  днів). Тому вірогідність виникнення «екстремальної» флуктуації в деякій точці дорівнює похідній функції в цій точці.

Нехай вірогідність отримання «екстремальної» флуктуації в кінці періоду релаксації рівна  $p$ . День закінчення періоду – перша точка, що має таку вірогідність. Звідси отримуємо, що для того, щоб знайти таку точку, треба розв'язати рівняння

$$f'(x) = p,$$

де  $f(x)$  – апроксимуюча крива.

Ось основні кроки пошуку довжини періоду релаксації. Природно, тут є деякі допущення, але, проте, аналізуючи таким чином графік, можна отримати приблизно довжину періоду релаксації.

### Хід роботи

Запускаємо програму командою >> **work7**.

Відкривається вікно для проведення розрахунків (рис. 3):

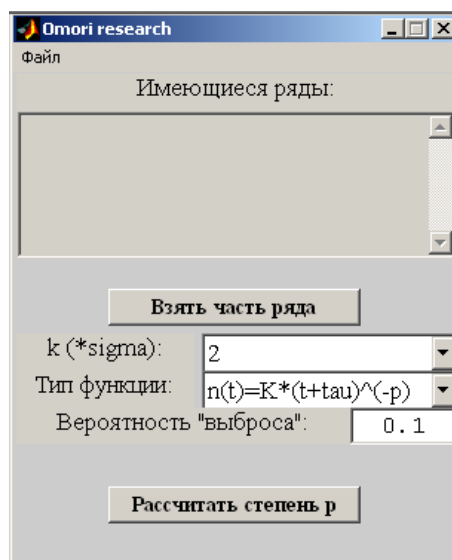


Рис. 3. Скріншот програмного продукту для дослідження після кризових релаксацій

У першому полі, « $k(*\sigma)$ », як завжди, вводимо коефіцієнт  $k$  множення середнього квадратичного відхилення, за допомогою якого знаходиться флуктуації, що виходять за межі інтервалу  $[\bar{x} - k\sigma, \bar{x} + k\sigma]$ .

Поле «тип функції» незмінне: там лише одне значення параметра.

У полі «Вірогідність викиду» встановлюємо вірогідність, за допомогою якої розраховуватимемо закінчення періоду релаксації. Пояснення вірогідності дане на початку.

Читаємо обраний ряд прибутковостей, наприклад, кризи 1987 року фондового ринку США за індексом Доу Джонса (рис. 4а). Для того, щоб узяти частку ряду, необхідно натискувати кнопку «Узяти частку ряду». В результаті відкриється вікно з рядом (рис. 4б).

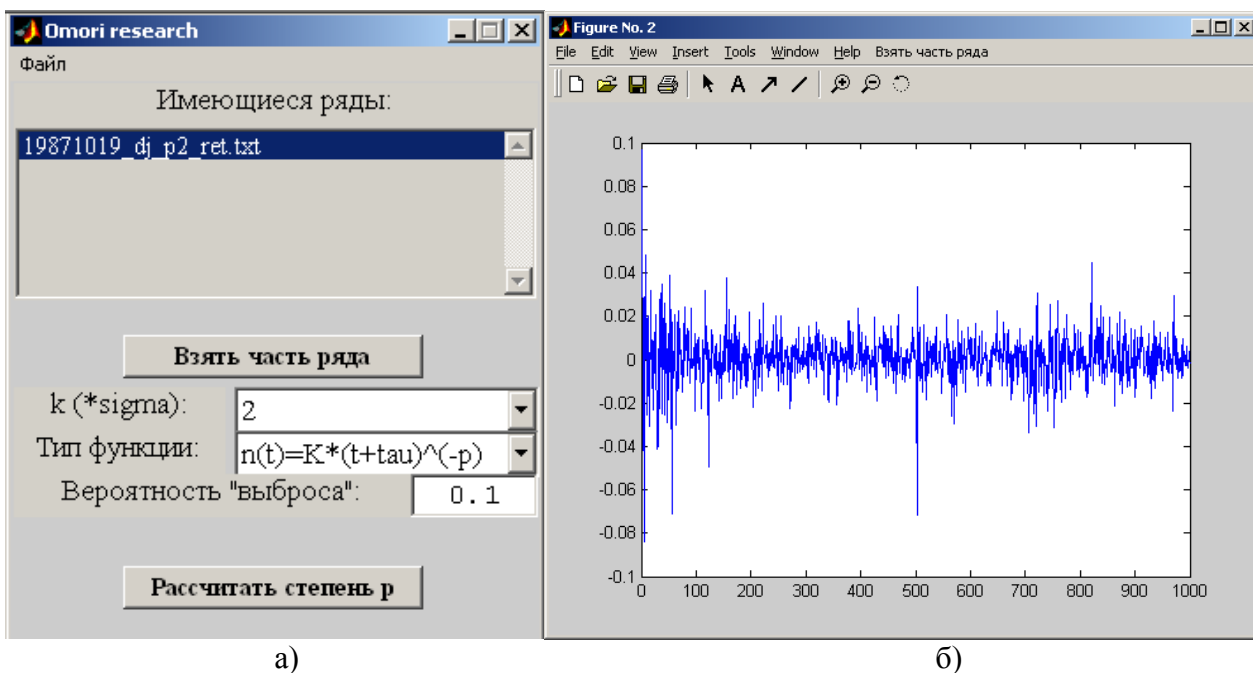


Рис. 4. Звантаження вхідного ряду (а) та вибір його частини (б)

Для виділення частини ряду вибираємо пункт меню «Взяти частку ряду» і мишкою вказуємо початок і кінець необхідного ряду. В результаті в списку рядів в основному вікні з'явиться ряд, який відповідає вирізаній частині (рис. 5а). Щоб продивитись цю частину ряду можна вибрати її і натиснути кнопку «Взяти частку ряду» (рис. 5б).

Далі вибираємо необхідний ряд для аналізу (наприклад, вирізаний) і натискуємо кнопку «Розрахувати ступінь р» (встановивши заздалегідь необхідні параметри). Відкриється одне вікно, в якому і проводиться робота.

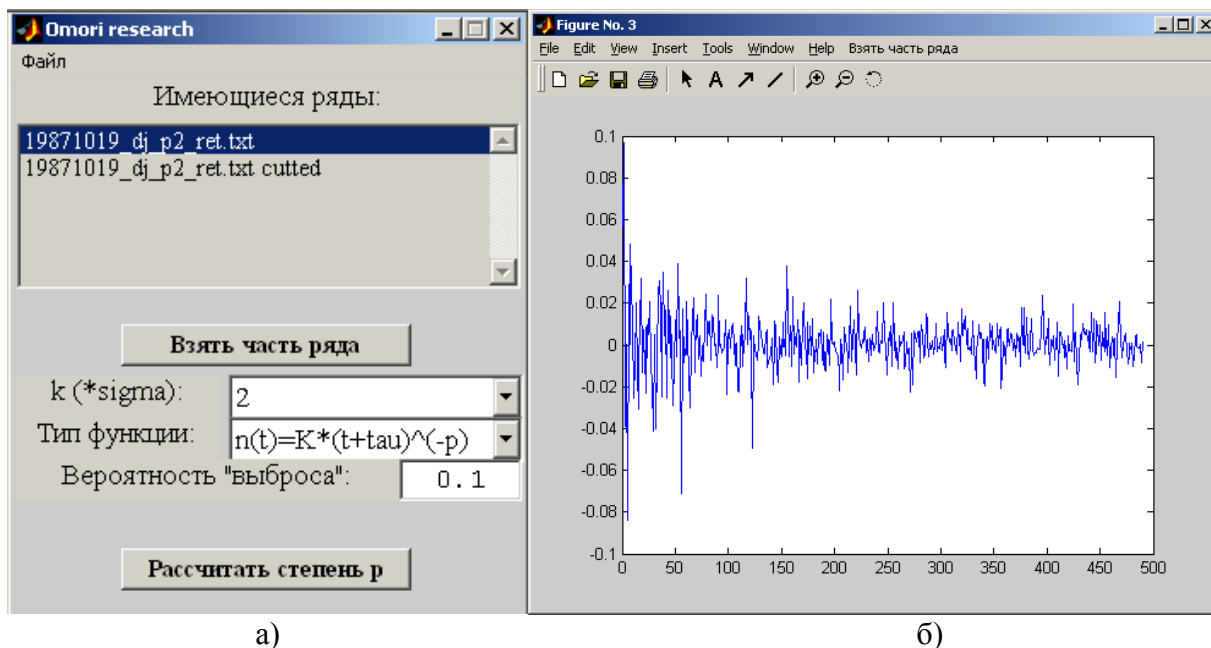


Рис. 5. Знаходження частини ряду (а) та його вигляд (б)

У цьому вікні, як і раніше, можна міняти масштаби (нормальний/логарифмічний) по осях і вибирати необхідну частину ряду.

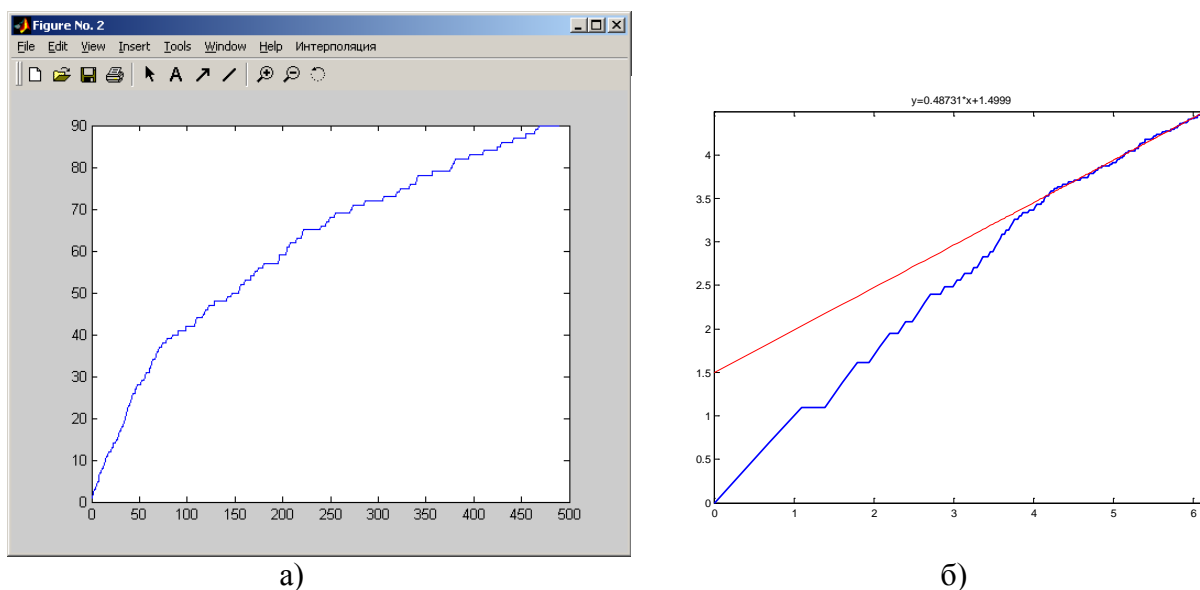


Рис. 6. Вікно для знаходження показника  $p$  (а) та прогноз закінчення релаксації за законом Омори

Для прогнозування встановлюємо подвійний логарифмічний масштаб і вибираємо ділянку ряду.

Після цього додатково до даного вікна (рис. 6а) відкриється ще одне.

Початкове вікно (заховано внизу) відображуватиме вибраний ряд накопичень і апроксимуючу криву. У заголовку графіка буде записана рівняння апроксимуючої прямої. Коефіцієнт, що цікавить нас, записаний біля  $x$  (рис. 6б).

Наприклад, для вирізані частини ряду взята для аналізу пряма ділянка в його кінці (рис. нижче). Отриманий коефіцієнт кута нахилу апроксимуючої прямої  $a = 0.48731$ .

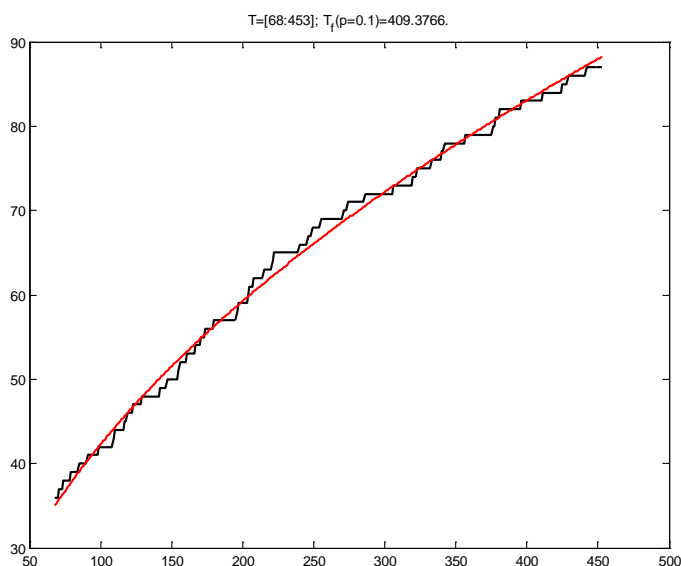


Рис. 7. Прогнозування дня закінчення релаксації

Для цього ж ряду друге вікно (буде першим після відкриття) показує вибрану для апроксимації частину ряду в нормальному масштабі і апроксимуючу криву. У заголовку

малюнка вказані інтервал, на якому був вибраний ряд ( $T = [68,453]$ , тобто дні з 68 по 453 після кризи або ж з початку ряду), і прогнозований день закінчення періоду релаксації:

$$T_f(p = 0.1) = 409.3766.$$

Тут  $f$  означає forecast.

Виходить, що період релаксації закінчився на 410-й день після кризи (вірогідність появи «екстремальної» флуктуації – 0.1 (рис. 7).

Додатково дослідження кризових явищ та посткризових станів складних систем можна досліджувати із залученням процедури рухомого вікна, спостерігаючи за зменшенням кількості аномальних відхилень із рухом вікна з часом. У цьому випадку залежність кількості відхилень від положення вікна відповідає закону того ж виду, проте в цьому випадку показник  $p$  буде від'ємним, оскільки у стані релаксації кількість відхилень та їх частота зменшуються.

Розрахунок показника  $p$  за формулою (1) проводиться після натиснення на кнопку «Рассчитать степень  $p$ », внаслідок чого буде відображено два вікна із відповідними графіками у лінійному та подвійному логарифмічному масштабі із вказанням значення показника. Додатково для підрахунку кількості аномальних відхилень можна використовувати критерій межі в одиницях середньоквадратичного відхилення, яка задається в полі « $k \cdot \sigma$ ».

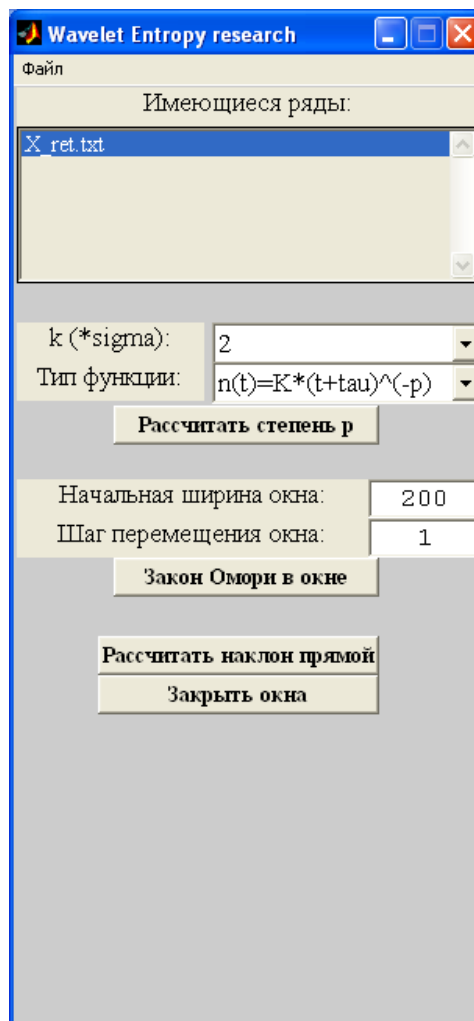


Рис. 8. Інтерфейс віконної процедури дослідження після кризових релаксацій

Процедура переміщеного вікна викликається натисненням на кнопку «Закон Омори в окне». Процедура використовує вже як вказані параметри, так і додатково ширину



досліджуваного вікна та крок його зміщення. Результати розрахунків процедури виводяться у вигляді, згаданому вище.

При розрахунку показника степеня у функціональній залежності (1) досить важливим є правильний вибір ділянки графіка, для якої будувати залежність. У випадку, коли програма автоматично не зможе вибрати таку ділянку, передбачено можливість її вибору користувачем за допомогою кнопки «Рассчитать наклон прямой» (рис. 8).

### **Завдання для самостійної роботи**

1. За отриманим у викладача завданням проведіть після кризовий аналіз часових рядів, які характеризують складні системи різної природи: геофізичні, біомедичні, фінансові.
2. Порівняйте результати, зробіть висновки щодо індивідуальної чи загальної природи досліджуваних явищ.

## **ЛАБОРАТОРНА РОБОТА № 11**

**Тема.** Кластерний аналіз.

**Мета.** Навчитися використовувати методи теорії графів для отримання характеристик внутрішньої структури економічної системи.

1. Побудова матриці відстаней на основі матриці крос-кореляцій.
2. Побудова мінімального остівного дерева.
3. Побудова ієрархічного дерева.
4. Порівняння одержаних результатів з кластерною структурою, отриманою методами теорії випадкових матриць.

### **Теоретичні відомості**

Останні роки група фізиків почала досліджувати фінансові та економічні системи, використовуючи інструменти та методологію, специфічні для фізичних наук. Такий вибір був мотивований кількома причинами. Перш за все, деякі фінансові та економічні системи є гарними екземплярами *складних систем*. По-друге, починаючи із 80-х років минулого століття велика кількість фінансових та економічних даних була записана у комп'ютерній формі і є легкодоступною для аналізу та тестування. Такі дані відбивають детальну інформацію про розглядувані процеси, і у випадку деяких бірж, містять інформацію про елементарні взаємодії, що проводяться на біржі (існують бази даних торгів та квот, де кожна транзакція, що виконується на біржі, документується). По-третє, аналіз та моделювання економічних і фінансових систем має базові теоретичні та прикладні аспекти. Базові аспекти стосуються моделювання систем, скомпонованих із кількох підмодулів, взаємодія яких зазвичай описується нелінійними моделями. Практичні аспекти стосуються жорсткого визначення та кращого виявлення присутнього в системі ризику у економічній діяльності.

На поточний момент опубліковано досить велику кількість робіт, що стосуються розглядуваного матеріалу, у математичних, економічних та фізичних журналах. Найбільш часто у статтях розглядається кілька питань:

- (i) вивчення статистичних властивостей зміни ціни акцій;
- (ii) новітній підхід до проблеми ціни опціонів;
- (iii) дослідження моделей фінансових ринків;
- (iv) порівняння між динамікою цін на фінансових ринках та швидкість зміни динаміки.

Приклади, наведені вище, не вичерпують весь спектр дослідницьких проблем, і є інші питання, які вимагають дослідження.

У даній лабораторній роботі буде розглянуто, як саме визначити наявність економічної інформації стосовно взаємодії різних рівнозначних економічних об'єктів (ринків, акцій тощо), що міститься у часових рядах цін акцій.

### Гіпотеза ефективного ринку

У основній стохастичній моделі динаміки ціни акції приймається, що натуральний логарифм ціни є процесом, подібним до процесу дифузії. Емпіричні дослідження визначення степеня зміни кореляції під час розвитку системи показують, що часова кореляція слабка або взагалі відсутня на часових інтервалах, що варіюються від одного торгового дня до кількох років. Відсутність довгочасової пам'яті у функції парної автокореляції логарифмів зміни цін відповідає гіпотезі про ефективність економічного ринку. На ефективних ринках уся інформація (минула, поточна та передбачення на майбутнє) доступна і негайно відбивається у ціні товарів.

Моделювання флуктуацій цінкових часових рядів у термінах випадкових процесів може бути на перший погляд парадоксальним, оскільки такий важливий індикатор, як ціна товару, є несуттєвим для випадкового процесу.

Розв'язок зазначеного парадоксу полягає в тому, що часовий ряд, багатий інформацією, є, насправді, незалежним від випадкових процесів. Таке твердження – один із висновків теорії складності. Внаслідок цього алгоритм дозволяє нам бути послідовними, одночасно заявляючи, що (1) часові цінні ряди є динамічними рядами, що містять досить великі обсяги інформації і (2) вони добре моделюються в термінах випадкових процесів.

### Отримання інформації з часової послідовності

Яким же чином можна отримати інформацію з часової послідовності цін товарів?

Основною ідеєю є дослідження кількох часових послідовностей одночасно; тобто, дослідження зміни коефіцієнта кореляції з плином часу прибутковостей пар акцій  $i$  та  $j$

$$Y_i = \ln P_i(t) - \ln P_i(t-1),$$

де  $i$  та  $j$  є порядковими номерами акцій, а  $P_i(t)$  є ціною закриття акції  $i$  в день  $t$ . Обчислення коефіцієнтів кореляції проводиться за наступною формулою

$$\rho_{ij} = \frac{\langle Y_i Y_j \rangle - \langle Y_i \rangle \langle Y_j \rangle}{\sqrt{(\langle Y_i^2 \rangle - \langle Y_i \rangle^2)(\langle Y_j^2 \rangle - \langle Y_j \rangle^2)}} \quad (1)$$

Кутові дужки  $\langle \dots \rangle$  – це середнє значення, розраховане для усіх днів розглядуваного часового періоду.  $\rho_{ij}$  визначає кореляцію, що спостерігається між логарифмом ціни акції  $i$  та логарифмом ціни акції  $j$ .  $\rho$  може змінюватись від -1 до 1.  $\rho = 1$  свідчить про існування повної кореляції між змінними  $Y_i$  та  $Y_j$ ,  $\rho = -1$  означає повну антикореляцію між  $Y_i$  та  $Y_j$ , а 0 спостерігається для некорельованих пар акцій.

Портфель із  $n$  акцій асоціюється із квадратною матрицею кореляцій розмірності  $n \times n$  яка є симетричною, з 1 на головній діагоналі. Обов'язкове число коефіцієнтів кореляції, що повністю описують ринок, рівне  $\frac{n(n-1)}{2}$ .

## Побудова матриці відстаней на основі матриці крос-кореляцій

Для отримання інформації з часових рядів, яку можна інтерпретувати, необхідно спочатку визначити метрику, в межах якої проводитиметься інтерпретація. В якості функції від коефіцієнтів кореляції можна використати узагальнену відстань, експериментальним визначенням якої є

$$d(i, j) = 1 - \rho_{ij}^2. \quad (2)$$

Замість вказаної можна використовувати також дещо видозмінену формулу:

$$d(i, j) = \sqrt{2(1 - \rho_{ij})}. \quad (2')$$

У такому випадку в дослідженнях  $d(i, j)$  чисельно задовольняє трьом аксіомам Евклідової метрики:

- (i)  $d(i, j) = 0$  якщо і лише якщо  $i=j$ ;
- (ii)  $d(i, j) = d(j, i)$ ;
- (iii)  $d(i, j) \leq d(i, k) + d(k, j)$ ,

і тоді  $d(i, j)$  може бути використана в якості Евклідової відстані для всіх практичних задач. Перша аксіома виконується внаслідок того, що  $d(i, j) = 0$  лише якщо кореляція повна ( $\rho = 1$ , що виникає у випадку, коли обидва економічних об'єкти виконують один і той же стохастичний процес). Друга аксіома виконується, оскільки матриця коефіцієнтів крос-кореляцій  $\rho$ , відповідно, матриця відстаней  $\mathbf{D}$  є симетричними за означенням. Третя аксіома виконується, оскільки вираз (2') еквівалентний Евклідовій відстані між двома векторами  $\tilde{Y}_i$  та  $\tilde{Y}_j$ , отриманими із часових рядів  $Y_i$  та  $Y_j$ , якщо використовувати кожний запис часового ряду як компонент вектора. Отриманий вектор мусить мати одиничну норму, а саме, він має бути отриманий відніманням від кожного запису середнього значення для всього ряду та нормалізацією його стандартним відхиленням.

Тепер, зважаючи на наведені міркування, можна дати практичне визначення метричної відстані між двома акціями певного портфеля. Однак, метрична відстань є недостатньою для однозначного визначення таксономії. Тому поглянемо на більш обмежуючий топологічний простір, який буде однозначно асоційовано з таксономією. Таким обмежуючим топологічним простором є субдомінантний ультраметричний простір, асоційований з матрицею Евклідових відстаней. Ультраметрична відстань визначається трьома аксіомами. Перші дві співпадають з описаними у визначенні вище, тоді як нерівність трикутника, (iii), замінюється більш сильною нерівністю, названою ультраметричною нерівністю

$$d(i, j) \leq \max\{d(i, k), d(k, j)\}. \quad (3)$$

В просторі елементів, де визначена метрика, завжди можна отримати субдомінантну ультраметрику. Приклад процедури отримання матриці субдомінантних ультраметричних відстаней з матриці Евклідових відстаней можна знайти у [2]. В ультраметричному просторі елементи розміщуються не вздовж однієї лінії, а в ієрархічному дереві. Набір акцій розглядуваного портфеля може інтерпретуватись як абстрактний простір,  $n$  елементів якого можуть бути з'єднані графом. У з'єднаному графі із взятими відстанями між усіма точками абстрактного простору мінімальне остівне дерево – МОД (minimal spanning tree – MST) має  $n-1$  ребро і найменшу довжину із довжин всіх остівних дерев, виходячи з суми відстаней між

двома точками. Структура ієрархічного дерева отримується з матриці ультраметричних відстаней та МОД, асоційованого з нею.

Мінімальне остівне дерево та асоційоване з ним ієрархічне дерево показують існування кластерів будь-яких активів ринків, відомості про які мають велике значення з економічної точки зору. Отримана таксономія дозволяє групувати економічні об'єкти, що є гомогенними у відношенні до економічної активності.

### Побудова мінімального остівного та ієрархічного дерев

Розглянемо систему індексів MSCI, яка включає 49 країн, з яких: 23 розвинених ([www.msci.com](http://www.msci.com)). Для побудови мінімального остівного дерева, асоційованого з Евклідовою відстанню, необхідно виконати наступні дії.

Спочатку усі елементи матриці відстаней  $D$  розташовуються у порядку неспадання (кожний наступний більший або рівний попередньому). Наприклад, маємо наступні вісім відстаней:

Spain-Italy	$d=0.61177$
France-Italy	$d=0.64137$
Nedherlands-France	$d=0.6462$
Germany-France	$d=0.65383$
France-Spain	$0.69175$
Argentina-Brazil	$d=0.73039$
Switzerland-Italy	$d=0.77672$
Belgium-France	$d=0.80568$
Mexico-Argentina	$d=0.8217$

Мінімальне остівне дерево послідовно будується шляхом з'єднання усіх елементів у граф, виходячи із мінімальних відстаней між країнами на ринку.

Почнемо із пари елементів з найменшою відстанню. У даному випадку це дві країни: **Spain** та **Italy** ( $d=0.61177$ ). Наступною найменшою за порядком відстанню є відстань між елементами **France** та **Italy** ( $d=0.64137$ ), тому додаємо до МОД елемент **France** та з'єднуємо його з елементом **Italy**. Таким чином, отримане дерево вже буде мати вигляд **France-Italy-Spain**. Наступною парою елементів у відсортованому списку є країни **Nedherlands-France**, включаємо елемент **Nedherlands** до дерева. Аналогічно додаємо країну **Germany**. Наступною парою країн з найменшою відстанню є **France-Spain**, проте обидві країни вже включено в мінімальне остівне дерево, тому для цієї пари нічого не робимо. Аналогічно продовжуємо процедуру, поки всі елементи не будуть включені в дерево.

Таким чином, в результаті перебору усіх відстаней, отримаємо мінімальне остівне дерево, що буде характеризувати взятий ринок, стосовно того, які фірми взаємодіють між собою тісніше, на скільки груп якого характеру розділений ринок і т.д.

### Хід роботи

У середовищі MatLab 6.5 завантажимо пакет для дослідження структури ринку, заданого часовими рядами його складових, змінивши шлях до програми та набравши у командному рядку **demo2** та натиснувши **Enter**. Внаслідок цього відкриється вікно програми (рис. 1).

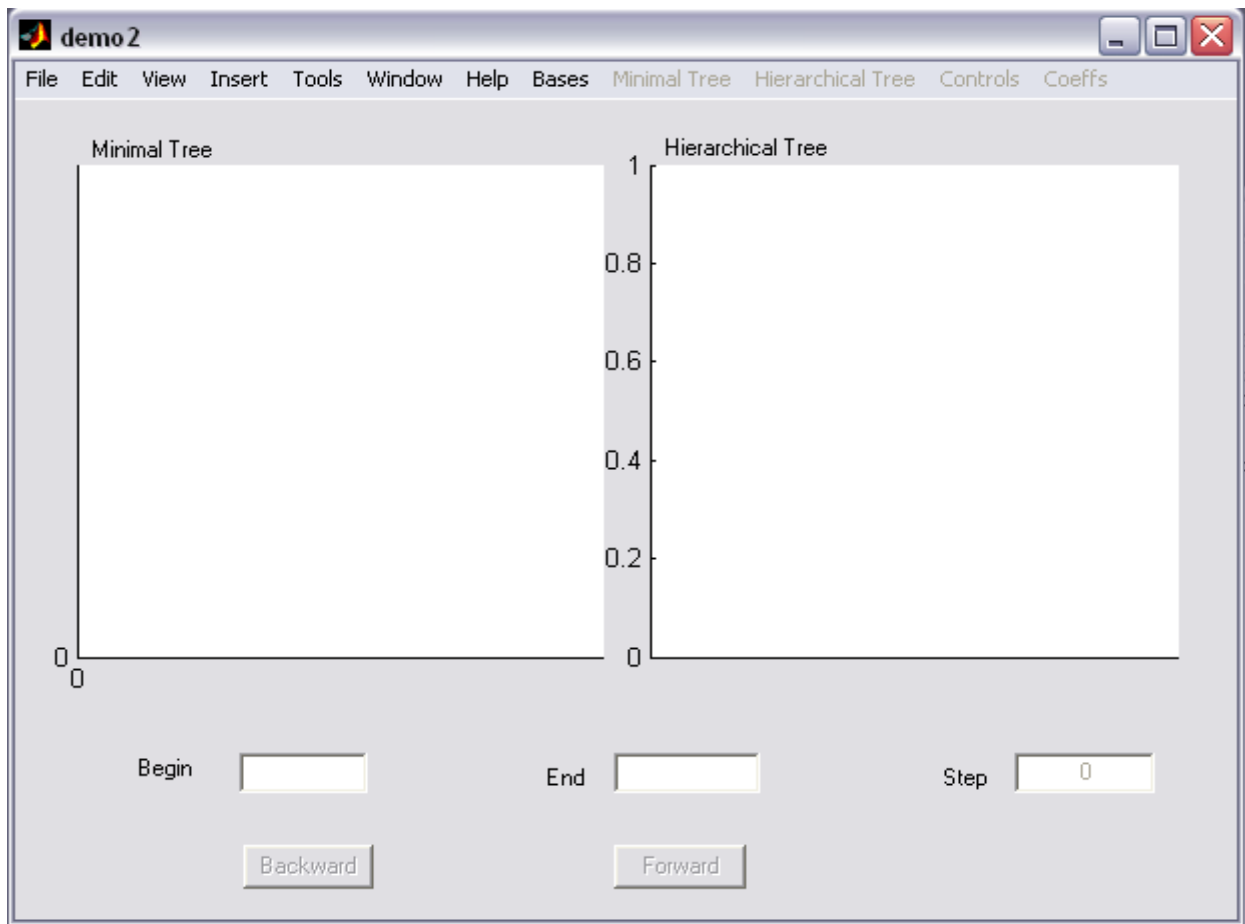


Рис. 1.

На рисунку видно три частини області користувача:

- меню, з якого викликаються дії, що необхідно виконати над досліджуваним об'єктом;
- область, де відображуються мінімальне та ієрархічне дерева системи;
- область, де задається часовий проміжок, за яким на екрані робляться всі обчислення.

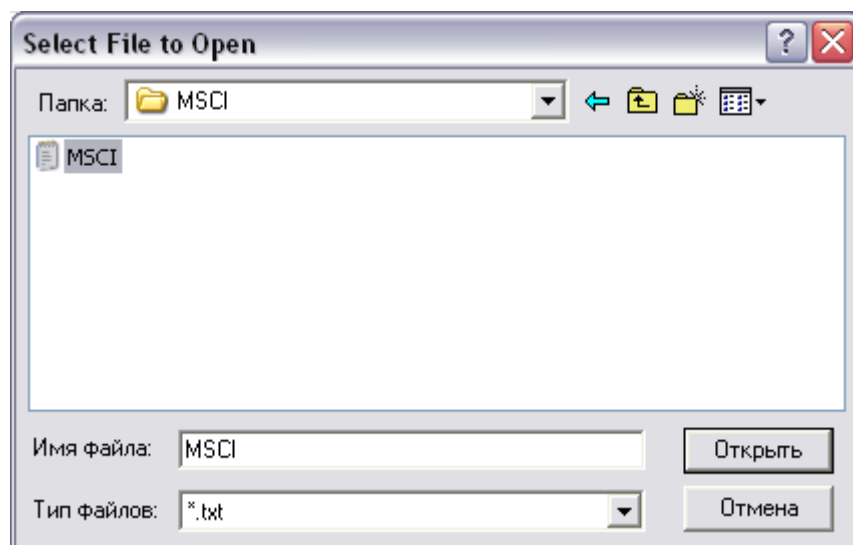


Рис. 2.

Для того, щоб розпочати роботу з програмою, необхідно вибрати та завантажити базу.

Для цього необхідно скористатися пунктом меню **Bases**→**Load** (бази – завантажити). За допомогою вікна для відкриття файлів, виберіть базу (наприклад, базу MSCI: в папці Bases вказати файл **msci.txt**).

Після цього Вам необхідно буде вказати проміжок часу, по якому вам потрібно побудувати дерева. Зразу після відкриття з'явиться вікно, де ви повинні вказати номер першого дня та довжину проміжку. Перший день може бути не менший 2-го. Наприклад, вас цікавить проміжок від 2 до 204. Для цього введіть наступне:

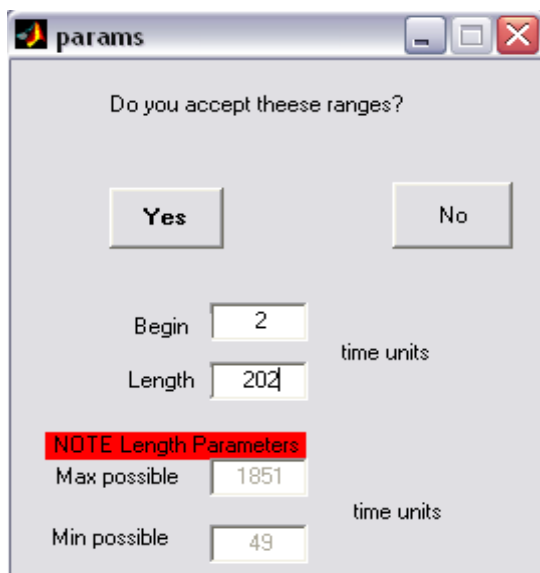


Рис. 3

Після введення даних, на екрані з'явиться 2 графи: граф мінімального остівного дерева та граф середнього дерева (рис. 4)

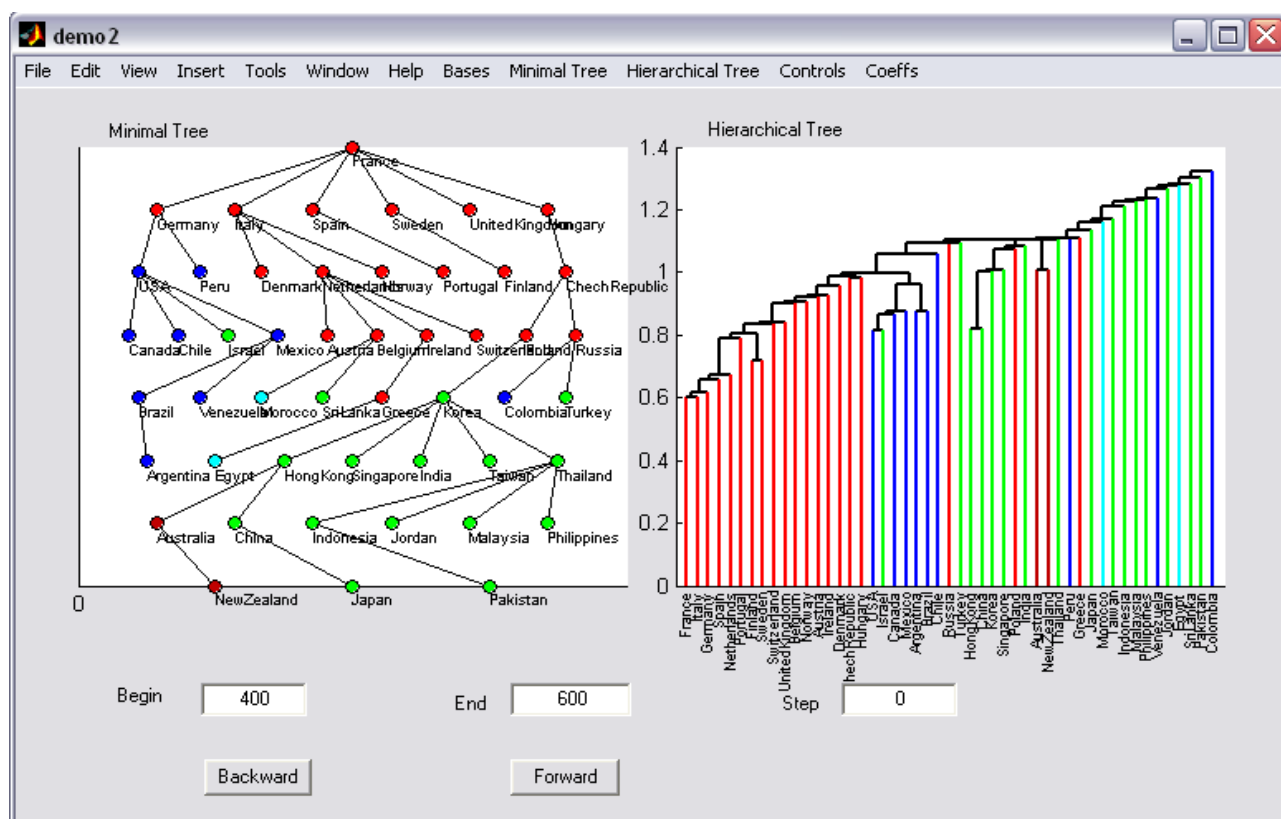


Рис. 4.





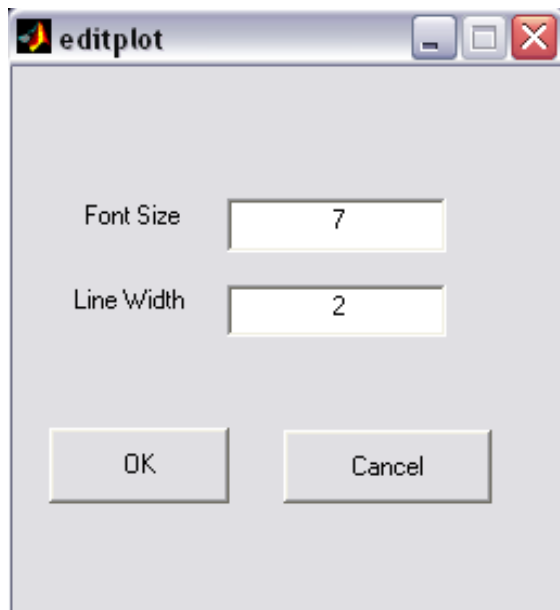


Рис. 6.

При зміні розмірів вікна, є можливість під коректувати висоту прямокутника, в якому зображене дерево. Для цього в лівому боці вікна існує смуга прокрутки, за допомогою якої можна задати необхідні розміри дерева (див. рис. 7).

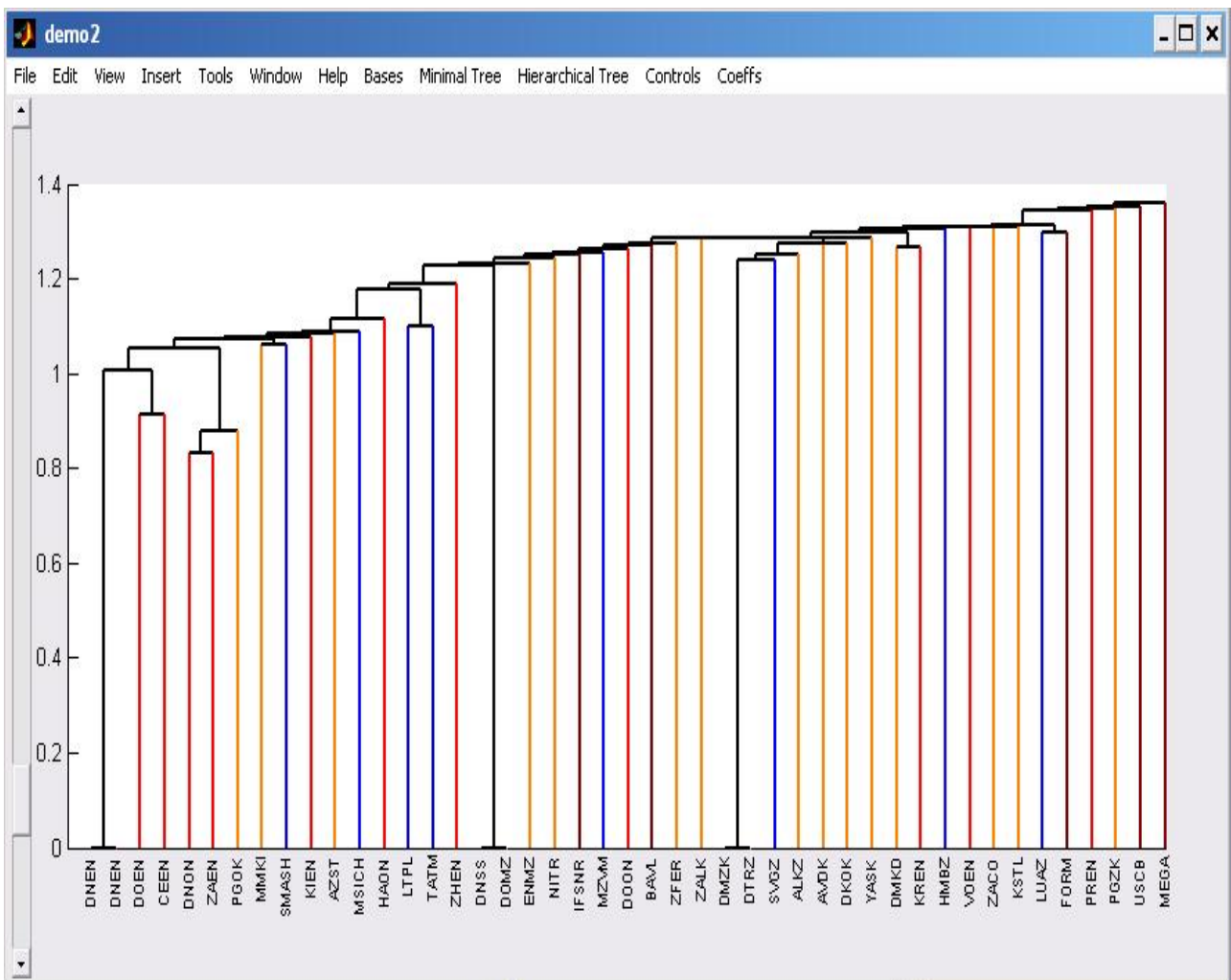


Рис. 7. Зміна висоти ієрархічного дерева

Ви повинні зберегти отримані зображення у графічному файлі. Перед цим необхідно прибрати з екрану елементи управління (поля введення інтервалу та кроку, слайдер зміни висоти графіку). Включити або виключити відображення елементів управління здійснюється опціями меню «Controls»:

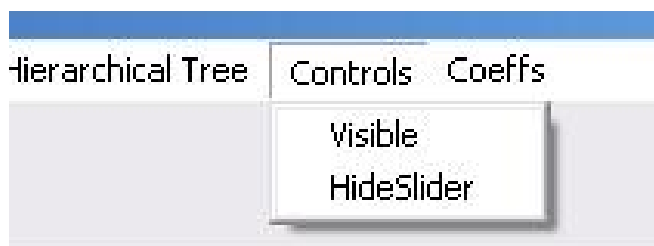


Рис. 8. Керування відображенням управляючих елементів.

Перший пункт меню «**Visible**» дозволяє вимкнути або вимкнути поля введення, які зображуються під графіками. Другий пункт меню «**HideSlider**» керує відображенням смуги прокрутки, яка знаходиться в лівому краю вікна і задає висоту зображень дерев. Повторний вибір опції включає або виключає зображення елементу у вікні.

Експорт зображення дерева можна зробити за допомогою стандартного меню **File**→**Export** (Файл – експортувати). На екрані Ви побачите діалогове вікно збереження (рис. 9).

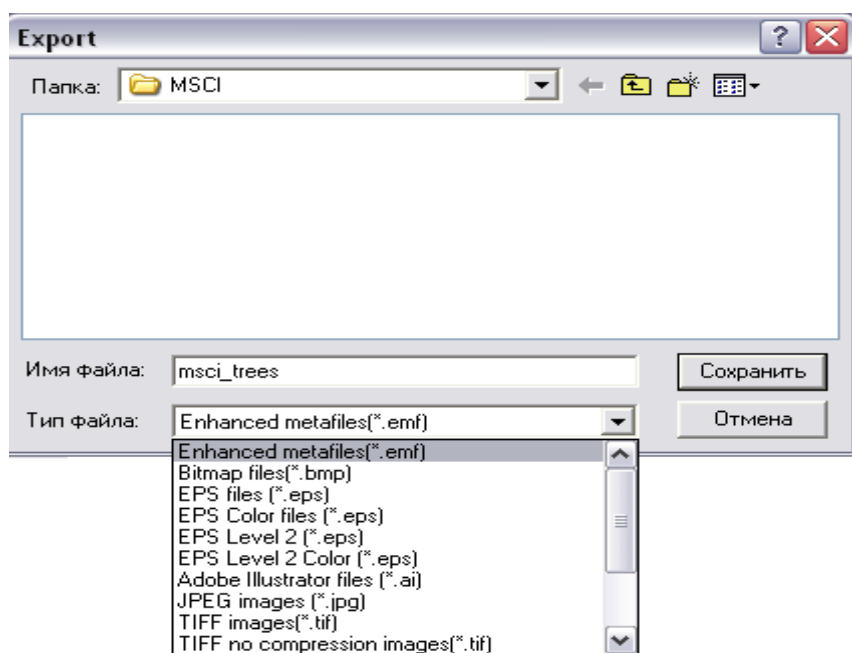


Рис. 9

Вам необхідно буде перейти в папку для збереження малюнків, вказати ім'я та тип графічного файлу.

Для того, щоб дізнатися, якому дню відповідає певний номер в базі, необхідно переглянути файл дат, який має розширення **.dt** та знаходиться в тій же папці, в якій і сама база. Наприклад, у випадку бази MSCІ цей файл називається **msci.dt**. Найзручніше переглядати його за допомогою вбудованого редактору Matlab. Для цього перейдіть до вікна Matlab. Виберіть опції меню **File**→**Open** або скористайтесь кнопкою на панелі **Open** (відкрити). За допомогою стандартного діалогового вікна виберіть папку, де зберігається Ваша база. В полі вибору типу файлів виберіть **All files (\*.\*)** (всі файли) та вкажіть на відповідний файл **.dt**. Ви побачите вміст файлу дат бази (рис. 10). У кожному рядку файлу

записана дата відповідного дня.

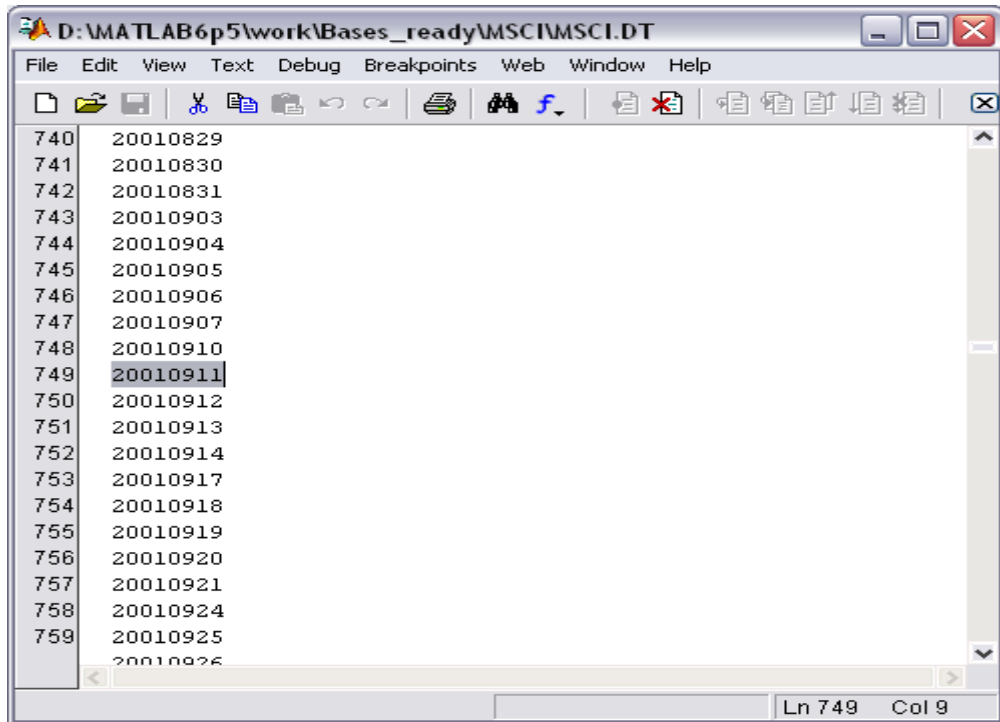


Рис. 10.

Вбудований редактор вказує номери кожного рядка, це допоможе швидко знайти номер дня потрібної дати. Наприклад, 11 вересня 2001 року відповідає 749-й день.

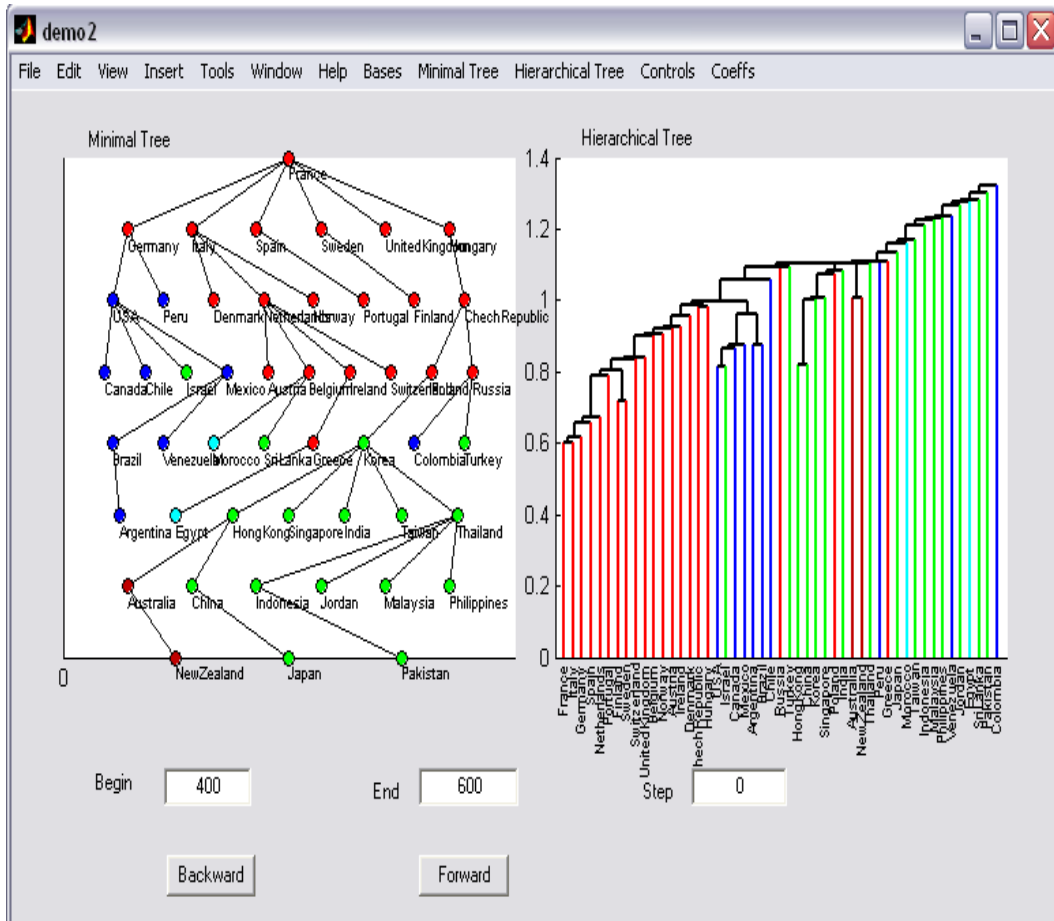


Рис. 11

При дослідженні кризових явищ за допомогою теорії графів, ми будемо переміщувати вибраний відрізок та спостерігати зміни зображень графів. У випадку, коли кінець вибраного проміжку дійде до дня кризи на ринку, який досліджується, Ви помітите зміни в зв'язках між елементами системи. Наприклад, при дослідженні індексу міжнародної інвестиційної привабливості MSCI, будемо спостерігати за Сполученими штатами на мінімальному та ієрархічному дереві.

Побудуємо дерева, вибравши проміжок 400-600 (рис.11). Будемо наближатися з довільним кроком до 749-го дня: 11 вересня 2001 року.

Зверніть увагу на американський кластер (зображений синім кольором). Сполучені штати, Канада та Мексика зберігають зв'язок на дереві до того часу, доки кінець вибраного проміжку не досягне 749-го дня. На малюнку зображена ситуація при виборі проміжку 556-756. На рис. 12 ми бачимо, що Сполучені штати втрачають зв'язки з іншими країнами американського кластеру, крім Канади.

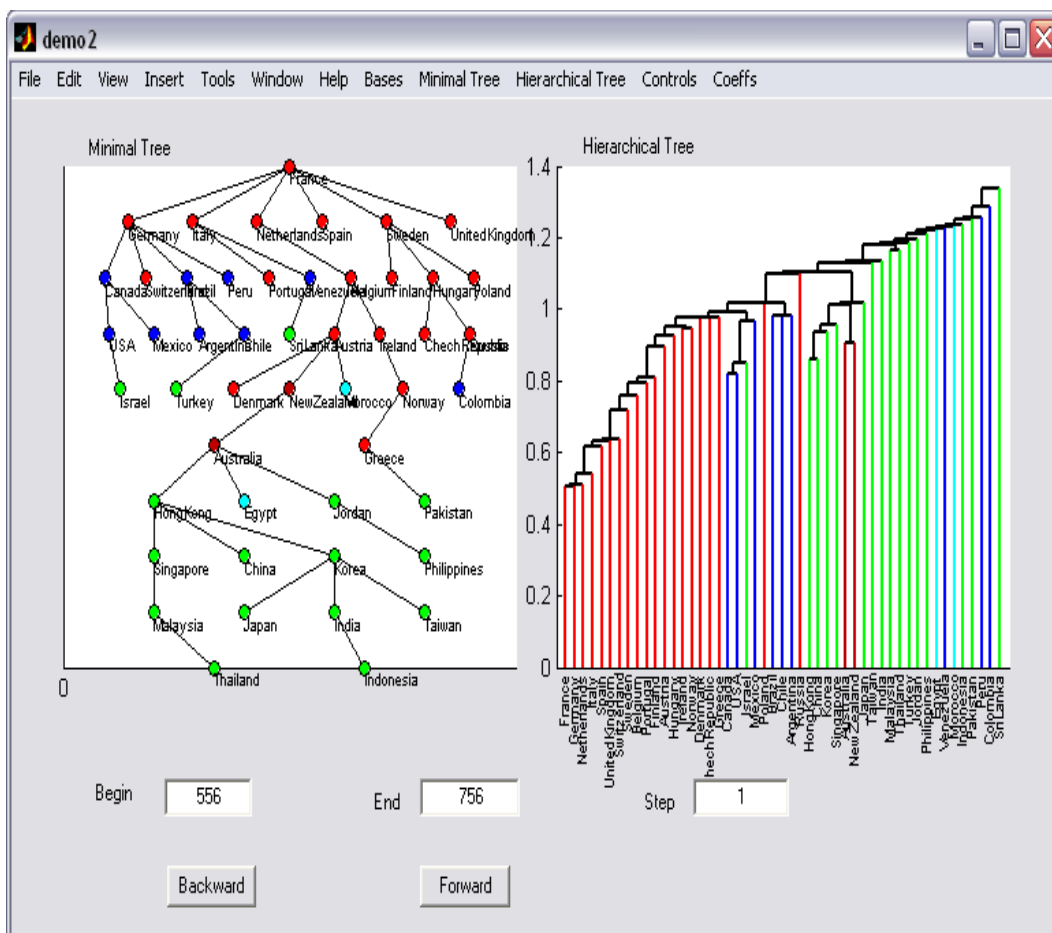


Рис. 12.

Таким чином можна досліджувати структуру і динаміку складних систем, відслідковувати їх зміни з часом, аналізувати вплив критичних, кризових і шокових явищ.

### Завдання для самостійної роботи

1. Виберіть в папці Bases відповідну базу даних, проведіть дослідження
  - кластерної структури, її динаміки, поясніть і інтерпретуйте результати;
  - порівняйте одержані результати з кластерною структурою, отриманою методами теорії випадкових матриць

Результати-звіт представте у вигляді текстового файлу.

### Контрольні питання

1. Поясніть основну ідею кластерного аналізу
2. Про що свідчить формування відповідних кластерів ?
3. Що можна сказати відносно поведінки кластерної структури в умовах кризи?
4. Як проявляються процеси глобалізації на кластерній структурі відповідних баз активів?

### ЛАБОРАТОРНА РОБОТА № 12

**Тема.** Структурні характеристики складної системи.

**Мета.** Навчитися використовувати елементи теорії графів для отримання структурних характеристик складної економічної системи.

1. Коефіцієнт кластеризації.
2. Середня довжина шляху на графі.
3. Діаметр графа.
4. Максимальна степінь вершин графа.

### Теоретичні відомості

Аналіз структури складної системи по її графам має суб'єктивний характер, оскільки важко визначити зміни в усій системі, розглядаючи її граф. Необхідно оперувати певними числовими характеристиками системи, по зміні яких можна робити висновок про стан системи. Такими характеристиками можуть виступати коефіцієнт кластеризації, діаметр графа, середня довжина шляху на графі. Розглянемо докладніше ці коефіцієнти та спосіб їх обчислення.

### Коефіцієнт кластеризації (транзитивності) графа

В суспільстві існують такі групи знайомих, коли кожен знає всіх інших в групі. Такі групи в теорії графів називають кластерами. Коефіцієнт кластеризації чисельно характеризує це явище в графі.

Існують різні способи визначення коефіцієнту кластеризації. Розглянемо деякі з них, які на сьогодні найбільш застосовні у практичних використаннях.

1) Візьмемо деяку вершину з графу та будемо обчислювати коефіцієнт кластеризації цієї вершини. До цієї вершини безпосередньо приєднані  $k_i$  вершин. Між цими вузлами існує  $E_i$  ребер. Максимально можлива кількість ребер, при якій ці вершини утворюють кластер, дорівнює

$$k_i(k_i + 1)/2.$$

**Коефіцієнт кластеризації вершини** є відношення між числом  $E_i$  ребер, які фактично існують між даними вершинами, до кількості ребер, при якому б ці вершини утворили кластер, дає величину коефіцієнта кластеризації вершини:

$$C_i = \frac{E_i}{E_{\max}} = \frac{2E_i}{k_i(k_i + 1)}.$$

Розглянемо приклад графу (рис. 1).

Будемо обчислювати коефіцієнт кластеризації елемента **France**. Безпосередньо з ним з'єднано 5 країн. Між 6 країнами максимально можлива кількість зв'язків буде

$E_{\max}=6 \cdot 5 / 2 = 15$ . Реально існує 8 зв'язків. Коефіцієнт кластеризації вузла **France** обчислюємо за формулою:

$$C_i = \frac{E_i}{E_{\max}} = \frac{2E_i}{k_i(k_i + 1)} = \frac{2 \cdot 8}{6 \cdot 5} \approx 0,53.$$

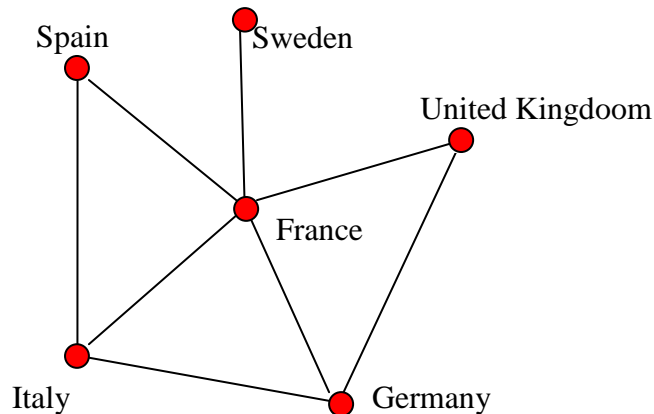


Рис. 1.

Коефіцієнт кластеризації графу обчислюється як середнє арифметичне коефіцієнтів кластеризації всіх вершин.

2) **Коефіцієнт транзитивності.** Коефіцієнт кластеризації обчислюється як відношення кількості трикутників (3 вершини, з'єднані одна з одною) до кількості з'єднаних пар ребер.

### Середня довжина шляху на графі

Між кожною парою елементів системи на довільному графі можливі різні шляхи. Наприклад, розглянемо приклад з попередньої роботи. Нехай дано підрахунки відстаней між трьома елементами системи у вигляді матриці відстаней:

	Spain	Italy	France
Spain	0	0.31	0.69
Italy	0.31	0	0.34
France	0.69	0.34	0

Нам необхідно знайти матрицю мінімальних відстаней між кожними двома елементами. Зверніть увагу на те, що відстань від елемента **France** до елемента **Spain** дорівнює 0.69, а якщо розглянути шлях **France-Italy-Spain**, то відстань цього шляху буде дорівнювати сумі відстаней **France-Italy** та **Italy-Spain**. В нашому випадку це  $0,31+0,34=0,65$ , що є менше, ніж безпосередньо відстань **France-Spain**. З цих даних ми можемо утворити матрицю найкоротших (оптимальних) відстаней в графі.

	Spain	Italy	France
Spain	0	0.31	0.65
Italy	0.31	0	0.34
France	0.65	0.34	0

Серед цих шляхів знаходиться середнє арифметичне, яке і буде наступним показником, за яким ми будемо слідкувати.

## Діаметр графа

Діаметр графа – це найдовша відстань у графі. Її будемо знаходити з матриці оптимальних шляхів, приклад знаходження якої був вище розглянутий.

В нашому випадку діаметром графа буде відстань 0,65.

## Максимальний степінь вершин графа

Степеном вершини графа називається кількість ребер, які безпосередньо виходять із даної вершини. Розглянемо приклад дерева (рис. 2).

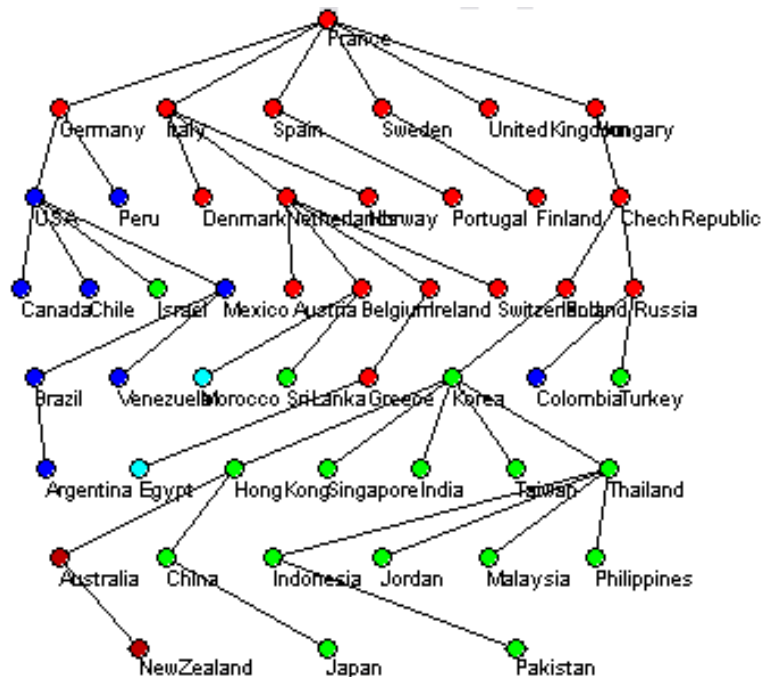


Рис. 2.

Найвищий на графі елемент – **France** має степінь 6, тому що безпосередньо з'єднаний з 6-ма країнами. Елемент **Russia** має степінь 3 і т. д. Найвищий степінь у елемента **France**. Тому показником параметру «*Найвищий степінь у графі*» буде число 6.

## Хід роботи

Аналогічно до попередньої лабораторної роботи, у середовищі MatLab 6.5 завантажимо пакет для дослідження структури ринку, набравши у командному рядку **demo2** та натиснувши **Enter**. Завантажуємо потрібну базу індексів та будуємо дерева на довільному проміжку.

Для обчислення коефіцієнтів складної системи необхідно скористатися меню **Coeffs** (коефіцієнти). Для перерахування всіх коефіцієнтів необхідно викликати опцію меню **Coeffs-All**. Після цього на екрані з'явиться вікно задання параметрів підрахунку (див. рис. 3).

Для підрахунку необхідно ввести довжину вікна (**Window size**), номер першого (**First day of the window**) та останнього (**Last day of the window beginning**) дня проміжку розрахунків, а також поріг для відкидання слабких зв'язків (**Threshold**).

Нехай, наприклад, ми маємо базу з 1853 днів. Нам потрібно прорахувати проміжок часу від 400-го до 800-го дня вікнами шириною 200 днів (див. рис.4)



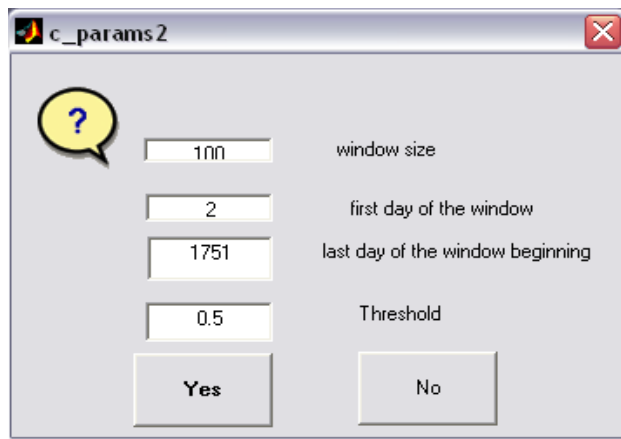


Рис. 3.

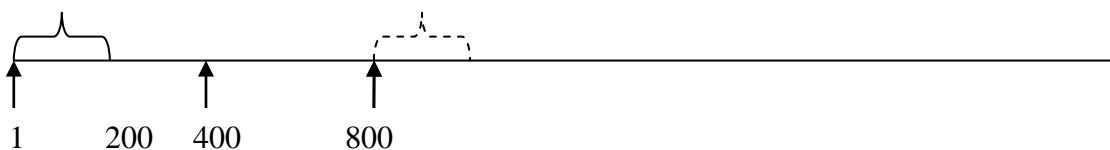


Рис. 4.

Довжина вікна (**Window size**) повинна бути 200, перший день (**First day of the window**) – 2, останній день (**Last day of the window beginning**) – 800. При цьому останній раз вікно буде захоплювати проміжок часу 800-1000.

Для обчислення коефіцієнтів кластеризації, необхідно створити так зване дерево найсильніших зв'язків. При цьому ребрами будуть вважатися ті зв'язки, які сильніші (знаходяться ближче один до одного), ніж встановлений поріг (**Threshold**). Цей поріг вам доведеться підібрати, якщо встановлений (0,5) не дасть необхідного результату.

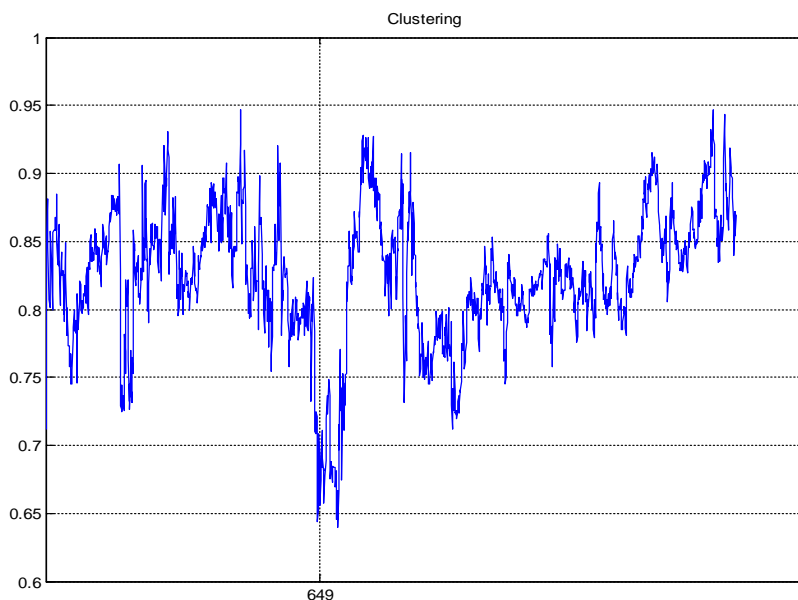


Рис. 5. Коефіцієнт кластеризації

Отже, після введення параметрів та натискання кнопки **Yes**, почнеться підрахунок коефіцієнтів. Після підрахунку на екран з'являться вікна з графіками відповідних коефіцієнтів. Ви можете зберегти потрібний графік (опція меню **File**→**Save** у вікні графіку),

або експортувати в графічний файл (опція меню **File**→**Export** у вікні графіку). Якщо ви ненароком закрили вікно з потрібним графіком, для його відображення можна скористатися іншими опціями меню **Coeffs**. До речі, після перерахунку коефіцієнтів, ви не зможете відновити старі графіки. Обов'язково збережіть результати Ваших обчислень.

Для прикладу розглянемо динаміку зміни коефіцієнтів під час кризи. На рис. 5. зображено графік коефіцієнту кластеризації для індексу MSCI.

Позначкою 649 вказаний день 11 вересня 2001 року, коли будівля Міжнародного Торгового Центру в Сполучених Штатах була атакована терористами.

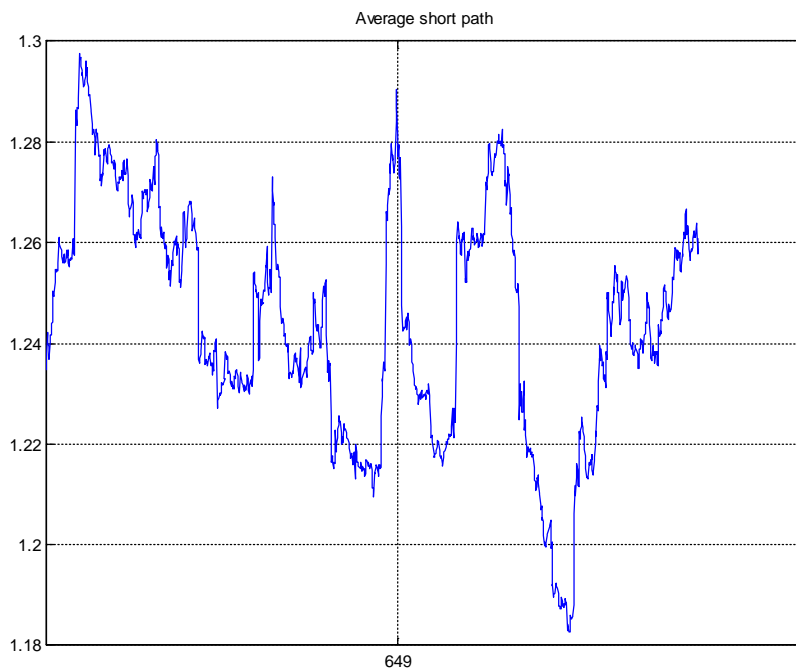


Рис. 6. Середня довжина шляху

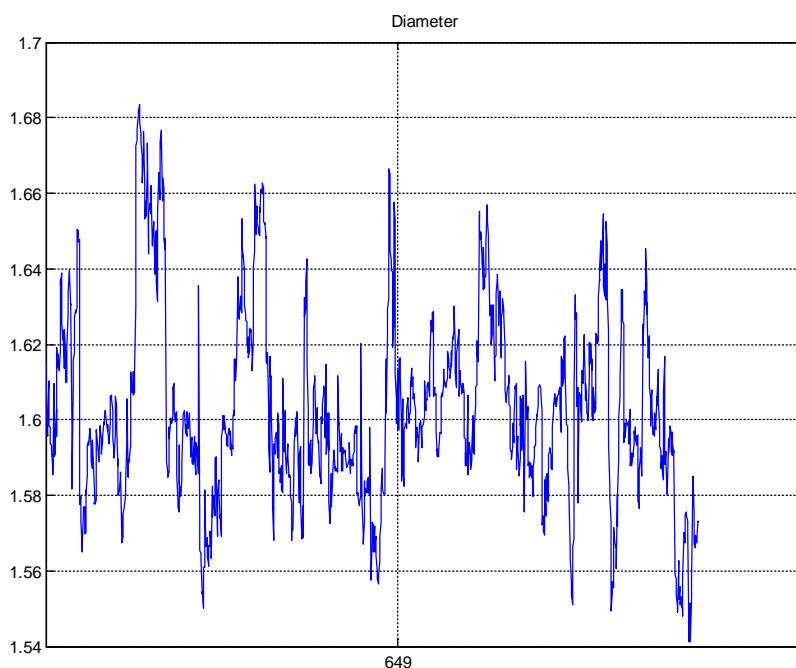


Рис. 7. Діаметр графу

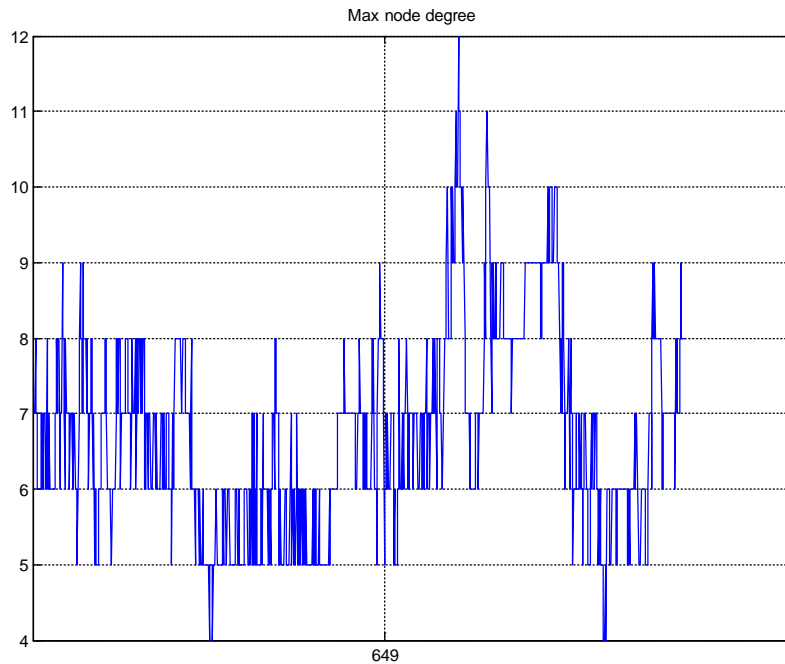


Рис. 8. Найвищий степiнь вершини

Оскільки розраховані параметри є найбільш широко вживаними при дослідженні структури складних мережеподібних систем, то відслідковуючи їх еволюцію, можна робити певні висновки про структурні зміни, наприклад, з метою передбачення небажаних з них.

#### Завдання для самостійної роботи

Виберіть в папці Bases відповідну базу даних, проведіть дослідження

- кластерної структури, її динаміки, поясніть і інтерпретуйте результати;
- порівняйте одержані результати з кластерною структурою, отриманою методами теорії випадкових матриць.

Результати-звіт представте у вигляді текстового файлу.

#### Контрольні питання

1. Поясніть основну ідею кластерного аналізу.
2. Про що свідчить формування відповідних кластерів ?
3. Що можна сказати відносно поведінки кластерної структури в умовах кризи?
4. Як проявляються процеси глобалізації на кластерній структурі відповідних баз активів?

### ЛАБОРАТОРНА РОБОТА № 13

**Тема.** Найпростіші мережі та мережні міри складності.

**Мета.** Навчитися використовувати елементи теорії графів для отримання спектральний і топологічних мір складності.

1. Мережі, моделі мереж.
2. Спектральні характеристики мереж.
3. Топологічні міри складності.

## Теоретичні відомості

Для сучасних складних систем характерна нерегулярність зв'язків і висока чисельність елементів, яка може досягати десятків і сотень тисяч. Таким системам та їх мережним моделям, які володіють нетривіальними топологічними властивостями, найбільше відповідає термін «комплексні». Комплексною мережею вважається система, яка

- (а) складається з великої кількості компонентів;
- (б) допускає «далекосяжні» зв'язки між компонентами;
- (в) володіє великомасштабною (у тому числі просторово-часовою) мінливістю.

Дана мережа є графом з досить великою кількістю вузлів різної природи, що характеризуються багатовимірним кортежем ознак і динамічно мінливими зв'язками; розподіл ознак вузлів і характеристик зв'язків може бути описаний ймовірнісною моделлю (багатомірним розподілом).

Основною причиною підвищення актуальності розробок у області теорії і практики комплексних мереж є результати сучасних досліджень реальних комп'ютерних, біологічних і соціальних мереж. Властивості багатьох реальних мереж істотно відрізняються від властивостей класичних випадкових графів з рівномірними зв'язками між вузлами, які донедавна розглядалися в якості їх базисного математичного модельного прототипу, і тому побудову їх моделей було запропоновано здійснювати з використанням зв'язних структур і степеневих розподілів.

У теорії комплексних мереж виділяють три основні напрямки:

- дослідження статистичних властивостей, які характеризують поведінку мереж;
- створення моделей мереж;
- прогнозування поведінки при зміні структурних властивостей мереж.

Комплексні мережі використовуються для моделювання об'єктів і систем, дослідження яких іншими способами (за допомогою спостереження або активного експерименту) недоцільні або неможливі. Комп'ютерні мережі відносяться до мереж, які постійно ростуть і розвиваються. Серед факторів, що впливають на зростання мережі в першу чергу необхідно відзначити розмір або протяжність локальної мережі, яка визначається відстанню між найвіддаленішими станціями, при якій в нормальному режимі роботи вузлів чітко розпізнаються колізії, і кількість об'єднаних у мережу комп'ютерів. Для Інтернет-мереж цей розмір називається діаметром мережі і складає приблизно 1 км відстані, що дозволяє отримати високу швидкість зв'язку та максимально можливий рівень сервісу. При зростанні мережі збільшується кількість колізій, різко падає її корисна пропускна здатність і швидкодія передавання сигналу. Обмеження мережі за довжиною є передумовою вибору структури мережі, розбиття її на окремі частини (сегменти), появи додаткових серверів з новою мережею зв'язків, проблеми генеруються в контексті технологій так званої «останньої милі». Спостерігається динаміка зростання мережі, своєрідна кластеризація, сервери виступають центрами утворених кластерів, відбувається просторове позиціонування компонент мережі у вигляді чітких ієрархічних структур.

Мережа розглядається як множина сегментів, кожен з яких закінчується точкою розгалуження або кінцевої вершиною мережі. Вершинами мережі є сервери, комутатори й кінцеві користувачі, загальну кількість яких позначимо  $N$ . Локальні комп'ютерні мережі є об'єктними прототипами графових структур і тому для їх дослідження застосовують методи теорії графів.

Моделювання мереж із використанням апарата теорії графів є важливим напрямком досліджень дискретної математики. В останні роки зросла зацікавленість дослідників до складних мереж з великою кількістю вузлів, зокрема до комп'ютерних мереж, структура яких нерегулярна, складна і динамічно розвивається в часі. Для таких мереж доводиться генерувати стохастичні графи з величезною кількістю вершин.

У загальному вигляді модель комп'ютерної мережі являє собою випадковий граф, закон взаєморозміщення ребер і вершин для якого задається розподілом ймовірностей.

У даний час найпоширенішими є два основних підходи до моделювання складних мереж:

- випадкові Пуассонівські графи та узагальнені випадкові графи;
- модель «тісного світу» Ватса і Строгатса та її узагальнення, еволюційна модель зростання мережі Барабаші й Альберт.

Перший передбачає генерацію випадкового графа із заздалегідь відомою кількістю вершин і заданими ймовірнісними властивостями. Його ще називають *графом Ердоша-Рені* зі сталою кількістю вершин  $N$ . Розподіл ступенів вузлів  $k$  для цього графа визначається формулою Пуассона  $P(k) = e^{-\langle k \rangle} \langle k \rangle^k / k!$ . Побудова графа здійснюється генеруванням, коли до  $N$  відокремлених вершин послідовно додаються ребра, що з'єднують випадковим чином довільні пари вершин. Початково граф складатиметься із сукупності малих вершин, які в процесі генерування з часом розростаються до гігантського кластера зв'язаних між собою вершин, число яких є скінченною частиною загальної кількості  $N$ . При генерації постійно зростає ймовірність зв'язування вершин, яка досягає з часом деякого критичного значення. В результаті процесу, який має характер фазового переходу, граф спонтанно розростається до гігантського кластера вершин, пов'язаних між собою, що нагадує конденсацію краплі води в перенасиченій парі.

*Модель Ваттса-Строгатца* є комп'ютерною моделлю тісного світу. Її побудова зводиться до наступного: розглядається одновимірний, замкнений у кільце, періодичний ланцюг, який складається із  $N$  вершин. Спочатку кожен вузол з'єднують з іншими сусідніми, які знаходяться від неї на відстані, не більшій за  $k$ , а потім кожне ребро з певною ймовірністю  $t$  перез'єднується з довільною вершиною, що призводить до трансформації регулярного ланцюга у граф тісного світу (рис. 1).

Оскільки в цій моделі кількість ребер є сталою, а ймовірності реалізації графів – різні, то вона зводиться до канонічного ансамблю графів і описує реально існуючі мережі, топологія яких не є ані цілком регулярною, ані цілком випадковою.

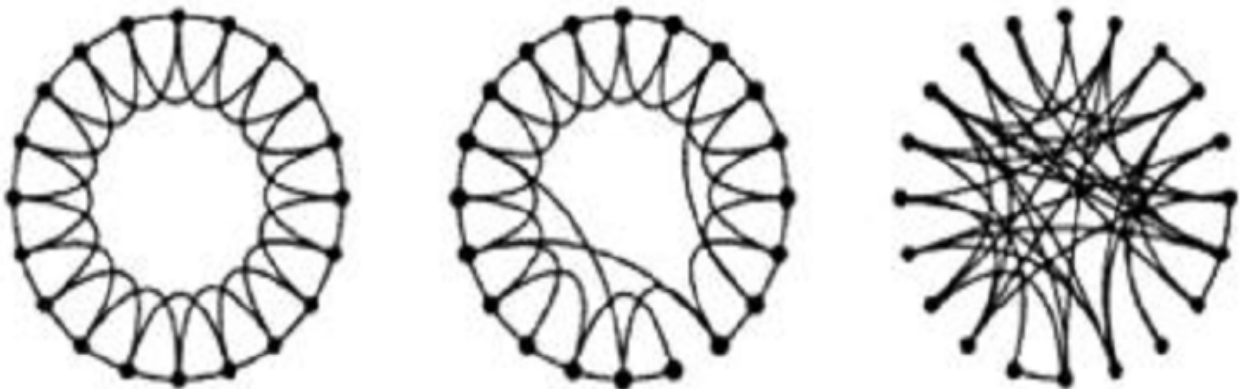


Рис. 1. Трансформація регулярного ланцюга у граф тісного світу й у випадковий граф

Більшість реальних графів підпорядковуються степеневому закону розподілу  $P(k)$ . Ці графи побудови мереж описуються моделлю переважного приєднання Барабаші-Альберт. Через далекоглядні взаємодії у системі не існує масштабу зміни характерних величин. Ріст і переважне приєднання є основними механізмами побудови безмасштабних (масштабно-інваріантних) мереж.

Нехай вузол  $i$  має  $k_i$  зв'язків і він може бути приєднаним (зв'язаним) до інших вузлів  $k_j$ . Ймовірність приєднання нового вузла до вузла  $i$  залежить від ступеня  $k_i$  вузла  $i$ . Величину  $W(k_i) = k_i / \sum_j k_j$  називають переважним приєднанням (preferential attachment). Не

всі вузли мають однакову кількість зв'язків, тому вони характеризуються функцією розподілу  $P(k)$ , що визначає ймовірність того, що випадково вибраний вузол має  $k$  зв'язків.

Для комплексних мереж функція  $P(k)$  відрізняється від розподілу Пуассона для випадкових графів. Для переважної більшості комплексних мереж спостерігається степенева залежність  $P(k) \propto k^{-\gamma}$ .

У попередній роботі ми вже ввели деякі з мір складності. Зараз же зробимо це систематично і покажемо, яким чином у межах єдиного алгоритму розрахувати і проаналізувати основні **спектральні** і **топологічні** властивості найпростіших графів.

Для аналізу мережі досліджують характеристики окремих вузлів (локальні), характеристики мережі в цілому (глобальні), та характеристики мережних підструктур. Числові показники деяких глобальних характеристик мережі можуть бути представлені у вигляді аналітичних узагальнень її локальних характеристик (наприклад – найменше, найбільше, середнє значення локального показника, взяте по всім вузлам). Окрім того, що глобальна характеристика може бути представлена у формі одного числа, це також може бути представлення у вигляді розподілу значень локальної характеристики вузлів по усій мережі.

У теорії графів і комбінаториці є багато теорем, при доказі яких застосовуються спектри графів, хоча вони і не зустрічаються у формулюванні теорем. Отже, використання спектрів грає роль досить важливого методу, який називається спектральним.

Спектром графу  $G$  називається множина власних значень матриці, що відповідає даному графу. Відомі декілька підходів встановлення зв'язку між графом  $G$  та його спектром. Для випадку регулярних графів (якими є графи часових рядів фондових індексів) можна показати, що різні види спектрів еквівалентні, тобто містять однакову кількість інформації про структуру графу  $G$ .

Одним із способів представлення графу у вигляді матриці є матриця суміжності. Матрицею суміжності графу, що складається з  $n$  вершин, називається квадратна матриця  $A$  порядку  $n$ , у якої  $a_{ij} = k$ , якщо вершини  $v_i$  та  $v_j$  суміжні кратності  $k$  і  $a_{ij} = 0$ , якщо вони несуміжні (зауважимо, що для графів часових рядів фондових індексів  $k=1$ ).

Матриця Лапласа (Laplacian matrix)  $K$  – також є одним видів подання графу. Вона може бути використана для розрахунку кількості остовних дерев для графу. Для знаходження матриці Кірхгофа використовують формулу:  $K=D-A$ , де  $D$  – діагональна матриця:

$$d_{ij} = \begin{cases} d_i, & i = j, \\ 0, & i \neq j, \end{cases}$$

де  $d_i$  – ступінь відповідної вершини графу. Отже,

$$k_{ij} = \begin{cases} d_i, & i = j, \\ -1, & (v_i, v_j) \in X, \\ 0, & (v_i, v_j) \notin X. \end{cases}$$

**Algebraic connectivity** (алгебраїчна зв'язність графу) – друге найменше власне значення матриці Лапласа. Це власне значення більше 0, тоді і тільки тоді, коли граф зв'язний. Величина цього значення відображає, наскільки добре граф пов'язує ці компоненти, і була використана при аналізі надійності та синхронізації мереж.

**Graph energy.** Енергія графу – це сума абсолютних значень власних значень матриці суміжності графу. Нехай  $G$  є граф з  $n$  вершинами. Передбачається, що  $G$  – простий, тобто він не містить петлі чи паралельних ребер. Нехай  $A$  – матриця суміжності графу  $G$  і  $\lambda_i$ ,  $i = 1, \dots, n$  – власні значення матриці  $A$ . Тоді енергія графу визначається як:

$$E(G) = \sum_{i=1}^n |\lambda_i|.$$

**Спектральний розрив** (spectral gap) – різниця між найбільшим і другим за величиною власного значення, надає інформацію про те, як швидко досягається синхронний стан.

**Спектральний радіус** є найбільшим за модулем власним значенням:

$$r(A) = \max_{\lambda \in \text{Spec}(A)} |\lambda|,$$

де  $\text{Spec}(A)$  – спектр власних значень матриці суміжності.

**Спектральний момент.** Для визначення  $k$ -ого спектрального моменту використовують матрицю суміжності.

$$m_k(A) = \frac{1}{n} \sum_{i=1}^n \lambda_i^k$$

де  $\lambda_i$  – власні значення матриці суміжності,  $n$  – вершини графу  $G$ .

Основні з **топологічних мір** складності описані нижче.

**Vertex eccentricity** (ексцентриситет вершини) – найбільша відстань між  $v$  і будь-якою іншою вершиною, тобто наскільки вершина віддалена від інших вершин графу.

У теорії графів мережному аналізі, центральність (*centrality*) вершини вимірює її відносну важливість у графі. Наприклад, де знаходиться впливова людина в соціальній мережі, як добре використовуються дороги в міській мережі. Існують чотири основні міри центральності: ступінь (*degree*), (*betweenness*), близькість (*closeness*) і власний вектор (*eigenvector*).

**Degree.** Історично першим і концептуально простим є ступінь центральності, який визначається як число зв'язків, що має вузол. Ступінь центральності вершини  $V$ , для даного графу  $G: = (V, E)$  з  $|V|$  вершин і  $|E|$  ребер, визначається як:

$$C_D(v) = \text{deg}(v).$$

**Betweenness** – є центральною мірою вершини в графі. Ця міра кількісно демонструє, скільки разів вузол виступає мостом найкоротшого шляху між двома іншими вузлами. Дана міра була введена як міра кількісного контролю міжособистісного спілкування в соціальній мережі і показує відносну кількість найкоротших шляхів, що проходять через даний вузол. Визначається, як сума по всім парам решти вузлів:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

де  $\sigma_{st}$  – загальна кількість найкоротших шляхів від вершини  $s$  до вершини  $t$ ;  $\sigma_{st}(v)$  – кількість шляхів, які проходять через  $v$ .

**Closeness centrality.** У зв'язкових графах існує природний показник відстані між усіма парами вузлів, який визначається довжиною їх найкоротших шляхів. Віддаленість (*farness*) вузла визначається як сума його відстаней до всіх інших вузлів, а його близькість (*Closeness*) визначається як обернена віддаленості. Таким чином, центральність вузла є нижчою його загальної відстані до всіх інших вузлів. Близькість може розглядатися як міра того, наскільки довго він буде приймати для розповсюдження інформації від секунд до всіх інших вузлів послідовно.



**Eigencentrality** – є мірою впливу вузла в мережі. Визначається спектральними властивостями. Для матриці суміжності  $A = (a_{v,t})$   $a_{v,t} = 1$ , якщо вершина  $v$  пов'язана з вершиною  $t$  і  $a_{v,t} = 0$ , тоді центральність власного вектора обчислюється за формулою:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t,$$

де  $M(v)$  – набір сусідів,  $\lambda = const$ . Завдяки перетворюванню дану формулу можна записати у векторній формі як рівняння на власні значення:  $AX = \lambda X$ .

Міра відстані  $d_{ij}$  визначається як довжина найкоротших шляхів між  $i$  та  $j$ .  $d_{max}$  – максимальна відстань, також відома як **діаметр**. Середня відстань (average distance) за всіма парами позначається як  $\bar{d}$  і дорівнює  $\frac{2}{n(n-1)}$  сум довжин найкоротших шляхів (індекс Вінера):

$$\bar{d} = \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j=i+1}^n d_{ij}.$$

Чим коротше шлях, тим менша середня відстань, більш надійним є граф.

**Кластеризація** вузла – це локальна характеристика мережі. Вона характеризує ступінь взаємозв'язку між собою найближчих сусідів даного вузла. Коефіцієнт кластеризації (або кластерності) даного вузла показує ймовірність того, що навмання обрані два її найближчих сусіда самі виявляються найближчими сусідами. Якщо усі найближчі сусіди вузла взаємопов'язані то коефіцієнт дорівнює 1, якщо ж зв'язків між ними немає, то дорівнює 0:

$$clust(i) = \frac{\Delta(i)}{\alpha(i)(\alpha(i)-1)/2}.$$

В чисельнику число  $\Delta(i)$  – реальна кількість трикутників (циклів довжини 3), приєднаних до цього вузла, а в знаменнику  $\alpha(i)(\alpha(i)-1)/2$  – їх максимально можлива кількість.

**Кластеризація мережі** визначається як середнє значення кластеризації по всім вузлам:

$$clust(G) = clust_{avg} = \langle clust(i) \rangle_i.$$

Таким чином, кластеризація мережі характеризує статистику циклів, а саме, трикутників в мережі. Це глобальна характеристика. Зазвичай реальні мережі характеризуються високою кластеризацією.

Це найважливіші локальні характеристики мережі, зрозуміло, існують і інші.

В рамках наших досліджень динамічної складності деякі з вищезазначених *глобальних топологічних характеристик* мереж, побудованих за такою методологією, розглядаються нами як міри динамічної складності системи.

Сукупність значень будь-якої *локальної характеристики*, впорядкована так само, як вершини мережі, що відображають стани системи у часі, представляє собою дуже цікаву структуру, що потребує окремих досліджень. Зокрема, нами проводилися розрахунки ентропійних і мультифрактальних показників для таких сукупностей ступенів вузлів, і можемо відзначити, що динаміка подібних показників, розрахованих для ковзних вікон, теж має певні специфічні закономірності, що дозволяє інтерпретувати їх в якості мір динамічної складності ринку.

## Хід роботи

Для аналізу спектральних і топологічних характеристик використовується програма **GraphsInWindow\_cmd.m**. Вона застосовна для аналізу так званих канонічних і модельних графів та графів, які ми отримуємо шляхом перетворення часового ряду у граф. В даній роботі розглянемо першу групу з них. Сюди відносяться канонічні графи: відрізок прямої (line), коло (circle) і зірка (star) та модельні: Ердоша-Рені (Erdos-Renyi), малого світу (small world) та переважного приєднання (prefattach)

В програмі, яка попередньо написана для обчислення мір складності за реальними графами, в цикл вставлено команди генерації графів, згідно моделей, а процедури розрахунку залишаються аналогічними. Треба вибрати потрібну модель, розкоментувати чи закоментувати потрібні рядки програми та запустити розрахунки.

Для цього необхідно:

1) створити папку з файлом **GraphsInWindow\_cmd.m** і \*.txt – файлом вихідного часового ряду;

2) задати файл вихідного логу logfile. У цю змінну треба записати ім'я файлу, в який буде записано результат. Цей файл буде створений в процесі роботи, а також будуть створені інші додаткові файли, початок імені у яких буде однаковим;

3) у рядку 6 обрати одну з шести перших моделей;

4) задати змінні wind та tstep (вікно і крок). В принципі крок можна задавати рівним 1, розмір вікна впливає на кількість елементів у модельному графі;

5) у рядку 37 обираємо тип мір для розрахунку: 1 – топологічні, 2 – спектральні. Відповідні функції **graphs\_in\_window\_topologic\_fun** або **graphs\_in\_window\_spectral\_fun** розраховують відповідні топологічні і спектральні властивості.

При цьому у циклі **for i=1:tstep:n-wind %** главный цикл (рядок 70 файлу **graphs\_in\_window\_spectral\_fun**) можна виправити кількість ітерацій, написавши замість *n* константу, яка задасть кількість ітерацій, які ви бажаєте виконати. Наприклад: `for i=1:20;`

6) у циклі треба розкоментувати рядки, які відповідають вибраній моделі;

7) запустити скрипт на виконання.

Наступним кроком роботи є розрахунок топологічних мір. Результати розрахунків також збережемо в окрему папку.

Проведемо тепер порівняльний аналіз спектральних і топологічних мір для канонічних і модельних графів.

Обравши для вхідного ряду будь-який часовий ряд (для канонічних і модельних графів це неважливо), отримаємо для деяких спектральних мір наступні результати (рис. 2, 2').

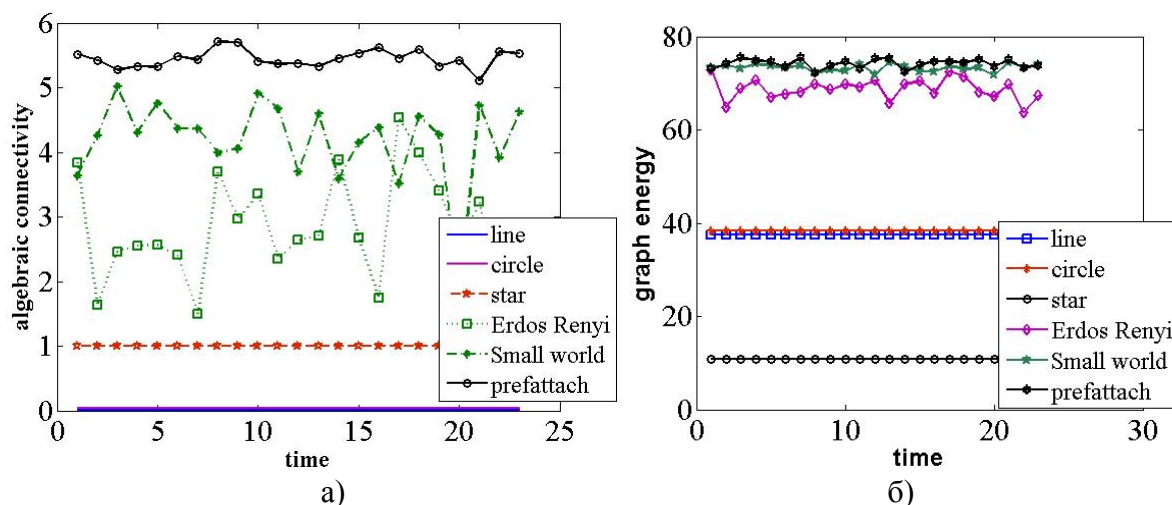


Рис. 2. Спектральні властивості канонічних і модельних графів (а) алгебраїчної зв'язності, (б) енергії графа

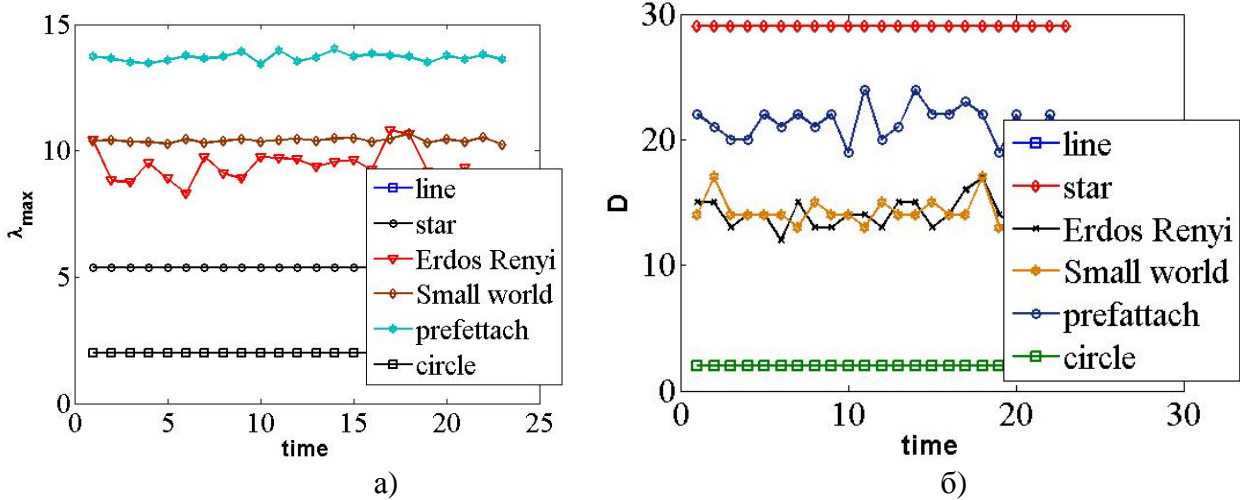


Рис. 2'. Спектральні властивості канонічних і модельних графів (а) максимального власного значення матриці суміжності та (б) максимального ступеня вершини

Логічно впорядкувати за мірами складності наші графи так, як це зроблено за мірою алгебраїчної зв'язності,  $\lambda_{\max}$  та  $D$ . Енергія графа, скоріш за все, не є адекватною мірою складності.

Деякі з топологічних мір представлені на рис. 3.

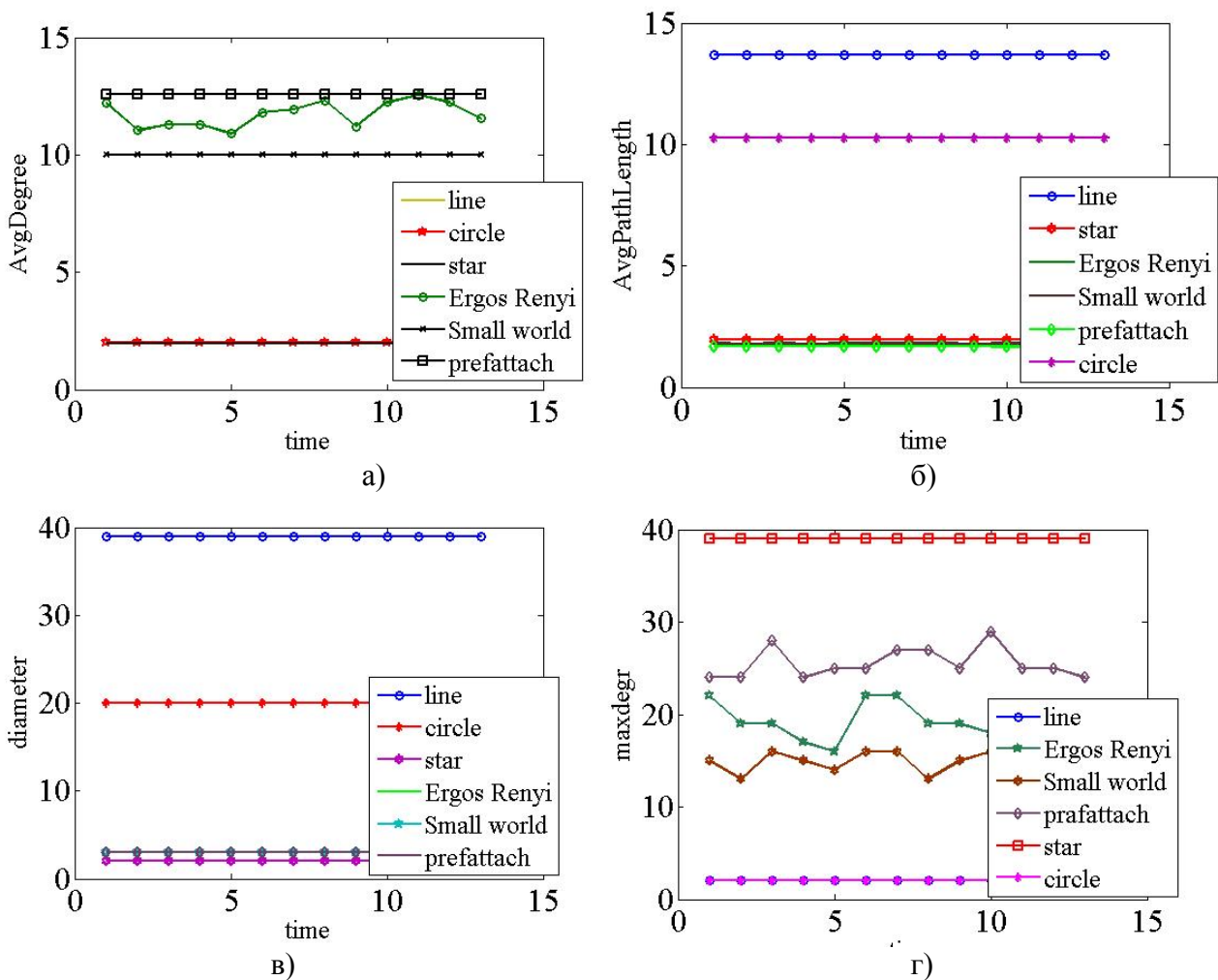


Рис. 3. Топологічні міри складності канонічних і модельних графів (а) середнього значення ступеня вершини, (б) шляху, (в) діаметра та (г) максимального ступеня вершини

Скоріш за все із наведених топологічних мір перевагу слід віддати першим двом і останній, яка співпадає з аналогічною, отриманою із спектральних розрахунків.

### Завдання для самостійної роботи

1. Проаналізуйте аналогічно інші з розрахованих мір складності як спектральних, так і топологічних.
2. Вкажіть і аргументуйте, які з них, на вашу думку, кількісно описують складність досліджуваних мереж?

## ЛАБОРАТОРНА РОБОТА № 14

**Тема.** Візуалізація мережних моделей.

**Мета.** Засвоїти техніку графічного відображення мереж.

1. Встановлення необхідного програмного забезпечення.

### Теоретичні відомості

**Встановлення програмного забезпечення.** Для візуалізації побудованих мереж нам знадобиться програма Gephi (<http://gephi.org>), яка є популярним сучасним інструментом дослідження та візуалізації графів. Для роботи даної програми треба встановити віртуальну машину java (JVM: <http://java.com/ru/download/index.jsp>).

Перед тим, як встановлювати Gephi, необхідно впевнитись у наявності на ПК Java Runtime Environment (середовища виконання програм Java).

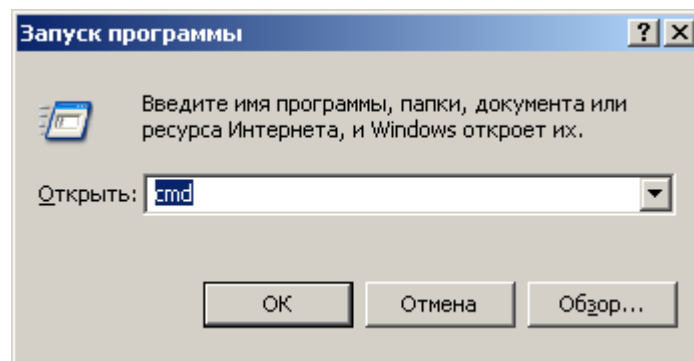


Рис. 1. Запуск консолі операційної системи

Найпростішим способом є використання командного рядка Windows. Для його запуску треба натиснути кнопки Win + R або виконати з системного меню ОС (кнопка «Пуск») пункт «Выполнить». У вікні, що відкриється, необхідно написати «cmd» (приклад подано на рис. 1, однак вікно може відрізнятись) та натиснути Enter, після чого відкриється консоль (або ж – командний рядок) операційної системи (рис. 2). Зрозуміло, що диск та шлях, прописані у вікні, можуть відрізнятись, однак це на даний момент абсолютно неважливо. Для перевірки наявності Java у операційній системі необхідно написати «java –version», після чого проаналізувати відповідь системи. Якщо середовище виконання Java встановлене, то у відповідь виведеться версія Java. Наприклад, на рис. 2 показано інформацію про встановлення віртуального середовища Java версії 1.7 оновлення 75.

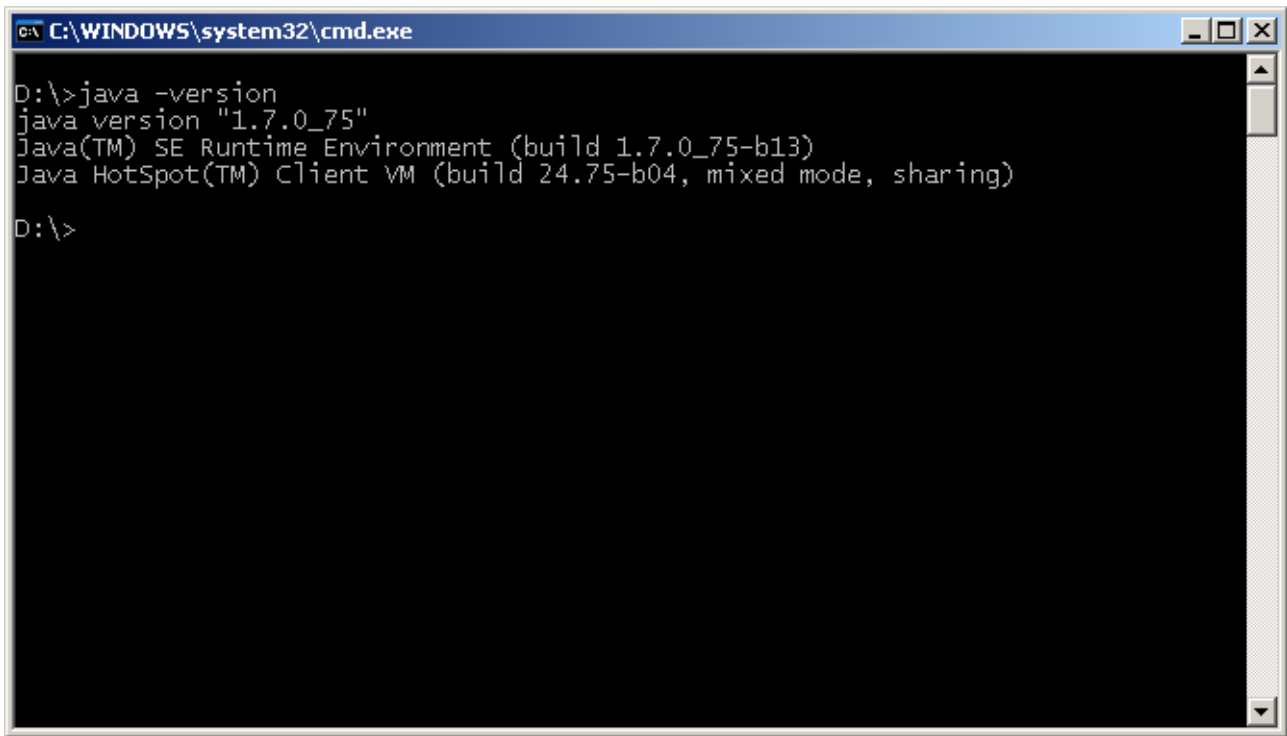


Рис. 2. Виведення версії віртуального середовища Java

У випадку, якщо середовище Java не встановлене, операційна система просто не розпізнає введеної команди і виведе відповідну інформацію про це. Іншим способом перевірки наявності середовища виконання програм Java може бути панель керування. Якщо зайти до панелі, то при встановленій Java повинен бути наявним відповідний елемент панелі керування (рис. 3).

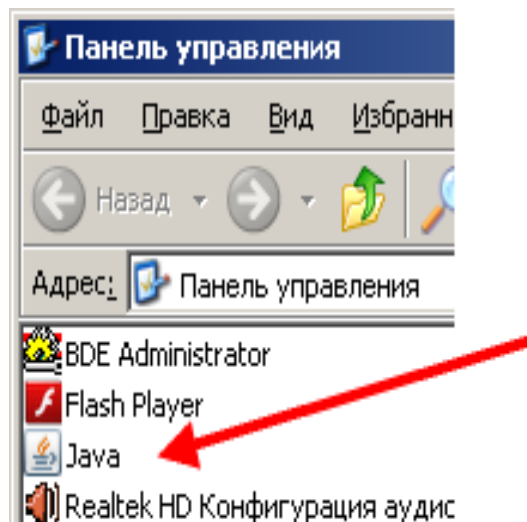


Рис. 3. Наявність елемента Java у панелі керування

Після активації елемента Java відкриється вікно, де на вкладці «General» необхідно натиснути кнопку «About», що дозволить побачити інформацію (рис. 4) про версію встановлених елементів Java (не лише середовища виконання програм, а й інших компонентів).



Рис. 4. Інформація про встановлене ПЗ Java, отримана за допомогою Java Control Panel

Зрозуміло, що якщо Java не встановлена, то відповідний елемент панелі керування буде відсутнім.

Для встановлення Java необхідно завантажити та запустити відповідний файл установки. Для виконання далі описаного матеріалу досить встановити лише Java Runtime Environment (JRE), тобто, достатньо буде встановити пакет, у назві якого наявна аббревіатура JRE. Приклад буде показано для установки JRE 1.7 update 75. Варто зауважити, що існують версії для 32- та 64-розрядних операційних систем Windows, тому при виборі файлу установки необхідно слідкувати за тим, яка версія ОС використовується. Щоправда, розробники зауважують про те, що якщо на комп'ютері у 64-розрядній версії Windows використовуються як 32- так і 64-розрядний браузер, то необхідно встановлювати обидві версії JRE (починаючи з 32-розрядної). Однак, для наших потреб досить і однієї версії.

Отже, у файлі:

- «jre-7u75-windows-i586.exe» 32-розрядна версія JRE 7;
- «jre-7u75-windows-x64.exe» 64-розрядна версія JRE 7;
- «jre-8u31-windows-i586.exe» 32-розрядна версія JRE 8;
- «jre-8u31-windows-x64.exe» 32-розрядна версія JRE 7.

Після вибору та запуску відповідної версії відкриється вікно привітання (рис. 5).



Рис. 5. Вікно привітання при установці Java JRE

Програма установки надає можливість змінити папку установки. Найкраще ставити JRE у папку по замовчуванню («C:\Program Files\Java\jre» або «C:\Program Files x32\Java\jre»), тому просто натискаємо «Install».

У наступному вікні, що відкриється, буде просто відображено рядок прогресу установки, після чого з'явиться останнє вікно з повідомленням про успішну установку JRE (рис 6). Після закриття вікна бажано перезавантажити операційну систему.



Рис. 6. Вікно успішної установки Java JRE

Для встановлення Gephi необхідно запустити файл установки «gephi-0.8.2-beta.setup.exe» (остання версія програми на момент написання поточного тексту). Процес установки коментарів не потребує і відображений на рис. 7. При установці ніякі параметри не налаштовувались (це непотрібно), а просто натискувалась кнопка «Next >>».



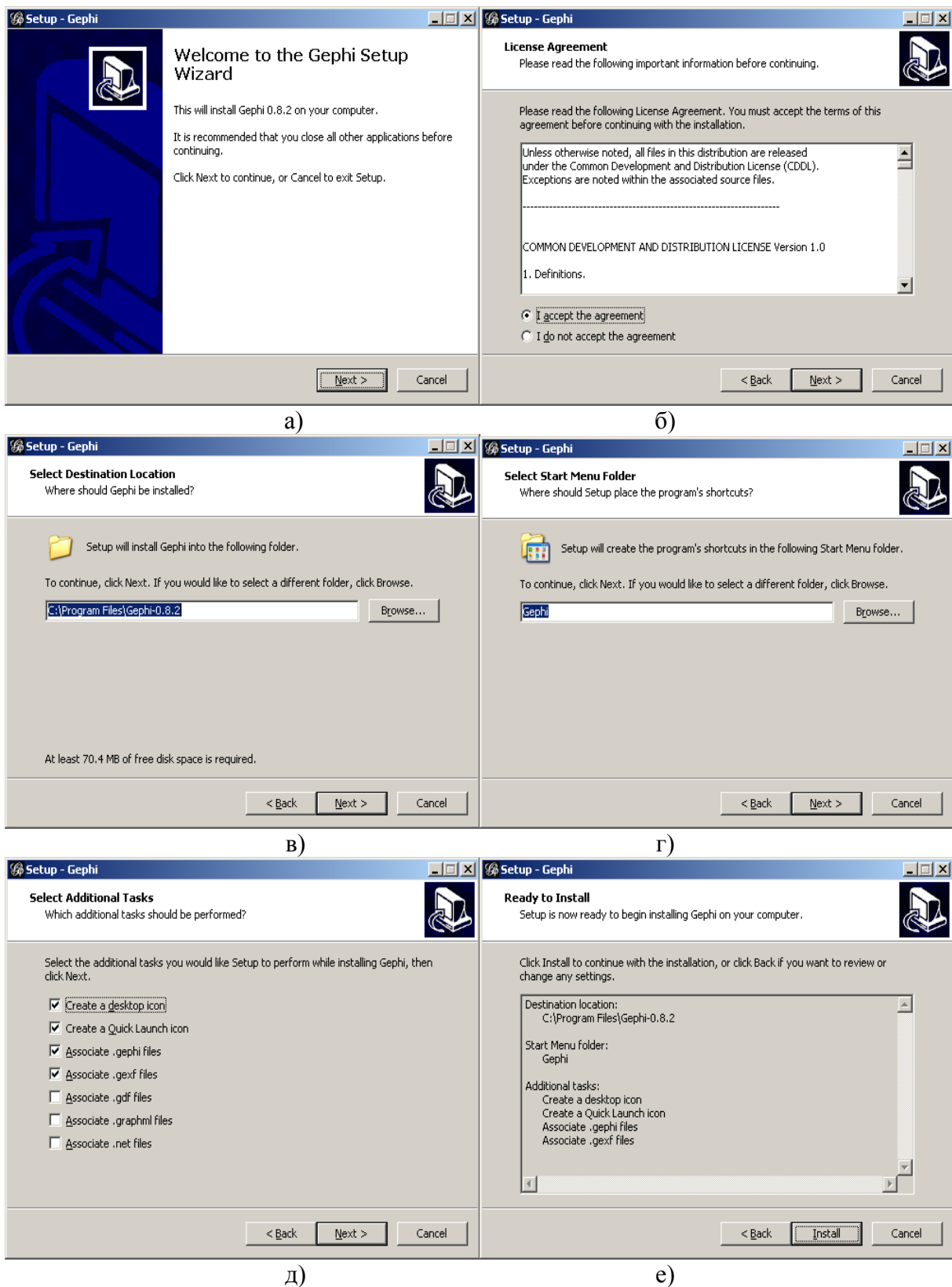


Рис. 7. Послідовність вікон при установці Gephі

На рис. 8 зображено останнє вікно установки Gephі, де по замовчуванню включена опція негайного запуску програми по закритті вікна «Launch Gephі». Для перевірки працездатності програми можна цю опцію залишити включеною, тоді, при умові успішної установки Gephі та наявності на ПК середовища Java, буде запущено програму Gephі.



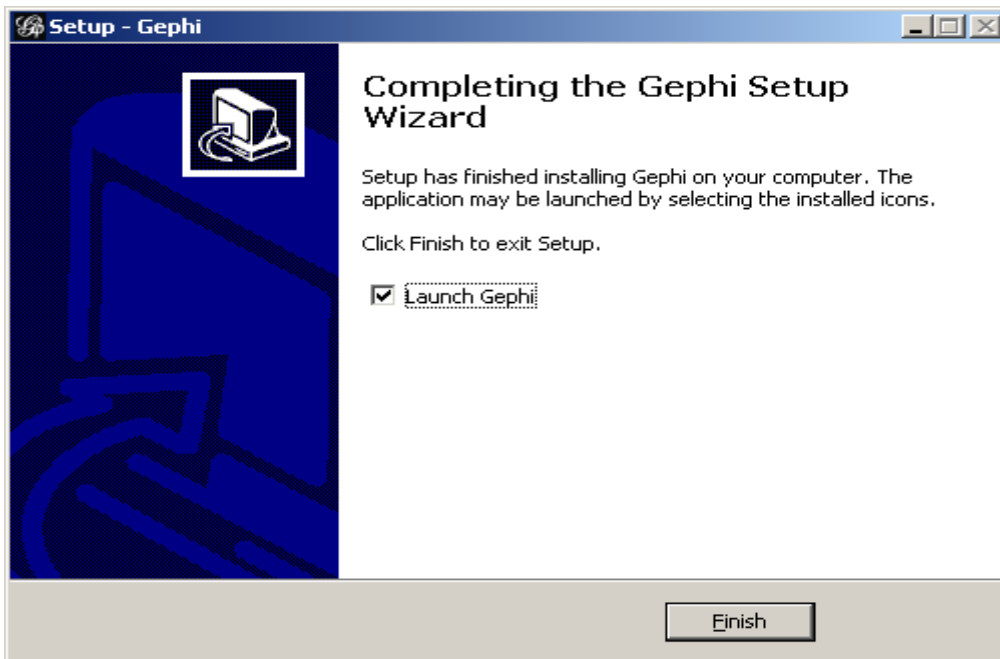


Рис. 8. Останнє вікно установки Gephi

### Хід роботи

В даній роботі досліджуються найпростіші графи та будуються їх зображення.

Наведемо приклад виконання роботи для найпростішого графа «лінія». Зображення даного графу наведено на рисунку 9.

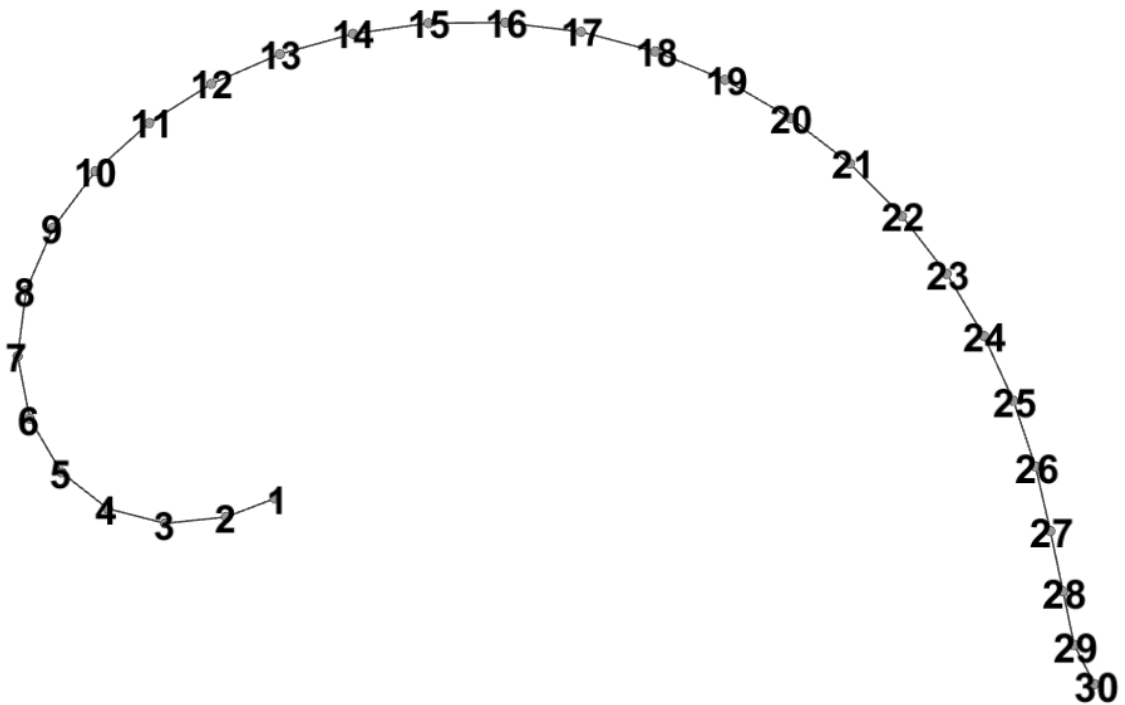


Рис.9. Зображення графу «Лінія»

Для генерації даних для системи Gephi в matlab розроблено скрипт, в якому задаються параметри графів і генерується вихідний файл, який далі завантажується в Gephi. Даний скрипт буде використовуватись нами в подальшому для генерації та дослідження більш складних графів.

Розглянемо параметри для генерації графу. Скрипт має назву **graphs\_for\_gephi.m**. Його необхідно відкрити системою matlab для редагування. Можна його скопіювати в свою робочу папку та перейти в Вашу робочу папку системою Matlab.

У перших рядках *m*-файлу задаються параметри:

1	% Graph from time series (canonical graphs, model graphs, time series: CRP, vg, hvg, RMT
2	matrices)
3	%filename='dax_91.txt';
4	%filename='world.txt';
5	outfile='model_type1.csv';
6	model_type=1; % 1- line, 2 - circle, 3 - star, 4 - Erdos Renyi, 5 - Small world, 6 - prefattach
7	% 7 - crp, 8 - visibility graph, 9 - horizontal vizibility, 10 - RMT matrix
8	
9	% для матричних БД в параметр infile передається txt-файл
10	t_beg=1000;% Номер дня початку фрагмента
11	t_end=1030; % Номер дня кінця фрагмента
12	% вирізання фрагмента «працює» для часових рядів і для матриць
13	%y_fragm=y(t_beg:t_end);
14	
15	n_rand=30; % для моделей! Кількість експериментів

В рядку 2 задаються вхідні текстові бази даних, вони в даній роботі не будуть використовуватись. Можна залишити незмінними.

У рядку 3 задається ім'я вихідного файлу, який буде створений цим скриптом. Цей параметр необхідно змінити, записавши ім'я файла, яке характеризує вибрану модель.

У рядку 6 задається вид моделі *model\_type*. В коментаріях наведені відповідні значення *model\_type* для кожної моделі. У нашому випадку вибираємо модель «лінія» (*line*, *model\_type=1*)

Інші параметри для даної роботи є несуттєвими, їх опис можна прочитати в коментаріях.

У рядку 15 задається кількість вузлів графу *n\_rand*. Для запуску скрипту на виконання натискаємо клавішу F5. При успішній генерації в поточній папці matlab має з'явитись файл, ім'я якого вказано у рядку 3 (*model\_type1.csv*).

В середовищі Gephi відкриваємо згенерований файл, використовуючи пункт меню **Файл**→**Открыть**→вибираємо необхідний файл→**Ок** (рис. 10). Після виконання команд в робочій області з'явиться граф. Робоча область Gephi також містить контекст в якому вказано кількість вузлів, ребер та тип графа:

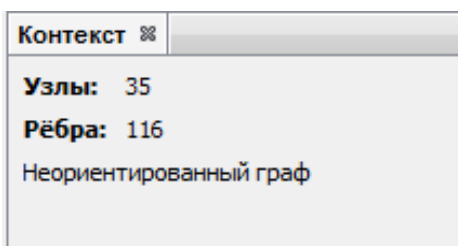
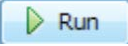


Рис.10. Контекст в Gephi

За допомогою декількох базових методів укладання можна змінити форму графа

(рис. 11). Для цього необхідно на лівій панелі вибрати один із алгоритмів та натиснути  для його запуску. Властивості алгоритму можна залишати за замовчуванням, або задавати власні.

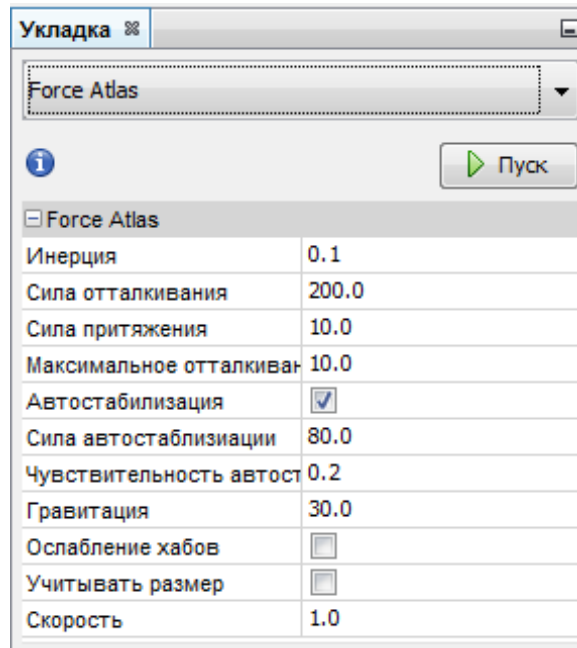


Рис.11. Зміна форми графа в Gephi

Принцип дії алгоритмів «Force-based» простий: пов'язані вузли притягують один одного, а не пов'язані вузлів відштовхують один від одного. На рис. 12 продемонстровано роботу цього алгоритму для обраного нами графа.

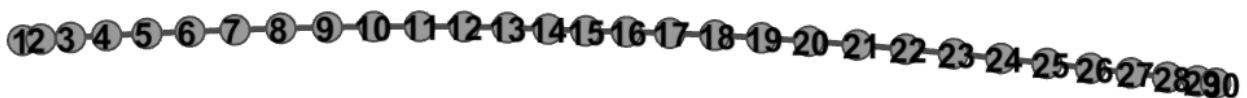











Рис.12. Приклад дії алгоритму Force Atlas для графу «лінія»

Робоча зона Gephi також має кілька інструментів для редагування зовнішнього виду графу, а саме:

-  - теплова карта, використовується для розфарбовування вузлів по градієнту в міру віддалення від обраного вузла (з урахуванням ваги ребер як відстані);
-  - найкоротший шлях, використовується для відображення найкоротшого шляху між двома вузлами;
-  - щітка, для розфарбовування вузлів і їх найближчих сусідів;
-  - заливка, використовується для розфарбовування вузлів;
-  - розмір, використовується для зміни розміру;
-  - олівець для малювання вузлів (ребер);
-  - показує імена вузлів (ребер);
-  - показує ребра;
-  - налаштування шрифта.

Техніку робот из цими інструментами рекомендуємо відпрацювати самостійно.

### Завдання для самостійної роботи

1. Для заданої кількості вузлів згенерувати перші 3 графи (лінія, коло, зірка), зобразити граф за допомогою gephi.
2. Використовуючи довідковий матеріал з сайту gephi.org, навчіться налаштовувати оптимальні зображення графів.
3. Використайте вбудовані міри складності та порівняйте з розрахованими нами у попередній роботі.

## ЛАБОРАТОРНА РОБОТА № 15

**Тема.** Графодинаміка складних систем.

**Мета.** Навчитися використовувати елементи теорії графів для отримання спектральних і топологічних мір складності систем, що характеризуються часовими рядами.

1. Методи перетворення часових рядів у графи:
  - рекурентні;
  - графи видимості;
  - кореляційні.
2. Розрахунок спектральних і топологічних властивостей змодельованих мереж.
3. Аналіз віконної динаміки графів (графодинаміка).
4. Мультиплексні графи і міри їх складності.

### Теоретичні відомості

У попередній роботі ми ввели поняття мір складності для найпростіших графічних об'єктів, поширених мережних моделей, порівняли деякі з мір складності. В даній роботі ми продемонструємо сучасні методи перетворення часових рядів, що характеризують складні системи, у мережу з подальшим дослідженням відповідних спектральних і топологічних мір складності. Крім цього, покажемо яким чином від простої мережі можна перейти до системи мереж – мультимережі і ввести для неї відповідні міри складності.

Ми також покажемо, що вказані міри можна співставляти з динамікою вихідного часового ряду і якщо вони є інформативними щодо можливих змін власне ряду, то їх можливо використовувати для побудови індикаторів характерної динаміки складних систем.

### Методи перетворення часових рядів у графи

Найбільш вживаними методами перетворення часових послідовностей у відповідні мережі є рекурентні, графи видимості та кореляційні.

Технологія рекурентних діаграм для візуалізації рекурентностей у фазовому просторі заснована на ідеї Анрі Пуанкаре щодо рекурентності фазового простору динамічних систем. Згідно з теоремою Такенса, еквівалентна фазова траєкторія, що зберігає структуру оригінальної фазової траєкторії, може бути відновлена з одного спостереження або часового ряду методом часових затримок:  $\hat{x}(t) = (u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau})$ , де  $m$  – розмірність вкладення,  $\tau$  – часова затримка (реальна часова затримка визначається як  $\tau \cdot \Delta t$ ). Рекурентна ж діаграма відображає наявні повторюваності у формі бінарної матриці  $R$ , де  $R_{i,j} = 1$ , якщо  $\bar{x}_j$  є сусіднім до стану  $\bar{x}_i$ , і  $R_{i,j} = 0$  у протилежному випадку. Сусідніми (або рекурентними) є стани  $\bar{x}_j$ , які потрапляють в  $m$ -вимірний окіл з радіусом  $\varepsilon$  і центром в  $\bar{x}_i$ . Зрозуміло, що параметри  $m$ ,  $\tau$  та  $\varepsilon$  є ключовими при проведенні рекурентного аналізу. Рекурентна

діаграма легко трансформується у матрицю суміжності, за якою розраховуються спектральні і топологічні характеристики графа.

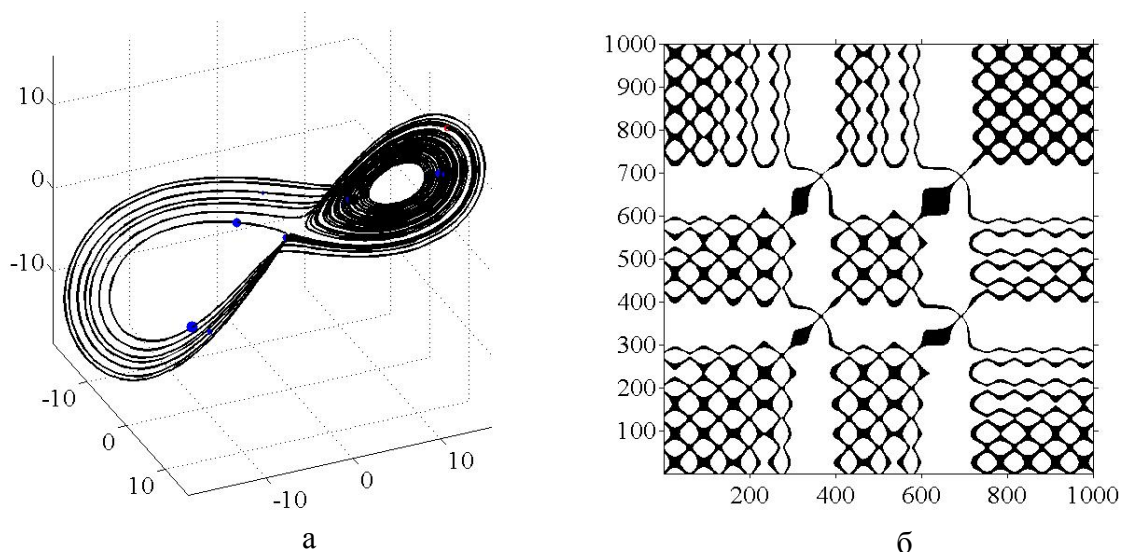


Рис. 1. Атрактор Лоренца (а) та його рекурентна діаграма (б).

Для фрагменту часового ряду фондового індексу довжиною у 2000 точок фазовий портрет і відповідна рекурентна діаграма представлені на рис. 2.

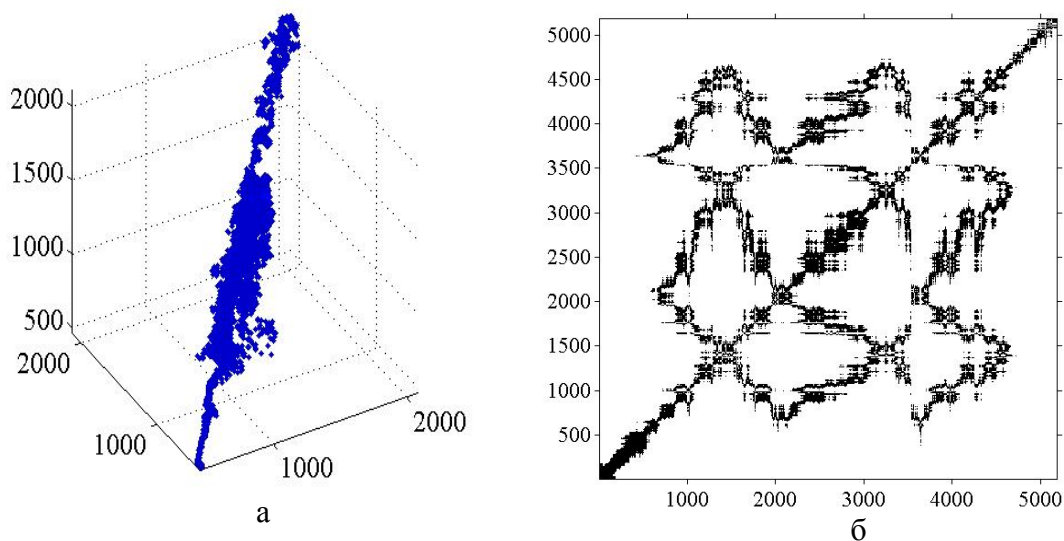


Рис. 2. Фазовий портрет фрагмента індексу S&P 500 (а) та його рекурентна діаграма (б).

Алгоритм графа видимості реалізується наступним чином. Візьмемо часовий ряд  $Y(t) = [y_1, y_2, \dots, y_n]$  довжини  $N$ . Кожну точку даних часового ряду можна розглядати як вершину в асоційованій мережі, а ребро буде з'єднувати дві вершини, якщо дві відповідні точки даних можуть «бачити» один одного з відповідної точки часового ряду (рис. 3). Формально два значення ряду  $y_a$  (на момент часу  $t_a$ ) і  $y_b$  (на момент часу  $t_b$ ) пов'язані, якщо для будь-якого іншого значення  $(y_c, t_c)$ , яке розміщене між ними (тобто,  $t_a < t_c < t_b$ ), задовольняється умова:  $y_c < y_a + (y_b - y_a) \frac{t_c - t_a}{t_b - t_a}$ .

Певна модифікація алгоритму (рис. 3б) називається горизонтальним графом видимості.

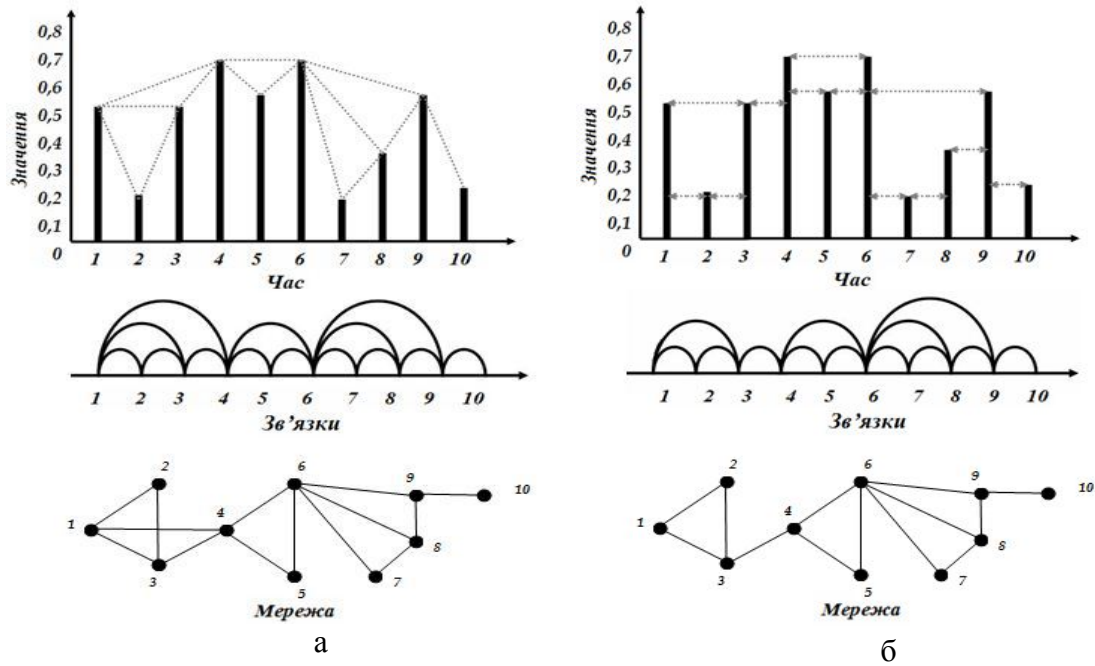


Рис. 3. Перетворення часового ряду у граф за алгоритмом графу видимості (а) та горизонтального графу видимості (б)

Для побудови кореляційних графів спочатку обчислюється логарифмічні прибутковості акції  $i=1, \dots, N$  за час  $\Delta t$ ,  $G_i(t) = \ln S_i(t + \Delta t) - \ln S_i(t)$ , де  $S_i(t)$  позначає ціну акції  $i$ . Оскільки різні акції мають різні рівні змінюваності (стандартні відхилення), визначимо нормалізовані прибутковості  $g_i(t) \equiv \frac{G_i(t) - \langle G_i \rangle}{\sigma_i}$ , де  $\sigma_i \equiv \sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}$  – стандартне відхилення  $G_i$ , а  $\langle \dots \rangle$  позначає середнє значення за період часу, що досліджується. Тоді обчислення матриці взаємних кореляцій  $C$  зводиться до формули:  $C_{ij} \equiv \langle g_i(t) g_j(t) \rangle$ .

На рис. 4а представлена матриця взаємних кореляцій між акціями компаній на фондовому ринку США за даними індексу S&P 500.

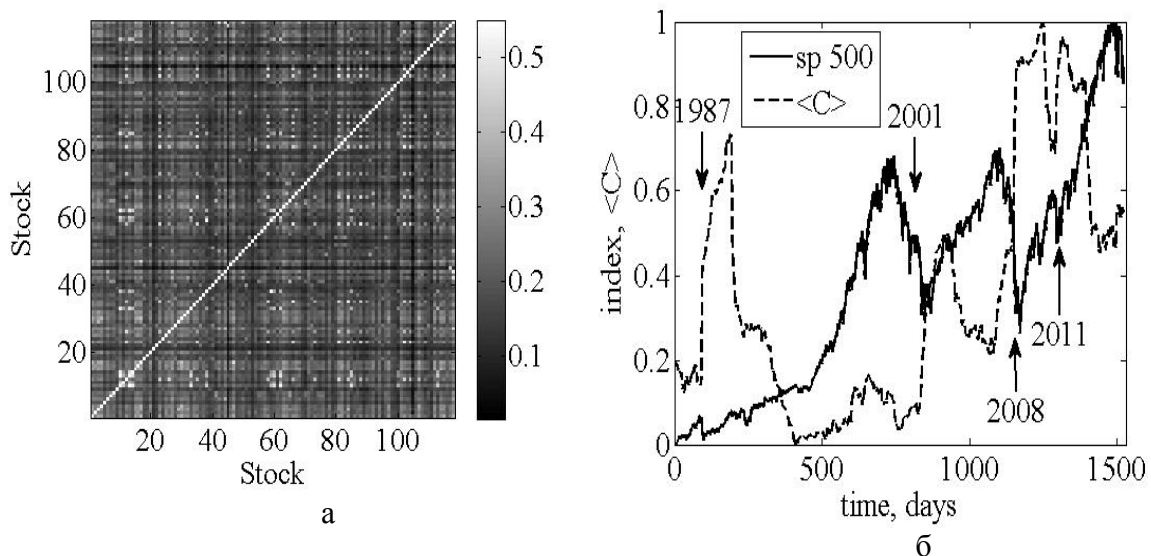


Рис. 4. (а) Матриця взаємних кореляцій (б) для індексу S&P 500 та (б) порівняльна динаміка самого значення індексу з середнім значенням коефіцієнта кореляції  $\langle C \rangle$



Видно, що активи на фондовому ринку помітно корельовані (рис. 4а). При цьому середні значення коефіцієнта кореляції сильно зростають у періоди відомих кризових явищ (вказані стрілками на рис. 4б).

Для побудови і аналізу властивостей кореляційного графа слід сформувати з кореляційної матриці матрицю суміжності. Для цього треба ввести величину, яка для поля кореляцій буде слугувати відстанню між корельованими агентами. Такою відстанню може слугувати залежна від коефіцієнта взаємної кореляції  $C_{ij}$  величина  $x(i, j) = \sqrt{2(1 - C_{ij})}$ . Так, якщо коефіцієнт кореляції між двома активами помітний, відстань між ними є малою, і, починаючи з деякої критичної величини  $x_{cr}$ , активи можна вважати зв'язаними на графі. Для матриці суміжності це означає, що вони є суміжними на графі. В протилежному випадку активи не є суміжними. При цьому обов'язковою умовою є зв'язність графа.

Мережі рідко бувають ізольованими. Тому необхідно враховувати міжмережну взаємодію, що можна реалізувати у рамках різних моделей. У даній роботі ми проведемо дослідження так званих мультиплексних мереж, особливості яких зводяться до фіксованої кількості вузлів у кожному шарі, але зв'язані вони між собою різними зв'язками.

Для мультиплексних мереж алгоритм графа видимості на прикладі двох шарів представлений на рис. 5а, а його візуалізація для трьох шарів на рис. 5б.

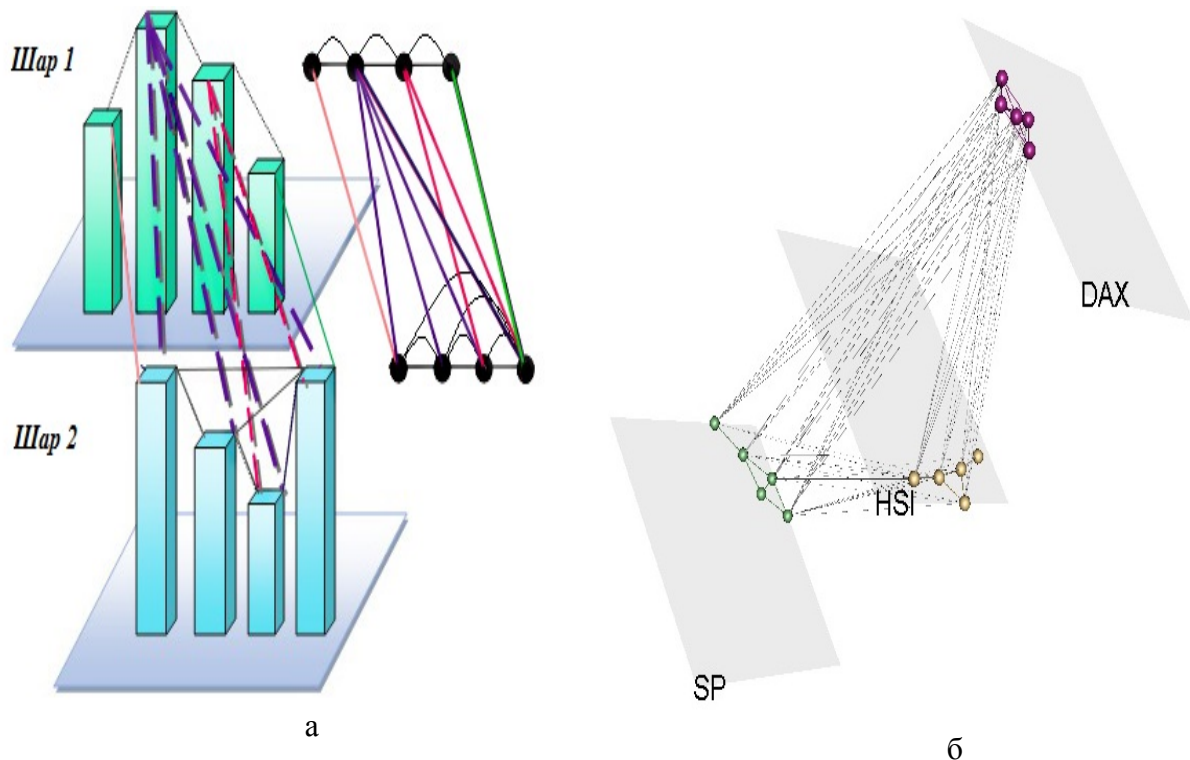


Рис. 5. (а) Модифікований алгоритм графів видимості та (б) візуалізація фрагменту з п'яти точок часових рядів для індексів фондових ринків США (SP), Китаю (HSI) і Німеччини (DAX)

Алгоритм і візуалізація рекурентного алгоритму відповідно зображена на рис. 6.

Мультишарова мережа (multilayer network) є пара  $M = (G, C)$ , де  $G = \{G_\alpha; \alpha \in \{1, \dots, M\}\}$  є сімейством графів (спрямованих чи ні, зважених або ні)  $G_\alpha = (X_\alpha, E_\alpha)$ , що називаються шарами і

$$C = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta; \alpha, \beta \in \{1, \dots, M\}, \alpha \neq \beta\}$$

є набором зв'язків між вузлами різних шарів  $G_\alpha$  та  $G_\beta$  при  $\alpha \neq \beta$ . Елементи  $C$  називаються

схрещеними (crossed), а елементи кожного  $E_\alpha$  - внутрішньошаровими (intralayer) зв'язками в  $M$  на відміну від елементів кожного  $E_{\alpha\beta} (\alpha \neq \beta)$ , що називаються міжшаровими (interlayer) зв'язками.

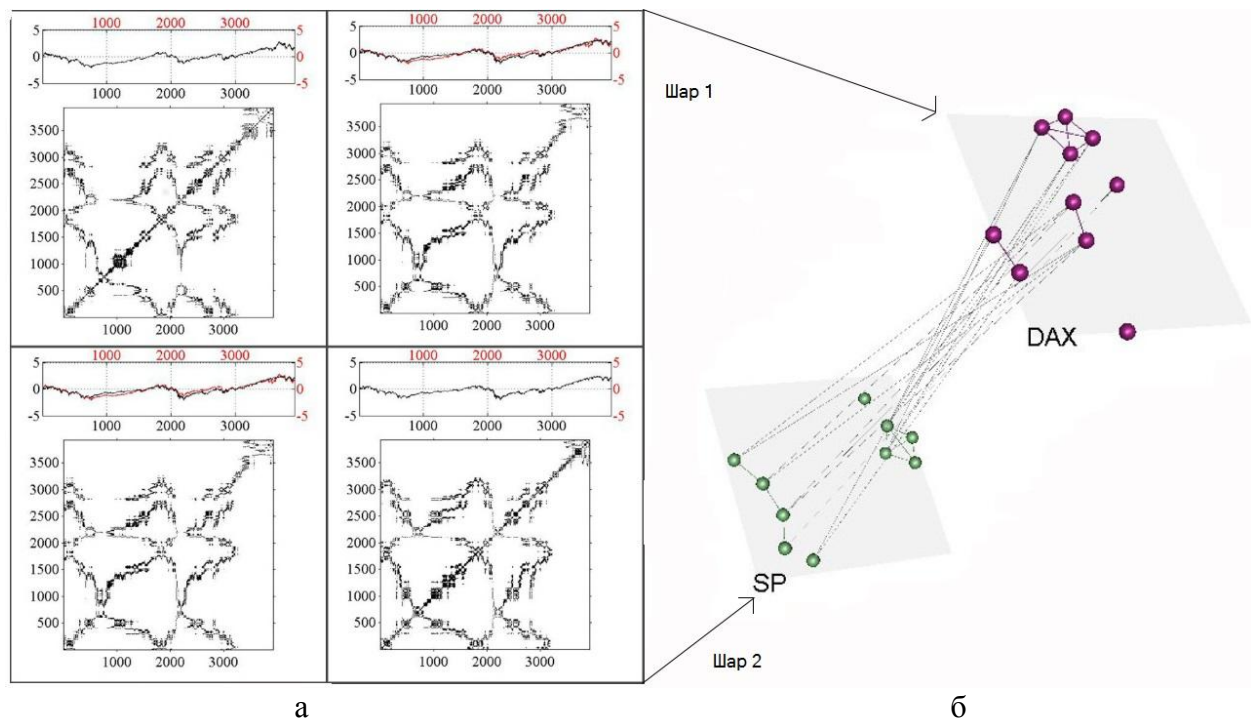


Рис. 6. Часові ряди індексів S&P 500 і DAX рекурентні діаграми (а) та зв'язки рекурентних точок у двох мультиплексних шарах (б)

Набір вузлів шару  $G_\alpha$  позначається  $X_\alpha = \{x_1^\alpha, \dots, x_{N_\alpha}^\alpha\}$ , а матриця суміжності – як  $A^{[\alpha]} = (a_{ij}^\alpha) \in R^{N_\alpha \times N_\alpha}$ , де

$$a_{ij}^\alpha = \begin{cases} 1, & (x_i^\alpha, x_j^\alpha) \in E_\alpha, \\ 0 & \end{cases}$$

для  $1 \leq i, j \leq N_\alpha$  та  $1 \leq \alpha \leq M$ . Для міжшарової матриці суміжності маємо  $A^{[\alpha, \beta]} = (a_{ij}^{\alpha\beta}) \in R^{N_\alpha \times N_\beta}$ , де

$$a_{ij}^{\alpha\beta} = \begin{cases} 1, & (x_i^\alpha, x_j^\beta) \in E_{\alpha\beta}, \\ 0 & \end{cases}$$

Мультиплексна (multiplex) мережа є частковим випадком міжшарової і містить фіксоване число вузлів, з'єднаних різними типами зв'язків.

Мультиплексні мережі характеризуються кореляціями різної природи, що дають можливість ввести додаткові до спектральних і топологічних міри, які назвемо мультиплексними.

### Мультиплексні міри

Оцінимо кількісно перекриття зв'язків між різними шарами. Середнє перекриття



зв'язків (average edge overlap) очевидно дорівнює

$$\omega = \frac{\sum_i \sum_{j>i} \sum_{\alpha} a_{ij}^{[\alpha]}}{M \sum_i \sum_{j>i} (1 - \delta_{0, \sum_{\alpha} a_{ij}^{[\alpha]}})},$$

і визначає число шарів, в яких наявний даний зв'язок. Його значення лежить на проміжку  $[1/M, 1]$  і дорівнює  $1/M$ , якщо зв'язок  $(i, j)$  існує тільки в одному шарі, тобто, якщо існує шар  $\alpha$  такий, що  $a_{ij}^{[\alpha]} = 1, a_{ij}^{[\beta]} = 0 \forall \beta \neq \alpha$ . Якщо всі шари ідентичні, то  $\omega = 1$ . Отже, ця міра може слугувати мірою когерентності вихідних часових рядів: високі значення  $\omega$  вказують на помітну кореляцію у структурі часових рядів.

Загальне (повне) перекриття (total overlap)  $O^{\alpha\beta}$  між двома шарами  $\alpha$  і  $\beta$  визначається як повне число зв'язків які є спільними між шарами  $\alpha$  і  $\beta$ :

$$O^{\alpha\beta} = \sum a_{ij}^{\alpha} a_{ij}^{\beta},$$

де  $\alpha \neq \beta$ .

Однією з ключових мір центральності (важливості) будь-якої мережі є ступінь вершини  $k$ , яка дорівнює кількості зв'язків вибраної вершини. Для мультиплексної мережі це вже вектор

$$k_i = (k_i^{[1]}, \dots, k_i^{[M]}),$$

де  $k_i^{[\alpha]}$  - ступінь вузла  $i$  в шарі  $\alpha$ , тобто,  $k_i^{[\alpha]} = \sum_j a_{ij}^{[\alpha]}$ , тоді як  $a_{ij}^{[\alpha]}$  - елементи матриці суміжності для шару  $\alpha$ . Специфіка векторного характеру ступеня вершини у мультиплексних мережах дозволяє вводити додаткові міжшарові характеристики. Однією з таких характеристик є перекриття ступеня вузла  $i$

$$o_i = \sum_{\alpha=1}^M k_i^{[\alpha]}.$$

Наступна міра кількісно описує міжшарові кореляції між ступенями вибраного вузла в двох різних шарах. Якщо вибрана з  $M$  шарів пара  $(\alpha, \beta)$  характеризується розподілами ступенів  $P(k^{[\alpha]}), P(k^{[\beta]})$ , так звана міжшарова взаємна інформація визначається за формулою:

$$I_{\alpha, \beta} = \sum \sum P(k^{[\alpha]}, k^{[\beta]}) \log \frac{P(k^{[\alpha]}, k^{[\beta]})}{P(k^{[\alpha]})P(k^{[\beta]})},$$

де  $P(k^{[\alpha]}, k^{[\beta]})$  - спільна ймовірність знайти вузол ступеня  $k^{[\alpha]}$  у шарі  $\alpha$  і ступеня  $k^{[\beta]}$  у шарі  $\beta$ . Чим більшими є значення  $I_{\alpha, \beta}$ , тим більш корельованими є розподіли ступенів двох шарів, а, отже, і структура асоційованих з ними часових рядів. Знайдемо також середнє значення  $I_{\alpha, \beta}$  за усіма можливими парами шарів - скалярну величину  $\langle I_{\alpha, \beta} \rangle$ , яка кількісно характеризує інформаційний потік у системі.

Величиною, яка кількісно описує розподіл ступеня вузла  $i$  між різними шарами, є ентропія мультиплексного ступеня:

$$S_i = -\sum_{\alpha=1}^M \frac{k_i^{[\alpha]}}{o_i} \ln \frac{k_i^{[\alpha]}}{o_i}.$$

Ентропія дорівнює нулю, якщо всі зв'язки вузла  $i$  знаходяться в одному шарі, і навпаки, має максимальне значення, коли вони рівномірно розподілені між різними шарами. Тобто, чим вищим є значення  $S_i$ , тим більш рівномірно зв'язки вузла  $i$  розподілені між шарами.

Подібною величиною є мультиплексний коефіцієнт участі  $P_i$  вузла  $i$ :

$$P_i = \frac{M}{M-1} \left[ 1 - \sum_{\alpha=1}^M \left( \frac{k_i^{[\alpha]}}{o_i} \right)^2 \right].$$

$P_i$  приймає значення на інтервалі  $[0,1]$  і визначає, настільки однорідно зв'язки вузла  $i$  розподілені серед  $M$  шарів. Якщо всі зв'язки вузла  $i$  лежать в одному шарі,  $P_i = 0$ , і  $P_i = 1$ , якщо вузол має точно задане число зв'язків у кожному з  $M$  шарів. Отже, чим більшим є коефіцієнт  $P_i$ , тим більш рівномірно розподілена участь вузла у мультиплексі. Коефіцієнт участі  $P$  усієї мультиплексної мережі визначається як середнє значення  $P_i$  за усіма вузлами:

$$P = 1/N \sum_i P_i.$$

Очевидно, величини  $S_i$  і  $P_i$  є дуже подібними.

### Хід роботи

Для побудованих описаними вище методами графів можна аналогічно до того, як це зроблено для канонічних і модельних мереж, розрахувати спектральні і топологічні властивості. У програмі **GraphsInWindow\_cmd.m** у рядку 6 необхідно обрати модель перетворення ряду у граф: 7 – рекурентну (crp), 8 – граф видимості (visibility graph), 9 – горизонтальний граф видимості (horizontal vizibility), 10 – кореляційний (RMT matrix).

Покажемо, що деякі з них слугують мірами складності системи, а динаміка їх змін дозволяє будувати передвісники кризових станів у складних системах.

Розрахунки проводились у такий спосіб. Обирався часовий проміжок (вікно), наприклад, два роки (приблизно 500 торгівельних днів), для нього будувались відповідні графи та розраховувались їх спектральні і топологічні властивості. Далі вікно зміщувалось з кроком, наприклад, один тиждень (5 торгівельних днів) і процедура повторювалась до вичерпання часових рядів.

Результати розрахунків для відновлених із часових рядів графів представлені на рис. 7-9. Для прикладу взято тільки по одній відповідно спектральній і топологічній мірі. Інші поводять себе подібним чином, або ж не являються мірами складності системи.

Зауважимо, що на фондовому ринку за період, охоплений досліджуваними часовими рядами, відбулися наступні відомі кризи: 1987, 2001, 2008, 2011 і 2015pp. На деяких графіках вказані кризи відмічені стрілками. Знаючи час настання кризи та співставляючи часовий ряд з динамікою певного показника, можна досліджувати його залежність від тих чи інших характерних змін на фондовому ринку: докризовий, кризовий та післякризовий періоди.

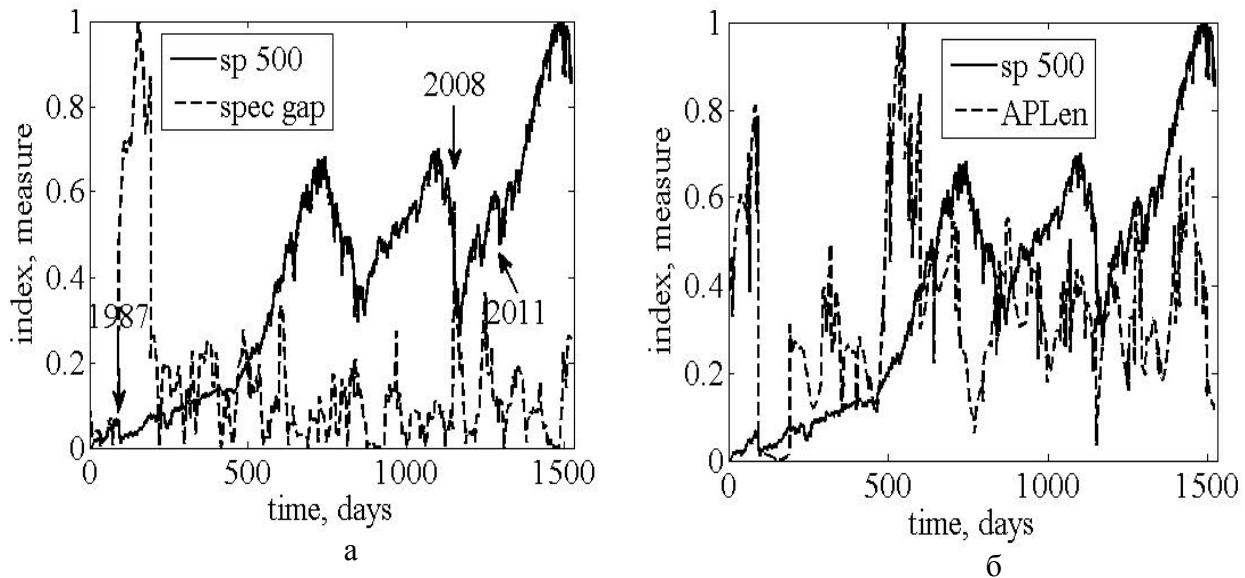


Рис. 7. Динаміка індексу S&P 500 та спектральної (а) і топологічної (б) мір складності. Граф будувався методом рекурентного аналізу

Так, з рис. 7 видно, що рекурентна мережа чутлива до процесів синхронізації, зростання кореляцій на фондовому ринку під час криз. Відповідно, спектральні міри стрімко зростають, а топологічні, навпаки, спадають, сигналізуючи про початок кризи.

Специфіка побудови графу видимості така, що у до кризовий період спектральні міри помітно зменшуються, зростаючи власне у кризовий період (рис. 8).

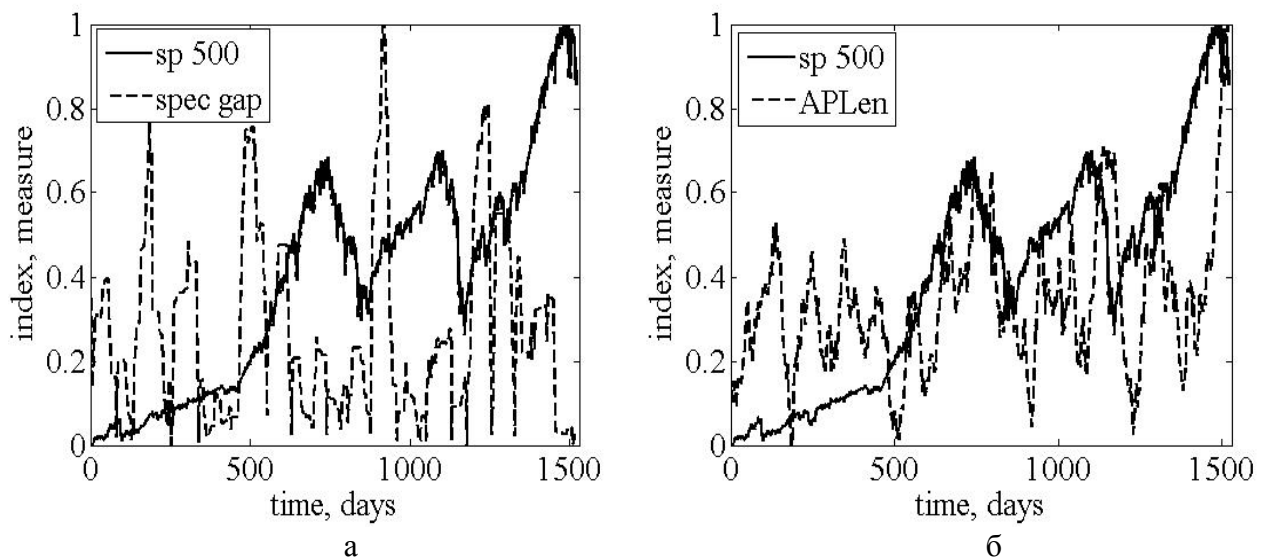


Рис. 8. Динаміка індексу S&P 500 та спектральної (а) і топологічної (б) мір складності. Граф будувався методом графу видимості

Топологічні міри при цьому поведуть себе антисиметрично до спектральних. Нарешті, кореляційні міри проявляють схожу до рекурентних динаміку, оскільки також чутливі до процесів кореляції і самоорганізації, які помітно зростають у періоди кризових станів (рис. 9).

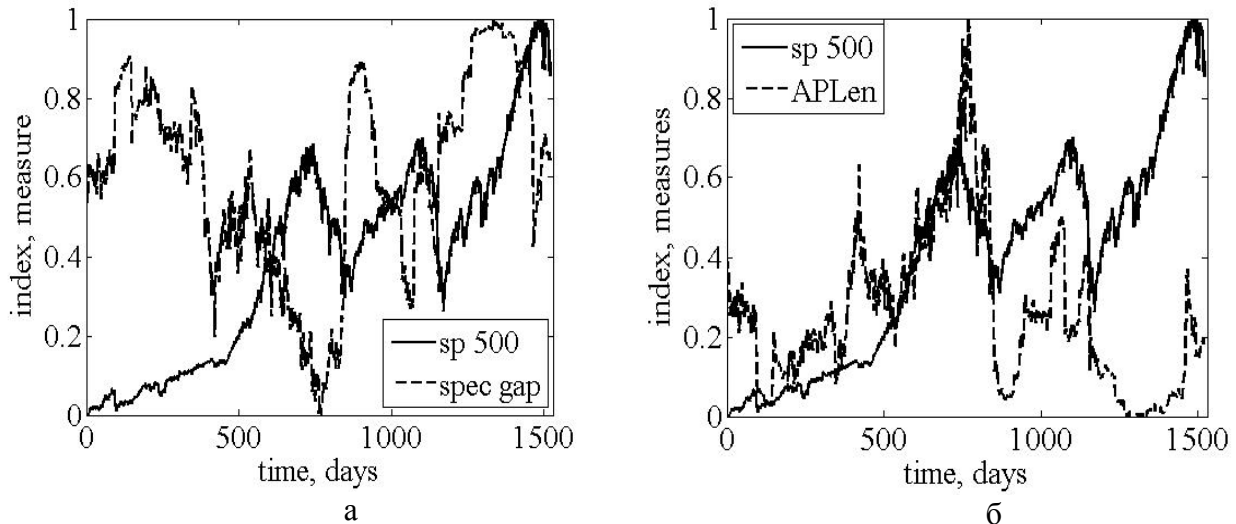


Рис. 9. Динаміка індексу S&P 500 та спектральної (а) і топологічної (б) мір складності. Граф будувався методом кореляційного аналізу

У програмі також передбачено збереження деяких виділених графічних об'єктів, що характеризують динаміку віконного процесу моделювання. Для цього імена filename (рядок 2) і filename\_index (рядок 6) повинні співпадати. Рядок 35 має бути таким: skip\_complex\_measures=1. Зберігаючи їх в окремій папці, ми можемо в подальшому створити відповідний **.avi** – файл, як це було зроблено у випадку візуалізації вейвлет-ентропії.

На рис. 10 представлено скріншот окремого кадру, який зберігається у процесі розрахунку за методом рекурентного мережного аналізу для індексу S&P 500 з вікном 500, кроком 5 днів.

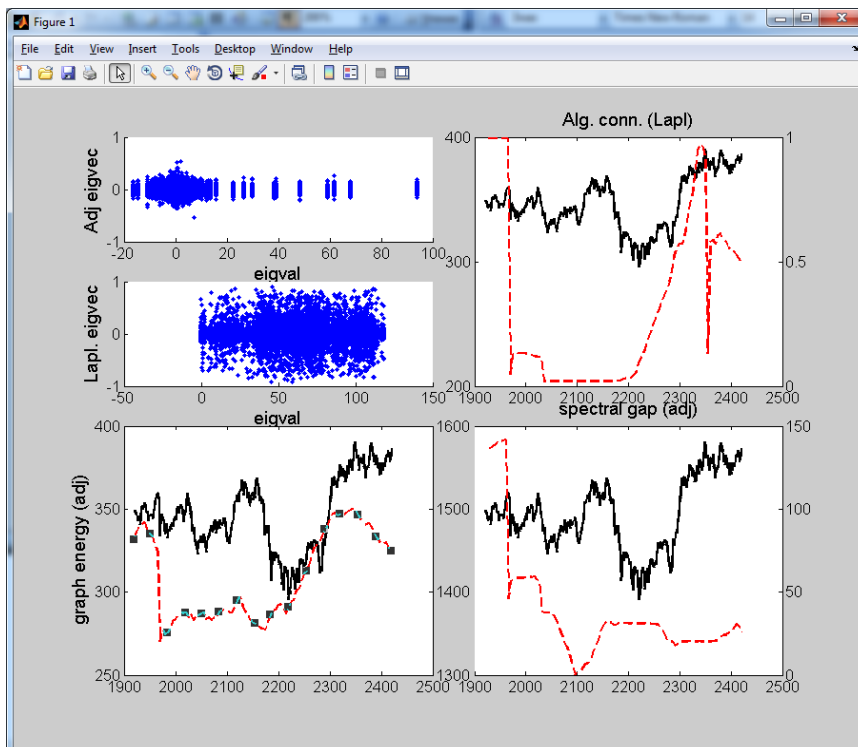


Рис. 10. Фрагмент графодинаміки індексу S&P 500, який візуалізує спектри матриць суміжності (adj) і Лапласа (Lapl) та обраних мір складності: алгебраїчної зв'язності (alg conn), енергії графа (graph energy) та спектрального розриву (spectral gap)

Як показали розрахунки, мультиплексні мережі надзвичайно чутливі до процесів

синхронізації, зростання кореляцій на фондовому ринку під час криз. Відповідно, спектральні міри стрімко зростають, а топологічні, навпаки, спадають, сигналізуючи про початок кризи. Але, на відміну від мономереж, у мультиплексних системах відчутні процеси десинхронізації. Це робить відповідні міри малоінформативними (рис. 11). Крім цього, зростання кількості шарів призводить до помітного зростання складності розрахунків.

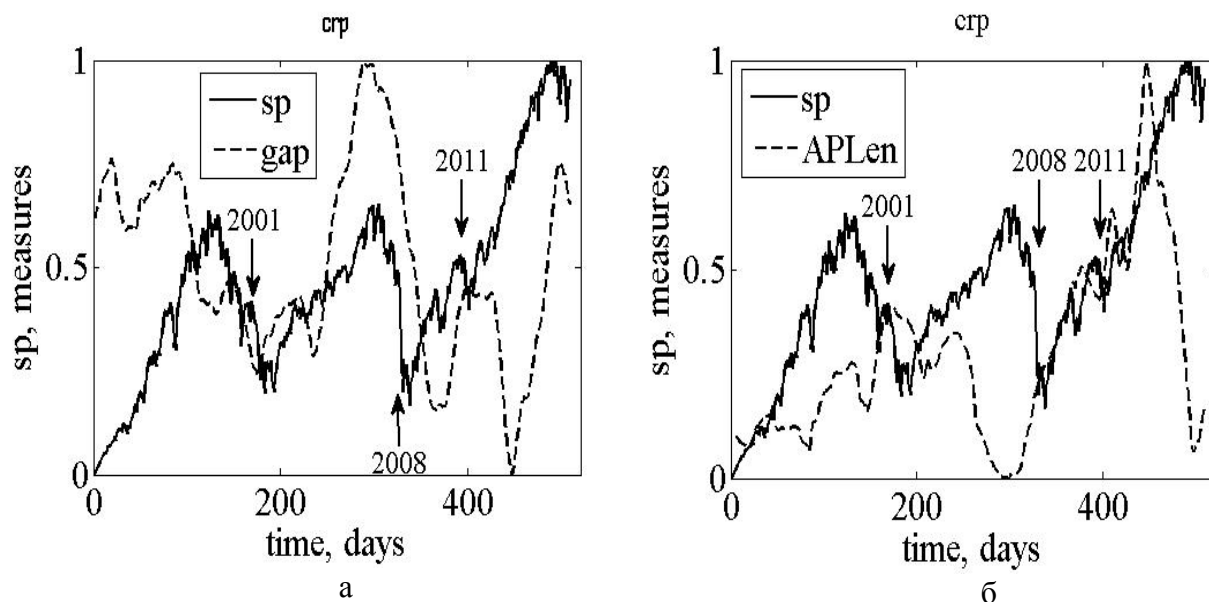


Рис. 11. Динаміка індексу S&P 500 та спектральної (а) і топологічної (б) мультиплексних мір складності. Граф будувався методом рекурентного аналізу

У цьому випадку мультиплексні міри зберігають свою як теоретичну так і практичну привабливість: слугують передвісниками кризових явищ і рахуються досить швидко навіть для помітної кількості шарів (рис. 12, 13).

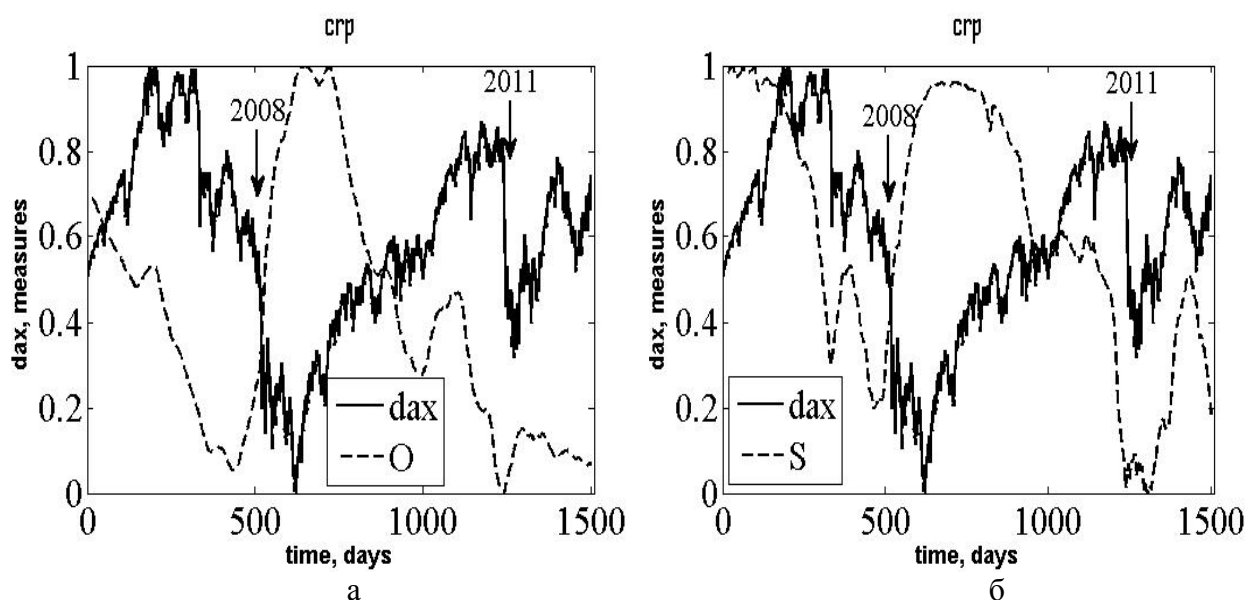


Рис. 12. Динаміка індексу DAX та мультиплексних мір повного перекриття (а) і ентропії (б). Граф будувався методом мультиплексного рекурентного аналізу

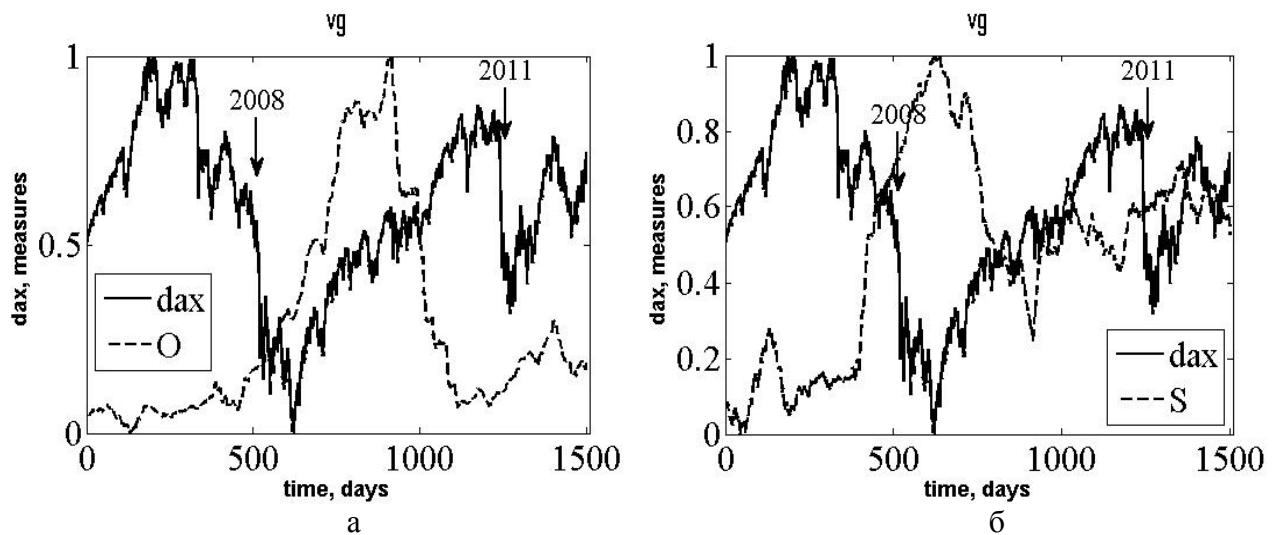


Рис. 13. Динаміка індексу DAX та мультиплексних мір повного перекриття (а) і ентропії (б). Граф будувався методом мультиплексного графу видимості

Таким чином, загальну схему моделювання мережних структур можна зобразити наступною схемою.

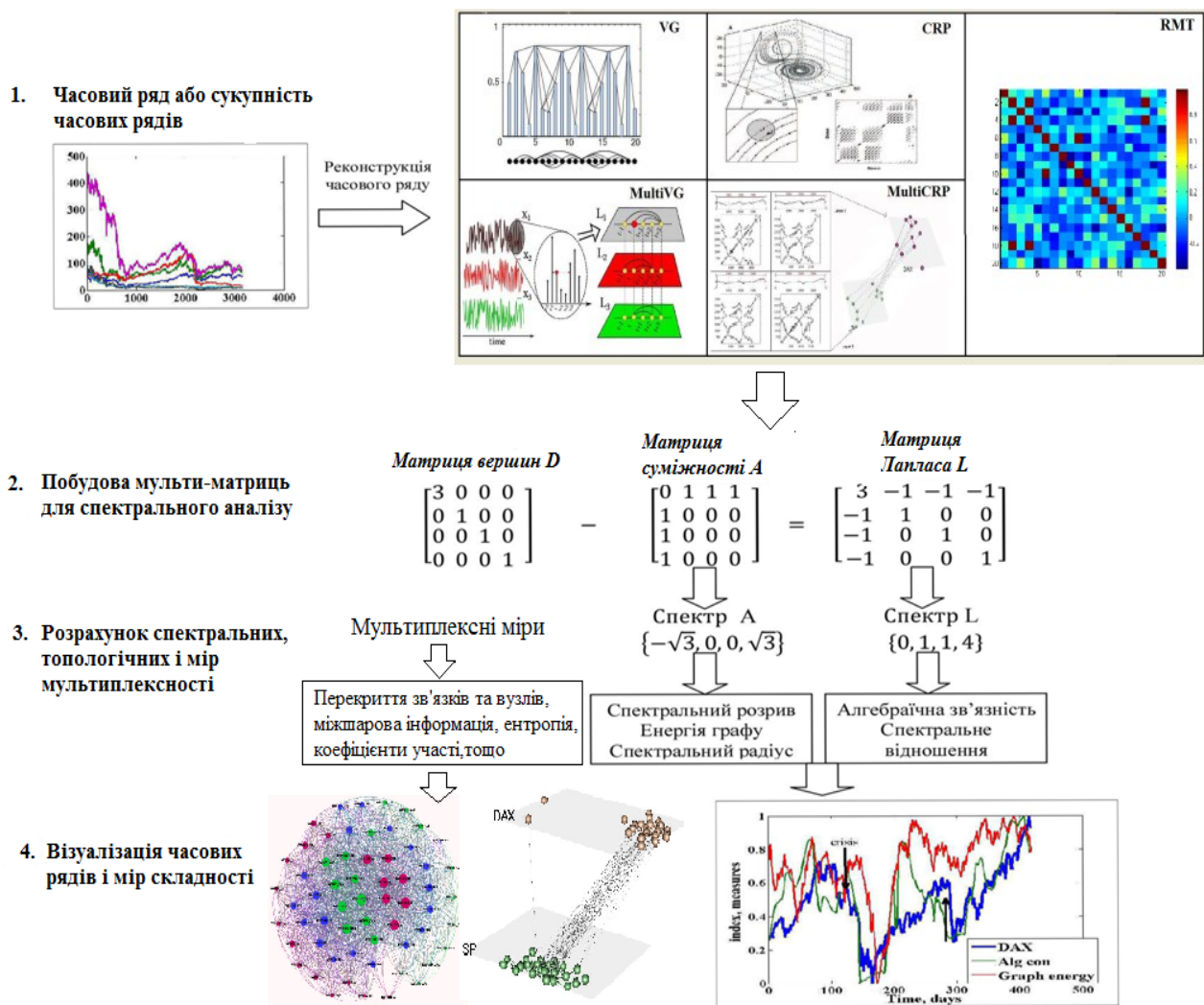


Рис. 14. Загальна схема моделювання мереж, включаючи і мультиплексні мережі

Загалом нами продемонстрована можливість дослідження складних систем у рамках мережної парадигми складності. Часовий ряд можна представити в еквівалентному вигляді – мережі, яка має широкий набір характеристик; як спектральних, так і топологічних. На прикладах відомих фінансових криз показано, що деякі з мережних мір можуть слугувати індикаторами-передвісниками кризових явищ і їх можна використовувати для можливого раннього попередження небажаних кризових явищ на фінансових ринках.

### Завдання для самостійної роботи

1. Для заданих часових рядів чи їх сукупності побудувати всі види мереж, дослідити їх графодинаміку, порівняти результати і зробити висновки щодо їх прогностичних можливостей.
2. Ті ж результати одержати для нормалізованих прибутковостей. Порівняти результати.
3. Як змінюються результати для не фінансових часових рядів. Чи спостерігається універсальність результатів?
4. Побудувати у середовищі Gephi фрагменти мереж різної природи, порівняти, зробити висновки.

## ЛАБОРАТОРНА РОБОТА № 16

**Тема.** Використання енергій вейвлет-коефіцієнтів 2-го та 3-го моментів при дослідженні криз і трендів.

**Мета.** Навчитися використовувати 2-й та 3-й момент енергій вейвлет-коефіцієнтів в дослідженнях фондового ринку.

### Теоретичні відомості

Дамо формальне визначення, яке лежить в основі вейвлет-аналізу. Інтегральним вейвлет-перетворенням функції  $f(t) \in L^2(\mathbb{R})$  називається вираз

$$W(a, b) = \frac{1}{|a|^{1/2}} \int_{-\infty}^{\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt,$$

де  $a, b \in \mathbb{R}, a \neq 0$ .

Функція  $\psi(t)$  називається вейвлетом (базисним або материнським вейвлетом). Параметр  $a$  визначає розмір вейвлета і називається масштабом (*scale*). Параметр  $b$  задає часову локалізацію вейвлета і називається зсувом (*shift*). Для функції  $\psi(t)$  виконуються основні дві умови. По перше, її середнє значення  $\psi(t)$  (тобто інтеграл  $\int_{-\infty}^{+\infty} \psi(x) dx$ ) дорівнює нулю. По-друге, функція швидко спадає при  $t \rightarrow \pm\infty$ .

Наступним кроком є розрахунок **2-го моменту** енергії вейвлет-коефіцієнтів, тобто знаходження значення функції  $W_f(x, a)$  та піднесення до квадрату, аналогічно для **3-го моменту** – піднесення до куба.

Основною перевагою використання вейвлет-аналізу є те, що вейвлет-аналіз дозволяє розкрити аспекти даних, які інші методи аналізу не вловлюють. Це пошук тенденцій, точок розривів і самоподібності.

### Хід роботи

Для розрахунку 2-го моменту енергії вейвлет-коефіцієнтів необхідно в середовищі



Матлаб підключити два програмні пакети – **RrseriesPeriodAnalysis** та **2\_moment\_wentr**. Для того, щоб розпочати розрахунки, необхідно у папці **2\_moment\_wentr** розмістити 2 файли, один зі значеннями обраного індексу та другий – із датами, які відповідають значенням індексу. Ці файли повинні бути у форматі **txt**. Варто звернути увагу на те, щоб дата була у форматі ДД.ММ.РР, а значення індексу – з крапкою, а не з комою після цілого значення. Далі необхідно відкрити файл **wentropy\_dyn\_data\_2moment.m** і в ньому задати необхідні параметри:

- рядок 5 – ім'я файлу зі значеннями індексу;
- 6 – ім'я файлу з датами, які відповідають значенням індексу;
- 21 – розмір вікна (500-600 – підбирається в залежності від динаміки ряду);
- 52 – діапазон значень, для яких необхідно провести розрахунки;
- 76 – значення масштабу (150-250).

```

64 -     ymax=size(POWER,1);
65 -     zmax=max(max(POWER));
66 -     axis([xmin xmax ymin ymax zmin zmax]);
67 -     %h=light;
68 -     set(hh,'FaceLighting','phong','FaceColor','interp',...
69 -         'AmbientStrength',0.5);
70 -     light('Position',[1 0 0],'Style','infinite');
71 -     light('Position',[1 0 0],'Style','infinite');
72 -     %colormap(hsv(128));
73 -     %set(h,'Position',[300,300,10]);
74 -     %set(h,'Color',[0,0,1]);
75 -     end%% 20111120
76 -     [SWT2,POWER2,Pj2]=wentropy_stepen(Y1,wtype,wetype,0,1,250,2,0);
77 -     axes(ax2);
78 -     hh=mesh(POWER2);
79 -     xmin=0;
80 -     ymin=0;
81 -     zmin=min(min(POWER2));
82 -     xmax=size(POWER2,2);
83 -     ymax=size(POWER2,1);
84 -     zmax=max(max(POWER2));
85 -     axis([xmin xmax ymin ymax zmin zmax]);

```

Рис. 1. Місце, де необхідно вказувати оптимальне значення масштабу

Вказавши необхідні параметри, слід зберегти ці параметри та запустити програму. Як тільки програма запуститься, необхідно встановити кут огляду та в командному вікні натиснуть **Enter**.

Даний програмний комплекс здійснює динамічну процедуру розрахунку 2-го моменту енергії вейвлет-коефіцієнтів. У результаті розрахунків отримаємо таку кількість файлів, яка відповідає обраному діапазону. Файли будуть у форматі **bmp**. Ці файли при побудові автоматично зберігаються у папці **2\_moment\_wentr**. Для зручності використання результатів розрахунків доцільно на основі отриманих файлів з графіками створити відеофайл. Це можна здійснити за допомогою програми **VirtualDub-1.9.11**. Bmp-файли доцільно перенести в окрему папку, присвоївши їй назву індексу, для якого проводились розрахунки. Запустивши програми **VirtualDub**, необхідно зайти у вкладку **File** і обрати **Open video file** – переходимо до папки з нашими bmp-файлами, натискаємо на перший з них, і далі – кнопку **Открыть**. В програмі з'являться графіки. Для того, щоб програма записала всі файли виконуємо наступні команди: **File**→**Preview input**. Після того, як програма пролистає всі файли, заходимо у вкладку **Video**→**Compression**→**Cinepek Codec by Radius**→**Ok**. Далі зберігаємо отриманий відео файл - **File**→**Save as AVI**. Необхідно із папки видалити bmp-файли і додати папку



**mplayer** та файли **MPLAYER.EXE** та **play.bat**. Відеофайл запускається файлом **play.bat**, зупиняється кнопкою пробілу.

Розглянемо результати розрахунку 2-го моменту вейвлет-коефіцієнтів для фондового індексу DJI на момент кризи 1929 року.

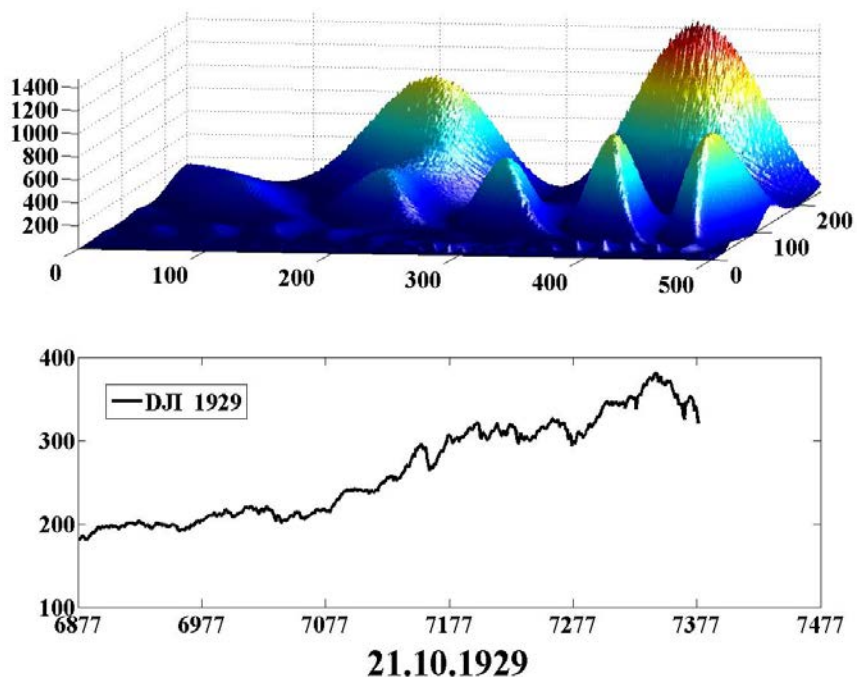


Рис. 2. Динаміка 2-го моменту енергії вейвлет-коефіцієнтів фондового індексу DJI за 7 днів до кризи 1929 року

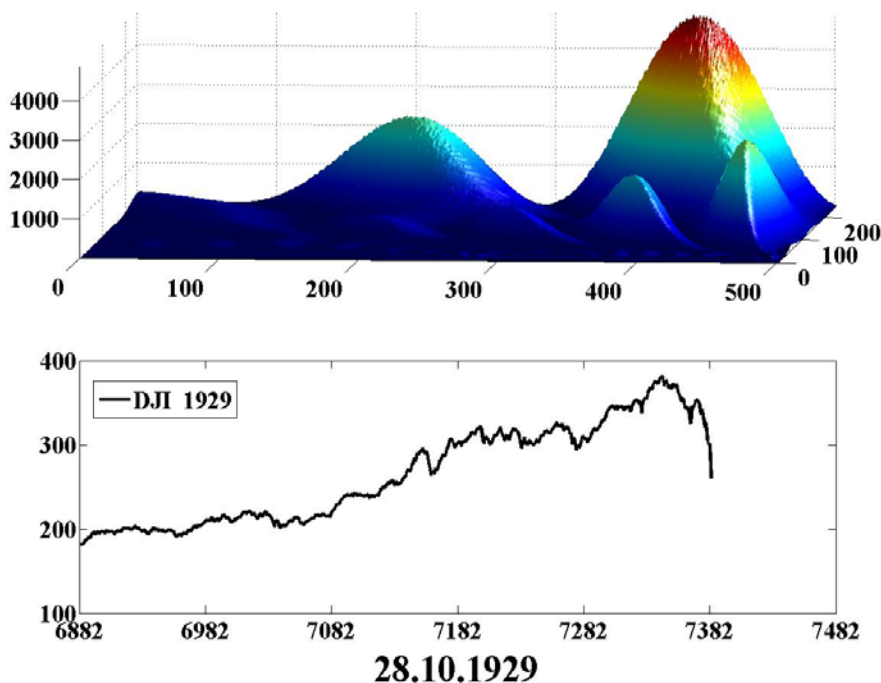


Рис. 2. Динаміка 2-го моменту енергії вейвлет-коефіцієнтів фондового індексу DJI в момент кризи 1929 року

Як бачимо, для передкризового періоду та власне кризи характерними є «три наростаючі хвилі». Це дає підстави використовувати даний показник в якості індикатора-передвісника криз.

Наступна частина роботи пов'язана з розрахунком 3-го моменту енергії вейвлет-коефіцієнтів. При цьому використовується продукт **3\_moment\_wentr**. Для того, щоб розпочати розрахунки, необхідно у папці **3\_moment\_wentr** розмістити 2 файли: один зі значеннями обраного індексу та другий – із датами, які відповідають значенням індексу, та відкрити файл **wentropy\_dyn\_data\_3moment.m** і в ньому задати необхідні параметри:

- рядок 5 – ім'я файлу зі значеннями індексу;
- 6 – ім'я файлу з датами, які відповідають значенням індексу;
- 23 – розмір вікна (170-200 – для часових значень індексу, та 250-350 – для денних значень індексу);
- 66 – діапазон значень, для яких необхідно провести розрахунки.

Вказавши необхідні параметри, треба зберегти ці параметри та запустити програму. Як тільки програма запуситься, необхідно встановити кут огляду та в командному вікні натиснуть **Enter**.

Створення відеофайлу розрахунків для 3-го моменту для обраного фондового індексу проводиться аналогічним чином, як це описано вище для 2-го моменту.

Даний програмний комплекс здійснює динамічну процедуру розрахунку 3-го моменту енергії вейвлет-коефіцієнтів.

На рис. 3 представлені результати розрахунку 3-го моменту енергії вейвлет-коефіцієнтів для денних значень фондового індексу UX (української біржі).

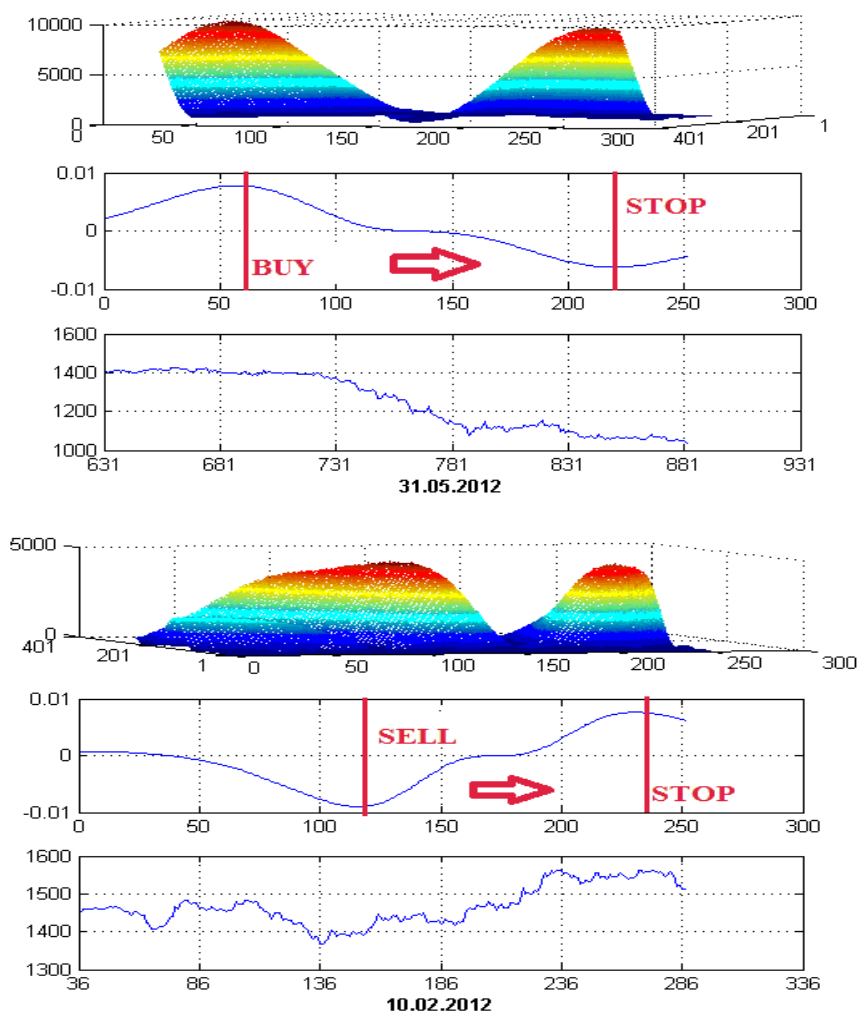


Рис. 3. Динаміка 3-го моменту енергії вейвлет-коефіцієнтів для щоденних значень фондового індексу UX

З рисунка видно, що після того як індикатор покаже максимальну відмітку, доцільно

увійти на ринок. Далі при досягненні індикатора мінімальної відмітки, тобто, коли ціна досягає свого мінімуму, вигідно зайняти позицію на купівлю акцій, таким чином придбавши їх за мінімальною ціною.

Після того, як індикатор покаже мінімальну відмітку, доцільно вийти з ринку. Видно, що при максимальній ціні варто зайняти позицію на продаж, тим самим отримати більший прибуток, продавши наявні акції за максимально високою ціною.

### Завдання для самостійної роботи

1. Провести дослідження 2-го моменту енергії вейвлет-коефіцієнтів для обраних часових рядів, підбравши такі оптимальні значення ширини вікна та масштабу, щоб для передкризових та кризових моментів були характерними «три наростаючі хвилі».
2. На основі дослідження 3-го моменту енергії вейвлет-коефіцієнтів на прикладі обраних часових рядів показати доцільність застосування даного показника в якості індикатора зміни тренду.

## ЛАБОРАТОРНА РОБОТА № 17

**Тема.** Вейвлет-аналіз нестационарних часових рядів.

**Мета.** Навчитися використовувати метод вейвлет-аналізу для отримання частотних характеристик початкового сигналу, а також застосовувати отримані результати для аналізу особливостей формування сигналу і прогнозування його майбутніх значень.

1. Вейвлет-аналіз.
2. Неперервне вейвлет-перетворення.
3. Аналіз статистик коефіцієнтів вейвлет-перетворення.

### Теоретичні відомості

Вейвлет-аналіз дає можливість робити локалізований *аналіз* – тобто, аналізувати обмежену область загального сигналу.

Розглянемо сигнал у вигляді синусоїди з невеликою неоднорідністю, що ледь видима (рис.1). Такий сигнал легко міг бути створений у реальному світі коливанням або шумом.

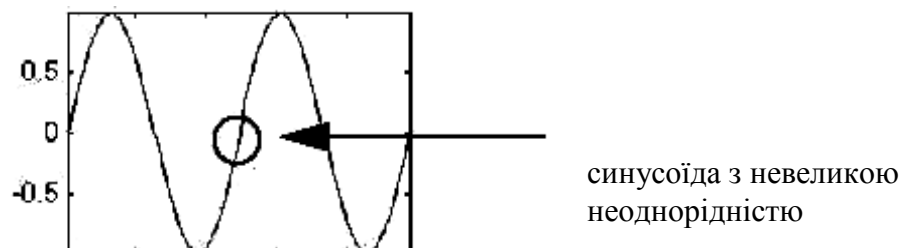


Рис.1. Сигнал з неоднорідністю

Графік Фур'є-коефіцієнтів цього сигналу не показує нічого особливо цікавого: плоский спектр із двома піками, що представляють єдину частоту. Однак, графік вейвлет коефіцієнтів ясно показує точне місце розташування неоднорідності (рис 2).

Вейвлет-аналіз дозволяє розкрити аспекти даних, які інші методи аналізу сигналу не вловлюють. Це пошук тенденцій, точок розривів і самоподібність. Крім того, він дозволяє побачити різні представлення даних на відміну від традиційних методів.

Слід відзначити, що вейвлетом називається функція  $\Psi(t)$ , у якій виконуються дві

основні умови. По-перше, її середнє значення  $\Psi(t)$  (тобто, інтеграл  $\int_{-\infty}^{\infty} \psi(x) dx$ ) дорівнює нулю. По-друге, функція  $\Psi(t)$  швидко спадає при  $t \rightarrow \pm\infty$ .

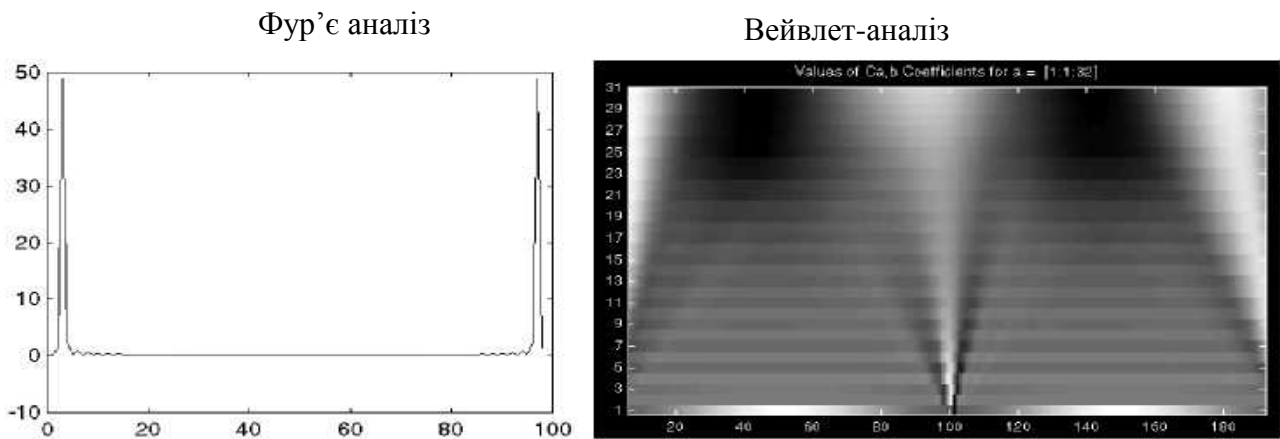


Рис. 2. Результати частотного аналізу синусоїди з невеликою неоднорідністю

Вейвлет має хвилеподібну форму. Якщо порівняти вейвлет з формою синусоїди, яка являється основою аналізу Фур'є, то видно, що синусоїди мають необмежену довжину – вони простягаються від мінус до плюс нескінченності. Синусоїди рівномірні і передбачувані, а вейвлет має тенденцію бути нерегулярним і асиметричним (рис. 3).

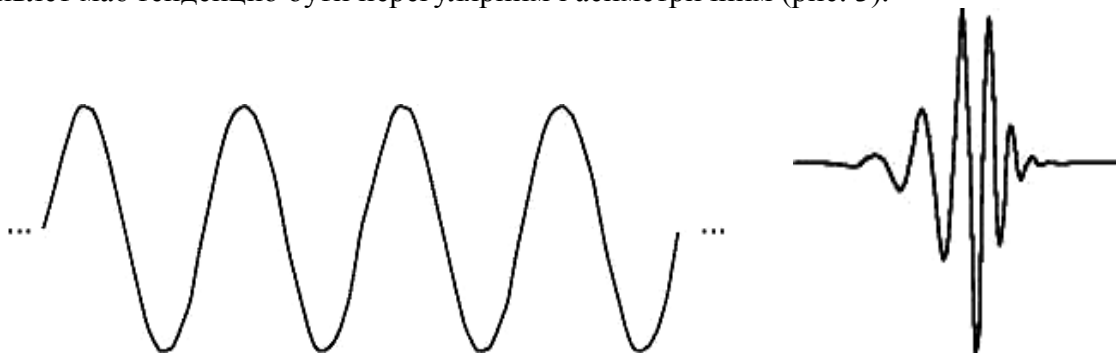


Рис. 3. Графік синусоїди і різновиду вейвлета Добеші – db(10)

Суть Фур'є-аналізу полягає у розкладі сигналу на синусоїди різних частот. Точно так само вейвлет-аналіз полягає в розбиванні сигналу на різновиди початкового вейвлета, які пересуваються і масштабуються.

Тільки дивлячись на зображення вейвлетів і синусоїди, можна інтуїтивно побачити, що сигнали з гострими змінами могли б бути краще проаналізовані нерегулярним вейвлетом ніж синусоїдою.

Математично процес Фур'є аналізу представлений перетворенням Фур'є:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt. \quad (1)$$

Результати перетворення - Фур'є-коефіцієнти  $F(\omega)$ , які при множенні на синусоїду частоти  $\omega$  приводять до складових синусоїдних компонентів початкового сигналу. Графічно процес виглядає наступним чином (рис. 4):

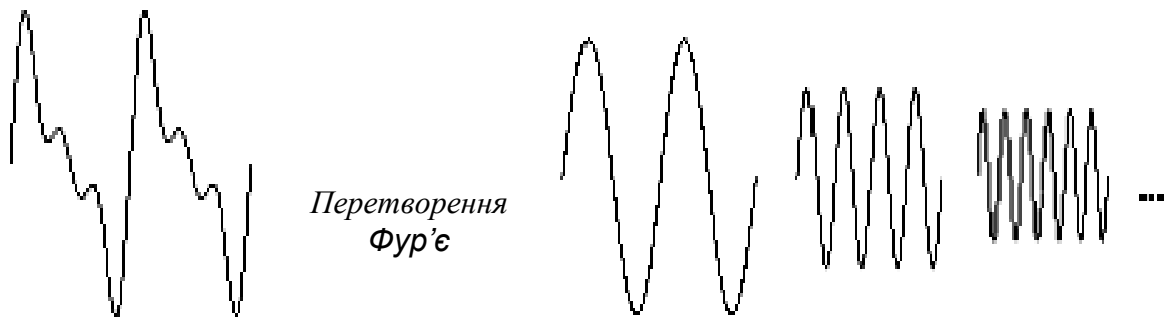


Рис. 4. Сигнал, розкладений на складові синусоїди різних частот у результаті Фур'є-аналізу

Точно так само неперервне вейвлет перетворення визначається як сума у часі сигналу, помноженого на масштабовані й переміщені модифікації вейвлет-функції  $\psi$ :

$$C(\text{scale}, \text{position}) = \int_{-\infty}^{\infty} f(t)\psi(\text{scale}, \text{position}, t)dt . \quad (2)$$

Результат неперервного вейвлет-перетворення - велика кількість вейвлет-коефіцієнтів  $C$ , що являються функцією від масштабу й позиції. Масштабування вейвлета просто означає його розтягування (або стискання).

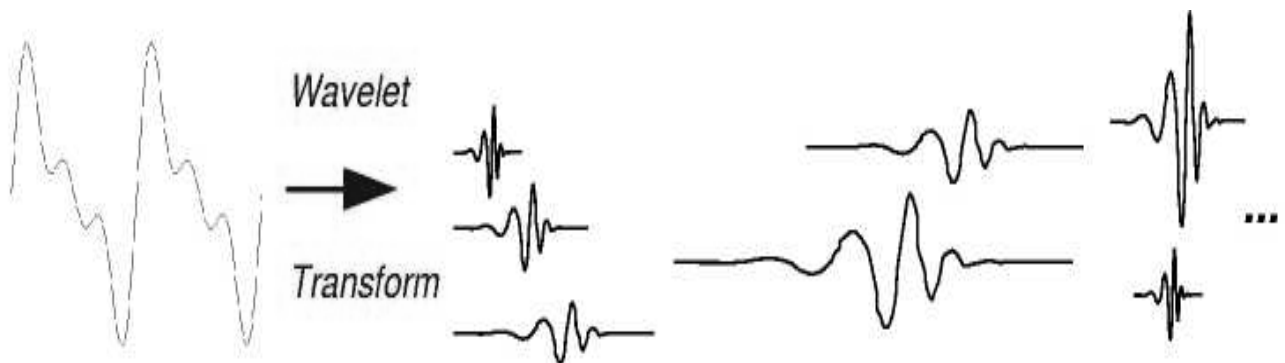


Рис. 5 Початковий сигнал, розкладений на складові сигнали на різних масштабах і позиціях

*Фактор масштабу* часто позначається буквою  $a$ . Якщо ми говоримо, наприклад, про синусоїди, ефект фактору масштабу дуже легко пояснює рис. 6.

Фактор масштабу працює точно так само й із вейвлетами. Чим менший фактор масштабу, тим більш «стиснутим» є вейвлет.

З діаграм видно, що для синусоїди  $\sin(\omega t)$  фактор масштабу пов'язаний (обернено пропорційно) з частотою  $\omega$ . Точно так само для вейвлет-аналізу, масштаб пов'язаний із частотою сигналу.

Переміщення вейвлета просто означає зсув початку вікна, для якого проводяться обчислення. Математично переміщення  $f(t)$  по  $x$ , де  $x$  – це час, представляється як  $f(t-x)$ .

Неперервне вейвлет-перетворення – сума за весь час сигналу, помноженого на масштабовані, переміщені модифікації вейвлета. У такий спосіб розраховуються вейвлет-коефіцієнти, що являються функцією від масштабу й положення.

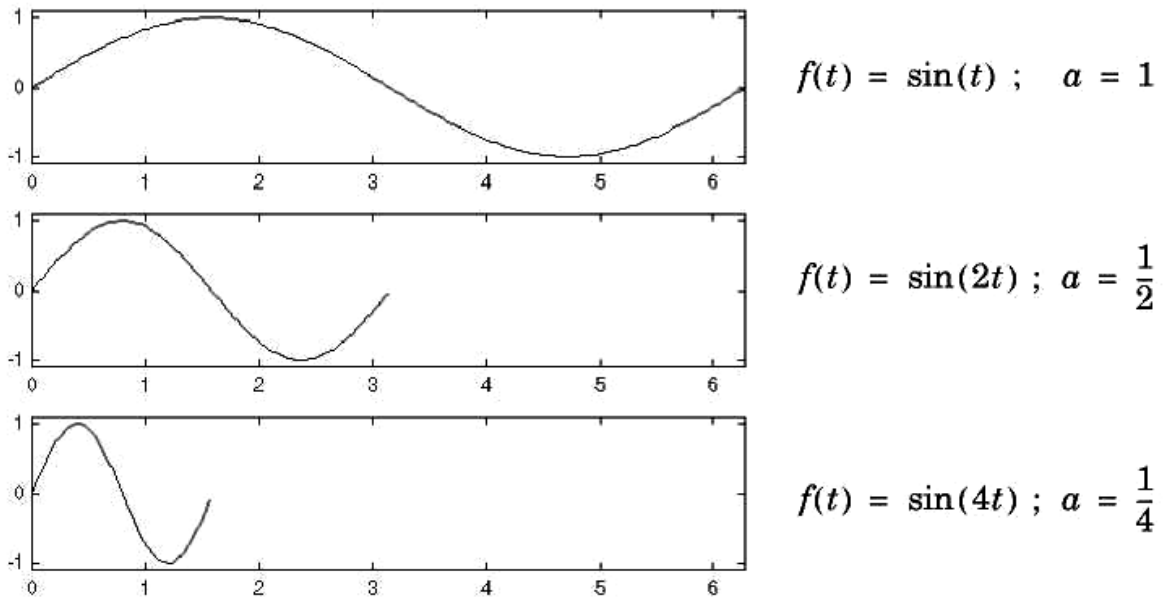


Рис 6. Масштабування синусоїди

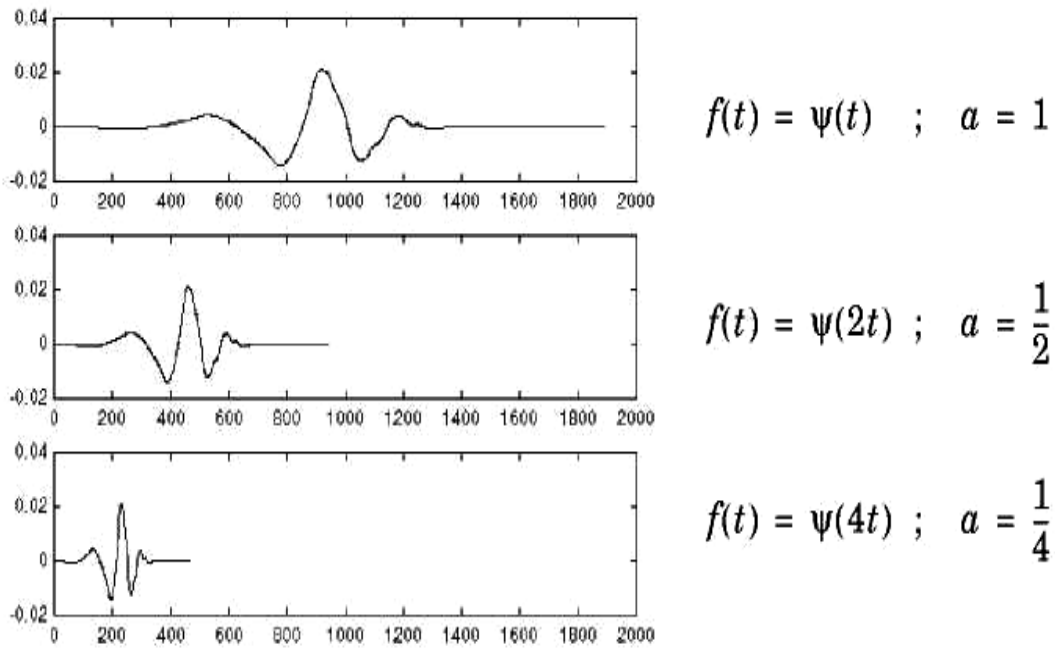
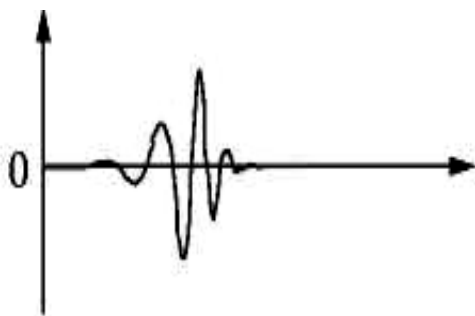
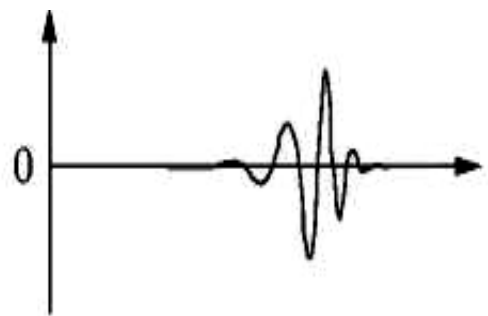


Рис. 7. Масштабування вейвлету



вейвлет-функція  $\psi(t)$

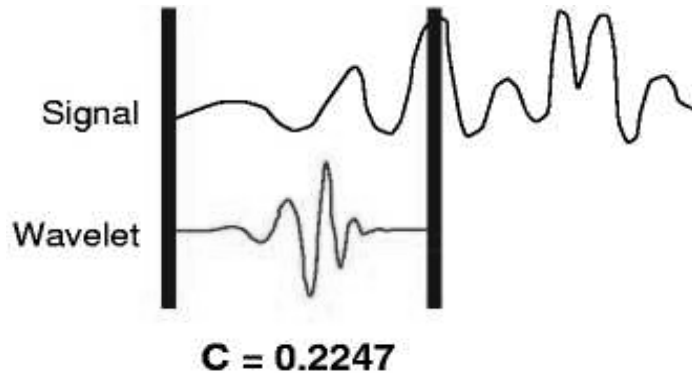


переміщена вейвлет-функція  $\psi(t-x)$

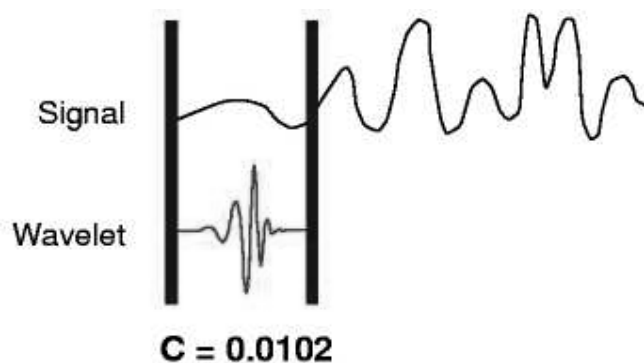
Рис. 8. Переміщення вейвлет-функції

Здійснення вейвлет-перетворення являє собою дуже простий процес, який можна передати п'ятьма кроками:

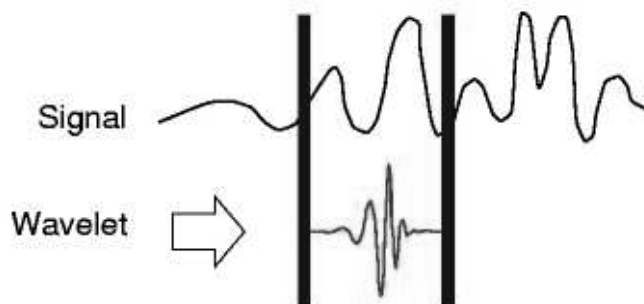
1. Береться вейвлет і порівнюється із початковою частиною вихідного сигналу.
2. Обчислюється значення  $C$ , яке показує, як сильно корелює вейвлет із частиною вихідного сигналу. Чим вище  $C$ , тим більша подібність. Точніше, якщо енергія сигналу й енергія вейвлета еквівалентні,  $C$  може інтерпретуватися як коефіцієнт кореляції.
3. Результати будуть залежати від форми вейвлета, що вибирається.



4. Вейвлет переміщується вправо й повторюються кроки 1 і 2, поки не буде досягнутий кінець сигналу.



5. Масштабується вейвлет і повторюються кроки 1-3.



6. Повторюються кроки 1-4 для всіх масштабів.

Результатом являються коефіцієнти, отримані на різних масштабах для різних частин сигналу.

Формальний запис визначеного алгоритму має наступний вигляд:

$$W_f(x, a) = \frac{1}{a} \int_{-\infty}^{+\infty} \overline{\Psi}\left(\frac{t-x}{a}\right) f(t) dt \quad (3)$$

Коефіцієнти неперервного вейвлет-перетворення дають точне уявлення про початковий сигнал у залежності від масштабу й часу.

Між масштабами вейвлета й частотою є відповідність :

– Низький масштаб  $a \rightarrow$  Зжатиий вейвлет  $\rightarrow$  Швидко змінюються деталі  $\rightarrow$  Висока частота  $\omega$ .

– Високий масштаб  $\rightarrow$  Розтягнутий вейвлет  $\rightarrow$  Повільна зміна, великі деталі  $\rightarrow$  Низька частота  $\omega$ .

Перевагою неперервного вейвлет-перетворення (CWT) у порівнянні з дискретним (DWT) є те, що воно може працювати на кожному масштабі: від початкового сигналу до деякого максимального масштабу, який визначається, виходячи з потреби в детальному аналізі й доступній обчислювальній потужності.

Процедура CWT також неперервна в термінах переміщення: впродовж обчислення вейвлет переміщується плавно по всій області аналізованої функції.

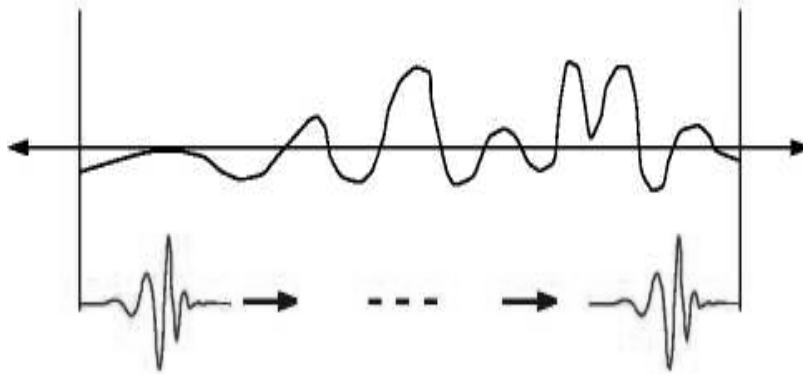


Рис. 8. Переміщення вейвлета вздовж аналізованого сигналу

Оскільки, динамічний ряд змінюється із часом, то його вейвлет-коефіцієнти також коливаються із частотою по відношенню до нульової частоти масштабу  $\omega_0$ . Отже, для аналізу цих флуктуацій можна використовувати дисперсії розподілів на різних масштабах. Для порівняння графіків дисперсій і моментів  $n$ -х порядків початкові данні нормалізуються:

$$X_i^* = \frac{X_i - \bar{X}}{\sigma_x}, \quad (4)$$

де  $X_i$  – це  $i$ -й елемент ряду динаміки ( $i = 1-n$ ),  $\bar{X}$  – середнє значення ряду,  $\sigma_x$  – дисперсія ряду,  $X_i^*$  – нормалізовані значення.

Дисперсія вейвлет-коефіцієнтів для вікна обчислюється за наступною формулою:

$$\sigma(j, w) = \frac{\sum_{k=i}^{i+w} [d_{j,k} - \langle d_{j,k} \rangle]^2}{w}, \quad (5)$$

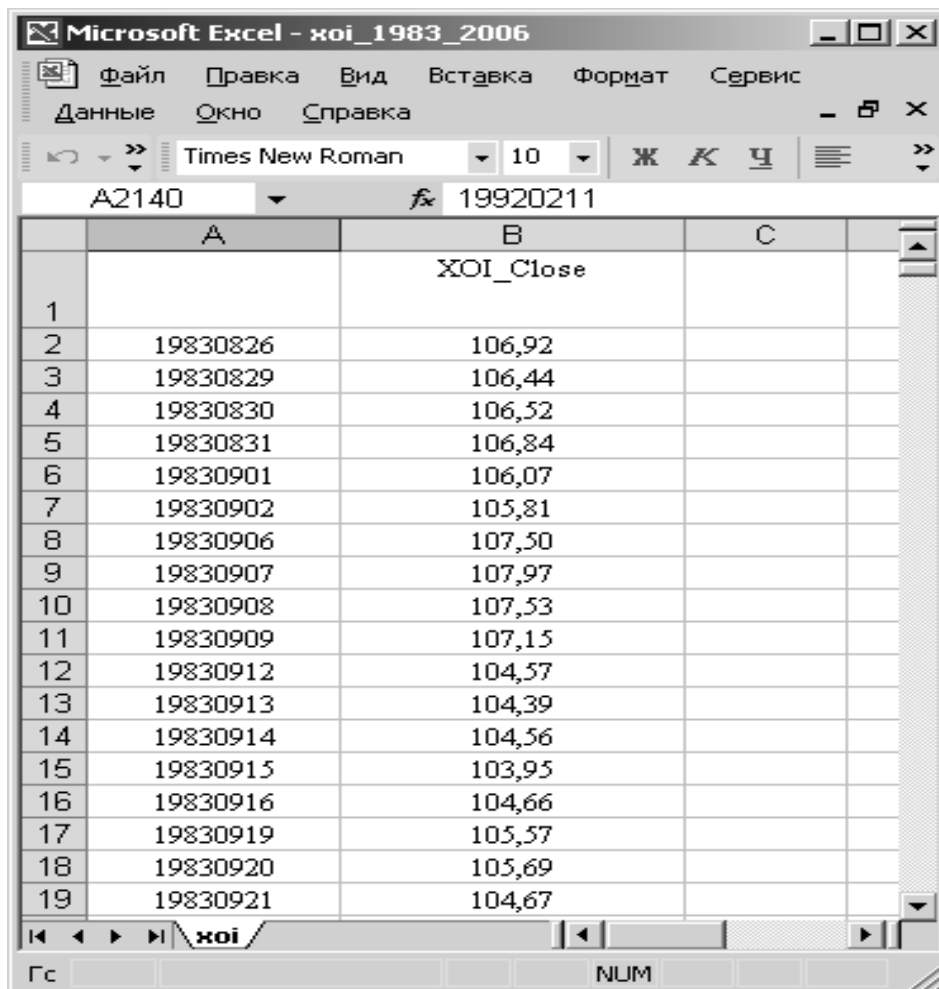
де  $w$  – означає число вейвлет-коефіцієнтів на рівні  $j$  у заданому часовому інтервалі;  $d_{j,k}$  – вейвлет-коефіцієнти;  $\langle d_{j,k} \rangle$  – середнє значення вейвлет-коефіцієнтів у вікні. Для дослідження флуктуацій початкового ряду також використовуються моменти вищих порядків.



## Хід роботи

Розрахунок вейвлет-коефіцієнтів, їх статистик здійснюється за допомогою функцій **dwvar\_norm**, **wvar\_norm** і **plots\_norm**. Ці функції мають знаходитися в каталозі **work**, або, якщо вони знаходяться в іншому каталозі, шлях до нього необхідно задати через меню **File→Set Path→Add Folder**.

Перед розрахунком в робочу область MatLab мають бути завантажені 3 масиви шляхом імпорту з файлу Excel через меню **File→Import Data**. Структура, яку повинен мати файл Excel, наведена на рис.9.



	A	B	C
1		XOI_Close	
2	19830826	106,92	
3	19830829	106,44	
4	19830830	106,52	
5	19830831	106,84	
6	19830901	106,07	
7	19830902	105,81	
8	19830906	107,50	
9	19830907	107,97	
10	19830908	107,53	
11	19830909	107,15	
12	19830912	104,57	
13	19830913	104,39	
14	19830914	104,56	
15	19830915	103,95	
16	19830916	104,66	
17	19830919	105,57	
18	19830920	105,69	
19	19830921	104,67	

Рис. 9. Структура файлу Excel, яка потрібна для розрахунку вейвлет-коефіцієнтів і їх моментів

Коли змінні вже завантажені в робочу область MatLab, для запуску розрахунку потрібно набрати у командному рядку **dwvar\_norm** і в дужках зазначити всі нижчепераховані параметри:

- ім'я масиву, який складається зі стовпця дат і стовпця значень;
- початкове значення розрахунків;
- кінцеве значення розрахунків.

Наприклад, проведемо аналіз протікання економічної кризи 1987 р. в США на основі динаміки індексу Dow Jones. Розрахуємо вейвлет-коефіцієнти і їх моменти для 3-х періодів: до кризи, під час кризи і після кризи. Для цього в командному рядку наберемо

```
>>d0=dwvar_norm(data,5258,6258);  
v0=dwvar_norm(data,5733,6733);  
p0=dwvar_norm(data,6258,6680);
```

Після закінчення розрахунку в робочій області з'явиться 3 змінні. Набравши в командній строчці `plots_norm(d0)`, отримаємо графічне зображення індексу Dow Jones до кризи, його вейвлет-коефіцієнтів, дисперсій, моментів 3-го, 4-го й 5-го порядку - всього виводиться 7 графіків. Відповідно вводиться функція і для отримання графічного зображення результатів розрахунків під час кризи і після кризи. Результати представлені на рис. 10-12.

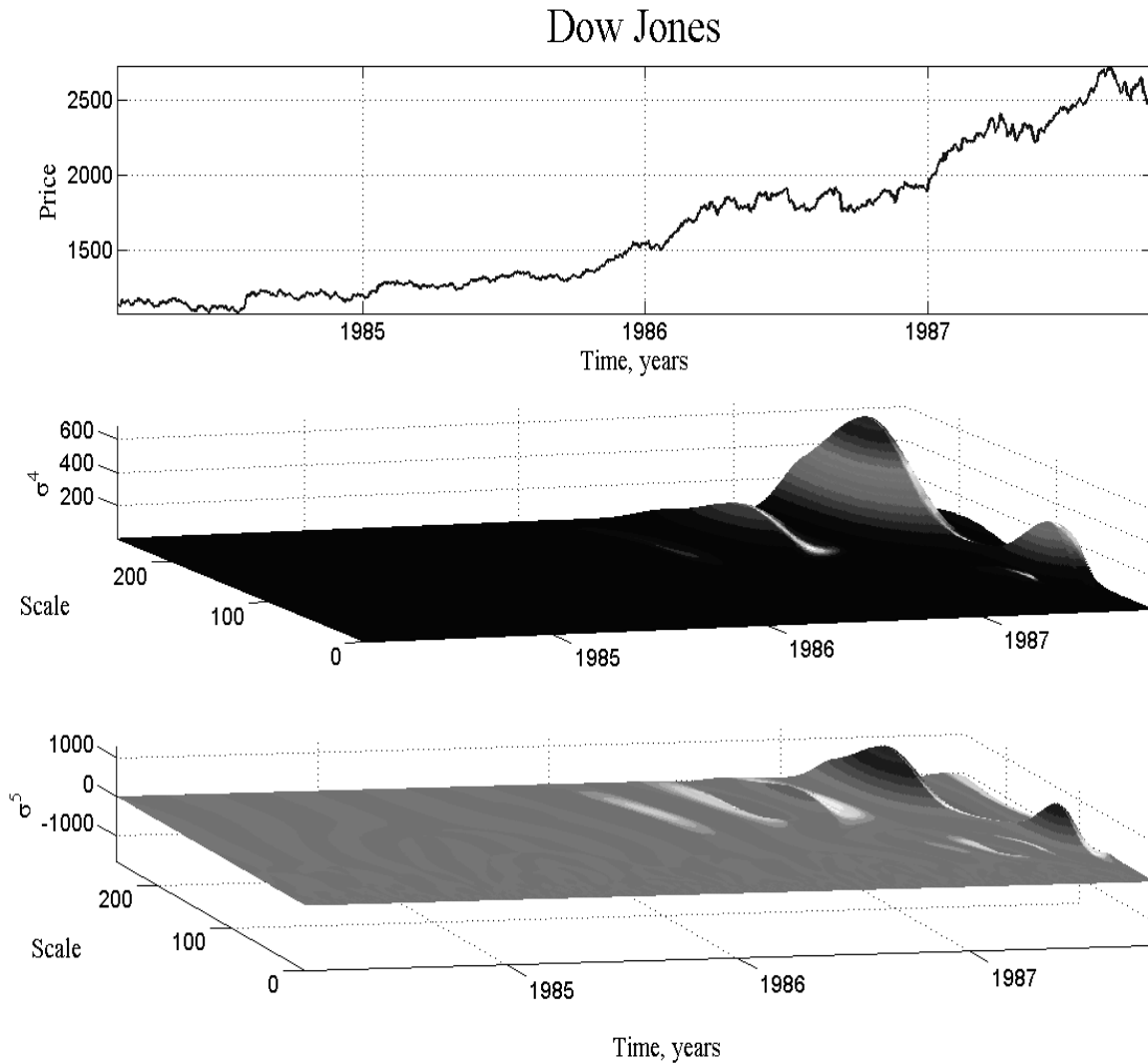


Рис. 10. Графік індексу Dow Jones, моментів 4-го й 5-го порядку його вейвлет-коефіцієнтів до кризи 19 жовтня 1987р.

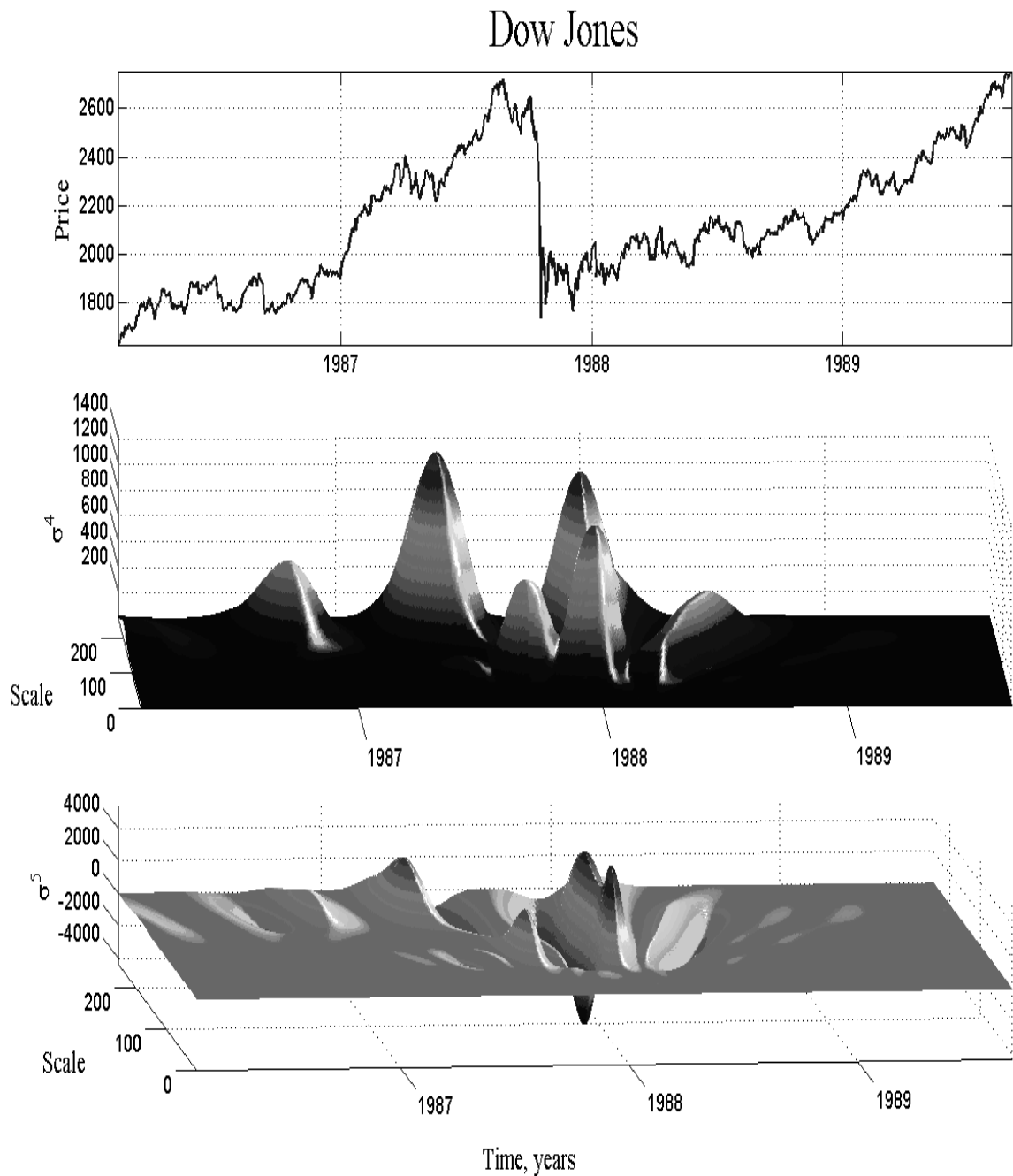


Рис. 11. Графік індексу Dow Jones, моментів 4-го й 5-го порядку його вейвлет-коефіцієнтів під час кризи 19 жовтня 1987р.

Порівняльний аналіз рис. 10 і 11 показав: формування і розв'язок економічної кризи 1987 р. в США відбувалися на масштабі 200-250. Причому перед самим початком кризи з'являються коливання на високих частотах (низькому масштабі), які можуть характеризувати ажіотажні дії дрібних економічних суб'єктів на біржах (наприклад, спекуляції дрібних трейдерів, пов'язані зі стрімким зростанням цін акцій і індексів фондових ринків). Цікавим фактом є те, що вхідні дані для 11 графіку не зачіпають моменту кризи, але показують перед кризою такі ж хвилеподібні коливання і на тих же частотах, що й на рис. 11.

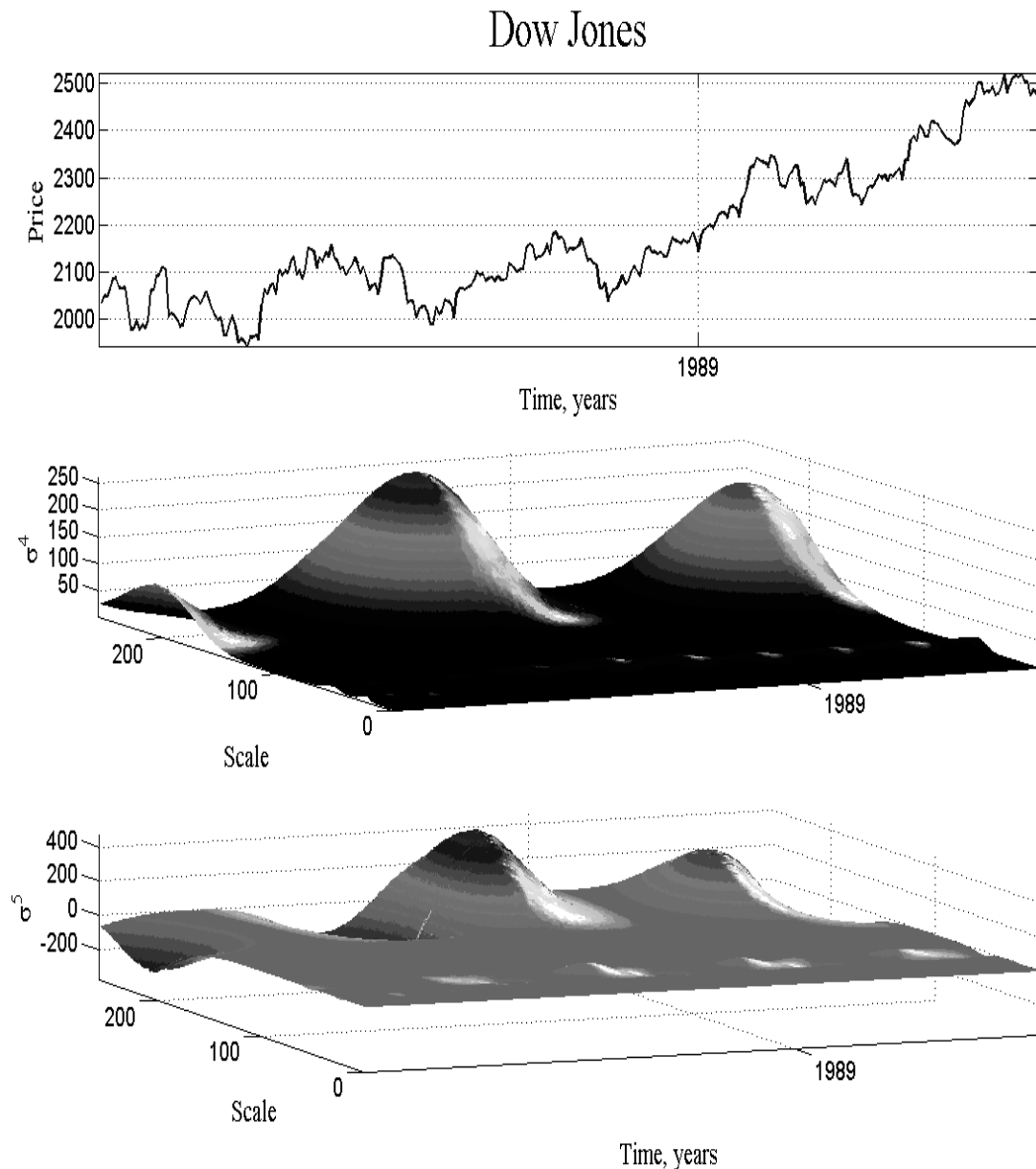


Рис. 12. Графік індексу Dow Jones, моментів 4-го й 5-го порядку його вейвлет-коефіцієнтів після кризи 19 жовтня 1987 р.

Графік, на якому зображено ситуацію після кризи (рис. 12), показує поступовий процес релаксації ринку після кризи і відновлення нормальної діяльності. Він характеризується рівномірними коливаннями індексу Dow Jones і поступовим відновленням попередньої динаміки, що відображається на графіку моментів 4-го і 5-го порядку вейвлет-коефіцієнтів.

### Завдання для самостійної роботи

Виберіть у папці Bases відповідну базу даних, бажано із кризовим явищем. Знайдіть інформацію про цю кризу, особливості її протікання. Виконайте 3 варіанти розрахунків:

- 1) докризовий стан;
- 2) під час кризи;
- 3) післякризовий стан.

Отримані графіки проаналізуйте.

Результати роботи представте у вигляді відповідно оформленої індивідуальної роботи.

### Контрольні питання

1. Поясніть основну ідею вейвлет-аналізу, його методику.
2. Для чого розраховуються дисперсії й моменти вейвлет-коефіцієнтів?
3. Які процеси проявляються на низьких і високих частотах вейвлет-коефіцієнтів?
4. Які спостерігаються зміни дисперсій і моментів вейвлет-коефіцієнтів під час кризи?

### ЛАБОРАТОРНА РОБОТА № 18

**Тема.** Використання методів нелінійної динаміки для дослідження динаміки складних соціально-економічних систем.

**Мета.** Набути практичних навичок розрахунку нелінійних динамічних характеристик складних систем з метою моніторингу і прогнозування поведінки складних систем.

1. Хаотичний процес
2. Вкладена розмірність, атрактор
3. Фрактальна розмірність
4. Кореляційна розмірність

### Теоретичні відомості

Сучасний фінансовий ринок – суттєво нелінійна система, дослідження якої неможливе у рамках існуючих лінійних парадигм.

Існує кілька конкуруючих підходів, які при побудові моделей використовують ідею нелінійності. Традиційні економетричні моделі є стохастичними (ARCH, GARCH і їх модифікації). Проте ті обмеження, які використовуються при побудові і підгонці моделі з метою зробити її придатною для практичного застосування, по суті справи, знищують ту унікальну внутрішню «складність», яка властива даному динамічному процесу і без якої він стає безликим представником абстрактних фінансових даних.

У зв'язку з цим, останнім часом інтенсивно розвивається альтернативний підхід до аналізу нелінійностей, а саме: підхід, що базується на теорії детермінованого хаосу, який пропонує вичерпне пояснення іррегулярній поведінці і аномаліям у системах, які, не будучи за своєю природою стохастичними, поведуться так само. Теорія хаосу пропонує абсолютно нові концепції і алгоритми для аналізу часових рядів, які можуть привести до глибшого і повнішого розуміння фінансових процесів, що вони відображать. Теорія динамічного хаосу вводить в традиційну мову фінансових аналітиків такі нові у цій сфері поняття, як фазовий простір, атрактор, ляпуновські експоненти, горизонт прогнозу, ентропії і розмірності, фрактальні статистики і інформаційні цикли і т.д.

### Основні властивості часових рядів

Аналіз взаємозв'язків економічних даних, представлених у вигляді часових рядів, є необхідною складовою частиною сучасних досліджень в області макроекономічної динаміки, перехідної економіки, економетрики фінансових ринків. В кінці 1980-початку 1990-х рр. було усвідомлено, що тільки врахування часової структури даних про реальні економічні процеси дозволяє адекватно відобразити їх в математичних і економетричних моделях. Усвідомлення цього факту привело як до ревізії багатьох макроекономічних теорій і побудов, так і до бурхливого розвитку специфічних методів аналізу таких даних. Знання цих методів і способів застосування їх до аналізу конкретних економічних процесів, є в даний час необхідною складовою підготовки економістів-аналітиків.

Розроблені для цілей економічних досліджень методи аналізу нестационарних випадкових процесів істотно відрізняються від тих, що знайшли широке застосування в

техніці і теорії управління прийомів роботи із стаціонарними випадковими процесами. У технічних додатках аналіз часових рядів використовується переважно для прогнозування і супроводжується значною часткою так званого «data mining». У сучасній економіці значно більше місце займає оцінювання динамічних моделей, що відображають короткострокові і довгострокові зв'язки між економічними змінними.

Вже в початковому курсі економетрики при вивченні автокореляції розглядаються співвідношення виду

$$x_t = \rho x_{t-1} + \varepsilon_t. \quad (1)$$

Значення деякої економічної змінної залежить від її ж значень в попередній момент часу, від її значень з лагом – зрушенням за часом на один крок назад. Включення змінних з лагом в економетричні співвідношення означає істотну зміну в «філософії» моделювання. Якщо в звичних економетричних моделях значення однієї змінної залежить від одночасних значень інших змінних, тобто від поточного стану економічної системи, та наявність змінних з лагом означає, що поведінка системи визначається не тільки її поточним станом, але і траєкторією, по якій система прийшла в цей стан.

З математичної точки зору економетрична модель такого типу є не функцію пояснюючих змінних, а функціоналом від траєкторії (траєкторій) економічних змінних. Тому елементами таких моделей є траєкторії: множина даних  $\{x_t, t \in T\}$ , де  $T$  – деяка скінчена або континуальна множина.

Моделювання залежностей виду  $y_t = f(x_t, \varepsilon_t)$ , де  $x$  і  $y$  – дані типи траєкторій, приводить до ситуації, коли звичні прийоми регресійного аналізу не дають прийнятних оцінок параметрів.

В курсі економетрики ряд змістовних задач приводить до рівнянь, які природно назвати авторегресійними (autoregressive). Зокрема, при усуненні автокореляції було одержано рівняння:

$$Y_t = \alpha + \beta Y_{t-1} + \gamma X_t + \varepsilon_t. \quad (2)$$

В ньому як пояснююча змінна з'явилася змінна із запізнюванням, тобто регресія змінної на саму себе – саме тому таке рівняння і назвали авторегресійним.

Крім того, в курсі економетрики розглядалися моделі з розподіленими лагами (distributed lag models), в яких пояснююча змінна також присутня із запізнюванням:

$$Y_t = \alpha + \beta Y_{t-1} + \gamma_0 X_t + \gamma_1 X_{t-1} + \dots + \varepsilon_t. \quad (3)$$

У математичних термінах можна записати:

$$Y_t = f(T_{t-\tau}, X_{t-\tau}; t \in [1, T], \tau \in [0, t]). \quad (4)$$

Важливо, що тепер у модель входить не тільки по одному спостережуваному значенню змінних  $X$  і  $Y$ , а сукупність спостережуваних значень (траєкторія). Причому по самому значенню ми не можемо вважати, що послідовні значення кожної із змінних незалежні між собою у статистичному значенні.

### Дослідження динаміки ринків методами хаотичної динаміки

Останні роки ознаменувалися інтересом, що зростає, до пошуку моделей нелінійної поведінки на фінансових ринках. Це пояснюється тим, що нелінійні моделі можуть

уловлювати дуже складні паттерни у фінансових даних. В той же час, як тепер уже добре усвідомлено, лінійний підхід не дозволяє врахувати і проаналізувати сильно нерегулярну поведінку, яка може демонструватися численними фінансовими активами. До початку 60-х років у нелінійних дисипативних динамічних системах в стаціонарному режимі спостерігали тільки періодичні і квазіперіодичні рухи. Проте в 1963 році у динамічній системі Лоренцем був знайдений дуже складний рух, який сприймався як хаотичний.

Для характеристики таких рухів ввели поняття "динамічний хаос". Слово "динамічний" означає, що відсутні джерела флуктуацій.

У статті математиків Рюеля і Такенса, опублікованої в 1971 році, був введений новий математичний образ динамічного хаосу - *дивний атрактор*. Слово "дивний" підкреслює дві властивості атрактора. Це, по-перше, незвичність його геометричної структури. Розмірність дивного атрактора є дробовою (фрактальною). По-друге, дивний атрактор – це притягуюча область для траєкторій з навколишніх областей. При цьому всі траєкторії усередині дивного атрактора динамічно нестійкі, що виражається у сильній (експоненціальній) розбіжності близьких у початковий момент траєкторій.

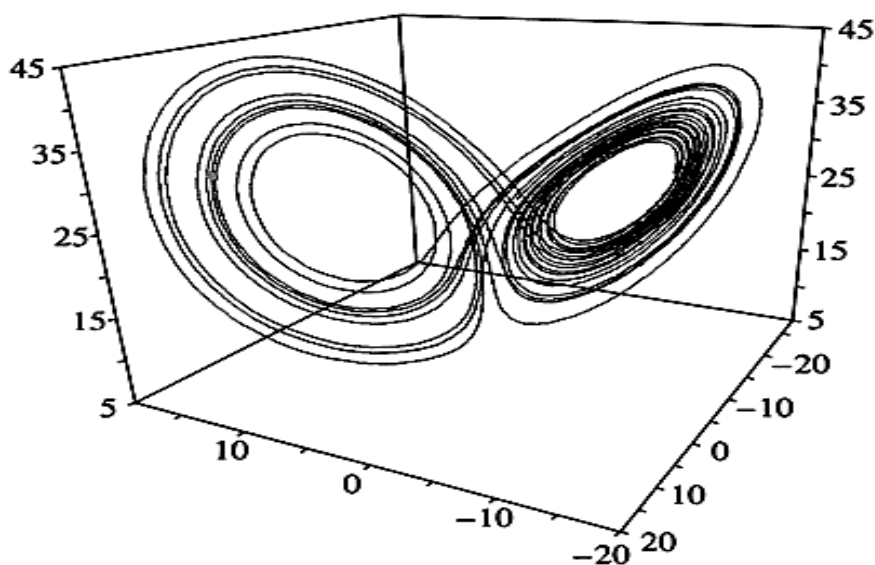


Рис. 1. Атрактор Лоренца

Для характеристики атракторів доцільно ввести поняття розмірності. Розмірність визначає кількість інформації, необхідної для задання координат точки, що належить атрактору, у рамках вказаної точності. Обговорювані у літературі визначення розмірності загалом розділяються на два типи: залежні тільки від метричних властивостей атрактора і, крім метрики, залежні від статистичних властивостей потоку, обумовлених динамікою. У типових випадках метричні розмірності приймають однакову величину, яку прийнято називати фрактальною розмірністю атрактора.

### Фрактальна розмірність дивного атрактора

Введемо визначення фрактальної розмірності  $D_f$  довільного атрактора у  $n$ -мірному фазовому просторі по Колмогорову-Хаусдорфу :

$$D_f = \lim_{\varepsilon \rightarrow 0} [\ln M(\varepsilon) / \ln(1/\varepsilon)]. \quad (5)$$

де  $M(\varepsilon)$  – мінімальне число  $n$ -мірних кубиків з ребром  $\varepsilon$ , необхідних для покриття атрактора. Для нетривіальних множин розмірність  $D_f$  може виявитися дробовою.

Зауважимо, що у формулі для обчислення фрактальної розмірності однаково важливі всі непорожні кубики. Це представляє серйозний недолік для дивних атракторів, оскільки вони просторово неоднорідні, тобто деякі області атрактора відвідуються частіше за інші. Потрібне знання дуже довгої траєкторії, щоб гарантувати попадання навіть дуже малоймовірних кубиків. Тому кожен непорожній кубик потрібно зважувати за допомогою відносної частоти, з якою він відвідується типовою траєкторією. Розмірності, що визначаються з урахуванням ймовірності відвідин траєкторією різних областей атрактора у фазовому просторі, називають імовірнісними.

Інформаційна розмірність визначається таким чином :

$$D_I = \lim_{\varepsilon \rightarrow \infty} [\ln I(\varepsilon) / \ln(1/\varepsilon)], \quad (6)$$

$$I(\varepsilon) = - \sum_{i=1}^{M(\varepsilon)} p_i \ln p_i. \quad (7)$$

Тут  $I(\varepsilon)$  – кількість інформації, необхідна для визначення стану системи у межах точності  $\varepsilon$ ,  $M(\varepsilon)$  – число кубиків із стороною  $\varepsilon$ , що покривають атрактор,  $p_i$  – ймовірність відвідин фазовою траєкторією  $i$ -го кубика. Оскільки для малих  $\varepsilon$   $I(\varepsilon) \approx D \ln(1/\varepsilon)$ , то  $D_I$  характеризує швидкість зростання інформації із зменшенням. Якщо атрактор просторово однорідний, то  $D_I = D_f$ , інакше  $D_I < D_f$ .

Ще одним представником класу ймовірнісних розмірностей є кореляційна розмірність  $D_C$ , що визначається співвідношенням :

$$D_C = \lim_{\varepsilon \rightarrow \infty} \left[ \ln \left( \sum_{i=1}^{M(\varepsilon)} p_i^2 \right) / \ln \varepsilon \right], \quad (8)$$

де  $p_i^2$  – ймовірність того, що пара точок атрактора належить  $i$ -му кубикі.

Кореляційну розмірність можна представити у вигляді:

$$D_C = \lim_{\varepsilon \rightarrow \infty} [\ln C(\varepsilon) / \ln \varepsilon], \quad (9)$$

$$C(\varepsilon) = \lim_{m \rightarrow \infty} \frac{1}{m^2} \sum \theta(\varepsilon - \rho(x_i, x_j)), \quad (10)$$

$$\theta(\alpha) = \begin{cases} 1, & \alpha \geq 0, \\ 0, & \alpha < 0, \end{cases} \quad (11)$$

$x_i$  – точки у фазовому просторі;  $\rho$  – відстань. Таким чином, розмірність  $D_C$  визначається значенням кореляційного інтеграла  $C(\varepsilon)$ , який характеризує відносне число пар точок  $x_i, x_j$ , віддалених на відстані

$$r_{ij} = \rho(x_i, x_j) \leq \varepsilon. \quad (12)$$

Відзначимо, що всі три розмірності, розглянуті вище, є окремими випадками узагальненої розмірності Реньї:

$$D_q = \lim_{\varepsilon \rightarrow \infty} [\ln I_q(\varepsilon) / \ln(1/\varepsilon)],$$

$$I_q(\varepsilon) = \frac{1}{1-q} \sum_{i=1}^{M(\varepsilon)} p_i^q \quad (13)$$



де  $I_q$  – інформація Реньї порядку  $q$ .

$$D_0 = \lim_{\varepsilon \rightarrow 0} \left[ \frac{1}{1-0} \ln \sum_{i=1}^{M(\varepsilon)} p_i^0 / \ln(1/\varepsilon) \right] = \lim_{\varepsilon \rightarrow 0} [\ln M(\varepsilon) / \ln(1/\varepsilon)] = D_f, \quad (14)$$

$$D_1 = \lim_{\varepsilon \rightarrow 0} \left[ - \sum_{i=1}^{M(\varepsilon)} p_i \ln p_i / \ln(1/\varepsilon) \right] = D_I, \quad (15)$$

$$D_2 = \lim_{\varepsilon \rightarrow 0} \left[ \frac{1}{1-2} \ln \sum_{i=1}^{M(\varepsilon)} p_i^2 / \ln(1/\varepsilon) \right] = \lim_{\varepsilon \rightarrow 0} \left[ \ln \sum_{i=1}^{M(\varepsilon)} p_i^2 / \ln(\varepsilon) \right] = D_C. \quad (16)$$

Для цілих  $q$  розмірність  $Dq$  має фізичний сенс. Великі додатні значення  $q$  підкреслюють найщільніші області у фазовому просторі, тоді як великі від'ємні значення  $q$  підкреслюють найбільш рідко відвідувані області. Таким чином, діапазон значень розмірностей  $Dq$  може розглядатися як характеристика ступеня просторової неоднорідності атракторів.

### Ентропія

Виконаємо далі розбиття фазового простору, що включає атрактор, на  $M(\varepsilon)$  непересічних  $n$ -мірних кубиків з ребром  $\varepsilon$ . Виконаємо  $m$  послідовних вимірювань, стежачи за фазовою траєкторією і через рівні проміжки часу  $\tau$  відзначаючи кубики  $s_i$ , в яких побудована траєкторія. При кожному незалежному випробуванні одержимо конкретну реалізацію у вигляді послідовності кубиків  $s_1, \dots, s_m$ . Нехай нам відома ймовірність  $P(s_1, \dots, s_m)$  появи всіх можливих послідовностей кубиків. Тоді ентропія Колмогорова визначається таким чином:

$$K_1 = \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{m \rightarrow \infty} \left[ \frac{1}{m\tau} \sum_{s_1, \dots, s_m} P(s_1, \dots, s_m) \ln P(s_1, \dots, s_m) \right]. \quad (17)$$

Характерний час, на який може бути передбачена поведінка системи, обернено пропорційний до ентропії Колмогорова. Якщо ентропія досягає нуля, то система стає повністю передбаченою. Так буде у разі регулярних процесів. Для істинно випадкових процесів ентропія необмежено велика. Ентропія системи у режимі дивного атрактора додатня, але має скінчене значення. Числове значення ентропії є кількісною характеристикою ступеня хаотичності системи.

Можна ввести поняття узагальненої ентропії Реньї як динамічного аналога узагальненої розмірності Реньї:

$$K_q = \lim_{\tau \rightarrow 0} \lim_{\varepsilon \rightarrow 0} \lim_{m \rightarrow \infty} \left[ \frac{1}{m\tau} \frac{1}{1-q} \ln \sum_{s_1, \dots, s_m} P^q(s_1, \dots, s_m) \right]. \quad (18)$$

В окремих випадках, при  $q=0$  одержуємо визначення топологічної ентропії  $K_0$ , при  $q=1$  – ентропії Колмогорова  $K_1$ , при  $q=2$  – кореляційної ентропії  $K_2$ .

### Спектр характеристичних показників Ляпунова

Ще однією важливою характеристикою атрактора є спектр характеристичних показників Ляпунова.

Нехай функція  $F: R^n \rightarrow R^n$  диференційовна,  $A$  – атрактор відображення  $x_{k+1} = F(x_k)$ .

Позначимо  $F^k$  суперпозицію  $F$  з собою  $k$  раз;  $DF^k(x)$  – матриця Якобі, тобто матриця  $n \times n$  частинних похідних  $F^k$ , обчислених у точці  $x$ . Нехай  $a_i(k, x)$  – модуль  $i$ -го власного значення  $DF^k(x)$ , причому  $a_1(k, x) \geq \dots \geq a_n(k, x)$ . Характеристичні показники Ляпунова атратора визначаються співвідношеннями:

$$\lambda_i = \lim_{k \rightarrow \infty} \frac{1}{k} \ln a_i(k, x). \quad (19)$$

При відповідних припущеннях ця границя існує і однакова для типових точок  $x$  на атраторі  $A$ . Подібне визначення може бути дане і для атратора автономної системи звичайних диференціальних рівнянь  $\dot{x} = F(x)$ . При цьому індекс  $k$  замінюється часом  $t$ . Впорядкована по убутанню послідовність чисел  $\lambda_1 \geq \lambda_2 \geq \lambda_3$  утворює спектр показників Ляпунова і дає корисну класифікацію атраторів.

Сума показників Ляпунова траєкторії  $x^0(t)$  характеризує швидкість зміни фазового об'єму в її околі. Режим дивного атратора реалізується тільки у дисипативних системах і характеризується наявністю у спектрі позитивних показників. Сума показників Ляпунова для дисипативних систем від'ємна. Якщо сума показників Ляпунова рівна нулю, то фазовий об'єм системи у часі не змінюється – система консервативна і атраторів не містить. У разі додатньої суми показників Ляпунова фазовий об'єм у часі наростає. З фізичної точки зору такий режим, як стаціонарний, нереальний.

Встановлений кількісний взаємозв'язок показників Ляпунова з ентропією Колмогорова. Доведено, що ентропія додатня у тому і лише у тому випадку, коли фазова траєкторія у середньому експоненціально нестійка на атраторі. Значить, спектр показників Ляпунова такої траєкторії має містити додатній показник. Явний вираз, що зв'язує ентропію Колмогорова з додатніми показниками Ляпунова, одержаний Песінім і в типових випадках виглядає достатньо просто:

$$K_1 = \sum_{\lambda_i > 0} \lambda_i, \quad (20)$$

тобто ентропія дорівнює сумі додатніх показників Ляпунова.

Для обчислення таких статистичних середніх, як розмірність, ентропія, спектр показників Ляпунова і інших характеристик атратора, необхідно мати безліч точок, що визначені у фазовому просторі розмірності і належать атратору. Число точок у розрахунках скінчене, але зобов'язане бути достатньо великим. Згідно формули

$$M \geq M_{\min} = 10^{2+0.4D}, \quad (21)$$

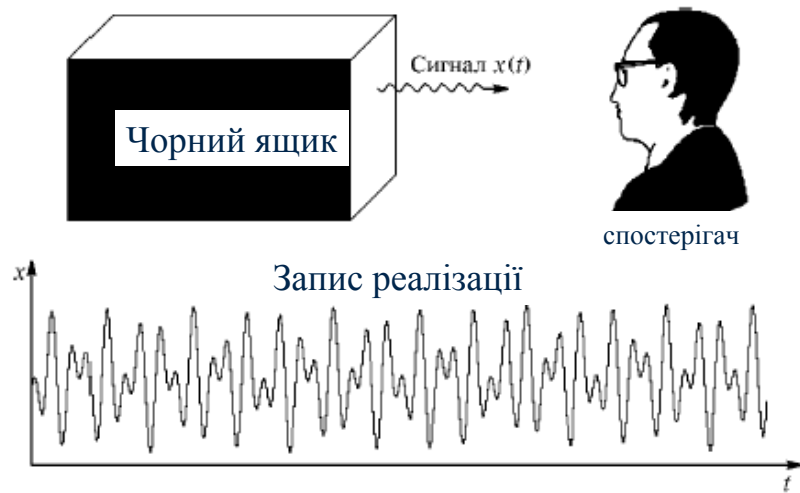
де  $D$  – розмірність атратора.

У разі, коли динамічна система задана дискретним оператором відображення, точки знаходяться автоматично після задання початкових умов. Якщо динамічна система задана системою диференціальних рівнянь, то у загальному випадку рішення може бути знайдене тільки чисельним інтегруванням системи на комп'ютері. Звичайно використовують метод Рунге-Кутта 4-го порядку, похибку задають, крок часу визначається конкретною системою і повинен бути вибраний порівняно з якнайменшим з її характерних часів.

Проте часто вимагається обчислити характеристики атратора деякої реальної системи, математична модель якої невідома. При цьому, як правило, невідома і розмірність її фазового простору. У цій ситуації ми маємо у своєму розпорядженні інформацію про поведінку у часі якої-небудь однієї з динамічних змінних. До того ж і інтервал часу експериментальної реалізації природно обмежений. Чи можна у таких умовах одержати

характеристики атрактора?

### Відновлення атрактора методом часової затримки координат (методом запізнення - delay-time reconstruction)



Шлях до розв'язання цієї проблеми був запропонований Такенсом. Ним доведено, що майже для всіх гладких динамічних систем по наявній часовій реалізації однієї спостережуваної динамічної змінної можна сконструювати новий атрактор, основні властивості якого будуть такими ж, як у початкового.

Нехай  $\epsilon$  часовий ряд експериментальних даних, що є відліками деякої фізичної величини:  $\{s_k\}_{k=0}^{M-1}$ . Якщо відомий крок за часом  $\Delta t$ , то і час  $t=k*\Delta t$ . Передбачається, що фізична величина  $s$  є однією із змінних динамічної системи. Система знаходиться у стаціонарному режимі, тобто фазова траєкторія проходить усередині атрактора.

Для відновлення атрактора Такенсом запропонований метод часової затримки координат. У  $n$ -мірному фазовому просторі будується послідовність точок вигляду:

$$x_k = (s_k, s_{k+\tau}, \dots, s_{k+(n-1)\tau}), \quad k = 0, \dots, m-1, \quad m = M - (n-1)\tau, \quad (22)$$

тут  $\tau$  – часова затримка,  $n$  – розмірність вкладення. Основний результат Такенса полягає у наступному. Якщо  $M \rightarrow \infty$ , то множина точок  $x_k \in R^n$  задає вкладення початкового атрактора майже при будь-якому виборі спостережуваної змінної, якщо  $n$  не менша подвоєної розмірності початкового атрактора.

Для оцінки характеристик реального досліджуваного атрактора можна обчислювати характеристики відновленого атрактора. З метою зменшення помилки, обумовленої скінченністю набору експериментальних точок  $\{s_k\}_{k=0}^{M-1}$ , необхідно проводити розрахунки при декількох різних значеннях  $M$  і  $n$  і добиватися незалежності одержуваних оцінок характеристик від  $M$  і  $n$  у межах заданої точності.

Для малих кроків за часом  $\Delta t$  значення  $s_k$  і  $s_{k+1}$  будуть близькими, тому велике значення приймає правильний вибір часової затримки  $\tau$ . Необхідно прагнути вибрати  $\tau$  так, щоб кореляція між  $s_k$  і  $s_{k+1}$  була по можливості мінімальною. Традиційний спосіб вибору часової затримки полягає в обчисленні автокореляційної функції часового ряду:

$$B(\tau) = \frac{1}{m} \sum_{k=0}^{m-1} (s_k - \bar{s})(s_{k+\tau} - \bar{s}), \quad m = M - \tau. \quad (23)$$

Затримка  $\tau$  вибирається рівною часу першого перетину нуля автокореляційною функцією. Другий спосіб заснований на обчисленні середньої взаємної інформації між двома вимірюваннями. Нехай дані дві множини вимірювань  $A$  і  $B$ . Взаємна інформація між елементом  $a_i$  множини  $A$  і елементом  $b_j$  множини  $B$  визначається як кількість інформації, яку мають вимірювання  $a_i$  і  $b_j$  по відношенню один до одного:

$$I_{a_i, b_j} = \ln \left[ \frac{P_{AB}(a_i, b_j)}{P_A(a_i)P_B(b_j)} \right]. \quad (24)$$

Якщо вимірювання незалежні, то взаємна інформація рівна нулю. Усреднюючи по всіх вимірюваннях, одержуємо:

$$I_{AB} = \sum_{a_i, b_j} P_{AB}(a_i, b_j) \ln \left[ \frac{P_{AB}(a_i, b_j)}{P_A(a_i)P_B(b_j)} \right]. \quad (25)$$

Замінюючи вимірювання  $a_i$  і  $b_j$  на  $s_k$  і  $s_{k+1}$  відповідно, одержуємо середню взаємну інформацію як функцію часової затримки  $\tau$ . Затримка  $\tau$  вибирається рівною часу першого мінімуму у взаємній інформації.

У разі модельних даних, коли нам відома розмірність  $n$  фазового простору динамічної системи і всі  $n$  координат кожної точки на аттракторі, кореляційну розмірність  $D_2$  аттрактора знаходять таким чином: розглянемо кореляційний інтеграл  $C(r)$ , який показує відносне число пар точок аттрактора, що знаходяться на відстані, не більшій  $r$ . Якщо виконується умова  $C(r) \propto r^{D_2}$ , то  $D_2$  вважають кореляційною розмірністю аттрактора. Прологарифмувавши останній вираз, отримуємо:

$$\ln C(r) \propto D_2 \ln r. \quad (26)$$

Звідки одержуємо оцінку розмірності аттрактора як тангенс кута нахилу прямої, що апроксимує графік кореляційного інтеграла  $C(r)$  у подвійному логарифмічному масштабі.

У разі експериментальних даних ми звичайно не знаємо розмірність фазового простору системи і маємо у своєму розпорядженні інформацію тільки про одну координату точок на аттракторі.

Тому всі розрахунки проводяться для декількох розмірностей  $N=1,2,3,\dots$  фазового простору. Для відновлення аттрактора використовується метод Такенса. При цьому кореляційна розмірність аттрактора  $D_2(n)$  спочатку зростає, але потім звичайно виходить на постійний рівень  $D_2(n) = D_2$ . Таким чином, одержують шукану кореляційну розмірність  $D_2$  аттрактора і оцінку розмірності фазового простору системи  $n \leq 2D_2 + 1$ . Якщо ж вихідний сигнал динамічної системи сильно зашумлений, то розмірність аттрактора постійно росте.

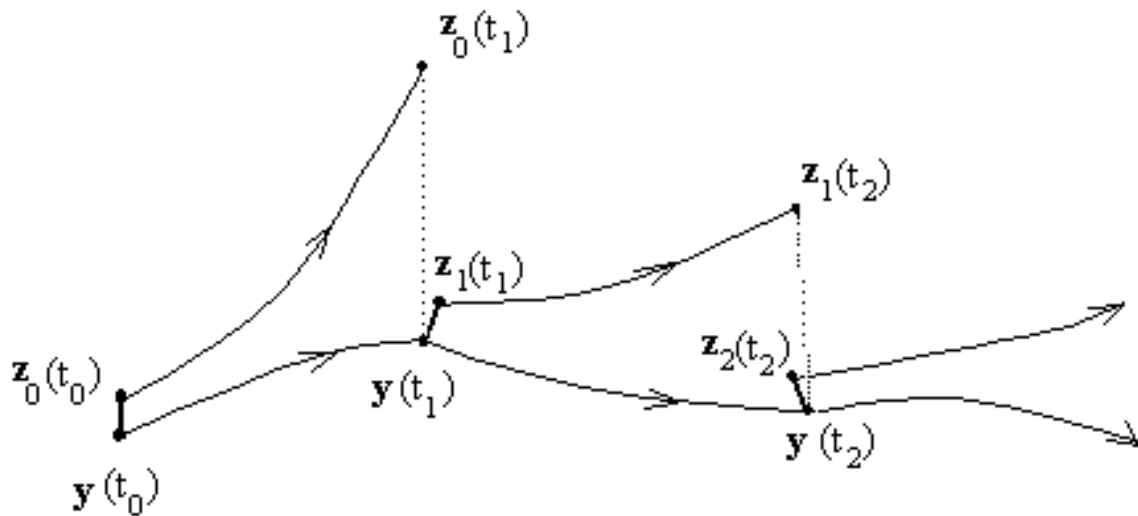
Кореляційна ентропія  $K_2$  може бути обчислена через кореляційний інтеграл, але розглядають не тільки його залежність від відстані  $r$ , але і від розмірності фазового простору  $n$ . При цьому вважають, що

$$C(r, n) \propto r^{D_2} \exp(-nK_2). \quad (27)$$

Звідки

$$K_2(r, n) = \ln \frac{C(r, n)}{C(r, n+1)}. \quad (28)$$

У 1985 році Вольф описав алгоритм знаходження найбільшого показника Ляпунова по часовому ряду даних. Алгоритм ґрунтується на спостереженні пари точок на атракторі.



Спостереження починається з першої точки даних  $y(t_0)$  і її найближчого сусіда  $z_0(t_0)$ , які знаходяться одна від іншої на відстані  $r_0$ . Ці дві точки зміщуються у часі з кроком  $\Delta t$ , поки відстань  $r'_0$  між ними не перевищить деяку величину  $\varepsilon$ . Перша точка даних  $y(t_1)$  зберігається, а новий сусід  $z_1(t_1)$  шукається такий, щоб відстань

$$r_1 = \|y(t_1) - z_1(t_1)\| \quad (29)$$

знову була меншою, ніж  $\varepsilon$ , і такою, що  $z_1(t_1)$  лежить так близько, наскільки це можливо, в тому ж напрямі, що й напрям від  $y(t_1)$  до  $z_0(t_1)$ .

Процедура продовжується до тих пір, поки прийнята за основу порівняння траєкторія не дійде до кінця часового ряду. Найбільший показник Ляпунова атрактора оцінюється як

$$\lambda_1 = \frac{1}{m \cdot \Delta t} \sum_{k=0}^{L-1} \ln \frac{r'_k}{r_k}, \quad (30)$$

де  $L$  – число кроків заміни і  $m$  – загальне число кроків за часом, протягом яких рухається траєкторія  $y$ .

### Хід роботи

Щоб зазначені нижче функції виконувалися в MatLab, необхідно через команду меню **File**→**Set Path** задати каталог з підкаталогами **OpenTSTOOL**.

Для початку необхідно завантажити в робочу область MatLab сигнал, який ви збираєтесь аналізувати, причому це має бути ряд прибутковостей. Здійснюється така операція за допомогою функції: **s = signal('ім'я файла','ascii')**.

Наприклад, **s = signal('xoi.txt','ascii')**. Або ж спочатку необхідно імпортувати ряд даних до робочої області, а потім набрати в Command Window: **s = signal(data)**.

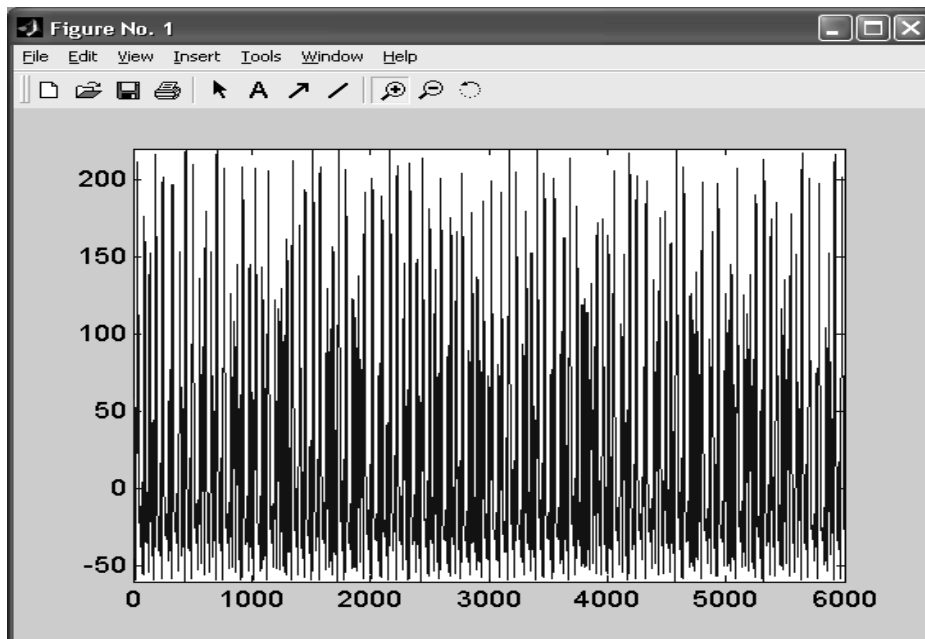
Результатом буде створення об'єкту сигналу **s** в робочій області і виведення в командному вікні інформації про сигнал:

Dlens : 5818

X-Axis 1 :

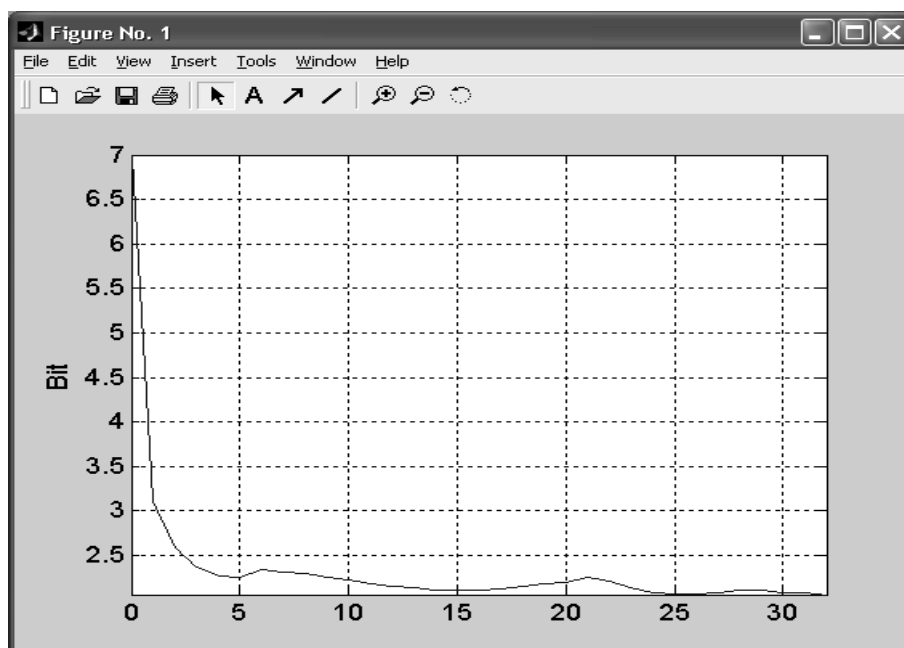
Name :  
Type :  
Attributes of data values :  
Comment :  
History :  
25-Nov-2006 17:03:08 : Imported from MATLAB workspace

Щоб зобразити сигнал, наберіть в командній строчці **view(s)**;



Далі визначається часова затримка як перший мінімум взаємної інформаційної функції (міра інформації Шеннона). Другий параметр в прикладі визначає кількість значень, які виведуться в розрахунку.

```
a = amutual(s,32);  
view(a);
```



Перший мінімум знаходиться в четвертому значенні, отже, це й буде часова затримка.

Для перевірки, чи дійсно ряд даних являється хаотичним, і визначення вкладеної розмірності треба задати наступну команду:

```
[E1, E2]=cao(s, maxdim, tau, NNR, Nref) .
```

s – вхідний сигнал;

maxdim – максимальна розмірність, до якої розглядається вкладення ряду;

tau – часова затримка;

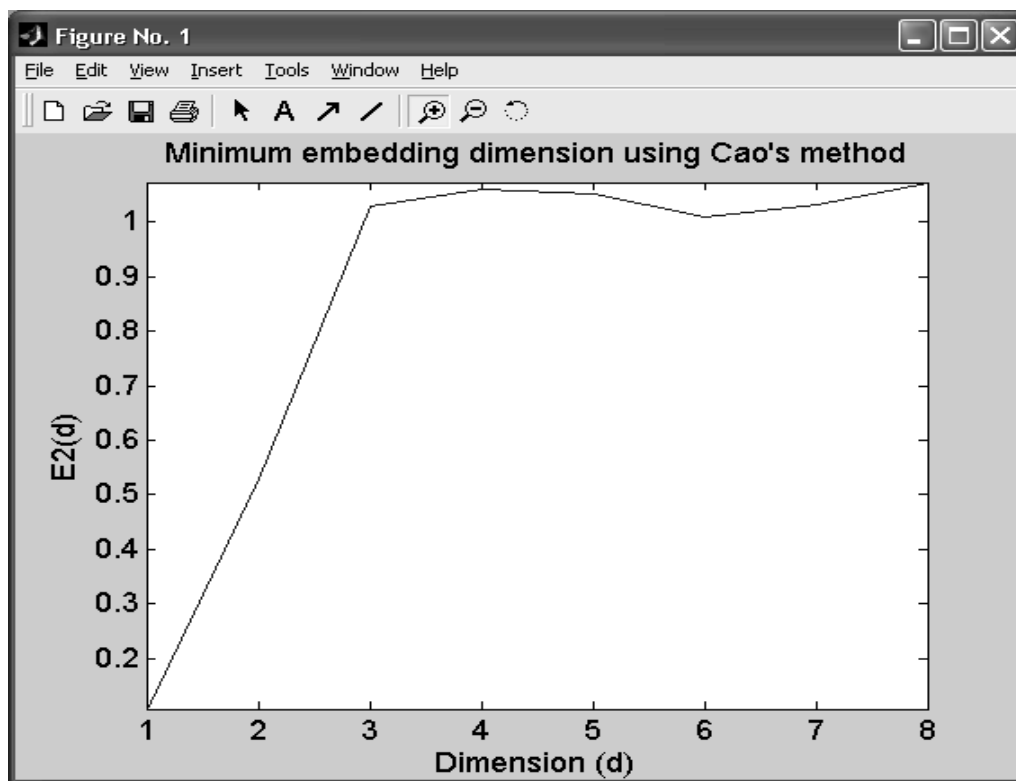
NNR – кількість найближчих сусідів;

Nref – кількість точок ряду даних, яка береться для аналізу (значення -1 – використовуються всі точки).

Наприклад:

```
[E1, E2]=cao(s,8,4,3,1000);
```

```
view(E2);
```



Сигнал є випадковим, якщо  $E2(d)$  наближене до одиниці при всіх значеннях  $d$ , тоді як хаотичний сигнал буде, якщо  $E2(d)$  приймає значення менше одиниці при малих значеннях  $d$ . Отже з отриманого значення видно, що вхідний сигнал являється хаотичним.

Команда `view(E1)` дасть зображення для пошуку вкладеної розмірності за методом Као.

Отже, графік функції  $E1(d)$  переходить в майже не змінюючу нахил пряму при досягненні значення  $d = 3$ . Значить мінімальною розмірністю вкладання для вхідного сигналу буде  $d = 3$ .

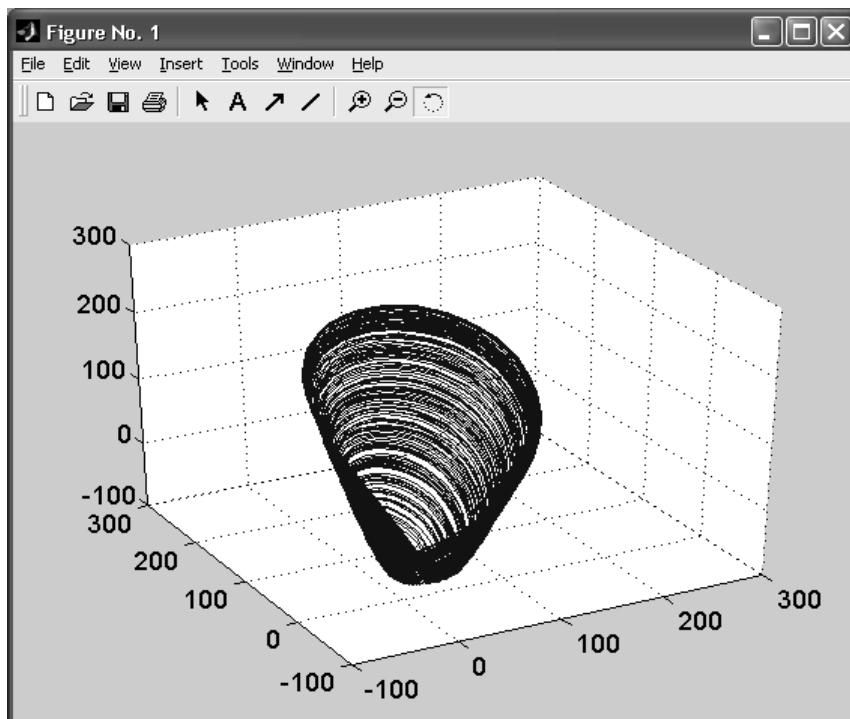
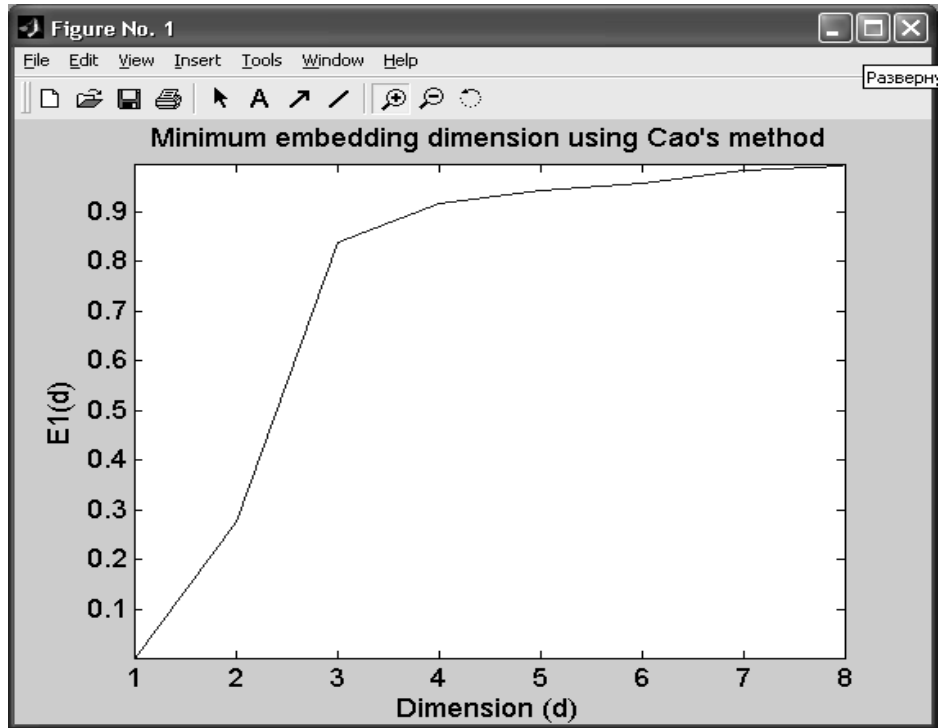
Зобразити графік вкладеного сигналу можна за допомогою наступних функцій

```
embed (s, dim, delay)
```

s - сигнал  
dim - вкладена розмірність  
delay – часова затримка

Наприклад:

```
e = embed(s, 3, 4);  
view(e);
```





Далі визначається кореляційна розмірність за допомогою функції **corrsum**:

**corrsum(pointset, query indices, range, exclude, bins)**

pointset – матриця, яка містить координати точок початкового сигналу у вкладеній розмірності;

query indices – індекси точок, на основі яких розраховується кореляційна розмірність (-1 - беруться до уваги всі точки)

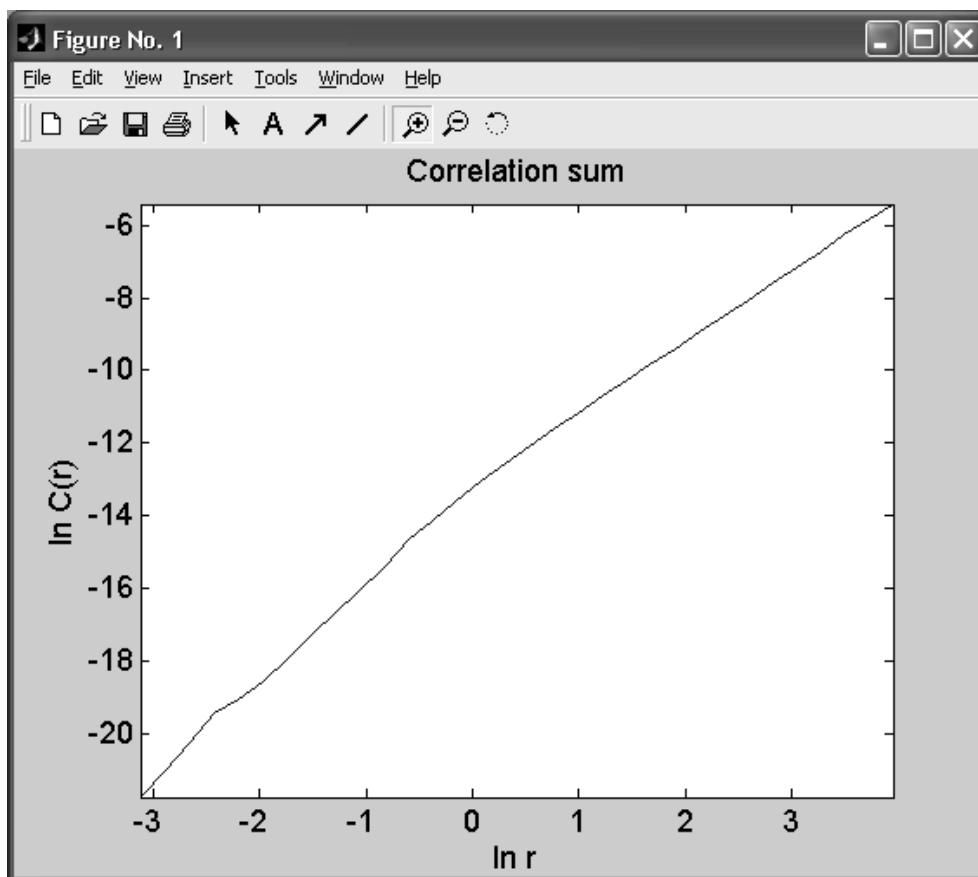
range – задається частина діаметру атарактора, в якій ведеться розрахунок кореляційної розмірності (задається від 0 до 1)

exclude – якщо точки матриці виходять за межі визначеної границі, їх індекси виключаються з розрахунків. Наприклад, якщо розглядається 124 точка, а exclude=3, то точки з індексами  $124-3=121$  і  $124+3=127$  виключаються з розрахунку. Якщо exclude = 0 точки виключаються з розрахунку автоматично, в ході розрахунків.

bins – значення дистанцій, для яких розраховується кореляційна сума

Наприклад:

```
view(corrsum(e, -1, 0.05, 40, 32));
```



Визначити нахил кривої кореляційної суми можна за допомогою оцінки Такенса (функція `takens_estimator`). Параметри в ній вводяться такі ж, як і для розрахунку кореляційної суми.

```
takens_estimator(e, -1, 0.05, 40)
```

Результат:

Using euclidian metric to calculated distances  
Number of reference points : 5993  
Upper bound for attractor size : 619.160418  
Maximal search radius : 15.479010

ans = 1.9150

Якщо отримане значення – ціле число, то аттрактор – «нормальний» геометричний об'єкт, такий як точка ( $D_C = 1$ ) або площина ( $D_C = 2$ ); якщо це значення не ціле, тоді аттрактор – «дивний» і система показує хаос.

Оцінюючи найбільшу експоненту Ляпунова, ми визначаємо горизонт прогнозу як проміжок часу, на якому траєкторії аттрактора ще не встигли розійтись:

```
largelyap(pointset, query indices, taumax, k exclude)
```

pointset – індекси точок, на основі яких проводиться розрахунок;

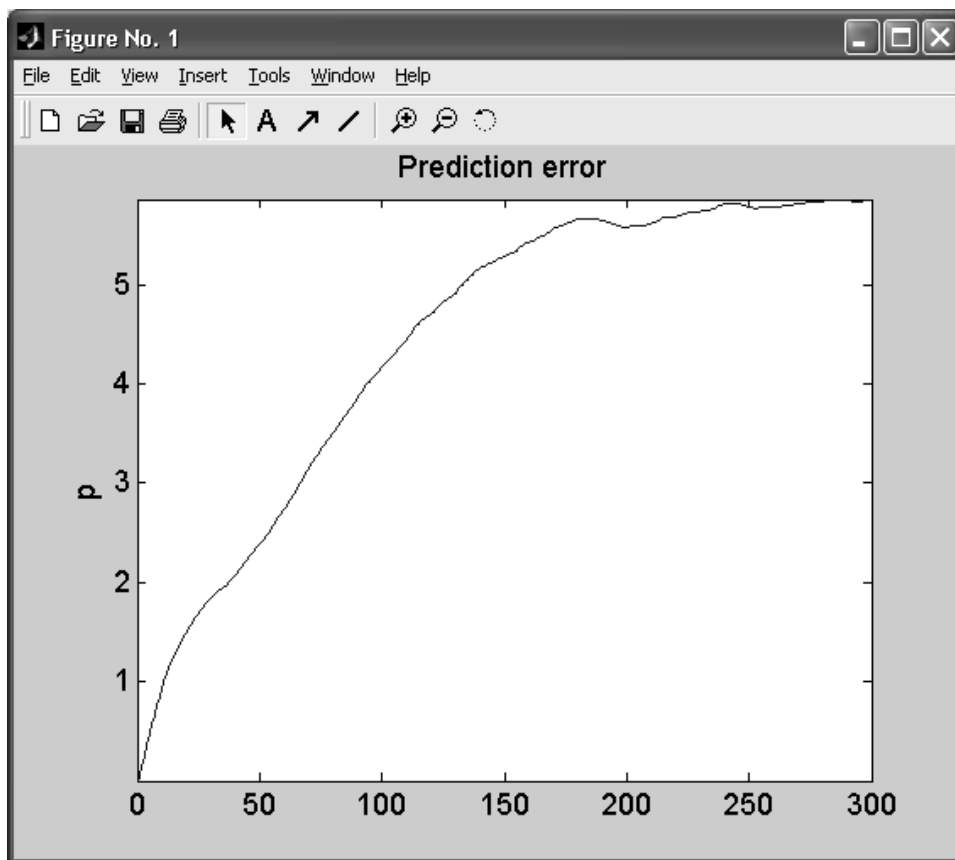
taumax – часова затримка;

k – кількість найближчих сусідів;

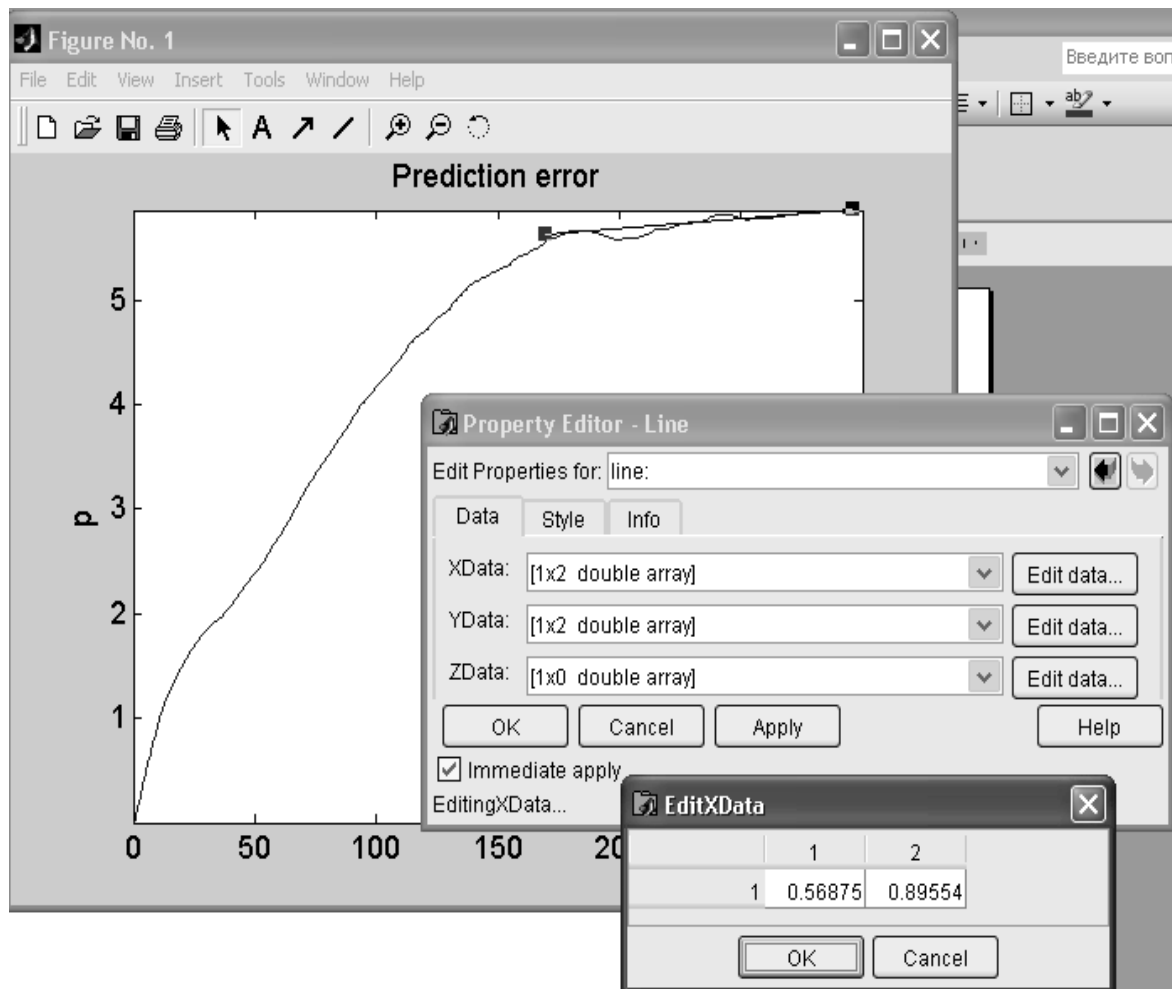
exclude – значення, які виключаються з розрахунку.

Візуалізація результатів проводиться за допомогою команди

```
view(largelyap(e, 1000, 300, 40, 2));
```



Для визначення часу прогнозу, на графіку для проміжку, де функція більш-менш вирівнюється, будується лінія тренду.



Для цієї лінії через вікно властивостей визначаються різниці між крайніми координатами x та y. Поділивши  $\Delta x$  на  $\Delta y$  ми визначимо горизонт прогнозу. Наприклад,  $(0,89-0,56)/(0,92-0,89) = 11$  днів – період прогнозування.

### Завдання для самостійної роботи

Дослідити нелінійні динамічні властивості вибраного економічного ряду.

### Контрольні питання

1. Поясніть, які відмінності хаотичного ряду динаміки від інших процесів.
2. Яким чином наявність або відсутність хаотичності може впливати на можливість прогнозування розвитку певного процесу?

## ЛАБОРАТОРНА РОБОТА № 19

**Тема.** Використання масштабно-залежних показників Ляпунова при дослідженні складних систем.

**Мета.** Навчитися застосовувати масштабно-залежні показники Ляпунова при аналізі об'єктів (систем) різної природи, заданих часовими рядами.

1. Задачі динаміки хаотичних систем
2. Показники Ляпунова
3. Масштабно-залежні показники Ляпунова (МЗПЛ) і міри складності
4. Інтегральна форма МЗПЛ, інтегральна міра складності
5. Використання мір складності при дослідженні кризових явищ

## Хід роботи

Повний опис динаміки хаотичних систем неможливий. Але можливим є опис типових властивостей.

Вводячи певні характеристики – ентропію, показники Ляпунова тощо, – отримують досить повну інформацію щодо властивостей динамічних систем.

В загальному випадку для дослідження поведінки в околі певної траєкторії використовуються показники Ляпунова (ПЛ).

Геометрично ПЛ характеризують ступінь розтягування або стиснення в системі вздовж стійких і нестійких напрямків.

Розглянемо дві близькі траєкторії  $x(t)$  та  $x_1(t) = x(t) + \varepsilon(t)$  (рис. 1).

Розглянемо функцію

$$\Omega(\varepsilon(0)) = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{|\varepsilon(t)|}{|\varepsilon(0)|}.$$

Тоді в залежності від напрямку вектора  $\varepsilon(0)$  функція  $\Omega$  буде приймати скінчений ряд значень  $\{\lambda_i\}$ ,  $i = 1, 2, \dots, n$ , які називаються показниками Ляпунова.

ПЛ являються мірою хаотичності, складності системи. Якщо є додатні показники, то поведінка динамічної системи буде хаотичною.

Подальші дослідження привели до поняття масштабно-залежних показників Ляпунова (Scale-Dependent Lyapunov Exponents, SDLE - МЗПЛ).

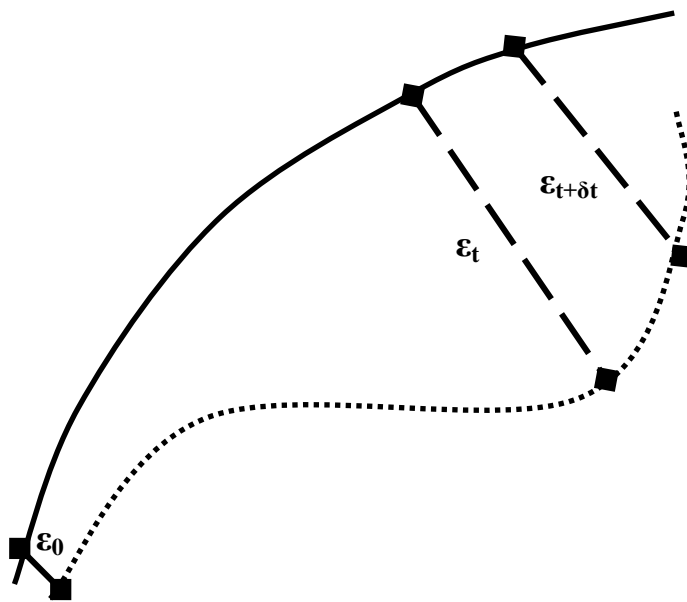


Рис. 1. Схематичне зображення двох довільних траєкторій в загальному високорозмірному просторі, з відстанями між траєкторіями в моменти часу 0, t, та t + delta t відповідно  $\varepsilon_0$ ,  $\varepsilon_t$ , та  $\varepsilon_{t+\delta t}$

МЗПЛ є поняттям, що впливає з концепції фазового простору великої розмірності. Припустимо, що задано тільки часовий ряд скалярних величин  $x[n] = x(1), x(2), \dots, x(n)$ . Можна реконструювати фазовий простір за допомогою технології вкладення лагових змінних.

Ця техніка є найбільш значним внеском теорії хаосу до практичного аналізу даних, тому що нетривіальні динамічні системи звичайно включають багато змінних стану і, отже,

повинні бути описані за допомогою багатовимірних станів простору (або фазового простору). Технологія вкладення полягає у створенні векторів виду:

$$V = [x(i), x(i + L), \dots, x(i + (m - 1)L)], i=1, \dots, N_p \quad (1)$$

де  $N_p = n - (m - 1)L$  – загальна кількість реконструйованих векторів,  $m$  – розмірність вкладення та  $L$  – час затримки. Значення параметрів  $m$ ,  $L$ ,  $N_p$  вибирається відповідно до певних критеріїв оптимізації.

Зокрема, значення  $L$  може бути визначене шляхом обчислення першого нуля автокореляційної функції, в той час як спільне визначення  $m$  та  $L$  може бути досягнуто за допомогою методу найближчого хибного сусіда.

Після належного реконструювання фазового простору розглянемо ансамбль траєкторій. Позначимо початкову відстань між двома близькими траєкторіями  $\varepsilon_0$ , а їх середню відстань в момент часу  $t$  та  $t + \Delta t$  для  $\varepsilon_t$  та  $\varepsilon_{t + \Delta t}$ , відповідно. Траєкторія поділу схематично показано на рис. 1. Ми можемо розглядати відношення між  $\varepsilon_t$  та  $\varepsilon_{t + \Delta t}$  при малих значеннях  $\Delta t$ . При  $\Delta t \rightarrow 0$ , маємо:

$$\varepsilon_{t + \Delta t} = \varepsilon_t e^{\lambda(\varepsilon_t) \Delta t}, \quad (2)$$

де  $\lambda(\varepsilon_t)$  є значенням залежного від масштабу показника Ляпунова (SDLE – МЗПЛ), що обчислюється таким чином:

$$\lambda(\varepsilon_t) = \frac{\ln \varepsilon_{t + \Delta t} - \ln \varepsilon_t}{\Delta t}. \quad (3)$$

Еквівалентно, ми можемо позначити цю величину, як  $d\varepsilon_t$

$$\frac{d\varepsilon_t}{dt} = \lambda(\varepsilon_t) \varepsilon_t dt. \quad (4)$$

При заданих даних часового ряду, найменше можливе значення  $\Delta t$  є часом дискретизації  $\tau$ .

Зауважимо, що класичний алгоритм обчислення показника Ляпунова  $\lambda_1$  базується на припущенні  $\varepsilon_t \sim \varepsilon_0 e^{\lambda_1 t}$  та оцінюванні  $\lambda_1$  як  $(\ln \varepsilon_t - \ln \varepsilon_0)/t$ . В залежності від  $\varepsilon_0$ , ця властивість може не виконуватись навіть для дійсно хаотичних систем. Це підкреслюється в схемі на рис. 1:  $\varepsilon_{t + \delta t}$  насправді може бути меншим, ніж  $\varepsilon_t$ . Великі труднощі з таким припущенням становить факт, що для будь-якого типу шуму  $\lambda_1$  завжди буде більше 0, що призводить до невірної оцінки сигналу як хаосу, а не як шуму. З іншого боку, рівняння (2) не включає ніяких припущень, крім того, що значення  $\Delta t$  є відносно малим.

Вказана процедура знаходження коефіцієнта Ляпунова дозволяє розділити різні типи сигналів:

- хаотичний сигнал:  $\lambda(\varepsilon) = const = \lambda_1$  ;
- хаотичний сигнал+стохастична складова:  $\lambda(\varepsilon) \propto -\gamma \ln \varepsilon$  ;
- фліккер-шуми (процеси  $1/f^\alpha$ ):  $\lambda(\varepsilon) \propto H \varepsilon^{-1/H}$  ;
- білий шум:  $\lambda(\varepsilon) = 0$ .

Як ми побачимо, хаос відповідає майже постійним значенням  $\lambda(\varepsilon)$  відносно діапазону  $\varepsilon$ .

Для обчислення SDLE, перевіряємо, чи виконується для пари векторів  $(V_i, V_i)$ , визначених рівнянням (1), наступна нерівність,

$$\varepsilon_k \leq \|V_i - V_j\| \leq \varepsilon_k + \Delta\varepsilon_k, k = 1, 2, 3, \dots, \quad (5)$$

де  $\varepsilon_k$  та  $\Delta\varepsilon_k$  є довільно вибрані невеликі значення відстаней, та

$$\|V_i - V_j\| = \sqrt{\sum_{w=1}^m (x_{i+(w-1)L} - x_{j+(w-1)L})^2}. \quad (6)$$

Геометрично нерівність (5) визначає оболонку у просторі високої розмірності (яка зводиться до кулі з радіусом  $\Delta\varepsilon_k$  при  $\varepsilon_k = 0$ ; в двовимірному випадку ця куля буде представлена колом, що описується рівнянням  $(x-a)^2 + (y-b)^2 = r^2$ , де  $(a, b)$  є центр кола, а  $r$  є його радіусом). Далі досліджуємо динаміку цих же пар векторів  $(V_i, V_j)$  всередині оболонки та здійснюємо усереднення по ансамблю за індексами  $i, j$ . Оскільки найбільшу цікавість викликають експоненціальні чи степеневі функції, припускаємо, що взяття логарифму та усереднення можуть бути змінені місцями, тому рівняння (3) матиме вигляд

$$\lambda(\varepsilon_t) = \frac{\langle \ln \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| - \ln \|V_{i+t} - V_{j+t}\| \rangle}{\Delta t}, \quad (7)$$

де  $t$  та  $\Delta t$  є цілими значеннями номеру виміру згідно інтервалу дискретизації, кутові дужки відповідають усередненню по індексам  $i, j$  всередині оболонки та

$$\varepsilon_t = \|V_{i+t+\Delta t} - V_{j+t+\Delta t}\| = \sqrt{\sum_{w=1}^m (x_{i+(w-1)L+t} - x_{j+(w-1)L+t})^2}. \quad (8)$$

Зазначимо, що початкова множина оболонок для обчислення SDLE грає роль початкових величин масштабів; в процесі еволюції динаміки скейли мають автоматично збігатись до величини властивого скейлу: такого, що визначають формули (3) і (4). Це питання буде більш яким після введення поняття характерного масштабу. Також відзначимо, що при аналізі хаотичного часового ряду при пошуку пар векторів всередині оболонки повинна виконуватись умова

$$|j - i| \geq t_{uncorrelated}, \quad (9)$$

де  $t_{uncorrelated}$  відповідає скейлу часу, за межами якого два вектори  $V_i$  та  $V_j$  більше не вважаються тангенціальними траєкторіями (тобто, поблизу орбітальних рухів, схожих на траєкторії двох авто, що прямують в одній смузі близько одне за одним) тієї ж траєкторії. Часто умови  $t_{uncorrelated} > (m - 1)L$  достатньо для ліквідації наслідків тангенціального руху та збіжності початкових скейлів до властивого скейлу.

Нарешті зауважимо, що

$$\Lambda(t) = \langle \ln \|V_{i+t} - V_{j+t}\| - \ln \|V_i - V_j\| \rangle \quad (10)$$

називається залежною від часу експоненціальною кривою. Оскільки  $\Lambda(t) = \ln \varepsilon_t - \ln \varepsilon_0$ , ми відразу ж бачимо, що SDLE відповідає локальному куту нахилу кривої виду:

$$\varepsilon_t = \varepsilon_0 e^{\Lambda(t)}. \quad (11)$$

Алгоритм обчислення SDLE є наступним:

(1). Довільно вибираємо параметри скейлу  $\varepsilon_k, \Delta\varepsilon_k, k = 1, 2, 3, \dots$ ; знаходимо значення  $m$

та  $L$  для реконструкції фазового простору зі скалярного часового ряду, використовуючи рівняння (1); також необхідно вибрати  $t_{uncorrelated}$ . Це є основні параметри для функції  $\lambda$  (http://www.gao.ece.ufl.edu/GCTH\_Wileybook/programs/lambda\_k\_curves/).

(2) Обчислюємо значення точок залежної від часу експоненти  $\Lambda(t)$ :

```
for i = 1:Np - tuncorrelated - Tmax
for j = i + tuncorrelated:NP - Tmax
check Inequality (5); if valid,
save  $\Lambda(t) = \ln V_{i+t} - V_{j+t}$ , for  $t = 0, 1, \dots, Tmax$ 
end
end
```

Фактично, у вкладених циклах по  $i$  та  $j$  відбувається пошук пари векторів, відстань між якими менша, ніж порогова для заданої оболонки (умова (5)). Внутрішнім циклом забезпечено те, що вектори знаходяться у часі не ближче, ніж  $t_{uncorrelated}$  відліків кроків дискретизації. Вектори, різниця між якими у часі менша  $t_{uncorrelated}$ , не будуть враховані, так як вважається, що це приблизно одна і та ж траєкторія.

При знаходженні пари векторів, що задовольняють умову (5), відбувається дослідження динаміки відстаней між парами векторів  $V_{i+t}$  та  $V_{j+t}$  відносно часу  $t$ . Фактично, для кожного значення  $t$  обчислюється вираз (10) та зберігається результат обчислення.

(3) Обчислюються локальні кути нахилу кривої SDLE  $\Lambda(k)$ . Зокрема, при значенні часу  $t = k\delta t$ , де  $\delta t$  є час дискретизації, параметр скейлу  $\varepsilon_t$  вибрано згідно з рівнянням (11), в той час як локальний кут нахилу кривої  $\Lambda(k)$  може бути визначено таким чином:

$$(\Lambda(k + 1) - \Lambda(k - 1)) / 2k\delta t .$$

Це рівносильно тому, що локальний кут нахилу може бути визначений на основі значення  $\ln \varepsilon_t$ , де  $\varepsilon_t$  задано рівнянням (11). Для покращення оцінки локального куту нахилу of  $\Lambda(k)$ , може бути використане фільтрування для позбавлення впливу локальних флуктуацій.

Іноді значно інформативнішою є інтегральна форма SDLE:

$$\ln \varepsilon_t = \ln \varepsilon_0 + \int_0^t \lambda(\varepsilon_t) dt .$$

### Хід роботи

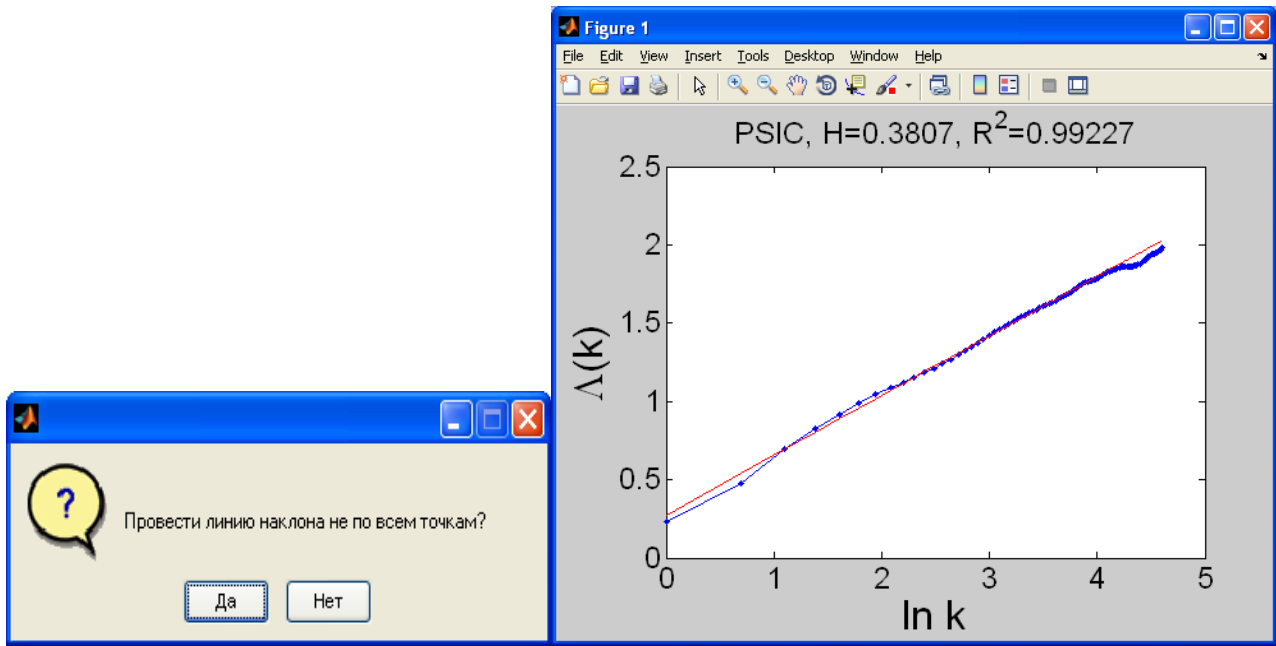
На першій стадії виконаємо розрахунки  $\Lambda(k)$ ,  $\lambda(\varepsilon)$ ,  $\ln \varepsilon_t$  для заданого часового ряду. Для цього використаємо програму **lambdademo.m**, що входить до пакету SDLE.

Візьмемо для прикладу ряд щоденних значень індексу дах з 1993 року (приблизно 5000 точок). У програмі **lambdademo.m** задамо параметри:

- рядок 33 – ім'я файла,
- 35-38 – параметри розрахунку,
- 42-43 – число оболонок і максимальне значення.

Буде розраховано  $\Lambda(k)$ ,  $\lambda(\varepsilon)$ ,  $\ln \varepsilon_t$ . При цьому для графіків, де можна провести лінію тренду, пропонується можливість це зробити (рис. 2).

**!!! Самостійно! Порівняйте результати з такими ж для перемішаного ряду. Зробіть висновки!**

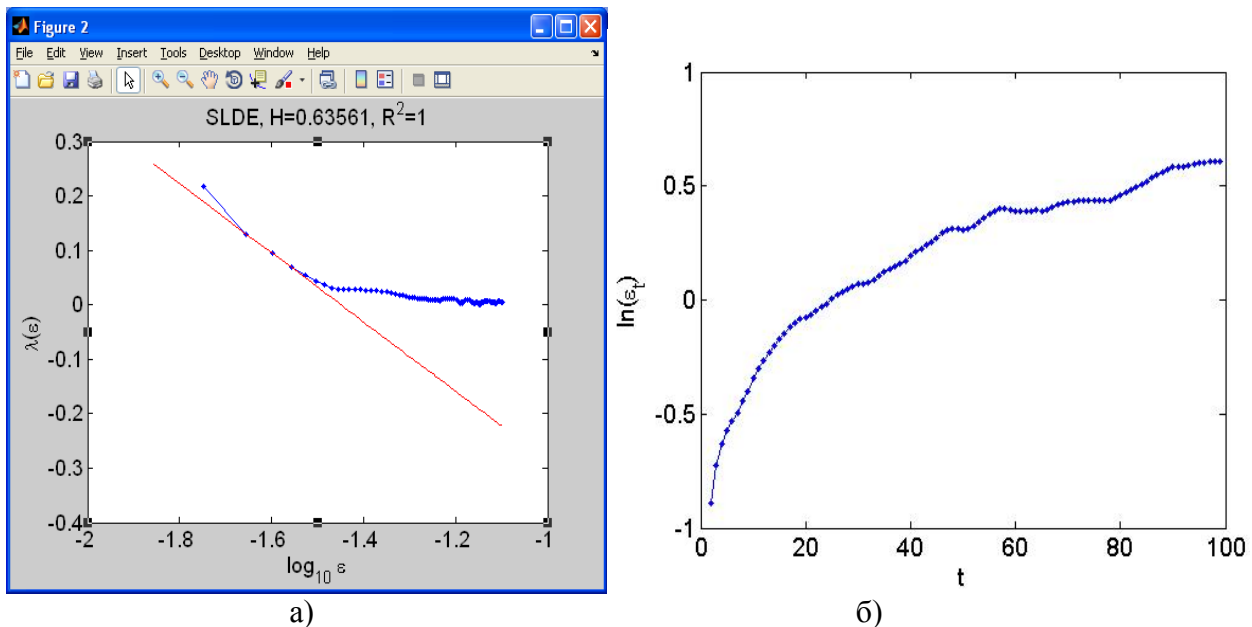


а) б)

Рис. 2. Послідовність визначення коефіцієнта Ляпунова (а-б),

Наступна частина роботи проводиться з віконними показниками. При цьому використовується продукт **SLDEinwindow.m**

Вибравши необхідні параметри, одержимо можливість зберегти у вигляді текстових файлів та графіків ряд віконних мір складності. При цьому зберігаються наступні віконні міри складності:  $\Delta \varepsilon$ ,  $\Delta \lambda$  (рис. 3а, розкид значень  $\varepsilon$  та  $\lambda$ ), а також віконна інтегральна міра **int compl** (площа під кривою на рис. 3б).



а) б)

Рис. 3. До розрахунку віконних мір складності (а)  $\Delta \varepsilon$ ,  $\Delta \lambda$  та інтегральної міри складності (б)

Дані міри можна співставити з вихідним рядом у відносному масштабі за допомогою програмного засобу **work2** та порівняти динаміку самого ряду і відповідних мір складності.

На рис. 4 представлені вказані міри для індексу фондового індексу Німеччини dax з 1993 року.



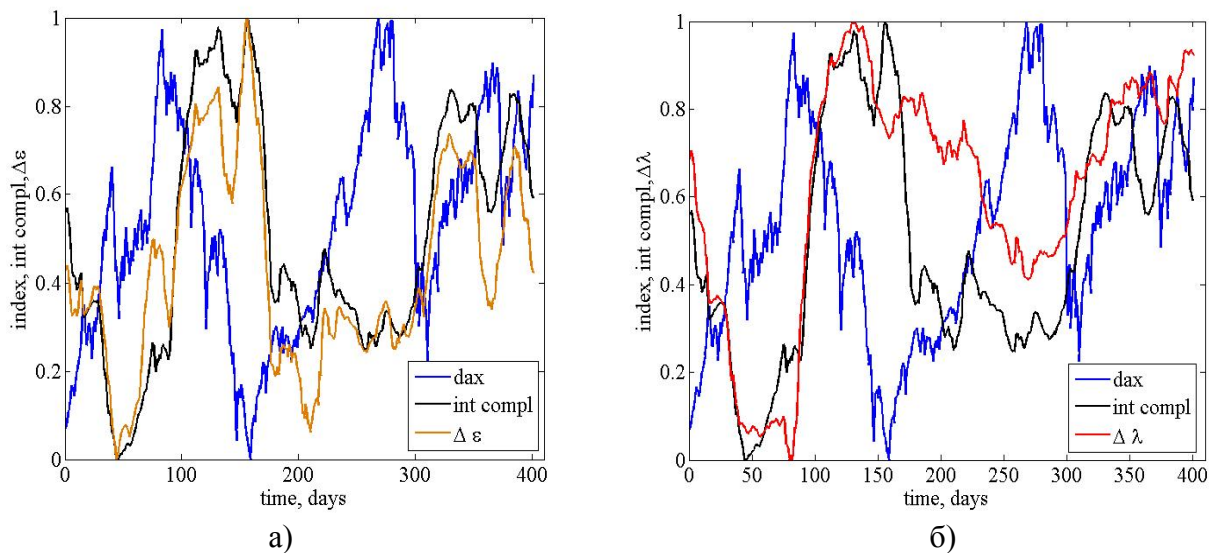


Рис. 4. Порівняльна динаміка нелінійних мір складності з динамікою вихідного часового ряду

З останніх рисунків видно, що у передкризові періоди складність системи падає, зростаючи у періоди власне кризи. Це означає, що можна будувати індикатори-передвісники, використовуючи введені міри складності.

### Завдання для самостійної роботи

1. Виберіть часовий ряд фондового індексу за період з 01.01.2004р.
2. В якості білого шуму візьміть перемішаний ряд.
3. Для дослідження детермінованого хаосу оберіть тестовий сигнал з відомим старшим коефіцієнтом Ляпунова (відомі атрактори: Ляпунова, Рюєля, логістичне відображення тощо).
4. Віконну динаміку мір складності дослідіть для фондового індексу з параметрами:  $wind=500$ ,  $step=25$ .
5. Поясніть результати, виявіть, яка з мір, на вашу думку, є кращою.

## ЛАБОРАТОРНА РОБОТА № 20

**Тема.** Дослідження процесів синхронізації в часових рядах.

**Мета.** Навчитись розраховувати значення різних мір синхронізації часових рядів та досліджувати їх динаміку при дослідженні складних сигналів.

### Теоретичні відомості

#### Синхронізація в складних системах

Явище синхронізації широко розповсюджене в науці, техніці і суспільстві. Тенденція до синхронної поведінки спостерігається в різних за суттю системах, таких як годинник, лазери, цвіркуни, пейсмейкери серця, нейрони, електронні генератори та аплодуючі глядачі в залі. Такі ефекти є універсальними; їх можна пояснити у рамках єдиного підходу, що ґрунтується на сучасних досягненнях нелінійної динаміки.

Постає питання про міру узгодженості двох сигналів, заданих дискретно. Однією з таких мір можна вважати синхронізацію.

Виявляють 2 види синхронізації: **повна синхронізація** сигналів (повна відповідність цих сигналів), **амплітудна синхронізація** (узгодженість амплітуд сигналів при рівномірній

зміні різниці фаз), та **фазова синхронізація** (узгодженість фаз сигналів при довільній амплітуді).

Для визначення фази сигналів використовується перетворення Гільберта, яке застосовується у фізиці. Комплексна частина дискретного перетворення Гільберта дає фазу дискретного сигналу

$$S(t) = s(t) + is_H(t) = A(t)e^{i\varphi(t)}, \quad s_H(t) = \pi^{-1} P.V. \int_{-\infty}^{\infty} \frac{s(\tau)}{t - \tau} d\tau,$$

де  $P.V.$  – позначає головне значення інтегралу Коші.

При практичному застосуванні цього методу визначення фази дискретного сигналу, з'являються проблеми з так званими нестационарними рядами. В таких рядах фаза за вищерозглянутим методом знаходиться невірно.

Для подолання цих труднощів, будемо використовувати генералізовану синхронізацію. Широка концепція генералізованої синхронізації була вперше запропонована в 1995 в роботі Рулкова. На основі цієї концепції розроблені методи визначення генералізованої синхронізації в експериментальних часових рядах, такі як: Amhold (1999), Quian Quiroga (2002), Rulkov (1995), Schiff (1996). Ці методи ґрунтуються на ідеї відновлення аттрактора методом часової затримки координат, запропонованої Такенсом (1981). Цей метод відновлює простір станів, використовуючи  $m$ -розмірні вектори, компонентами яких є послідовні значення сигналів  $X_n = \{x_n, x_{n+\tau}, \dots, x_{n+(m-1)\tau}\}$ ,  $Y_n = \{y_n, y_{n+\tau}, \dots, y_{n+(m-1)\tau}\}$ ,  $m$  – розмірність фазового простору,  $\tau$  – часовий проміжок. Якість реконструкції простору станів залежить від вибору показників  $m$  та  $\tau$ . За теоремою Такенса, розмірність  $m$  повинна бути достатньо великою, щоб визначити аттрактор, але обмеженою, щоб забезпечити швидкість обчислень. Нами реалізовано геометричний метод визначення  $m$ . Ця процедура базується на мінімізації числа хибних сусідів (як і у методі рекурентних діаграм). Параметр затримки  $\tau$  було обрано за домовленістю, як перший нуль автокореляційної функції сигналів.

За означенням, генералізована синхронізація існує між двома динамічними системами  $X$  і  $Y$  коли стан залежної системи  $Y$  є функцією стану системи  $X$ :  $Y=F(X)$ . Якщо  $F$  неперервна, дві точки, що лежать близько одна до одної на аттракторі  $X$ , мають відповідати 2-м близьким точкам на аттракторі  $Y$  (рис. 1). Важливою рисою генералізованої синхронізації є те, що 2 синхронізовані часові ряди можуть мати дуже несхожий вигляд. Це є перевагою цього методу в порівнянні з іншими, в яких вимагається хоча б деяка схожість поведінки часових рядів для з'ясування взаємодії.

### Міри синхронізації. Фазова синхронізація

Фазова синхронізація традиційно визначається як *захоплення фаз* ( $\varphi_x(t) - \varphi_y(t) = \text{const}$ ) або, у випадку хаотичних систем, *зчеплення фаз* ( $\varphi_x(t) - \varphi_y(t) < \text{const}$ ) двох коливних сигналів.

В подальшому будемо використовувати *усереднену фазову когеренцію*:

$$R = \left| \frac{1}{N} \sum_{j=1}^N e^{i[\varphi_x(t_j) - \varphi_y(t_j)]} \right|,$$

де  $\varphi_x(t)$  – функція фази сигналу  $x$  у часі, а  $\varphi_y(t)$  – функція фази сигналу  $y$  у часі.

Середня фазова когеренція обчислювалась усередненням по  $j$  на вікні довжини  $N$  з переміщенням вікна у часі. В результаті можна досліджувати динаміку синхронності у часі.

Для обчислення фазової синхронізації двох сигналів  $y(t)$  і  $x(t)$  у часі  $\varphi(t) = \frac{s_1(t)}{s(t)}$

необхідно визначити їх фази  $\varphi_y(t)$  і  $\varphi_x(t)$ . Для цього ми використали методикку, запропоновану

в теорії аналітичного сигналу. Миттєва фаза визначається так:

$$s_1(t) = \frac{1}{\pi} P.V. \int_{-\infty}^{+\infty} \frac{s(t')}{t-t'} dt',$$

де  $s_1(t)$  обчислюється за допомогою *перетворення Гільберта*.

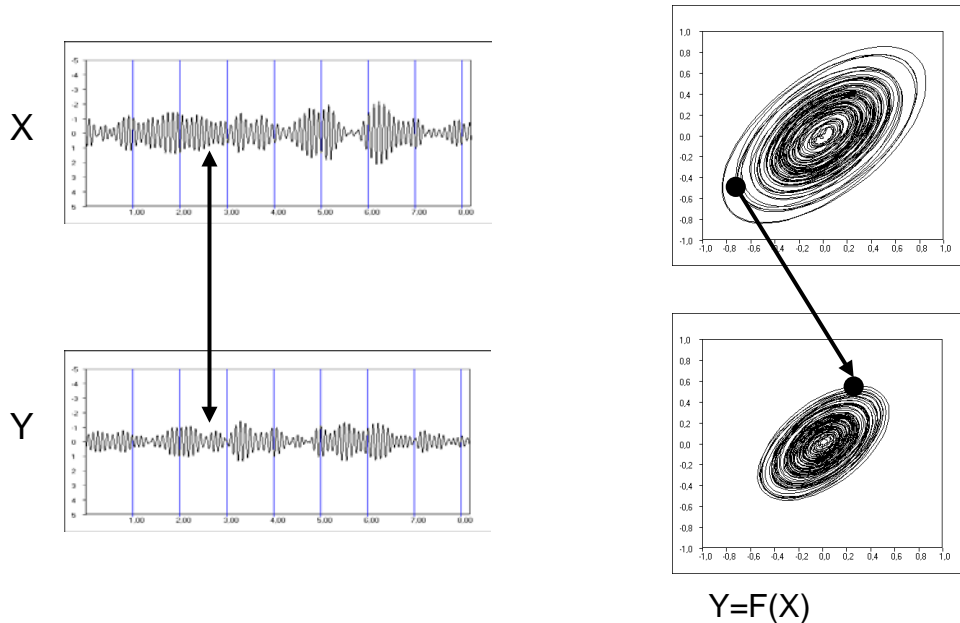


Рис. 1. Схема генералізованої синхронізації

### Міра синхронізації між максимумами сигналів

При застосуванні вищезгаданих методик до реальних сигналів, які містять в собі як шум, так і коливання в широкому діапазоні, виникають ряд проблем, для вирішення яких необхідно здійснити попередню фільтрацію даних по певній частоті. Метод, про який ідеться далі, має можливість вирішити ці проблеми без застосування фільтрації. Цей метод полягає у відтворенні фази сигналу по сусіднім максимальним точкам (пікам коливань). Точки максимумів вважаються точками, в яких фаза кратна  $2\pi$ . Алгоритм аналізу наступний: знаходяться локальні точки максимумів сигналів (максимум на інтервалі вибраної довжини  $w$ ). Ця довжина дає можливість налаштуватись на певну частоту, яка є найбільш інформативною в сигналі. Оскільки на інтервалі  $w$  має бути тільки один максимум, то пульсації високої частоти не впливають на фазу сигналу.

По отриманій таким чином фазі обчислюється середня фазова когеренція, яка і вважається мірою синхронізованості сигналу.

### Хід роботи

Для аналізу часових рядів на синхронізацію будемо користуватись двома програмними продуктами, які знаходяться, відповідно, у папках `c:\matlab6p5\work\sinchron` та `c:\matlab6p5\work\Synchro`. Вони використовують дещо різні види кількісних показників синхронізації. Тому розглянемо їх окремо.

Запустіть матлаб та перейдіть у папку з пакетом **sinchron**, або вкажіть шлях до пакету. Для запуску скористайтесь командою **start**. Після успішного запуску Ви побачите робоче вікно пакету **sinchron** (рис.2).

Пакет дає можливість завантажити 2 сигнали для аналізу їх синхронності. Є можливість налаштувати деякі параметри обчислення:

- довжина рухомого вікна, в якому обчислюються міри синхронізації;
- крок, з яким дані пропускаються для швидкості обчислень.

Довжина вікна для алгоритму синхронізації максимумів. Ця довжина впливає на частоту піків, по якій аналізується синхронність сигналів.

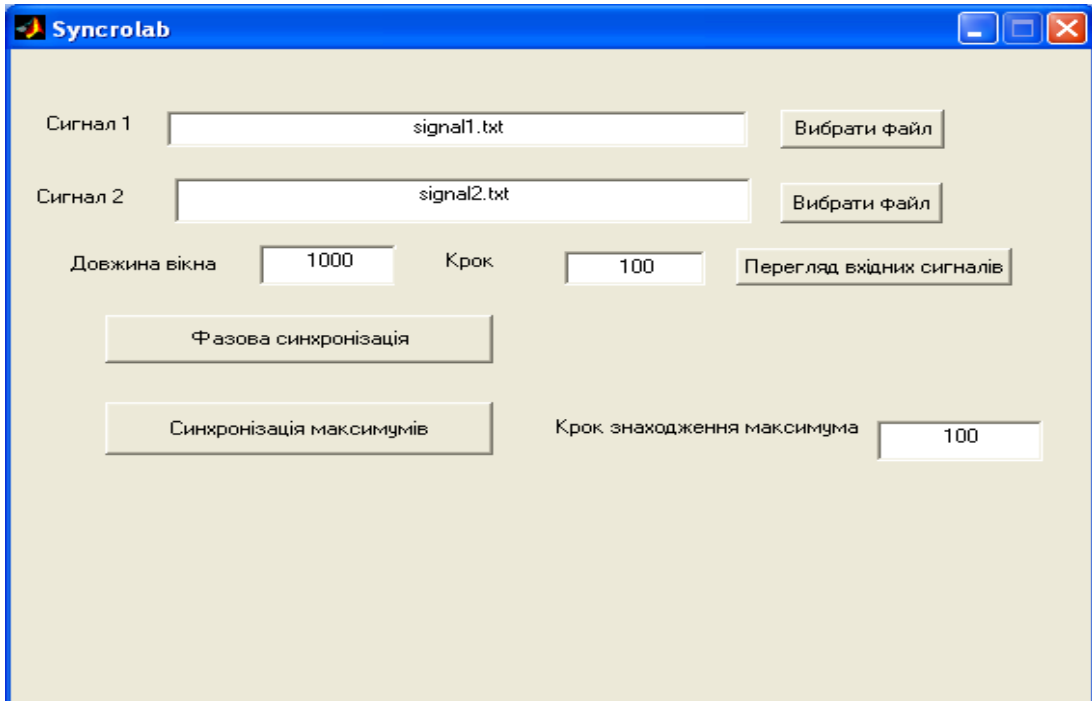


Рис. 2. Робоче вікно пакету **synchron**

Завантажити дані можна за допомогою кнопок «Обзор» для відповідних рядів. Виберіть пару рядів та завантажте їх в пакет. Для перегляду виду сигналів можна скористатися кнопкою «Перегляд», яка відобразить сигнали на графіку (рис. 3).

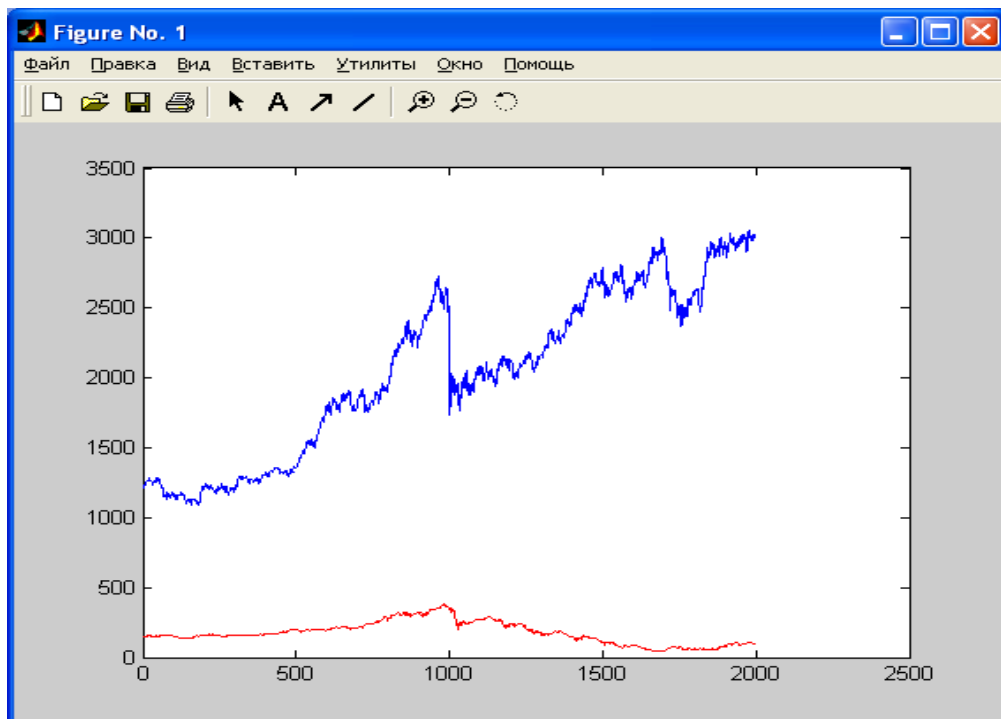


Рис. 3 Вигляд сигналів, для яких проводиться аналіз синхронізаційних властивостей

Для аналізу синхронізації натисніть кнопку, яка відповідає виду синхронізації. Наприклад, «Фазова синхронізація» або «Синхронізація максимумів».

Для дослідження процесів синхронізації візьмемо динамічні ряди, які містять кризу. Довжина рядів повинна співпадати. Наприклад, на рисунках 4-5 наведені приклади розрахунків фазової синхронізації та синхронізації максимумів.

Мета пошукової роботи в тому, щоб проаналізувати синхронізацію між різними парами кризових даних з метою прогнозування та попередження цих небажаних явищ.

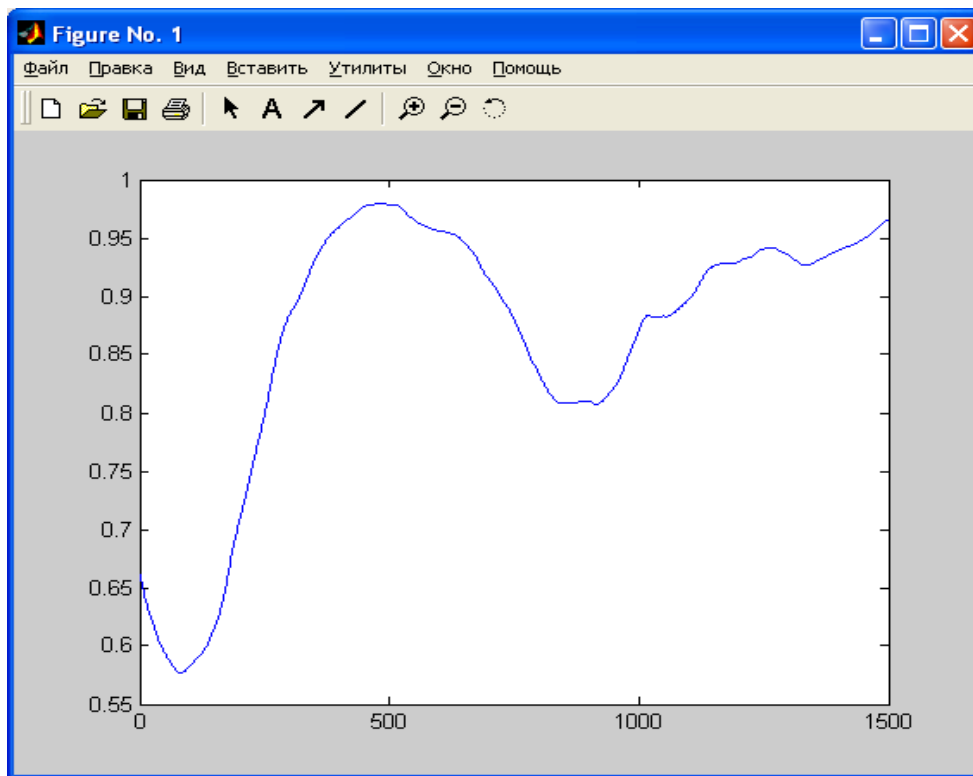


Рис. 4 Фазова синхронізація між кризами dj1929 та dj 1987

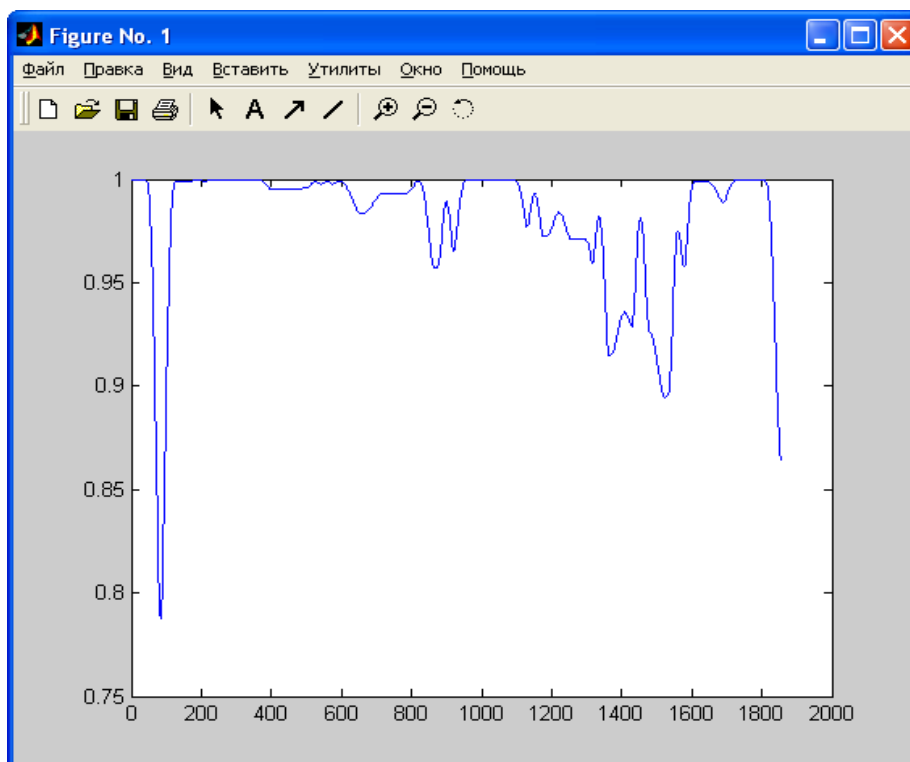


Рис. 5 Синхронізація максимумів вибраних рядів

При виконанні роботи необхідно досліджувати синхронізацію обома реалізованими методами. Як було зазначено вище, метод синхронізації максимумів вимагає налаштування розміру вікна, на якому вибирається максимум. При дослідженні здійсніть перегляд сигналу та виберіть цей розмір в залежності від періоду найбільш інформативних коливань в сигналі. Розмір вікна має бути приблизно рівним півперіоду коливань.

Пакет **Synchro** використовує лінійні і нелінійні міри синхронізації. У якості лінійної міри використовується функція автокореляції. Нехай ми маємо два незалежно знайдені дискретні часові ряди  $x_n$  і  $y_n, n = 1, \dots, N$ . Тоді функцію автокореляції можна визначити з виразу

$$c_{xy}(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \left( \frac{x_i - \bar{x}}{\sigma_x} \right) \left( \frac{y_{i+\tau} - \bar{y}}{\sigma_y} \right),$$

де  $\bar{x}$  і  $\sigma_x$  позначені середнє і варіація, а  $\tau$  – часовий лаг. Автокореляція дає міру лінійної синхронізації між  $x$  і  $y$ . Її абсолютне значення знаходиться на проміжку від нуля (синхронізація відсутня) до одиниці (максимальна синхронізація). При цьому функція симетрична:  $c_{xy}(\tau) = c_{yx}(\tau)$ .

Щодо нелінійної взаємозалежності існують декілька мір. Нехай з часових рядів, які характеризують дві системи  $x$  і  $y$ , реконструйовані вектори затримки  $x_n = (x_n, \dots, x_{n-(m-1)\tau})$  і  $y_n = (y_n, \dots, y_{n-(m-1)\tau})$ , де  $n=1, \dots, N$ ,  $m$  – розмірність вкладення, а  $\tau$  визначає величину часового лагу. Позначимо далі  $r_{n,j}$  і  $s_{n,j}, j = 1, \dots, k$ , часові індекси  $k$  найближчих сусідів  $x_n$  і  $y_n$ , відповідно.

Для кожного  $x_n$ , середнє квадратичне Евклідової відстані до його  $k$  сусідів визначене, як

$$R_n^{(k)}(X) = \frac{1}{k} \sum_{j=1}^k (x_n - x_{r_{n,j}})^2,$$

а  $Y$  – обумовлена середня квадратична Евклідова відстань, що визначається шляхом заміни найближчих сусідів еквівалентними за часом найближчими сусідами  $y_n$

$$R_n^{(k)}(XY) = \frac{1}{k} \sum_{j=1}^k (x_n - x_{s_{nj}})^2.$$

Якщо скупчення точок  $\{x_n\}$  має середній радіус  $R(X) = (1/N) \sum_{n=1}^N R_n^{(N-1)}(X)$ , то  $R_n^{(k)}(XY) \approx R_n^{(k)}(X) \ll R(X)$ , якщо системи сильно корельовані, і  $R_n^{(k)}(XY) \approx R(X) \gg R_n^{(k)}(X)$ , якщо вони незалежні. Відповідно ми можемо визначити міру взаємозалежності  $S^{(k)}(XY)$  як

$$S^{(k)}(XY) = \frac{1}{N} \sum_{n=1}^N \frac{R_n^{(k)}(X)}{R_n^{(k)}(XY)}.$$

Оскільки  $R_n^{(k)}(XY) \geq R_n^{(k)}(X)$  за побудовою, маємо

$$0 < S^{(k)}(XY) \leq 1.$$

Малі значення  $S^{(k)}(XY)$  вказують на незалежність між  $X$  і  $Y$ , тоді як великі значення вказують на синхронізацію (досягаючи максимуму при  $S^{(k)}(XY) \rightarrow 1$ ).

Можна ввести і іншу міру нелінійної взаємозалежності

$$H^{(k)}(XY) = \frac{1}{N} \sum_{n=1}^N \log \frac{R_n^{(k)}(X)}{R_n^{(k)}(XY)}.$$

Ця міра є нульовою, якщо  $X$  і  $Y$  цілком незалежні, при цьому додатною, якщо близькість в  $Y$  має на увазі також близькість в  $X$  для рівних у часі партнерів. Вона буде від'ємною, якщо близькі пари в  $Y$  будуть відповідати переважно віддаленим парам в  $X$ .

Ще однією нормалізованою (на відміну від  $S$ ) мірою синхронізації є величина

$$N^{(k)}(XY) = \frac{1}{N} \sum_{n=1}^N \frac{R_n(X) - R_n^{(k)}(XY)}{R_n(X)}.$$

Вона (як і  $H$ ) може бути слабо негативною і, як правило, більш стійка міра, ніж  $S$ .

Протилежні взаємозалежності  $S(YX), H(YX)$  і  $N(YX)$  визначаються за аналогією і, взагалі кажучи, не дорівнюють  $S(XY), H(XY)$  і  $N(XY)$ , відповідно. Вказана асиметрія – їх головна перевага над іншими нелінійними мірами, зокрема і фазовою синхронізацією, описаною в першій частині роботи. Ця асиметрія може надати інформацію відносно різних аспектів динамічних властивостей кожного з об'єктів.

У вікні команд наберемо команду **synchro\_gui2**. У відповідних вікнах інтерфейсу користувача (рис. 6) вибираємо часові ряди для аналізу і параметри розрахунку.

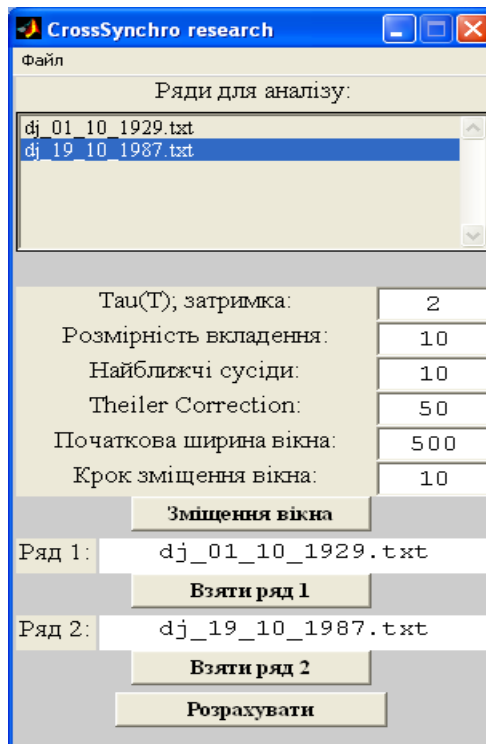


Рис. 6 Інтерфейс пакету **Synchro** для вибраних кризових рядів

**Увага!** Особливістю розрахунку вказаних мір є вимога до вхідних даних, які не повинні бути великими за модулем числами. Тому, якщо виникають проблеми, треба

працювати не з оригінальними, а з нормалізованими даними. Нормалізацію легко реалізувати за допомогою відомого пакету **work2**.

В результаті розрахунків параметрів, які відповідають рис. 6, будуть сім графіків:  $S(YX)$ ,  $H(YX)$ ,  $N(YX)$ ,  $S(XY)$ ,  $H(XY)$ ,  $N(XY)$  і  $C$  (автокореляція), аналіз загальних і частинних залежностей яких для різних сигналів і входить до одного із завдань роботи.

### Контрольні запитання

1. Поясніть природу явища синхронізації в складних системах різної природи.
2. Які міри синхронізації ви знаєте? На яких принципах вони побудовані?
3. Дайте інтерпретацію одержаним результатам. Чи можна використовувати міри синхронізації для побудови передвісників кризових явищ? Якщо можна, то які з них?

## ЛАБОРАТОРНА РОБОТА № 21

**Тема:** Агентне моделювання фінансових ринків.

**Мета:** Ознайомитись з програмним пакетом AgentsCL, дослідити економічний процес на прикладі випадкової торгівлі.

1. Концепція агентного моделювання.
2. Моделі агентів.

### Теоретичні відомості

Концепція агентного моделювання є однією з найсучасніших в моделюванні ринків. Традиційні кібернетичні моделі використовують концепцію «чорної скриньки», для якої нам відомі вхідні та вихідні дані, але процес всередині системи невідомий. Моделювання ринку на мікрорівні дозволяє зазирнути всередину «чорної скриньки» та більш детально дослідити механізми роботи ринку.

Під агентами розуміють об'єкти, які співіснують між собою в певному середовищі та взаємодіють між собою задля досягнення певної мети. В даній лабораторній роботі під агентами будемо розуміти учасників біржових торгів – трейдерів.

Агенти у подальшому – це комп'ютерні програми або підпрограми, які автоматично виконують дії, торгуючи на штучному ринку, тобто, продають чи купують активи. В першій роботі ми будемо використовувати випадкові агенти, які випадковим чином подають заявки на купівлю чи продаж, в наступних роботах буде здійснене порівняння результатів торгівлі випадкових трейдерів з моделюванням більш інтелектуальних стратегій агентів. Головна мета першої роботи з агентного моделювання – навчитись проводити симуляцію торгів за допомогою пакету AgentsCL та аналізувати результати моделювання, такі як динаміка модельної ціни та статистика розподілу активів між агентами.

### Хід роботи

Для агентного моделювання в пакеті Matlab розроблено програмний пакет AgentsCL. Для його запуску необхідно змінити папку системи Matlab на кореневу пакету AgentsCL. Пакет записаний в папці C:\Work\AgentsCL, для початку роботи необхідно відкрити закладку менеджера файлів та перейти в папку AgentsCL. Вміст цієї папки Ви побачите у менеджері файлів та на наступному рисунку.



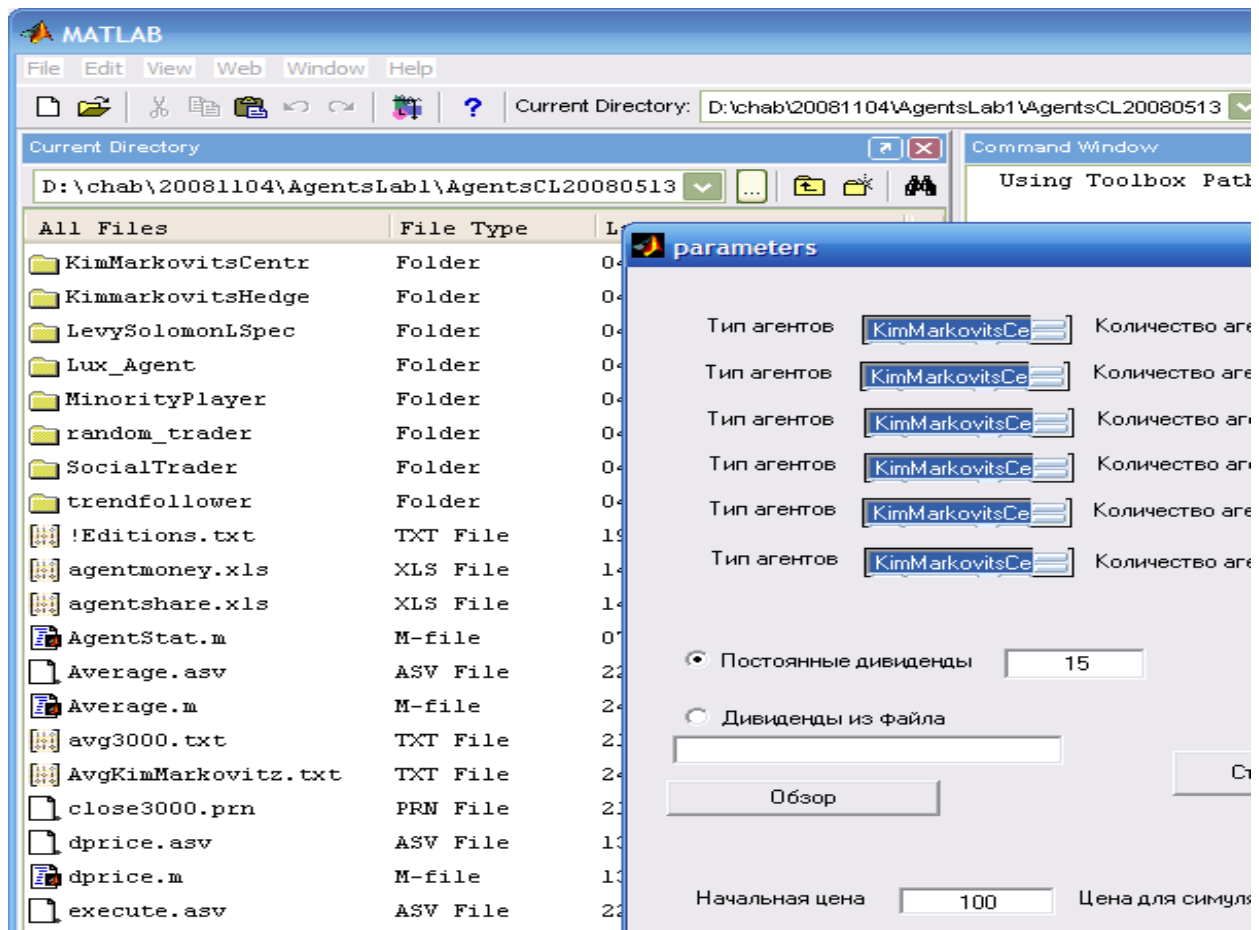


Рис. 1. Вміст кореневої папки пакету AgentsCL

Поряд з файлами ми бачимо ряд папок, які містять моделі агентів. Кожна папка відповідає одному виду агентів та включає наступні файли:

Defsettings.m – цей скрипт задає властивості агента по замовчуванні.

Settings.m – скрипт для налаштування властивостей агента.

Settings.fig - діалогове вікно на лаштування агенту.

OrderModel.m – скрипт, який проводить аналіз ситуації та подає заявки на придбання чи продаж активу.

### Деякі моделі поведінки агентів, які розроблені в пакеті AgentsCL

Розглянемо вибрані моделі агентів.

**Випадковий трейдер (random\_trader)** Заявки такого трейдера генеруються випадково у межах, які задані налаштуваннями агента.

**Трейдер-максимізатор**, який працює по стандартній схемі максимізації прибутку (див. Levi-Levi-Solomon).

**Тренд-трейдер**, який слідує руху курсу (trend\_follower). Трейдери цього виду купують актив при зростанні ціни, а при пониженні на певний відсоток продають весь куплений портфель. Це відповідає стратегії «биків» на ринку.

Моделі агентів нескладно множити власноручно. Ви можете самостійно розробити модель власного агента, запрограмувавши для нього умови купівлі та продажу активу. Для більш детальної інформації див. тексти розроблених моделей.

Для запуску середовища необхідно задати параметри моделювання. Для цього, знаходячись в папці **C:\Work\AgentsCL**, введіть команду **parameters**. З'явиться діалогове вікно налаштування параметрів симуляції.

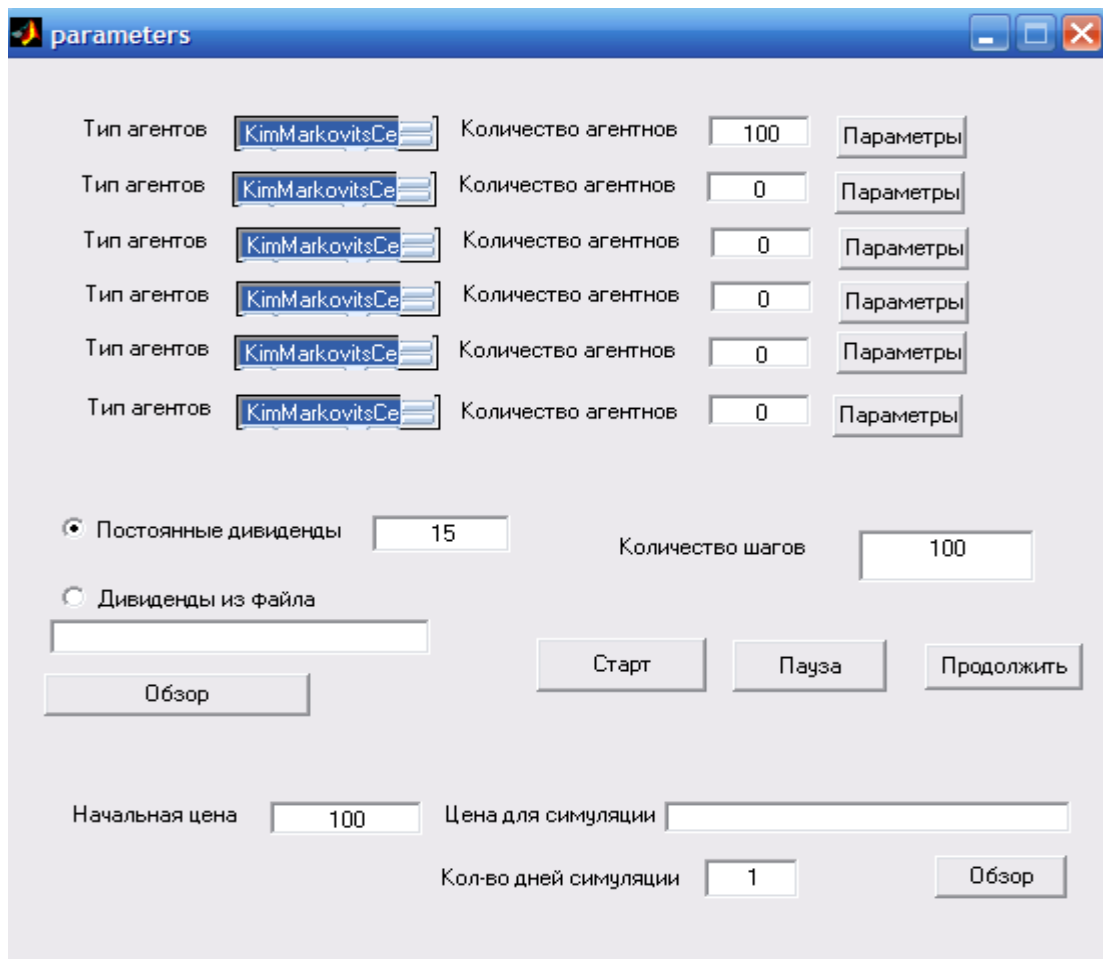


Рис. 2. Вікно налаштування симуляції ринку

Існують наступні *можливості налаштування*:

- задання типу, кількості та специфічних параметрів агентів;
- задання часового проміжку (кількості одиниць часу для моделювання). В нашій моделі одиниця часу рівна часу однієї торговельної сесії;
- задання файлу дивідендів, які будуть впливати на торгівлю акціями.

Після вибору типу агента можна задати кількість таких агентів та їх параметри. Параметри задаються натисканням кнопки «**Параметры**» навпроти виду агентів, який налаштовується. Для кожного виду агентів з'являється своє вікно налаштувань. Наприклад, на наступному рисунку, зображене вікно налаштувань випадкового агента («random trader»).

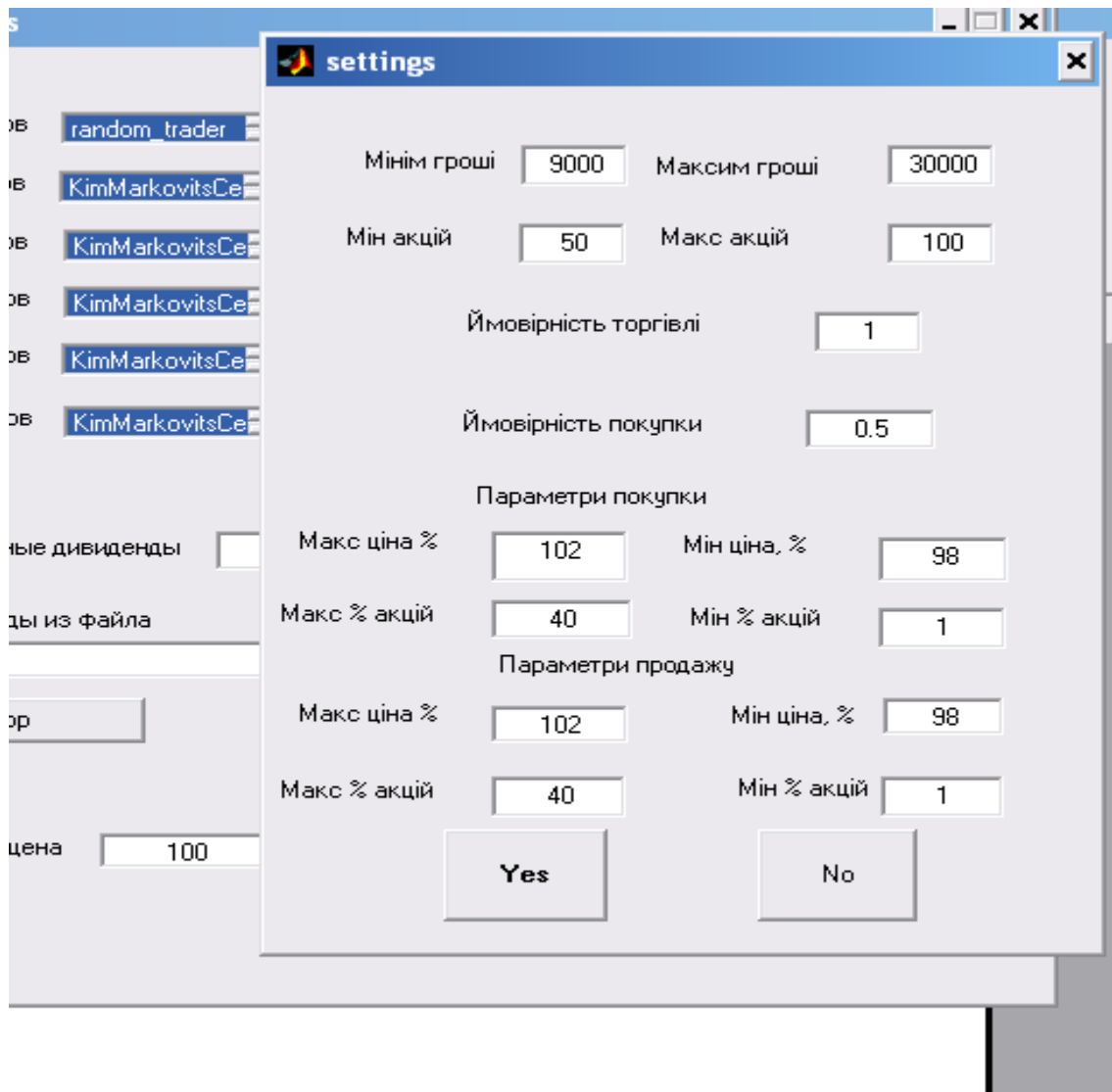


Рис. 3. Налаштування параметрів агента

Багато з цих параметрів спільні для усіх агентів. Для агентів типу «випадковий агент» («random trader») можна задати наступні параметри.

– **Максимальна і мінімальна кількість грошей агента.** Для агентів з цієї групи кількість грошових одиниць буде згенерована випадковим чином рівномірно в указаному інтервалі;

– **Максимальна і мінімальна кількість акцій.** Аналогічно кількість наявних акцій у агентів цієї групи буде генеруватись як випадкові цілі числа, рівномірно розподілені у вказаному інтервалі.

– **Ймовірність торгівлі.** Агент в кожен момент часу приймає рішення, торгувати, чи не торгувати. Випадковий трейдер це рішення «приймає» випадково із ймовірністю (від 0 до 1), яка вказується у відповідному полі.

– **Ймовірність покупки.** Якщо прийняте рішення про активність на ринку, то приймається рішення про покупку або продаж. Випадковий трейдер це питання вирішує випадково із вказаною ймовірністю.

Окремо для покупки і продажу можна задати параметри ціни та об'єму, які будуть вказуватись в заявці агента. Ці параметри інформують відносно діапазонів цін та об'ємів в процентах від наявної ціни на ринку. Наприклад,

- **Макс ціна %** – максимальна ціна в заявці у процентах від наявної;
- **Мін ціна %** – мінімальна ціна в заявці у процентах від наявної;
- **Макс акцій** – максимальна кількість акцій, шт

– **Мін акцій** – мінімальна кількість акцій, шт.

Ці параметри заявки можна окремо задати для заявки на продаж та на купівлю.

Спробуйте вибрати декілька видів агентів та задати їх кількість. Якщо ви не натискаєте на кнопку «**параметри**», для цього виду агентів приймаються налаштування по замовчуванню.

Після вибору агентів спробуйте змоделювати перші 100 кроків моделювання. Задайте значення 100 в полі введення «» та натисніть кнопку «**Старт**». Перед вами з'явиться смуга прогресу, яка відображає процес моделювання. За нею можна судити, скільки часу залишилось до кінця процесу моделювання. Після закінчення симуляції на екрані ви побачите графік зміни середньої ціни активу, яка сформувалась в процесі моделювання. Під час моделювання середовище зберігає результати торгів (середню максимальну та мінімальну ціну та об'єми торгів) у файлі **simulation.xls**, а також кількість грошей та акцій кожного з агентів (**agentmoney.xls** та **agentshare.xls** відповідно). Всі файли є текстовими, що дає можливість конвертувати результати в потрібний формат чи імпортувати в будь-яке середовище для подальшого аналізу.

Радимо робити копії цих файлів для кожного етапу моделювання, тому що старі файли будуть замінені новими. Тому **скопійуйте файли simulation.xls, agentmoney.xls та agentshare.xls в окрему папку.**

Результати моделювання можна імпортувати в Microsoft Excel для подальшого редагування. Для цього достатньо завантажити необхідний файл в табличний процесор. Комірки, які містять числа, необхідно перетворити на числовий формат, замінивши точку на кому. Це можна здійснити за допомогою опції меню «**Правка – Заменить**». В полі введення «**Найти**» задати точку (символ «.»), а в полі «**Заменить на**» задати кому (символ «,»), після чого натиснути кнопку «**Заменить все**».

На рис. 4 зображені результати моделювання, відкриті в табличному процесорі Microsoft Excel. Структуру файлу **simulation.xls** пояснює таблиця 1.

Таблиця 1

Структура файлу **simulation.xls**.

№ поля	Назва поля	Опис
	n	Номер проміжку часу (етапу торгів), параметр часу
	open	Ціна відкриття торгів (перша угода)
	max	Максимальна ціна за етап торгів
	min	Мінімальна ціна за етап торгів
	average	Середня ціна за етап торгів
	close	Ціна закриття (остання угода)
	Price2	Ціна векторної моделі ціноутворення
	volume	Об'єм торгів за етап
	AverageMoney	Середня к-сть грошей агентів
	MinMoney	Мінімальна кількість грошей в агентів
	MaxMoney	Максимальна кількість грошей в агентів
	AverageShare	Середня к-сть акцій агентів
	MinShare	Мінімальна кількість акцій агентів
	MaxShare	Максимальна кількість акцій агентів

	A	B	C	D	E	F	G	H	I	J
1	n	Open	max	min	average	close	Price2	volume	AverageMo	Min
2	1	9.866623€	1.039344€	9.365083€	9.954953€	1.018036€	100	381	2.030121€	7.5€
3	2	9.941138€	1.050753€	9.168938€	9.960127€	1.050101€	1.002964€	327	2.030121€	7.5€
4	3	9.927392€	1.055674€	9.645841€	1.005579€	1.055674€	9.956680€	408	2.030121€	8.2€
5	4	1.007976€	1.021094€	9.275146€	9.960023€	9.946074€	1.018582€	389	2.030121€	8.2€
6	5	9.944278€	1.040885€	9.719585€	1.009336€	1.040885€	1.049134€	328	2.030121€	8.2€
7	6	1.010157€	1.053696€	9.607964€	1.007894€	1.026411€	1.070110€	353	2.030121€	7.2€
8	7	9.947214€	1.018227€	9.456751€	9.917902€	1.018227€	1.076152€	338	2.030121€	7.2€
9	8	9.906847€	1.025016€	9.223090€	9.736987€	1.001691€	1.079572€	388	2.030121€	7.2€
10									2.030121€	5.7€
11									2.030121€	7.5€
12									2.030121€	6.3€
13									2.030121€	6.3€
14									2.030121€	6.3€
15									2.030121€	5.0€
16									2.030121€	4.5€
17									2.030121€	4.5€
18									2.030121€	3.3€
19									2.030121€	2.0€
20									2.030121€	2.0€
21	20	1.038449€	1.120400€	9.483159€	1.037301€	1.116443€	1.224922€	373	2.030121€	2.0€
22	21	1.039503€	1.095687€	9.623300€	1.048005€	1.051869€	1.237120€	346	2.030121€	2.0€
23	22	1.052537€	1.091475€	1.005571€	1.052268€	1.071467€	1.274056€	359	2.030121€	7.6€
24	23	1.052123€	1.183133€	1.035999€	1.086896€	1.183133€	1.284202€	237	2.030121€	7.6€
25	24	1.020074€	1.170952€	1.020074€	1.108240€	1.152406€	1.300504€	272	2.030121€	7.6€
26	25	1.108140€	1.243173€	1.046211€	1.127007€	1.243173€	1.316692€	355	2.030121€	7.6€
27	26	1.123083€	1.198077€	1.038073€	1.132873€	1.189458€	1.371298€	334	2.030121€	7.6€
28	27	1.132483€	1.244625€	1.096536€	1.156191€	1.223129€	1.412452€	302	2.030121€	7.6€
29	28	1.161370€	1.279745€	1.121753€	1.109744€	1.271312€	1.430630€	250	2.030121€	7.6€

Рис. 4. Результат моделювання торгів, збережені в файлі **simulation.xls**.

Процес моделювання торгів проходить наступним чином:

- всі агенти подають заявки;
- автоматично формуються пари взаємних заявок – угоди;
- здійснюється розрахунок агентів згідно з угодами;
- обчислюється статистика проведених торгів.

На кожному етапі торгів ці дії повторюються, таким чином у відповідних файлах зберігається інформація про результати торгів (файл **simulation.xls**, див. табл. 1) та фінанси кожного агента (файли **agentmoney.xls** та **agentshare.xls** відповідно).

Для аналізу даних програмою **work2** необхідно підготувати та зберегти ряд, який підлягатиме обробці. Для цього необхідно скопіювати ряд (наприклад, стовпець з позначкою **close** із файлу «**simulation.xls**») на чистий лист, замінити знак «кома» на знак «точка» як роздільник цілої та дробової частини (в текстових файлах Matlab у якості такого розділювача використовується знак «точка») та зберегти лист як текстовий файл. В подальшому такий файл можна відкрити пакетом **work2**.

Також аналізу мають підлягати розподіл багатства між агентами в процесі торгів, а також розподіл видів активів (акцій, готівки) між агентами. Дані про баланс готівки та акцій кожного агента записуються в файли **agentmoney.xls** (готівка) та **agentshare.xls** (акції). Структура цих файлів – це таблиця, стовпці якої – номери агентів, а рядки – час (етап

торгівлі). Таким чином, можна переглянути баланс кожного агента, побудувати графік зміни кількості певного активу або експортувати ряд для подальшого аналізу в текстовий файл, аналогічно попередньому прикладу.

Для аналізу розподілу повного багатства агентів за даними двох вищезгаданих таблиць (таблиці готівки **agentmoney.xls** та таблиці акцій **agentshare.xls**) побудувати аналогічну таблицю повного багатства агентів. Для цього необхідно скористатись таким засобом Microsoft Excel як абсолютні та відносні посилання. Для зручності скопіюємо вміст обох таблиць як 2 листи (**Готівка** та **Акції**) в одну книгу та створимо лист **«Багатство»**. Також нам буде необхідна таблиця ціни в кожний момент часу, яку можна взяти, створивши лист **«Ціна»** та скопіювавши на цей лист вміст файлу **«simulation.xls»**. Багатство агента  $k$  в момент часу  $t$  необхідно обчислити за формулою:

$$V_k(t) = m_k(t) + a_k(t) \cdot p(t),$$

де  $V_k(t)$  – повне багатство агента  $k$  в момент часу  $t$  (значення комірки таблиці **«Багатство»** з координатою стовпця  $k$  та рядка  $t$ );

$m_k(t)$  – кількість готівки агента  $k$  в момент часу  $t$  (значення комірки таблиці **«Готівка»** з координатою стовпця  $k$  та рядка  $t$ );

$a_k(t)$  – кількість акцій агента  $k$  в момент часу  $t$  (значення комірки таблиці **«Акції»** з координатою стовпця  $k$  та рядка  $t$ );

$p(t)$  – ціна акцій в момент часу  $t$  (стовпець ціни на листі **«Ціна»**).

Таким чином можна обчислити таблицю багатства кожного агента в будь-який момент торгів. Формула та результат таких обчислень зображена на рис. 5 (наприклад, для комірки B2 ця формула має такий вигляд: **=Готівка!B2+Акції!B2\*Ціна!\$E2**).

	A	B	C	D	E	F	G	H	I	J	K
1	n	1	2	3	4	5	6	7	8	9	
2	1	=Готівка!B2+Акції!B2*Ціна!\$E2			3,30E+04	3,55E+04	3,22E+04	2,90E+04	3,45E+04	3,05E+04	1,78E
3	2	3,38E+04	2,63E+04	2,35E+04	3,30E+04	3,55E+04	3,22E+04	2,90E+04	3,44E+04	3,05E+04	1,79E
4	3	3,39E+04	2,63E+04	2,36E+04	3,32E+04	3,56E+04	3,24E+04	2,92E+04	3,45E+04	3,06E+04	1,80E
5	4	3,38E+04	2,63E+04	2,35E+04	3,31E+04	3,55E+04	3,23E+04	2,91E+04	3,44E+04	3,06E+04	1,79E
6	5	3,40E+04	2,64E+04	2,36E+04	3,32E+04	3,56E+04	3,24E+04	2,92E+04	3,45E+04	3,06E+04	1,80E
7	6	3,39E+04	2,64E+04	2,36E+04	3,32E+04	3,56E+04	3,23E+04	2,92E+04	3,44E+04	3,06E+04	1,80E
8	7	3,38E+04	2,63E+04	2,34E+04	3,30E+04	3,54E+04	3,22E+04	2,90E+04	3,43E+04	3,06E+04	1,79E
9	8	3,37E+04	2,62E+04	2,32E+04	3,29E+04	3,53E+04	3,21E+04	2,87E+04	3,43E+04	3,05E+04	1,77E
10	9	3,36E+04	2,59E+04	2,30E+04	3,28E+04	3,52E+04	3,17E+04	2,85E+04	3,42E+04	3,05E+04	1,80E
11	10	3,37E+04	2,60E+04	2,33E+04	3,30E+04	3,54E+04	3,19E+04	2,88E+04	3,43E+04	3,06E+04	1,81E
12	11	3,37E+04	2,60E+04	2,32E+04	3,30E+04	3,53E+04	3,18E+04	2,87E+04	3,43E+04	3,05E+04	1,80E
13	12	3,37E+04	2,60E+04	2,31E+04	3,29E+04	3,52E+04	3,18E+04	2,86E+04	3,43E+04	3,02E+04	1,79E
14	13	3,39E+04	2,61E+04	2,35E+04	3,31E+04	3,54E+04	3,20E+04	2,90E+04	3,43E+04	3,03E+04	1,81E
15	14	3,39E+04	2,62E+04	2,37E+04	3,32E+04	3,55E+04	3,22E+04	2,92E+04	3,43E+04	3,03E+04	1,82E

Рис. 5. Обчислення повного багатства

### Завдання для самостійної роботи

1. Виберіть тип агенту **«Випадковий» (RandomTrader)**, задайте кількість таких агентів 20 та початкові параметри агентів. Змоделуйте 1000 етапів торгів цієї групи агентів.
2. Зробіть копії файлів, які містять результати торгів (файли **simulation.xls**, **agentmoney.xls**, **agentshare.xls**) в окрему папку. Зафіксуйте параметри агентів, при яких здійснювалось

моделювання.

3. Розгляньте динаміку зміни ціни при таких налаштуваннях агентів (графік, вміст файлу **simulation.xls**), зробіть висновок про наявність рівноважної ціни при даних налаштуваннях агентів.
4. За допомогою текстового процесору Excel збережіть часовий ряд цін закриття (**close**) для подальшого аналізу у **work2**, задавши символ «точка» як розділювач цілої та дробової частини числа.
5. Для кожного моменту модельного часу обчисліть повне багатство кожного агента як суму готівки та добуток кількості акцій на ціну акцій в даний момент часу. Це можна зробити, використовуючи абсолютні та відносні посилання, працюючи з вищезгаданими файлами в Microsoft Excel.
6. Збережіть ряд повного багатства всіх агентів на початку моделювання (в момент часу 1) та в кінці (в момент часу 1000) як ряд чисел для подальшого аналізу у **work2**.
7. За допомогою **work2** обчисліть прибутковості ряду ціни моделі та проаналізуйте ряд на розподіл прибутковостей. Зробіть висновки про наявність чи відсутність так званих «довгих хвостів» та характер розподілу.
8. Проаналізуйте ряд повного багатства агентів на розподіл, побудуйте діаграми розподілу багатства на початку торгів та в кінці.
9. Змініть початкову ціну моделювання та зробіть декілька експериментів з цим параметром. Спробуйте віддаляти або наближати початкову ціну до рівноважної, яка отрималась в минулих експериментах. Зберігайте результати кожного експерименту в окремій папці, обов'язково фіксуйте параметри моделювання. Зробіть висновок про динаміку модельної ціни в цих експериментах.
10. Змініть середні параметри початкового багатства агентів (кількість грошей, кількість акцій). Зробіть висновки про вплив цих параметрів на динаміку ціни та на існування рівноважної ціни.
11. Додайте групу агентів з аналогічними параметрами, але зробіть більшою ймовірність купівлі для нових агентів (так звані «бики» на ринку). Дослідіть вплив нових агентів на динаміку ціни.

## ЛАБОРАТОРНА РОБОТА № 22

**Тема.** Використання принципу невизначеності Гейзенберга для прогнозування кризових явищ в складних системах.

**Мета.** Навчитись розраховувати економічні аналоги фізичних величин з метою управління їх невизначеністю в умовах критичних та кризових явищ у складних системах різної природи.

1. Принцип невизначеності Гензенберга і його економічний аналог.

### Теоретичні відомості

Нестабільність глобальних фінансових систем щодо звичайних і природних збурень сучасного ринку та наявність погано передбачуваних фінансових криз свідчать в першу чергу про кризу методології моделювання, прогнозування та інтерпретації сучасних соціально-економічних реалій.

Доктрина єдності наукового методу стверджує, що для вивчення подій у соціально-економічних системах застосовні ті ж методи і критерії, що і при вивченні природних явищ. Однак очевидно, що їх схильність до помилок вносить елемент невизначеності в будь-які соціально-економічні дії. Подібне не відбувається в області природних явищ. І цю різницю Дж. Сорос пропонує враховувати, вводячи концепцію рефлексивності. Якщо поділ між

сукупністю тверджень і сукупністю фактів призводить до зміненого сприйняття реальності, це має відбиватися (рефлексувати) і на сукупності фактів. Такий зв'язок і повинна виражати концепція рефлексивності. Іншими словами, рефлексивність є двостороннім механізмом зворотного зв'язку, що впливає не тільки на твердження (оцінюючи їх істинність), але й на факти (вводячи в хід подій елемент невизначеності). Глобальна криза 2008 року на думку Сороса підтвердила його парадигму: рефлексивність фінансових ринків породила невизначеність, не піддається кількісній оцінці і привела до подій, які неможливо завчасно прорахувати.

Нещодавно ми запропонували нову парадигму моделювання складних систем, засновану на ідеях і уявленнях квантової, в тому числі і релятивістської, механіки. Було показано, що використання при описі соціально-економічних процесів квантово-механічних аналогій, включаючи принцип невизначеності, поняття оператора і квантову інтерпретацію вимірювальних процедур, має великі перспективи, враховуючи дискусії, які носять перманентний характер. Цікавим і перспективним також є міждисциплінарний підхід, який поєднує теорію складних мультиплексних мереж (включаючи квантові). Вивчення взаємодії їх структури і динаміки відкриває нові можливості побудови ефективних моделей складних систем (як природних, так і штучних): від роботи мозку до квантової теорії гравітації.

Метою даної роботи є аналіз фундаментальних фізичних понять і констант з точки зору досягнень сучасної теоретичної фізики, пошук адекватних їм і корисних аналогів в соціально-економічних явищах і процесах, а також їх використання для можливого раннього попередження небажаних кризових явищ на фінансових ринках.

**Про природу і взаємозв'язки основних фізичних понять.** До вихідних, строго не визначених фізичних понять зазвичай відносять час, відстань і масу, вважаючи, що шляхом тих чи інших процедур вимірювання їм можуть бути поставлені у відповідність певні числові значення. У такому разі інші фізичні величини, наприклад, швидкість, прискорення, імпульс, сила і та ін., можуть бути виражені і визначені через три зазначені вище основні поняття з використанням відповідних фізичних законів.

Підкреслимо, що без базових фізичних понять не обходиться жодна сучасна фізична теорія, включаючи релятивістську і квантову фізику. Проте ми хотіли б звернути увагу на наступне.

Як показав Ейнштейн в загальній теорії відносності, наявність неоднорідно розподілених мас призводить до викривлення 4-мірного простору-часу, в результаті чого «декартові» координати простору Мінковського  $(x, y, z, ict)$ , ( $i = \sqrt{-1}$  – уявна одиниця,  $c$  – швидкість світла у вакуумі,  $t$  – час), стають криволінійними. Таким чином існування мас у нашому світі може бути описано і геометричною мовою.

Якщо від глобальних макроявищ перейти до мікросвіту, де діють закони квантової фізики, то ми приходимо до того ж самого висновку про пріоритетну роль просторово-часових координат у визначенні всіх інших фізичних величин, включаючи масу.

Скористаємося відомим співвідношенням невизначеності Гейзенберга, яке є фундаментальним наслідком аксіом нерелятивістської квантової механіки і має вигляд:

$$\Delta x \cdot \Delta v \geq \hbar / 2m_0, \quad (1)$$

де  $\Delta x$  і  $\Delta v$  – середньоквадратичні відхилення координати  $x$  і швидкості  $v$  частинки з масою (спкою)  $m_0$ ,  $\hbar$  – постійна Планка. Вважаючи можливими вимірювання величин  $\Delta x$  і  $\Delta v$  в умовах, коли їх добуток досягає мінімуму, з (1) отримуємо:

$$m_0 = \hbar / (2 \cdot \Delta x \cdot \Delta v), \quad (2)$$

тобто маса частки виражається через невизначеності її координати і її швидкості – похідною за часом від тієї ж координати.



## Динамічні особливості економічних вимірів, економічний аналог співвідношення невизначеностей Гейзенберга

Характерною рисою основних фізичних законів є те, що для їх опису у формулах використовуються константи, котрі як і самі закони, залишаються незмінними протягом принаймні останніх  $\sim 10^{11}$  років. До таких констант відносяться гравітаційна постійна, швидкість поширення світла у вакуумі, постійна Планка.

Якщо говорити про економічні закони, що в принципі засновані на результатах динамічних вимірювань, як фізичного (наприклад, кількості тих чи інших матеріальних ресурсів), так і економічного (наприклад, тарифи) характеру, то тут ситуація дещо інша. Адекватність формалізмів, що використовуються для їх математичного опису повинна піддаватися постійній перевірці та необхідній корекції. Це пов'язано з тим, що вимірювання – це завжди порівняння з чимось, прийнятим за еталон, проте в економіці постійних еталонів не існує в принципі. Таким чином, економічні виміри в своїй першооснові відносні і мають локальний у часі, просторі та інших соціально-економічних координатах характер.

Саме з цих причин важливе значення для оцінки стану, тенденцій та перспектив розвитку економік (глобальних, регіональних, національних) грає аналіз часових рядів, що породжуються динамікою фондових індексів, курсів валют, цін на спотових ринках.

Нехай існує сукупність з  $M$  часових рядів довжиною  $N$  відліків кожен, що стосуються одного і того ж проміжку часу  $T$ , з однаковим мінімальним кроком часу  $\Delta t_{\min}$ :

$$X_i(t_n), t_n = \Delta t_{\min} n; n = 0, 1, 2, \dots, N-1; i = 1, 2, \dots, M. \quad (3)$$

Щоб привести всі ряди до безрозмірного і однакового (з точністю до адитивної постійної) подання, нормалізуємо їх:

$$x_i(t_n) = \ln X_i(t_n), t_n = \Delta t_{\min} n; n = 0, 1, 2, \dots, N-1; i = 1, 2, \dots, M. \quad (4)$$

Будемо вважати, що кожен новий ряд  $x_i(t_n)$  представляє собою одновірну траєкторію деякої фіктивної частинки з номером  $i$ , координата якої реєструється через кожний проміжок часу  $\Delta t_{\min}$ , і оцінимо середньоквадратичні відхилення її координати та швидкості на деякому часовому вікні  $\Delta T$ :

$$\Delta T = \Delta N \cdot \Delta t_{\min} = \Delta N, 1 \ll \Delta N \ll N. \quad (5)$$

«Миттєва» швидкість  $i$ -ої частки в момент часу  $t_n$  визначається співвідношенням:

$$v_i(t_n) = \frac{x_i(t_{n+1}) - x_i(t_n)}{\Delta t_{\min}} = \frac{1}{\Delta t_{\min}} \ln \frac{X_i(t_{n+1})}{X_i(t_n)}, \quad (6)$$

її дисперсія  $D_{v_i}$ :

$$D_{v_i} = \frac{1}{(\Delta t_{\min})^2} \left( \left\langle \ln^2 \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n, \Delta N} - \left( \left\langle \ln \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n, \Delta N} \right)^2 \right), \quad (7)$$

а середньоквадратичне відхилення  $\Delta v_i$ :

$$\Delta v_i = \sqrt{D_{v_i}} = \frac{1}{(\Delta t_{\min})} \left( \left\langle \ln^2 \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} - \left( \left\langle \ln \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} \right)^2 \right)^{\frac{1}{2}}, \quad (8)$$

де  $\langle \dots \rangle_{n,\Delta N}$  означає середнє значення для часового вікна довжиною  $\Delta T = \Delta N \cdot \Delta t_{\min}$ .

В якості оцінки дисперсії координати  $i$ -ї частки використаємо дисперсію  $D_{\Delta x_i}$  випадкової величини  $\ln(X_i(t_{n+1})/X_i(t_n))$ :

$$\left\langle \ln^2 \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} - \left( \left\langle \ln \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} \right)^2. \quad (9)$$

Тоді для середньоквадратичного відхилення  $\Delta x_i$  з урахуванням (9) отримуємо:

$$\Delta x_i = \sqrt{D_{\Delta x_i}} = \left( \left\langle \ln^2 \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} - \left( \left\langle \ln \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} \right)^2 \right)^{\frac{1}{2}}. \quad (10)$$

За аналогією з (1) для траєкторії фіктивної частинки можна записати співвідношення невизначеностей у вигляді:

$$\Delta x_i \cdot \Delta v_i \sim h / m_i, \quad (11)$$

або, з урахуванням (7) та (10):

$$\frac{1}{\Delta t_{\min}} \left( \left\langle \ln^2 \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} - \left( \left\langle \ln \frac{X_i(t_{n+1})}{X_i(t_n)} \right\rangle_{n,\Delta N} \right)^2 \right) \sim \frac{h}{m_i}, \quad (12)$$

де  $m_i$  – економічна «маса»  $i$ -го ряду,  $h$  – величина, що є аналогом економічної «постійної» Планка.

На відміну від фізичної постійної Планка  $\hbar$ , величина  $h$ , взагалі кажучи, може залежати від історичного періоду часу, для якого взяті ряди, положення і величини інтервалу усереднення (економічні процеси по різному протікають під час кризи та рецесії), від номера ряду  $i$  та ін.

Узагальнимо співвідношення (12) на випадок, якщо економічні виміри на інтервалі часу  $T$ , в результаті яких були отримані ряди (4), проведені з кроком  $\Delta t = k \cdot \Delta t_{\min}$ , де  $k \geq 1$  – деяке задане ціле позитивне число. З формальної точки зору це означало б, що у вихідних рядах (4) викинуті всі члени, за винятком членів з номерами  $n=0, k, 2k, 3k, \dots$ . У результаті співвідношення (12) перетворюється до виду:

$$\frac{1}{k \Delta t_{\min}} \left( \left\langle \ln^2 \frac{X_i(t_{n+k})}{X_i(t_n)} \right\rangle_{n,\Delta N} - \left( \left\langle \ln \frac{X_i(t_{n+k})}{X_i(t_n)} \right\rangle_{n,\Delta N} \right)^2 \right) \sim \frac{h}{m_i} \quad (13)$$

і містить залежність від  $k$ .

Як у випадку реальної фізичної частинки, так і для її формального економічного аналогу, будь-який вимір впливає на результат. Тому статистичні властивості «розріджених»

рядів, на основі яких і було отримано співвідношення (13), повинні залежати від того, чи проводилися реальні виміри в проміжних точках.

Таким чином, узагальнюючи все вище сказане, ми повинні вважати відношення  $h/m_i$  в правій частині (13) певною невідомою функцією номера ряду  $i$ , розміру вікна усереднення  $\Delta N$ , часу  $\bar{n}$  (центру вікна усереднення) та кроку часу спостереження (реєстрації)  $k$ .

Щоб отримати для цієї функції хоча б і наближений вираз, але такий, що явно обчислюється, і простежити характер залежностей, постулюємо для правої частини (13) наступний модельний вигляд:

$$\frac{h}{m_i} \cong \frac{\tau(\bar{n}, \Delta N_\tau) \cdot H_i(k, \bar{n}, \Delta N_H)}{\Delta t_{\min} \cdot m_i}, \quad (14)$$

де  $m_i$  – безрозмірна економічна маса  $i$ -го ряду,

$$m_i^{-1} = \langle \phi_i(n, 1) \rangle_{(0 \leq n \leq N-2)}, \quad (15)$$

$$\tau(\bar{n}) = \frac{\langle \phi_i(n, 1, \Delta N_\tau) \rangle_{(\bar{n} - \Delta N_\tau/2 < n < \bar{n} + \Delta N_\tau/2), (1 \leq i \leq M)}}{\langle \langle \phi_i(n, 1, \Delta N_H) \rangle_{(\bar{n} - \Delta N_\tau/2 < n < \bar{n} + \Delta N_\tau/2), (1 \leq i \leq M)} \rangle_{\bar{n}}}, \quad (16)$$

– локальний коефіцієнт стиснення ( $\tau(\bar{n}) < 1$ ) або розтягування ( $\tau(\bar{n}) > 1$ ) фізичного часу, що дозволяє ввести поняття неоднорідного економічного часу (для однорідного випадку  $\tau(\bar{n}) = 1$ ),

$$H_i(k, \bar{n}) = \frac{\langle \phi_i(n, k, \Delta N_H) \rangle_{\bar{n} - \Delta N_H/2 < n < \bar{n} + \Delta N_H/2}}{\langle \phi_i(n, 1, \Delta N_H) \rangle_{\bar{n} - \Delta N_H/2 < n < \bar{n} + \Delta N_H/2}}, \quad (17)$$

– безрозмірний коефіцієнт порядку одиниці, що вказує, для заданих  $i$  і  $\bar{n}$ , на відмінності в залежності дисперсії  $D_{\Delta x_i}$  (див. (9) з урахуванням випадку  $k \geq 1$ ) від закону  $D_{\Delta x_i} \sim k$ ,

$$\phi_i(n, k, \tilde{N}) = \frac{1}{k} \left( \ln^2 \frac{X_i(t_{n+k})}{X_i(t_n)} - \left( \langle \ln \frac{X_i(t_{n+k})}{X_i(t_n)} \rangle_{n, \tilde{N}} \right)^2 \right) \quad (18)$$

(в останній формулі індекс  $\tilde{N} = N$ ,  $\Delta N_\tau$ ,  $\Delta N_H$  вказує на параметри по  $n$  відповідно до формул (15-17)), вікна усереднення  $\Delta N_\tau$ ,  $\Delta N_H$  вибираються з урахуванням умов:

$$k_{\max} < \Delta N_\tau < \Delta N_H < N. \quad (19)$$

Відповідно до визначень (16, 17) для коефіцієнтів  $\tau(\bar{n})$  і  $H_i(k, \bar{n})$  мають місце умови нормування:

$$\langle \tau(\bar{n}) \rangle_{\bar{n}, N} = 1; \quad H_i(1, \bar{n}) = 1, \quad (20)$$

а множник  $1/\Delta t_{\min}$  у правій частині (14) можна розглядати як інваріантну складову

економічної постійної Планка  $h$ :

$$\bar{h} = 1 / \Delta t_{\min}, \quad (21)$$

Як видно, величина  $\bar{h}$  має природну розмірність зворотного часу.

Зазначимо також, що можна ввести середню економічну масу всієї сукупності рядів (або будь-якої виділеної групи рядів) за формулою:

$$m^{-1} = M^{-1} \sum_{i=1}^M m_i^{-1}. \quad (22)$$

Отримані на базі рядів (3) співвідношення (4, 13-21) допускають і інші варіанти інтерпретації.

### Хід роботи

Розрахунки описаних вище параметрів реалізовано у вигляді скрипта **heisenberg\_mass\_main.m**. Це означає, що у папці, де будуть зберігатися результати розрахунків крім вказаного скрипта повинні бути необхідні часові ряди та додатковий скрипт **sglazh.m** для згладжування даних. Для апробації запропонованого підходу були вибрані спеціально підготовлені дві групи часових рядів щоденних значень індексів відомих фондових ринків. До першої групи входять часові ряди, які включають індекс Доу-Джонса (djia) за більш ніж 100-літній період та фрагменти цього ж індексу, що містять найбільш потужні фінансові кризи 1929, 1987 і 2008 рр. Довжина фрагментів однакова – 2000 точок, при цьому власне пік кризи (точка, з якої починається стрімке падіння фондового індексу) приходить на середину ряду – 1000-ну точку. Знаючи час настання кризи та співставляючи часовий ряд з динамікою певного показника, можна досліджувати його залежність від тих чи інших характерних змін на фондовому ринку: докризовий, кризовий та післякризовий періоди. До другої групи входять індекси понад 30-ти фондових ринків світу.

На рис. 1а наведено динаміку нормалізованих значень фрагментів індексу Доу-Джонса для найвідоміших криз, а на рис. 1б відповідні залежності економічної маси  $m$ . Розрахунки економічної маси проводились у рамках алгоритму ковзного вікна: величина маси розраховувалась для підряду певної довжини (вікна), наприклад, 500 точок (приблизно 2 роки), потім зміщувалось з певним кроком пропорційним  $\Delta t_{\min}$  і процедура продовжувалась до вичерпання часового ряду. З рис. 1б видно, що у докризовий і власне кризовий періоди маса помітно зменшується, відновлюючись у післякризовий період. При цьому її поведінка універсальна для криз різної природи. Для кризи 1929 р. післякризового відновлення не спостерігається оскільки ця криза тривала понад шість років (Велика Депресія).

Цікаво тепер співставити динаміку вихідного часового ряду з розрахованим показником – в нашому випадку з економічною масою. Якщо в період кризи вибраний показник поводить себе характерним чином, наприклад, зростає чи спадає, то його можливо використовувати для ідентифікації або попередження кризи.

На рис. 2а для порівняння наведено динаміку індексу Доу-Джонса під час кризи 2008 р. і відповідної економічної маси  $m$ . Видно випереджальний характер зменшення маси ще до настання кризи. Аналогічна картина спостерігається і для інших кризових станів.

На рис. 2б таке співставлення часового ряду і відповідної маси проведено для індексу Доу-Джонса з 02.01.1900 по 31.12.2015 рр. (всього 29030 значень). Відмічено додатково до вже згаданих криз також найбільш помітні кризи 1914, 1946, 1962, 1973 та 2001 рр. При цьому найменшим значенням мас відповідають найбільші кризи.

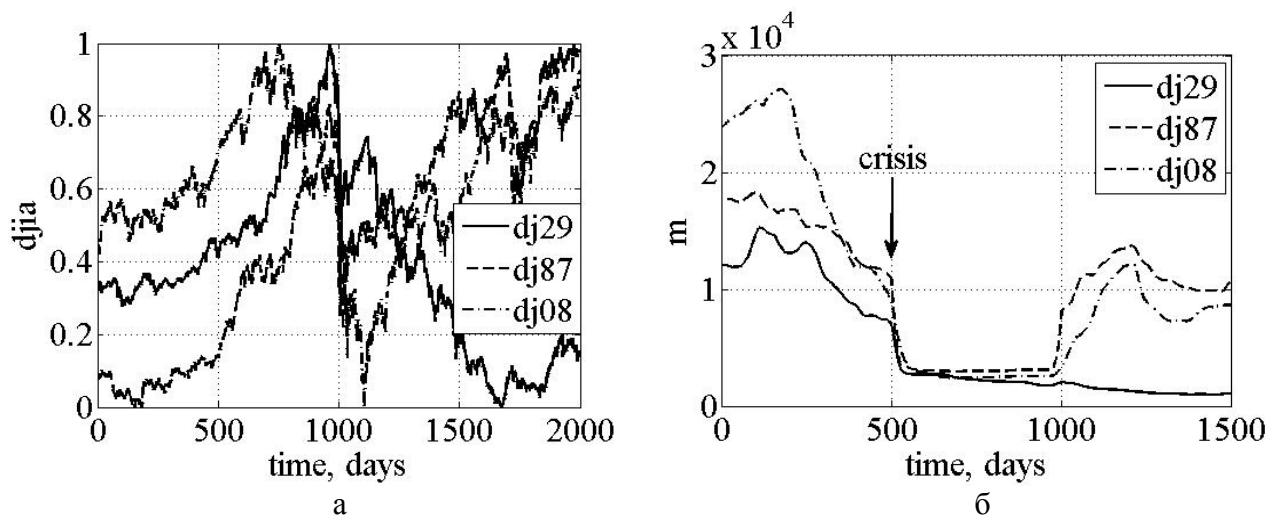


Рис. 1. Динаміка індексу Доу-Джонса djia (а) та економічної маси  $m$  (б) для найбільш потужних криз на фондовому ринку 1929 (dj29), 1987 (dj87) та 2008 (dj08) рр.

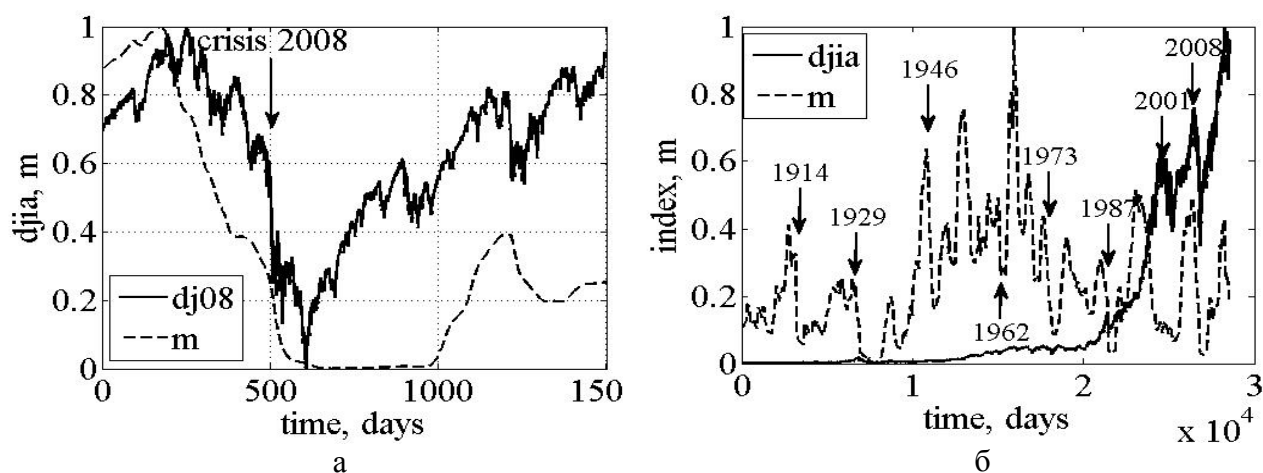


Рис. 2. Порівняльна динаміка нормалізованих часових рядів індексів Доу-Джонса та відповідних економічних мас

На рисунку 3 наведені результати розрахунків для наступних індексів: sp (повна назва s&p500) – популярного індексу «широкого» (враховує 500 компаній для розрахунку) ринку США, dax (deutscher aktienindex) – основного фондового індексу Німеччини, bsesn (s&p bse sensex) – Індії, ssec (sse composit index) - континентального Китаю.

Стрілками вказані останні три кризи: глобальна світова 2008-2009 рр., боргова криза 2011 р. та китайська криза, яка почалася в червні 2015 р. та продовжилася двома хвилями: у кінці вересня та грудня. Очевидно, що для глобальної кризи 2008-2009 рр. величина мас практично не залежить від індексу і має найменше значення на всьому проміжку спостереження. Боргова криза відчутно вразила країни Європи. Тому найменшу масу має індекс dax. Нарешті, індекс китайського ринку характеризується найменшою з мас в період поточної кризи 2015 р. Аналогічні результати отримано і для інших індексів другої групи часових рядів обраної бази даних.

Таким чином економічна маса часового ряду може слугувати індикатором кризових явищ. По-перше, чим меншою є маса, тим відчутнішою є криза. По-друге, зменшення маси випереджає падіння фондового індексу, що дозволяє використати цей факт у якості індикатора-передвісника кризи, що насувається.

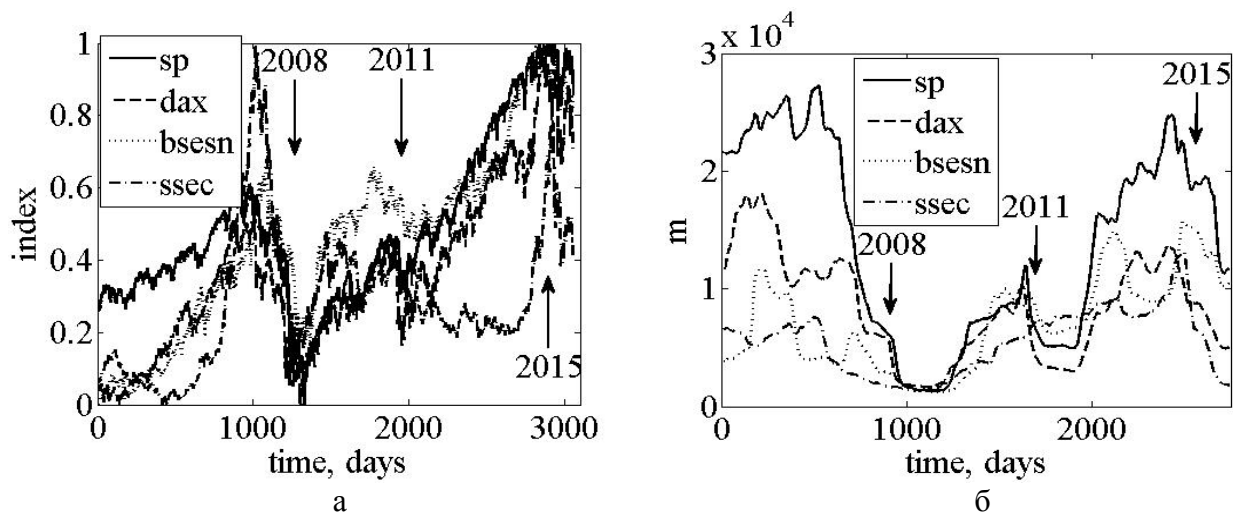


Рис. 3. Типова динаміка фондових індексів за період з 2004 по 2015 рр. (а) та відповідних економічних мас (б).

### Завдання для самостійної роботи

1. Проведіть аналогічні дослідження для часових рядів інших фінансових активів, наприклад, валютних, спотових тощо.
2. Знайдіть і дослідіть складні системи неекономічної природи: фізичні, біологічні, штучні. Порівняйте результати і зробіть висновки.

## Список рекомендованої літератури

1. Agaev A., Kuperin Yu.F. Multifractal Analysis and Local Hoelder Exponents Approach to Detecting Stock Markets Crashes // e-print: <http://arXiv:cond-mat/0407603>.
2. Agent-based Models of Financial Markets [електронний ресурс] / E. Samanidou, E. Zschischang, D. Stauffer, T. Lux // arxiv:physics/0701140v1 – 11 Jan 2007. – <http://arxiv.org/>
3. Albert R., Barabasi A.-L. Statistical Mechanics of Complex Networks, Rev. Mod. Phys. – 2002. –V.74. – P.47-97. [Електронний ресурс] – Режим доступу: [arXiv.org/cond-mat/0106096](http://arXiv.org/cond-mat/0106096).
4. Amaral L. A. N., Ottino J. M. Augmenting the Framework for the Study of Complex Systems // Eur.Phys.J. 2004, v.B38. – P.147-162.
5. Andersson M. K. On the Effects of Imposing or Ignoring Long Memory When Forecasting // Working Paper Series in Economics and Finance, 1998. – № 225.
6. Araujo T., Louca F. The geometry of crashes – a measure of the dynamics of stock market crises // arXiv:cond-mat/0506137.
7. Bak P. How Nature Works. The Science of Self-organized Criticality. Oxford, Oxford University Press, 1997.
8. Barabasi A.-L. Emergence of scaling in random networks / A.-L. Barabasi, R. Albert // Science. – 1999. – Vol. 286. – P. 509 – 512.
9. Barrat A. Dynamical processes on complex networks / Barrat A., Barthelemy M., Vespignani A. // Cambridge University Press, 2008. – 347 p.
10. Battena J., Ellis 3. Scaling Relationships of Gaussian Processes. School of Accounting & Finance Deakin University, 2001.
11. Bianconi G. Interdisciplinary and physics challenges in network theory / G. Bianconi // Europhysics Letters. – 2015. – Vol. 11, № 5. – P.1-7.
12. Black F., Jones R. C. Simplifying portfolio insurance // The Journal of Portfolio Management, V.14, 1987. – P.48-51.
13. Boccaletti S., Latora V., Moreno Y., Chavez M., Hwang D.-U. Complex networks: Structure and dynamics, Phys. Rep. – 2006, – V.424. – P.175-209.
14. Bollerslev T. Generalized autoregressive conditional heteroskedasticity / T. Bollerslev // Journal Econometrics. – 1986. – V. 31. – P. 307–327.
15. Boltzmann L. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen // Sitzber. Acad. Wiss. Wien. 1872, Bd. 66.– S. 275-376.
16. Borland L. Long-range memory and nonextensivity in financial markets // Econophysics news, 2005. – V .36. – №6. – P.228-231.
17. Brock,W., Dechert, D., Sheinkman, J., LeBaron, B. (1996). A test for Independence Based on the Correlation Dimension. Econometric Reviews, 15. – P. 197-235.
18. Clausius R. Über verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie //Ann. Phys. Folge 2, 1865. – Bd. 125.- S. 353-400.
19. Coronello C., Tumminello M, Lillo F., Micciche S., Mantegna R.N. Sector identification in a set of stock return time series traded at the London Stock Exchange // ArXiv:cond-mat/0508122
20. Costa M. Multiscale entropy analysis of biological signals [Electronic resource] / Costa M., Goldberger A. L., Peng C.-K. – Mode of access : <http://dx.doi.org/10.1103/PhysRevE.71.021906/>.
21. Costa M. Multiscale entropy analysis of physiologic time series [Electronic resource] / Costa M., Goldberger A. L., Peng C.-K. – Mode of access : <http://dx.doi.org/10.1103/PhysRevLett.89.068102>.
22. Cox J. C., Ross R. A., Rubinstein M. Option pricing: a simplified approach Journal of Financial Economics, 1976.3. – P. 229-263.

23. Danilchuk G. Dynamics of graph spectral entropy in financial crisis / G. Danilchuk, V. Soloviev // Socio-economic aspects of economics and management. – Taunton, MA, USA. – 2015. – Vol. 2. – P. 227-234.
24. Dario G. Perez, Luciano Zunino, Mario Garavaglia, Osvaldo A. Rosso. Wavelet entropy and fractional Brownian motion time series, arXiv:physics/0501105 v1.
25. Devaney Robert L. An Introduction to Chaotical Dynamical Systems? Sec.Ed. Addison-Wesley, Reading, Mass., 1989.
26. Donner R. V. Recurrence-based time series analysis by means of complex network methods / R. V. Donner, M. Small, J. F. Donges, N. Marwan et.al. // [Электронный ресурс] – Режим доступа: arXiv:1010.6032v1 [nlin.CD] 25 Oct 2010.
27. Drozd S., Grummer F., Ruf F., Speth J, Dynamics of correlations in the stock market // arXiv:cond-mat/0103605
28. Dynamic Behaviours of Mix-game Models and Its Application [электронный ресурс] / C. Gou //arxiv:physics/0504001 – 29 Aug 2005. –<http://arxiv.org/>
29. Eckmann J.-P., Ruelle D. Fundamental limitations for estimating dimensions and Lyapunov exponents in dynamical systems // Physica, 1992. – №56. – P.185-187.
30. Eckmann, S. O. Kamphorst, D. Ruelle, Recurrence plots of dynamical systems // Europhys. Lett., 5, 1987. – P. 973-977.
31. Eichengreen B., Rose A., Wyplosz Ch. Exchange market mayhem. The antecedents and aftermaths of speculative attacks // Economic Policy. October 1995. - P. 249–312.
32. Elwakil A. S., Soliman A. M., Mathematical models of the twin-T, wien-bridge and family of minimum component electronic chaos generators with demonstrative recurrence plots, Chaos Solit. Fract. 10 (8) (1999) 1399-1411.
33. Engle R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation // Econometrica, 1982. –V.50. – P. 987-1008.
34. Erdos P. On the evolution of random graphs / P. Erdos, A. Renyi // Publications of the Mathematical Institute of the Hungarian Academy of Sciences. – 1960. – Vol. 5. – P. 17 – 61.
35. Erken A., Yalcin Karatepe Y. Evaluating and forecasting banking crises through neural network models: An application for Turkish banking sector // Expert Systems with Applications, 33, (2007). – pp. 809–815.
36. Fama E. F. and Miller M. H., The Theory of Finance. New York: Holt, Rinehart and Winston, 1972.
37. Farmer J. D. Information dimension and the probabilistic structure of chaos. Z. Naturforsch. 37, 1304-1325 (1982).
38. Frankel J. A., Rose A. K. Currency Crashes in Emerging Markets: Empirical Indicators // NBER Working Paper. № 5437, 1996. – Cambridge, Massachusetts, MIT Press.
39. Gephi [Электронный ресурс]. – Режим доступа : <https://gephi.org/>
40. Gilmore C. G. An examination of nonlinear dependence in exchange rates, using recent methods from chaos theory. // Global finance journal N12, 2001. – P.139-151.
41. Gilmore C.G. A new test for chaos. // Journal of economic behavior and organization. N22, 1993. – P.209-237.
42. Giuliani A., Manetti C., Hidden peculiarities in the potential energy time series of a tripeptide highlighted by a recurrence plot analysis: a molecular dynamics simulation, Phys. Rev. E 53 (6) (1996) 6336-6340.
43. Granger C. W. J. An Introduction to Long Memory Time Series Models and Fraction Differencing / Granger C. W. J., Joyeux R. // Journal of Time Series Analysis. – 1980. – № 1 (1). – P. 15–29.
44. Grassberger P., Procaccia I. Characterization of strange attractors. Phys. Rev. Lett. 50, 346-349 (1983).



45. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors. *Physica D* 9, 189-208 (1983).
46. Grech D., Mazur Z. Can One Make Any Crash Prediction in Finance Using the Local Hurst Exponent Idea? // arXiv:cond-mat/0311627
47. Halvin S., Cohen R. Complex networks. Structure, robustness and function / Halvin S., Cohen R. // Cambridge University Press, 2010. – 238 p.
48. [http:// www.msci.com](http://www.msci.com)
49. <http://www.santafe.edu>
50. Hurst H. E. Long Term Storage Capacity of Reservoirs // Transactions of the American Society of Civil Engineers, 1951. – № 116. – P. 770-799.
51. Ivanov P. Ch., Hausdorff J. M., Halvin S. et.al. Levels of Complexity in Scale-Invariant Neural Signals // arXiv:cond-mat/0409545
52. Joshua S. Physiological time-series analysis using approximate entropy and sample entropy / Joshua S., Richman J., Moorman R. // *Am. J. Physiol. Heart Circ. Physiol.* 278: H2039-H2049, 2000.
53. Kaminsky G., Lizondo S., Reinhart C. M. The leading indicators of currency crises. IMF Staff paper No. 45. International Monetary Fund, 1998. – P.1-48.
54. Kaminsky G., Reinhart C. Financial Crises in Asia and Latin America: Then and Now // AEA Papers and Proceedings. № 98, 1998.
55. Kantelhardt J. W., Zschiegner S. A., Koscielny-Bunde E. et.al. Multifractal Detrended Fluctuation Analysis of Nonstationarity Time Series // e-print – arXiv:physics/0202070
56. Kaplan, D. T. (1994) Exceptional Events as Evidence of Determinism // *Physica D*, 73. – P. 38-44.
57. Kim G., Markowitz H. M. Investment rules, margin and market volatility // *Journal of Portfolio Management*, V. 16, 1989. – P.45-52.
58. Kurths J., Schwarz U., Sonett C. P., Parlitz U. Testing nonlinearity in radiocarbon data, *Nonlinear Processes Geophys.* 1(1) (1994) 72-75.
59. Kwapien J., Drozd S., Speth J. Alternation of different fluctuation regimes in the stock market dynamics // arXiv:cond-mat/0306608
60. Lacasa L. From time series to complex networks: The visibility graph / L. Lacasa, B. Luque, F. Ballesteros et.al. // *PNAS.* -2008. – V. 105, No 13. – P. 4972-4975.
61. LeBaron B. The time series properties of an artificial stock market / B. LeBaron, W. Arthur, R. Palmer // *Journal of Economic Dynamics and Control*, V23, 1999. – P. 1487-1516.
62. LeBaron B., Some Relations Between Volatility and Serial Correlations in Stock Market Returns. Working Paper, February 1990.
63. Levy M., Levy H., Solomon S. A microscopic model of the stock market: Cycles, booms, and crashes // *Economics Letters*, V.45, 1994. – P.103-111.
64. Liu Y., Gopikrishnan P., Cizeau P., Meyer M., Peng C.-K., Stanley H. E. The statistical properties of the volatility of the price fluctuations // arXiv:cond-mat/9903369
65. Lorenz E. N., Deterministic Nonperiodic Flow // *J. Atmos. Sci.*, 1963, 20, P. 130-141. (Перевод: Лоренц Э. Детерминированное непериодическое течение. – В сб.: Странные аттракторы // Под ред. Я. Г. Синая и Л.П.).
66. Lux T., Marchesi M. Scaling and criticality in a stochastic multi-agent model of a financial market // *Nature*, V.397, 1999. – P.498-500.
67. Mandelbrot B. *Fractals: Form, Chance and Dimension* / B. Mandelbrot. – Freeman, San Francisco, 1977.
68. Mandelbrot B., Statistical Methodology for Non-Periodic Cycles: From the Covariance to R/S Analysis. *Annals of Economic Social Measurement* 1, 1972.

69. Mandelbrot B.B. When Can Price Be Arbitraged Efficiently? A limit to the Validity of the Random Walk and Martingale Models // *The Review of Economics and Statistics*, 1971. – № 53 (1). – P. 225-236.
70. Manetti C., Giuliani A., Ceruso M.-A., Webber C. L., Zbilut J. P., Recurrence analysis of hydration effects on nonlinear protein dynamics: multiplicative scaling and additive processes, *Phys. Lett. A* 281 (5-6) (2001) 317-323.
71. Manimaran P., Panigrani P. K., Parikh J. C. Wavelet analysis and scaling properties of time series – arXiv:nlin.CD/0412046
72. Mantegna R. N. Hierarchical structure in financial markets. – *Eur. Phys. J. B.* 1999, v.25. p. 193–197.
73. Mantegna R. N., Stanley H. E. *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge: Cambridge University Press, 2000. – 257 p.
74. March T. K., Chapman S. C., Dendy R. O., Recurrence plot statistics and the effect of embedding, *Physica D* 200 (1-2) (2005) 171-184.
75. Markowitz H. M. *Simsript: A Simulation Programming Language*. Prentice Hall, Englewood Cliffs NJ, 1963.
76. Markowitz H. M., Portfolio Selection. *Journal of Finance* 7, 1952.
77. Marwan N., Romano M., Thiel M., Kurths J. Recurrence plots for the analysis of complex system. *Physics Reports* 438 (2007) 237 – 329.
78. Marwan N., Thiel M., Cross recurrence plot based synchronization of time series, *Nonlinear Processes Geophys.* 9 (3/4) (2002) 325-331.
79. Marwan N., Trauth M. H., Vuille M., Kurths J., Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods, *Clim. Dynam.* 21 (3-4) (2003) 317-326.
80. Maslov S. Measures of globalization based on cross-correlations of word financial indices // *Physica A*, 2001, v.301. – P.397-406.
81. Maturana, Humberto, and Francisco Varela, *The Tree of Knowledge*, Shambala, Boston, 1987.
82. Mehta M. L. *Random Matrices*. Academic Press, Boston, 1991.
83. Nelson D. B. Conditional Heteroscedasticity in Asset Pricing / D. B. Nelson // *Econometrica*. – 1991. – V. 59. – P. 347–370.
84. Newman M. E. J. The Structure and Function of Complex Networks / M. E. J. Newman // *SIAM Review*. – 2003. – Vol. 45, N 2. – P. 167 – 256.
85. Newman M., Watts D., Barabási A.-L. *The Structure and Dynamics of Networks*, Princeton University Press. – 2006. – 456 p.
86. Nichols J. M., Trickey S. T., Seaver M., Damage detection using multivariate recurrence quantification analysis // *Mech. Syst. Signal Process*, 20 (2), 2006. – P. 421-437.
87. Niemira M., Saaty T. An Analytic Network Process model for financial-crisis forecasting // *International Journal of Forecasting*, 20, 2004. – P. 573– 587.
88. Osbome, M. F. M. «Brownian Motion in the Stock Market», in P.Coother, ed., *The Random Character of Stock Market Prices*. – Cambridge: MIT Press, 1964.
89. Ott E., Grebogi C., Yorke J.A., Theory of First Order Phase Transitions for Chaotic Attractors of Nonlinear Dynamical Systems // *Phys. Lett*, 1989, A135. – P. 343-348.
90. Pancham S., Evidence of the Multifractal Market Hypothesis Using Wavelet Transforms. Florida International University, 1994.
91. Peng C.-K., Halvin S., Hausdorff J. M. Fractal Mechanism and Heart rate dynamics // *J. of Electrocardiology*. – 1995. vol. 28 – P. 59-64.
92. Perez D., Zunino L., Garavaglia M., Rosso O., «Wavelet entropy and fractional Brownian motion time series», arXiv:physics/0501105 v1.
93. Pincus S. M. Approximate entropy as a measure of system complexity / S. M. Pincus //

- Proceedings of the National Academy of Sciences of the United States of America. – № 88 (1991). – P. 2297–2301.
94. Plerou V., Gopikrishnan P., Rosenow B., Amaral L. A. N., Guhr T., Stanley H.E. Random matrix approach to cross correlations in financial data. – *Phys.Rev.E* 2002, v.65, N 12. – P.126-142.
95. Plerou V., Gopikrishnan P., Rosenow B., Amaral L. A. N., Stanley H.E. Econophysics: Financial time series from a statistical physics point of view // *Physica A* . – № 279. – 2000. – P. 443-456.
96. Quian Quiroga R. Wavelet entropy: a measure of order in evoked potentials / Quian Quiroga R., Rosso O. A., Başar E. // *Electroencephalogr Clin Neurophysiol Suppl.* – 1999. – № 49. – P. 299–303.
97. Rachev S. T., Weron A., Weron R. CED Models for Asset Returns and Fractal Market Hypothesis. *Mathematical and Computer Modelling* №29, 1999.
98. Renyi A. *Probability Theory.*–Amsterdam: North-Holland, 1970.
99. Richman J. S. Physiological time series analysis using approximate entropy and sample entropy [Electronic resource] / Richman J. S., Moorman J. R. // *Am. J. Physiol.* – 2000. – 278 (6). – P. H2039-H2049. – Mode of access : <http://ajpheart.physiology.org/cgi/content/abstract/278/6/H2039>.
100. Rosso O. A, Figliola A. «Order/disorder in brain electrical activity», *Revista Mexicana deFisica* 50 (2) 149-155, Abril 2004.
101. Ruelle D., Takens F. On the nature of turbulence // *Comm. Math. Phys.*, V. 20, 1971. – P. 167-192.
102. Sachs J., Tornell A., Velasco A. Financial crises in emerging markets: The lesson from 1995 // *Brooking Papers on Economic Activity*, V.1,1995. – P.147-198.
103. Sample entropy analysis of neonatal heart rate variability / Douglas E. Lake, Joshua S. Richman, M. Pamela Griffin, J. Randall // *Am. J. Physiol. Integr. Comp. Physiol.* – V. 283. – 2002. – P. 789–797.
104. Sapsin V. M. Heisenberg uncertainty principle and economic analogues of basic physical quantities. / V. M. Sapsin, V. M. Soloviev // *Computer Modelling and New Technologies.* – 2011. – Vol. 15, № 3. – P.21-26.
105. Sello S. Wavelet entropy as a measure of solar cycle complexity / S. Sello // *arXiv:physics/0501105.* – V. 1.
106. Shannon C. E. A mathematical theory of communications // *Bell Systems Tech. J.*, 1948. – V. 27. – P.623-656.
107. Sharpe W. F. Capital Asset Price: A Theory of Market Equilibrium under Conditions of Risk//*Journal of Finance*, V. 29 (3),1964. – P. 425-442.
108. Sheinkman, J., LeBaron, B. Non-linear dynamics and stock returns // *Journal of Business*, V.62, 1989. – P. 311-327.
109. Soloviev V. Quantum econophysics – problems and new conceptions / V. Soloviev, V. Sapsin, I. Stratiychuk // *Вісник КНУТД.* – 2012, № 5. – С.243-248.
110. Sornette D. «Stock Market Crashes, Precursors and Replicas», *J. Phys. I France*, V.6, 1996. – P. 167-175.
111. Sornette D. Critical market crashes // *Physics Reports*, V. 378, 2003. – P. 1-98.
112. Struzik Z. R., Local Effective Hoelder Exponent Estimation on the Wavelet Transform Maxima Tree, in *Fractals: Theory and Applications in Engineering*, Eds: M. Dekking, J. L´evy V´ehel, E. Lutton, C. Tricot. – Springer Verlag, 1999. – P. 93–112.
113. Theiler J. Estimating the fractal dimension of chaotic time series. *Lincoln Lab. J.* 3, 63-86 (1990).
114. Theiler J. Estimating the fractal dimension of chaotic time series // *Lincoln Lab. J.*, V. 3, 1990. – P.63-86.

115. Tsallis C. Nonextensive Statistics: Theoretical, Experimental and Computational Evidences and Connections / C. Tsallis // *Brazilian Journal of Physics*. – Vol. 29. – №. 1. – March, 1999.
116. Tsallis, C. Entropic nonextensivity: a possible measure of complexity // *Chaos, Solitons, & Fractals*, V. 13, 2002. – P.371-391.
117. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics // *Journal of Statistical Physics*, V. 52, 1988. – P.479-487.
118. Vaga T. The Coherent Market Hypothesis // *Financial Analysts Journal*, V. 46, №6, 1991. – P.36-49.
119. Watts D. J. Collective dynamics of «small-world» networks / D. J. Watts, S. H. Strogatz // *Nature*. – 1998. – Vol. 393. – P. 440 – 442.
120. Wavelet entropy of stochastic processes / Zunino L., Perez D. G., Garavaglia M., Rosso O. A. // e-print: arXiv:physics/0603144v1 [physics.data-an] 17 Mar 2006.
121. Weidlich W. *Socio-Dynamics: A Systematic Approach to Mathematical Modelling in the Social Science*. Taylor and Francis, London, 2002. (російський переклад Вайдліх В. М. Социодинамика. Системный подход к математическому моделированию в социальных науках. – М. : Едиториал УРСС, 2005. – 480 с.).
122. Wolf A., Swift J. B., Swinney H. L., Vastano J. A. Determining Lyapunov exponents from a time series // *Physica D*, V. 16, 1985. – P.285-317.
123. Zbilut J.P., Zaldívar-Comenges J.-M., Stozzi F. Recurrence quantification based Liapunov exponent for monitoring divergence in experimental data // *Phys. Lett. A*, V. 297, 2002. – P. 173-181.
124. Автономов В.С. Методологические проблемы современной экономической науки // *Вестник РАН*, т. 76, №3, 2006. – С. 203-208.
125. Бак П. Самоорганизованная критичность / П. Бак, К. Чен // *В мире науки*, № 3, 1991. – С. 16-24.
126. Бокс Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. – М. : Мир, 1974. – Вып. 1, 2.
127. Бурлачков В. Экономическая наука и эконофизика: главные темы диалога // *Вопросы экономики*, №12, 2007. – С. 111-122.
128. Владимиров В. А., Воробьев Ю. Л., Малинецкий Г. Г., Посашков А. С. и др. *Управление риском. Риск, устойчивое развитие, синергетика*. – М. : Наука, 2000 – 432 с.
129. Гальчинський А. С. Методологія складних систем [Електронний ресурс] / А. С. Гальчинський // *Економіка України*. – 2007. – № 8. – С. 4–18. – Режим доступу : [http://tourlib.net/statti\\_others/galchynsky.htm](http://tourlib.net/statti_others/galchynsky.htm).
130. Дербенцев В. Д. Синергетична парадигма дослідження складних фінансово-економічних систем / В. Д. Дербенцев // *Моделювання та інформаційні системи в економіці*. Зб. наук. праць. Вип. 74. – К. : КНЕУ, 2006. – С.8-20.
131. Дербенцев В. Д., Ганчук А. А., Соловйов В. М. Моделювання критичних та кризових явищ на фінансових ринках // *Фінансова система України*. Зб. наук.праць. – Острог: Вид-во «Національний університет «Острозька академія», 2006. – Вип.8. – Ч.3. – С.297-304.
132. Дербенцев В. Д., Сердюк О. А., Соловйов В. М. Передвісники критичних явищ в складних економічних системах // *Зб. наук. статей «Моделювання нелінійної динаміки економічних систем»*. – Донецьк, ДонНУ, 2005.– № 1.– С.5-13.
133. Евин И. А. Введение в теорию сложных сетей. / Е. И.Евин // *Математические основы и численные методы моделирования*. – 2010. – Т.2, №2. – С.121-141.
134. Занг В.-Б. Синергетическая экономика. Время и переменны в нелинейной экономической теории / В.-Б. Занг ; пер. с англ. Н. В. Островской под ред. В. В. Лебедева и В. Н. Разжевайкина. – М. : Мир, 1999. – 335 с. : ил.

135. Клеменс М. П. Прогнозирование в макроэкономике / М. П. Клеменс, Д. Ф. Хэндри // Обозрение прикладной и промышленной математики. – т. 3. – вып. 6. – 1996. – С. 859-898.
136. Князева Е. Н. Законы эволюции и самоорганизации сложных систем / Е. Н. Князева, С. П. Курдюмов. – М. : Наука, 1994. – 238 с.
137. Колмогоров А. Н. Новый метрический инвариант транзитивных динамических систем и автоморфизмов пространства Лебега. ДАН СССР, т.119, с.861-864, 1958.
138. Кроновер Р. М. Фракталы и хаос в динамических системах. Основы теории. Москва : Постмаркет, 2000. – 352с.
139. Кузнецов С. П. Динамический хаос (курс лекций). – М. : Издательство Физико-математической литературы, 2001. – 296 с.
140. Ландау Л. Д. Квантовая механика. Нерелятивистская теория / Л. Д. Ландау, Е. М. Лифшиц // – М : Наука, 1973. – 750 с.
141. Ландэ Д. В., Снарский А. А., Безсуднов И. В. Интернетика: Навигация в сложных сетях: модели и алгоритмы. – М. : Книжный дом «ЛИБРИКОМ», 2009. – 264 с.
142. Лиховидов В. Н. Фундаментальный анализ мировых валютных рынков: методы прогнозирования и принятия решений. – Владивосток, 1999. –234 с.
143. Малинецкий Г. Г. Теория самоорганизации. На пороге IV парадигмы / Г. Г. Малинецкий // Компьютерные исследования и моделирование. – 2013. – Т. 5. – № 3. – С. 315–366.
144. Малинецкий Г. Г., Курдюмов С. П. Нелинейная динамика и проблемы прогноза // Вестник РАН, 2001, т.71, №3. – С.210-232
145. Малинецкий Г. Г., Потапов А. Б., Подлазов А. В. Нелинейная динамика: Подходы, результаты, надежды. – М. : КомКнига, 2006 – 280 с.
146. Мандельброт Б. Фрактальная геометрия природы / Б. Мандельброт. – Москва–Ижевск : Институт компьютерных исследований, 2002. – 656 с.
147. Маршалл Д. Ф., Бансал В. К. Финансовая инженерия. Пер. с англ. – М. : ИНФРА-М, 1998.
148. Матеріали інформаційного сайту OANDA з питань історичних курсів валют [Електронний ресурс]. – Режим доступу : <http://www.oanda.com/lang/ru/currency/historical-rates>.
149. Миркин Я. М. Ценные бумаги и фондовый рынок. – М. : Перспектива, 1995.
150. Монтес М. Ф., Попов В. В. Азиатский вирус или голландская болезнь? Теория и история валютных кризисов в России и других странах. Пер. с англ. – М. : АНХ при Правительстве РФ, Дело, 1999.
151. Мочерний С. Синергетичний підхід в економічному дослідженні // Економіка України. – 2001. – №5. – С.44-51.
152. Нельсон Р., Уинтер С. Эволюционная теория экономических изменений / Академия народного хозяйства при Правительстве РФ / М. Я. Каждан (пер.с англ.). – М. : Дело, 2002.
153. Нікольський Ю. В. Дискретна математика / Ю. В. Нікольський, В. В. Пасічник, Ю. М. Щербина. – Львів : «Магнолія – 2006», 2009. – 432 с.
154. Новое в синергетике: Взгляд в третье тысячелетие. / Сб. статей под ред. Г. Г. Малинецкого, С. П. Курдюмова. – М. : Наука, 2002. – 478 с.
155. Осипов А. И. Энтропия и ее роль в науке / А. И. Осипов, А. В. Уваров // Соревский образовательный журнал. – 2004. – Том 8. – № 1. – С. 70–79.
156. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка. Пер. с англ.– М. : Мир, 2000. – 333 с.
157. Петров А. А., Поспелов И. Г., Шананин А. А. Опыт математического моделирования экономики. – М. : Энергоатомиздат, 1996. – 544 с.
158. Піскун О. В Кореляційні та спектральні властивості світового валютного ринку /

- О. В. Піскун, О. А. Сердюк, Ю. В. Триус // Фінансова система України. Збірник наукових праць. – Острог : Видавництво «Національний університет «Острозька академія», 2006. – Вип. 8. – Ч.3. – С. 449-454.
159. Полтерович В. М. Кризис экономической теории : доклад на научном семинаре отделения экономики и ЦЭМИ РАН [Электронный ресурс] / В. М. Полтерович. – М., 1997. – Режим доступа : <http://5fan.info/ujgmerbewbewujgmer.html>.
160. Попков В. В. Экономифизика и эволюционная экономика – перспективное направление исследований / В. В. Попков, Д. Б. Берг // Всероссийская Интернет-конференция «Проблемы Экономифизики и эволюционной экономики». 2005. [http:// www.ephes.ru/](http://www.ephes.ru/)
161. Пригожин И. От существующего к возникающему: Время и сложность в физических науках / И. Пригожин. – Перевод с английского. Серия «Синергетика: от прошлого к будущему». Изд.3 URSS, 2006. – 296 с.
162. Пригожин И. Порядок из хаоса: Новый диалог человека с природой / И. Пригожин, И. Стенгерс. – М., 1986.
163. Рибчинська О. М. Неревверсивні міри складності / О. М. Рибчинська, В. М. Соловійов, Д. М. Чабаненко // В колект. монографії «Інформаційні технології та моделювання в економіці: на шляху до міждисциплінарності». – Черкаси: Брама-Україна, 2013. – С. 100-108.
164. Рудый К. В. Финансовые кризисы: теория, история, политика. – М. : Новое знание, 2003.– 399 с.
165. Сапцин В. М. Релятивистская квантовая экономифизика. Новые парадигмы в моделировании сложных систем / В. М. Сапцин, В. Н. Соловьев. – Черкасы : Брама-Украина, 2009. – 69 с.
166. Сергеева Л. Н. Моделирование поведения экономических систем методами нелинейной динамики (теории хаоса). – Запорожье : ЗГУ, 2002. – 227с.
167. Сергеева Л. Н. Нелинейная экономика: модели и методы / Л. Н. Сергеева ; науч. редактор д. э. н., проф. Ю. Г. Лысенко. – Запорожье : Полиграф, 2003. – 218 с.
168. Сердюк О. А. Моделювання передвісників кризових явищ фінансових ринків // Проблеми і перспективи розвитку банківської системи України: Збірник наукових праць. Т. 18. – Суми: УАБС НБУ, 2006. – С. 315-321.
169. Синергетичні та економіфізичні методи дослідження динамічних та структурних характеристик економічних систем : монографія / В. Д. Дербенцев, О. А. Сердюк, В. М. Соловійов, О. Д. Шарапов. – Черкаси : Брама-Україна, 2010. – 300 с.
170. Складні мережі / Ю. Головач, О. Олемський, К. фон Фербер [та ін.] // Журнал фізичних досліджень. – 2006. – Т. 10, №4. – С. 247 – 289.
171. Соловійов В. М. Використання масштабно-залежних показників Ляпунова для дослідження складності фінансово-економічних систем / В. М. Соловійов, І. О. Стратійчук // Наука і економіка, науково-теоретичний журнал Хмельницького економічного університету, 2012. – №4 (28), т2. – С.88-93.
172. Соловійов В. М. Ентропія Тсалліса і неекстенсивні міри складності економічних систем / В. М. Соловійов, О. А. Сердюк // В колект. монографії «Моделі оцінки і аналізу складних соціально- економічних систем».-Х. : ІД «ИНЖЕК», 2013. – С. 146-157.
173. Соловійов В. М. Кількісні методи оцінки складності в прогнозуванні соціально-економічних систем / В. М. Соловійов, К. В. Соловійова // В колект. монографії: «Прогнозування соціально-економічних процесів: сучасні підходи та перспективи». Бердянськ. – 2012. – с.141-155.
174. Соловійов В. М. Рекурентні міри як метод кількісної оцінки складності / В. М. Соловійов, А. В. Батир // Вісник КНУТД, 2012, №5. – с.254-257.
175. Соловійова В. В. Порівняльний аналіз динаміки фондового ринку України з використанням фрактальних мір складності / В. В. Соловійова, В. М. Соловійов, К. В. Соловійова // Вісник Черкаського університету, сер. «Економічні науки», 2012. – №33

(246). – С.51-58.

176. Сорос Д. Мировой экономический кризис и его значение. Новая парадигма финансовых рынков / Д. Сорос; пер. с англ. К. Вагнера. – М. : Манн, Иванов и Фарбер, 2010. – 272 с.

177. Статистика індексів світового фондового ринку [Електронний ресурс]. – Режим доступу : <http://finance.yahoo.com>.

178. Федер Е. Фракталы. Пер с англ.– М. : Мир, 1991. – 254 с.

179. Фейгенбаум М. Универсальность в поведении нелинейных систем // Успехи физических наук. – 1983. – т. 141. – Вып. 2 – С. 342-347.

180. Хакен Г. Информация и самоорганизация. Макроскопический поход к сложным системам / Г. Хакен ; пер. с англ. Ю. А. Данилова. – М. : Мир, 1991. – 240 с. : ил.

181. Хакен Г. Синергетика / Г. Хакен. – М. : Мир, 1980. – 408 с.

182. Чернавский Д. С. О проблемах физической экономики / Д. С. Чернавский, Н. И. Старков, А. В. Щербаков // Успехи физических наук. – 2002. – Т. 172. – № 9. – С. 1045–1066.

183. Шарапов О. Д., Дербенцев В. Д., Соловйов В. М. Моделювання явищ самоорганізації в фінансово-економічних системах // Економіко-математичне моделювання. Вісник ТАНГ. Вип. 14. – Тернопіль : ТАНГ, 2003. – №3. – С. 104-110.

184. Шарковский А. Н., Майстренко Ю. А., Романенко Е. Ю. Разностные уравнения и их приложения. – К. : Наукова думка, 1986. – 280 с.



## ЗМІСТ

Передмова .....	3
ЛАБОРАТОРНА РОБОТА № 1. Аналіз флуктуацій часового ряду.....	6
ЛАБОРАТОРНА РОБОТА № 2. Фрактальний і мультифрактальний аналіз часового ряду.....	15
ЛАБОРАТОРНА РОБОТА № 3. Використання рекурентного аналізу для моделювання і прогнозування нелінійних динамічних властивостей складних систем.....	28
ЛАБОРАТОРНА РОБОТА № 4. Кількісний аналіз рекурентних діаграм .....	37
ЛАБОРАТОРНА РОБОТА № 5. Дослідження процесів самоорганізації в складних системах.....	42
ЛАБОРАТОРНА РОБОТА № 6. Інформаційні методи оцінки складності економічних систем.....	58
ЛАБОРАТОРНА РОБОТА № 7. Ентропійний аналіз часового ряду.....	67
ЛАБОРАТОРНА РОБОТА № 8. Визначення індексу фрактальності.....	81
ЛАБОРАТОРНА РОБОТА № 9. Дослідження асиметрії (необоротності) часу складних сигналів.....	84
ЛАБОРАТОРНА РОБОТА № 10. Дослідження процесів післякризової релаксації складних систем.....	90
ЛАБОРАТОРНА РОБОТА № 11. Кластерний аналіз.....	96
ЛАБОРАТОРНА РОБОТА № 12. Структурні характеристики складної системи.....	107
ЛАБОРАТОРНА РОБОТА № 13. Найпростіші мережі та мережні міри складності....	112
ЛАБОРАТОРНА РОБОТА №14. Візуалізація мережних моделей.....	120
ЛАБОРАТОРНА РОБОТА № 15. Графодинаміка складних систем.....	128
ЛАБОРАТОРНА РОБОТА № 16. Використання енергій вейвлет-коефіцієнтів 2-го та 3-го моментів при дослідженні криз і трендів.....	139
ЛАБОРАТОРНА РОБОТА № 17. Вейвлет-аналіз нестационарних часових рядів.....	143
ЛАБОРАТОРНА РОБОТА № 18. Використання методів нелінійної динаміки для дослідження динаміки складних соціально-економічних систем.....	153
ЛАБОРАТОРНА РОБОТА № 19. Використання масштабно-залежних показників Ляпунова при дослідженні складних систем.....	167
ЛАБОРАТОРНА РОБОТА № 20. Дослідження процесів синхронізації в часових рядах.....	173
ЛАБОРАТОРНА РОБОТА № 21. Агентне моделювання фінансових ринків.....	180
ЛАБОРАТОРНА РОБОТА № 22. Використання принципу невизначеності Гейзенберга для прогнозування кризових явищ в складних системах.....	187
Список рекомендованої літератури.....	195