

Міністерство освіти і науки України  
Запорізька обласна державна адміністрація

Шифр: „ ..... ”

Тема роботи: „Кластеризація часових рядів показників банків”

Напрямок: .....

Виконала:

ст.гр.ФЕУ-113м Шлендер К.О.

2017 рік

## ЗМІСТ

ВСТУП.....	
РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ КЛАСТЕРИЗАЦІЇ ЧАСОВИХ РЯДІВ ПОКАЗНИКІВ БАНКІВ.....	
1.1 Теоретичні засади аналізу часових рядів показників банків	
1.2 Забезпечення порівнянності рівнів часових рядів	
РОЗДІЛ 2 КЛАСТЕРИЗАЦІЯ ЧАСОВИХ РЯДІВ ПОКАЗНИКІВ БАНКІВ	
2.1 Дослідження часових рядів показників банків	
2.2 Аналіз отриманих результатів	
РОЗДІЛ 3	
ВИСНОВКИ.....	
ПЕРЕЛІК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ.....	

## ВСТУП

З метою визначення фінансового стану окремих банків та аналізу функціонування банківської системи в цілому важливим є дослідження динаміки (трендів) заданих показників діяльності банків упродовж певних часових інтервалів.

Особливо актуальним це завдання є в періоди високої ринкової нестабільності, коли, з одного боку, можуть виникати загальносистемні тенденції, характерні для більшості банків (наприклад, вплив коштів), а з другого, – протилежні тенденції можуть компенсувати одна одну та спричиняти недостатню інформативність узагальненого тренду. Важливо те, що загальносистемний тренд є середньозваженим результатом, на який найбільший вплив мають великі системні банки, що, в свою чергу, негативно впливає на репрезентативність загальносистемного тренду.

Таким чином, об'єктом аналізу має бути сукупність трендів та визначення ключових, схожих за своїми характеристиками, тенденцій та їхній вплив на загальносистемну тенденцію.

# РОЗДІЛ 1

## ТЕОРЕТИЧНІ ОСНОВИ КЛАСТЕРИЗАЦІЇ ЧАСОВИХ РЯДІВ ПОКАЗНИКІВ БАНКІВ

### 1.1. Теоретичні засади аналізу часових рядів показників банків

Часовий ряд — це послідовність впорядкованих у часі числових показників, що характеризують рівень стану і зміни досліджуваного явища. Аналіз часових рядів — сукупність математико-статистичних методів аналізу, призначених для виявлення структури часових рядів і для їх прогнозування. Сюди належать, зокрема, методи регресійного аналізу. Виявлення структури часового ряду необхідно для того, щоб побудувати математичну модель того явища, яке є джерелом аналізованого часового ряду.

Прогноз майбутніх значень часового ряду використовується для ефективного прийняття рішень.

Кластерний аналіз — це багатовимірна статистична процедура, яка виконує збір даних, що містять інформацію про вибірку об'єктів і потім упорядковує об'єкти в порівняно однорідні групи — кластери (Q-кластеризація, або Q-техніка, власне кластерний аналіз). Основна мета кластерного аналізу — знаходження груп схожих об'єктів у вибірці. Громадські явища можна вивчати в двох розрізах: у статичному і динамічному. Ряди розподілу, які отримують в результаті зведення та групування статистичних даних, відносяться до одного періоду або моменту часу і вивчаються в статиці. Тут час бере участь у пасивній формі. Якщо ж дані відносяться до різних періодів або моментів, великий інтерес викликає порівняння даних у часі, яке набуває тут вирішальне значення.

Відомо, що будь-яке явище може бути правильно зрозуміло, якщо його вивчати в русі та розвитку. При вирішенні будь-якого питання, при

аналізі будь-якого явища важливо знати, як воно виникло, розвивалося і розвивається. Тільки за цих умов можна вирішити питання про перспективи його розвитку.

У процесі розвитку змінюються розміри, склад, обсяг, структура конкретних суспільних явищ. Тому одним з найважливіших завдань статистики є вивчення цих змін: процесу їх розвитку, їх динаміка. Це завдання статистика вирішує шляхом побудови й аналізу тимчасових рядів.

Тимчасові ряди мають величезне значення для виявлення і вивчення складаються закономірностей у розвитку явищ економічного, політичного і культурного життя суспільства.

Існують дві основні мети аналізу часових рядів: визначення природи ряду і прогнозування (проорокування майбутніх значень часового ряду по теперішнім і минулим значенням). Обидві ці цілі вимагають, щоб модель ряду була ідентифікована і, більш-менш, формально описана. Як тільки модель визначена, ви можете з її допомогою інтерпретувати представлені дані (наприклад, використовувати у вашій теорії для розуміння сезонного зміни цін на товари, якщо займаєтеся економікою). Не звертаючи уваги на глибину розуміння і справедливості теорії, ви можете екстраполювати потім ряд на основі знайденої моделі, тобто передбачити його майбутні значення.

Часові ряди досліджуються з різними цілями. В одному ряді випадках буває достатньо отримати опис характерних особливостей ряду, а в іншому ряді випадків потрібне не тільки передбачати майбутні значення часового ряду, а й управляти його поведінкою. Метод аналізу часового ряду визначається, з одного боку, цілями аналізу, а з іншого боку, ймовірнісної природою формування його значень.

Найпоширеніші методи аналізу часових рядів:

– спектральний аналіз - дозволяє знаходити періодичні складові часового ряду

- кореляційний аналіз - дозволяє знаходити суттєві періодичні залежності і відповідні їм затримки (лаги) як всередині одного ряду (автокореляція), так і між кількома рядами. (Кроскореляції)

- моделі авторегресії і ковзного середнього - моделі орієнтовані на опис процесів, що виявляють однорідні коливання, порушені випадковими впливами. Дозволяють передбачати майбутні значення ряду.

- багатоканальні моделі авторегресії і ковзного середнього - моделі застосовуються в тих випадках, коли є кілька корельованих між собою часових рядів. У них є коливання, порушені однією причиною. Дозволяють передбачати майбутні значення ряду.

- сезонна модель Бокса-Дженкінса - застосовується, коли часовий ряд містить явно виражений лінійний тренд і сезонні складові. Дозволяє передбачати майбутні значення ряду. Модель була запропонована у зв'язку з аналізом авіаперевезень.

- прогноз експоненціально зваженим ковзаючим середнім - найпростіша модель прогнозування часового ряду. Застосовна в багатьох випадках. У тому числі, охоплює модель ціноутворення на основі випадкових блукань.

Зазвичай при практичному аналізі часових рядів послідовно проходять такі етапи:

- величину інвестиційних вкладень підприємства у звітному періоді та рівень їх фінансування за рахунок внутрішніх фінансових джерел;

- обсяг фінансових ресурсів, які підприємство залучило із зовнішніх джерел;

- графічне подання і опис поведінки часового ряду.

- виділення та видалення закономірних складових часового ряду, що залежать від часу: тренда, сезонних і циклічних складових.

- виділення та видалення низько- або високочастотних складових процесу (фільтрація).

- дослідження випадкової складової часового ряду, що залишилася після видалення перерахованих вище складових.
- побудова (підбір) математичної моделі для опису випадкової складової і перевірка її адекватності.
- прогнозування майбутнього розвитку процесу, представленого часовим рядом.

Всякий часовий ряд включає два обов'язкових елемента: по-перше, час  $t$ , по-друге, конкретне значення показника, або рівень ряду.

Часові ряди розрізняються за такими ознаками: за часом — моментні та інтервальні. Інтервальний ряд — послідовність, в якій рівень явища відносять до результату, накопиченому або знову виробленому за певний інтервал часу. Такі, наприклад, такі ряди показників обсягу продукції підприємства по місяцях року, кількості відпрацьованих людиною днів по окремих періодах (місяцях, кварталах, півріччях, роках, тощо) і т. д. Якщо ж рівень ряду характеризує досліджуване явище в конкретний момент часу, то сукупність рівнів формує моментний ряд.

## **1.2. Забезпечення порівнянності рівнів часових рядів**

Найважливішою умовою правильного формування часових рядів є порівнянність рівнів, що утворюють ряд. Рівні ряду, що підлягають вивченню, повинні бути однорідні за економічним змістом і враховувати сутність досліджуваного явища і його мету.

Статистичні дані, представлені у вигляді часових рядів, повинні бути порівняні по території, колу охоплених об'єктів, одиницям вимірювання, моменту реєстрації, методикою розрахунку, цінами, достовірності. Неспівмірність за територією виникає в результаті змін кордонів країн, регіонів, господарств тощо. Для приведення даних до порівнянного вигляду проводиться перерахунок колишніх даних з урахуванням нових кордонів.

Цілісність охоплення різних частин явища — найважливіша умова порівнянності рівнів ряду. Вимога однакової цілісності охоплення різних частин досліджуваного об'єкта означає, що рівні ряду за окремі періоди повинні характеризувати різноманітність заходів того чи іншого явища по одному і тому ж колу, які входять до складу його частин. Наприклад, при характеристиці динаміки врожайності овочевих культур у регіоні по роках не можна в одні роки враховувати тільки сільськогосподарські підприємства, а в інші — всі категорії господарств, у тому числі присадибні ділянки сільських жителів і сади, городи городян.

При визначенні порівнюваних рівнів ряду необхідно використовувати єдину методику їх розрахунку. Особливо часто ця проблема виникає при міжнародних порівняннях.

Неспівмірність показників, що виникає в силу неоднакового застосовуваних одиниць виміру, сама по собі очевидна. З різницею застосовуваних одиниць вимірювання доводиться зустрічатися при вивченні динаміки: виробничих ресурсів, коли вони представляються то у вартісному, то в трудовому обчисленні; енергетичних потужностей (кВт-год, л.с.); атмосферного тиску і т. д.

Труднощі припорівнянні даних по моменту реєстрації виникають через сезонні явища. Чисельність худоби в домашніх господарствах через економічну доцільність відмінна взимку і влітку, тому рівні при порівнянні повинні ставитися до певної дати щорічно.

При аналізі показників у вартісному вираженні слід враховувати, що з плином часу відбувається безперервна зміна цін. Причин цього процесу безліч — інфляція, зростання витрат, ринкові умови (попит і пропозиція) і т. д. У цьому зв'язку при характеристиці вартісних показників обсягів продукції в часі має бути усунуто вплив трансформаційних змін цін. Для вирішення цього завдання кількість продукції, виробленої в різні періоди, оцінюють в цінах одного періоду, які називають фіксованими або в визначених в статистичних органах — порівнянними цінами.



Широке використання в статистичних дослідженнях вибіркового методу вимагає враховувати достовірність кількісних і якісних характеристик досліджуваних явищ в динаміці. Різна репрезентативність вибірки за періодами внесе суттєві похибки в величини рівнів ряду.

Однією з умов порівнянності рівнів інтервального ряду, крім рівності періодів, за які наводять дані, є однорідність етапів, у межах яких показник підпорядковується одному закону розвитку. У цих випадках проводять періодизацію часових рядів, типологічну угруповання в часі. Всі вищеназвані обставини слід враховувати при підготовці інформації для аналізу змін явищ у часі (динаміці).

Зазвичай, метою прикладного статистичного аналізу часових рядів є побудова моделі ряду, за допомогою якої можна пояснити поведінку ряду і здійснити прогноз на майбутні періоди.

Аналіз часового ряду починається з побудови і вивчення його графіка. Якщо нестационарність часового ряду очевидна, то спочатку необхідно виокремити його нестационарну складову. Процес виокремлення тренду та інших компонент ряду, що призводять до порушення стаціонарності, може проходити в декілька етапів. На кожному з них розглядається ряд залишків, отриманий у результаті вирахування з вихідного ряду підібраної моделі тренду, або результат різницевого і інших перетворень ряду. Крім графіків, ознаками нестационарності часового ряду можуть служити автокореляційна функція, що прямує не до нуля (за винятком дуже великих значень лагів) і наявність яскраво виражених піків на низьких частотах у періодограмі. За допомогою автокореляційної функції досліджують також внутрішні зв'язки між елементами часових рядів.

У вибіркових дослідженнях найпростіші числові характеристики описової статистики (середнє, медіана, дисперсія, стандартне відхилення, коефіцієнти асиметрії й ексцесу) звичайно дають достатньо інформативне уявлення про вибірку. Графічні методи зображення й аналізу вибірок при

цьому грають лише допоміжну роль, дозволяючи краще зрозуміти локалізацію і концентрацію даних, їхній закон розподілу.

Роль графічних методів при аналізі часових рядів цілком інша. Табличне представлення часового ряду й описових статистик частіше за все не дозволяє зрозуміти характер процесу, у той час як за графіком часового ряду можна зробити досить багато висновків. Надалі вони можуть бути перевірені й уточнені за допомогою розрахунків.

Після того, як вихідний процес максимально наближений до стаціонарного, можна приступити до вибору різноманітних моделей отриманого процесу. Мета цього етапу — опис і урахування надалі аналізу кореляційної структури аналізованого процесу. Модель може вважатися підбраною, якщо залишкова компонента ряду є процесом типу, як правило, «білого шуму». Після підбору залишки аналізуються для перевірки адекватності моделі та побудови надійних інтервалів.

Останнім етапом аналізу часового ряду може бути прогнозування його майбутніх (екстраполяція) або відновлення пропущених (інтерполяція) значень і визначення точності цього прогнозу на базі підбраної моделі. Добре підібрати математичну модель вдається не для всякого часового ряду. Нерідко буває і так, що для опису підходять відразу декілька моделей. Неоднозначність вибору моделі може спостерігатися як на етапі виділення детермінованого компонента ряду, так і при виборі структури ряду залишків. Тому досить часто розробляють декілька прогнозів, зроблених за допомогою різних моделей.

## РОЗДІЛ 2

### КЛАСТЕРИЗАЦІЯ ЧАСОВИХ РЯДІВ ПОКАЗНИКІВ БАНКІВ

#### 2.1 Дослідження часових рядів показників банків

Позначимо  $z^t$  значення заданого показника  $Z$  на певний момент часу  $t=1, \dots, T$ , для заданого банку  $n=1, \dots, N$ , де  $T$  є загальною кількістю моментів часу досліджуваного періоду;  $N$  – загальною кількістю банків.

У разі абсолютної природи показника  $Z$ , аби мати можливість порівняння різних банків, сформуємо відносний показник виду усереднене за часом значення показника  $i$ -го банку.

За змістом показник  $z$  відображає відносний рівень показника в кожний окремий момент часу щодо середнього рівня за заданий період та змінюється навколо 1. Для лаконічності у подальшому не використовуватимемо верхню риску.

Визначимо міру відстані між часовими послідовностями значень досліджуваного показника для банків  $i$  та  $j$ , або інакше, міру несхожості двох часових послідовностей як:

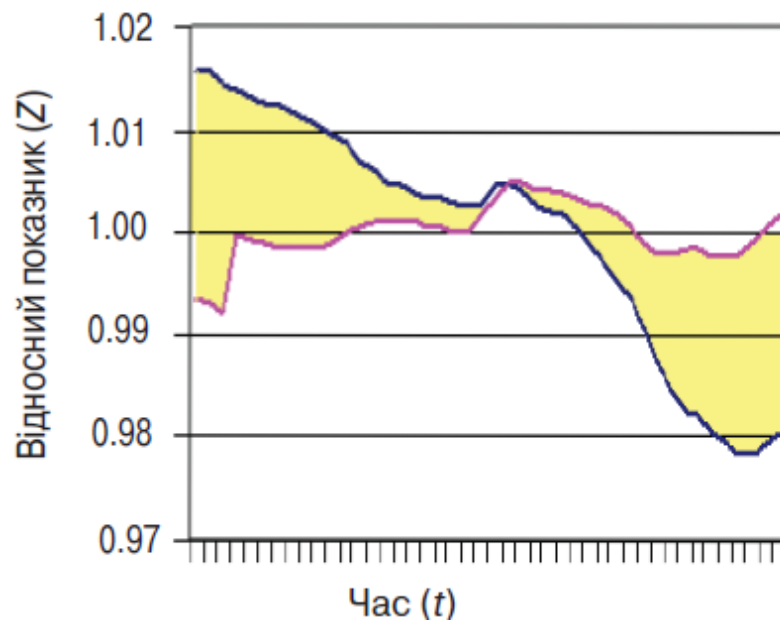
$$d(i, j) = . \quad (2)$$

Ця відстань пропорційна квадрату площі між кривими трендів (див. графік 1) та дорівнюватиме нулю лише в разі рівності трендів. Відстань буде тим більшою, чим більші відмінності трендів між собою.

Зауважимо, що запропонована метрика відстані жодним чином не використовує часової природи досліджуваних показників. Точнішим, але водночас і складнішим підходом для порівняння часових послідовностей є, наприклад, порівняння їхніх спектральних характеристик, що додатково дало б змогу ідентифікувати наявність циклічності показників.

Зобразити взаємне розташування сукупності банків з урахуванням визначених формулою (2) відстаней можемо у вигляді ненаправленого графа

[1], [2]. Його вершина відповідатиме окремому банку, а ребра – відстані між банками (див.схему). Такий граф є повним – для кожного банку визначено відстань до кожного іншого банку. Тому для наочності варто залишити лише певну кількість найближчих банків (із найбільш схожими трендами). Оскільки для кожного банку визначено найближчі банки, такий граф точно відобразить всі банки.



**Графік 1. Визначення відстані між часовими послідовностями різних банків як площі між кривими**

Задача кластеризації банків у сенсі аналізу часової динаміки показника  $Z$  полягає в пошуку близьких у сенсі (2) трендів банків та визначення характерних однорідних груп банків – кластерів банків. При цьому відкритим залишається питання кількості кластерів, на які слід розподіляти досліджувані об'єкти. Кількість таких кластерів не може бути великою. Водночас кластери повинні у повному обсязі відображати характерну кількість відокремлених груп близьких об'єктів.

Математичне визначення оптимальної кількості кластерів має таку послідовність. Визначимо внутрішньокластерну відстань  $k$ -го кластера як суму всіх відстаней у межах цього кластера:

$$w_k = d(i, r_k) \quad (3)$$

де  $r_k$  – центр  $k$ -го кластера. На схемі три кластери  $k = 1, 2, 3$  позначено прямокутниками. Їхні внутрішньокластерні відстані відповідають сумарній довжині ребер усередині цих кластерів.

Загальна внутрішньокластерна відстань є сумою відстаней, визначених у (3):

$$w(K) = (4)$$

де  $K$  – загальна кількість кластерів.

Зі збільшенням  $K$  ця функція спадає, оскільки відстань до центрів кластерів зі збільшенням їхньої кількості зменшується. У граничному випадку, коли кількість кластерів дорівнює кількості банків (тобто кожен кластер складається з одного банку):

$$w(K = N) = 0.$$

Визначимо міжкластерну відстань як суму відстаней між центрами кластерів:

$$b(K) = d(r_k, r_0) \quad (5)$$

де  $r_0$  – центр системи. На схемі міжкластерна відстань відповідає сумі довжин ребер позначених штрих-пунктиром. Ця функція зростає зі збільшенням кількості кластерів та у граничному випадку, коли кількість кластерів дорівнює кількості банків

$$b(K = N) = w(K = 1).$$

Визначення оптимальної кількості кластерів  $K$  полягає в мінімізації загальної внутрішньокластерної та максимізації міжкластерної відстані, що досягається у граничному випадку.

Отже, оптимальна кількість кластерів має бути обмежена з інших міркувань, наприклад, тією кількістю, що забезпечує найбільше зменшення загальної внутрішньокластерної відстані за умови однакового зростання

міжкластерної відстані. В ролі такого критерію може бути використано мінімум функції:

$$w(K) - b(K), (6)$$

який досягається при рівності  $w(K), b(K)$ .

Остаточо питання щодо кількості необхідних кластерів треба вирішувати експертним шляхом, враховуючи при цьому специфіку конкретного застосування. Так, у випадку аналізу часової поведінки показників поділ на три категорії є інтуїтивним: висхідного, сталого (із певним визначенням сталості) та низхідного трендів. Перевірку правильності такої категоризації, визначення необхідності й можливості подальшого “подрібнення” категорій аналітик може здійснити зокрема за допомогою пропонованої методики.

Побічним результатом її застосування є загальна метрика несхожості всіх трендів банків між собою, а саме загальна внутрішньокластерна відстань  $w(K = 1)$ . Ця відстань для різних показників  $Z$  є виміром нестабільності показника протягом досліджуваного періоду. Великий вплив на величину  $w(K = 1)$  мають тренди зі значними відхиленнями від загальносистемного тренду. Знівелювати цей вплив можна шляхом порівняння міжкластерної відстані для певної кількості сформованих кластерів, наприклад,  $w(K = 5)$ . У такому випадку для дуже віддалених трендів утворюватимуться окремі кластери, а отже, їхній вплив на загальну внутрішньокластерну відстань буде значно меншим. Велика різниця між  $w(K = 1)$  та  $w(K = 5)$  свідчить про те, що деякі банки мають значні відмінності трендів показника.

Щоденні балансові дані досліджуваних банків для уникнення випадкових флуктуацій даних перед здійсненням подальшого аналізу були згладжені методом ковзного середнього. Розмір вікна обирався з міркувань найменшого викривлення даних, щоне змінювало тижневі тренди та завдяки симетричності не викривляло дані.

Для згладжених часових послідовностей було побудовано сукупності відстаней між банками відповідно до (2).

Кластеризація здійснювалася за методом “найближчих сусідів” (K-means) [3] зі збільшенням кількості кластерів  $K$  з 1 до 10. Сутність цього методу кластеризації полягає у випадковому виборі початкових центрів кластерів та послідовному повторенні двох кроків:

- визначення приналежності банку до найближчого кластера;
- уточнення центрів кластерів як банків, що забезпечують найменшу внутрішньокластерну відстань.

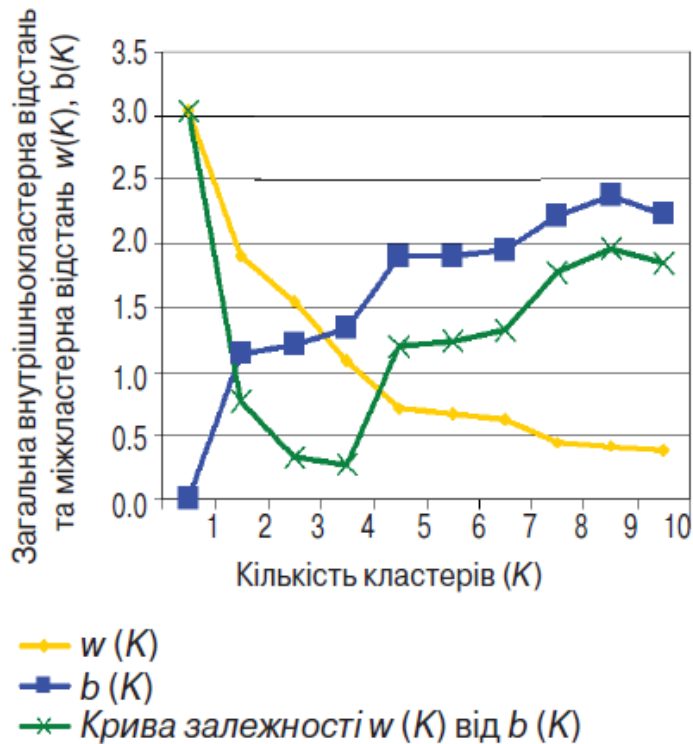
Процедура припиняється після досягнення стійкого стану – незмінності центрів кластерів. Після цього остаточно вимірювалися загальна внутрішньокластерна та міжкластерна відстані, визначені відповідно в (4) та (5).

Для уникнення можливого потрапляння в локальні мінімуми загальної внутрішньокластерної відстані описана в попередньому параграфі процедура повторювалася багато разів, а в ролі остаточного результату кластеризації обиралася конфігурація з найменшою загальною внутрішньокластерною відстанню.

Задача кластеризації банків у сенсі аналізу часової динаміки показника  $Z$  полягає в пошуку близьких у сенсі (2) трендів банків та визначення характерних однорідних груп банків – кластерів банків. При цьому відкритим залишається питання кількості кластерів, на які слід розподіляти досліджувані об’єкти. Кількість таких кластерів не може бути великою. Водночас кластери повинні у повному обсязі відображати характерну кількість відокремлених груп близьких об’єктів.

Остаточно питання щодо кількості необхідних кластерів треба вирішувати експертним шляхом, враховуючи при цьому специфіку конкретного застосування. Перевірку правильності такої категоризації, визначення необхідності й можливості подальшого “подрібнення” категорій аналітик може здійснити зокрема за допомогою пропонованої методики.

З метою перевірки правильності роботи алгоритмів було здійснено візуалізацію даних.



**Графік 2. Залежність сумарної внутрішньокластерної відстані  $w(K)$  та міжкластерної відстані  $b(K)$  від кількості кластерів**

Поряд із графіками часових послідовностей показників побудовано графі відстаней між трендами банків, аналогічні зображеному на схемі. При цьому здійснено два типи кольорового маркування вершин графів:

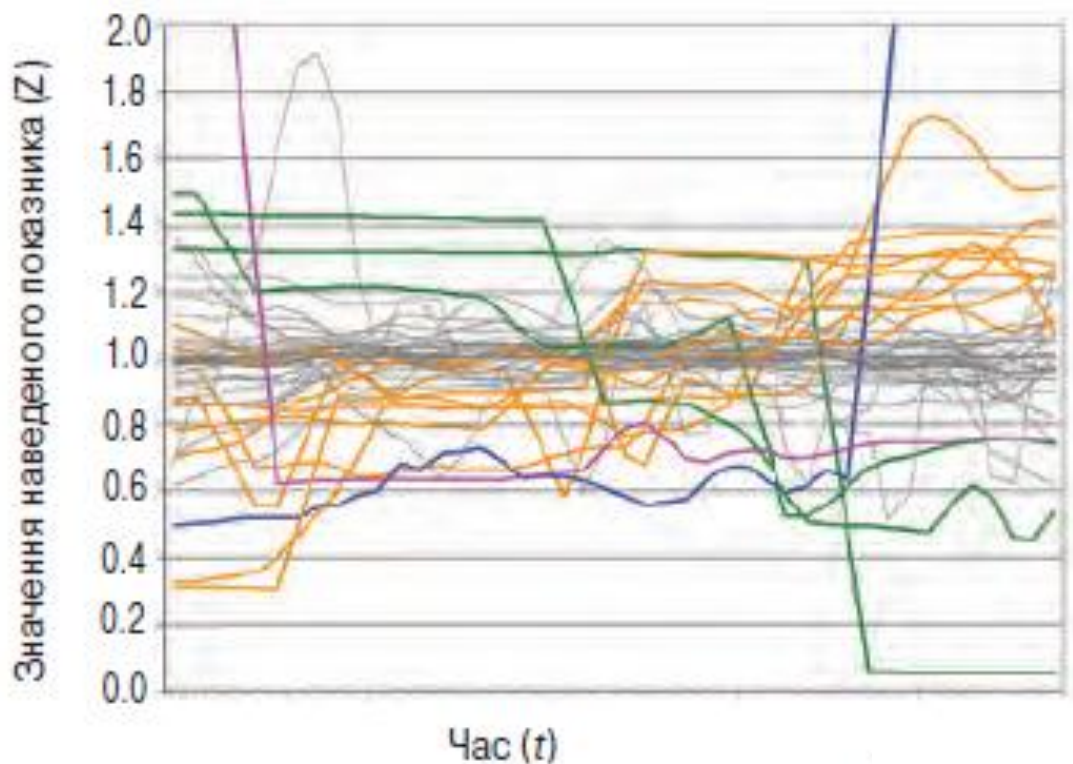
- уточнення центрів кластерів як банків, що забезпечують найменшу внутрішньокластерну відстань
- різними кольорами позначено банки з різних кластерів;
- поруч позначено категоризацію трендів на три категорії: висхідного, сталого та низхідного трендів. При цьому під сталим трендом маємо на увазі зміну показника в межах 5%, а у висхідних та низхідних трендах було додано по дві підкатегорії: зміни на 5–10% та зміни на понад 10%.

Остаточна перевірка роботи алгоритму здійснювалася шляхом маркування часових послідовностей кольором кластера, до якого ці тренди належать.



## 2.2 Аналіз отриманих результатів

Розглянемо результати аналізу на прикладі показника залишків у банках гривневих строкових коштів юридичних осіб. Часову динаміку зазначених залишків за досліджуваний період різних банків наведено на графіку 3. Граф відстаней, аналогічний зображеному на схемі, на якому вершини відповідають окремим банкам, довжина ребер – відстаням між трендами банків, а колір маркування – зміні залишків за досліджуваний період, зображено на графіку. Цей граф свідчить про те, що загальносистемний сталий тренд (зображений великим сірим колом) став результатом двох протилежних тенденцій, що були притаманні банкам. При цьому домінували банки зі зростанням залишків (зображені червоними колами).



**Графік 3. Кластерний аналіз часової динаміки залишків гривневих строкових коштів юридичних осіб 46 найбільших банків України**

Для визначення оптимальної кількості кластерів було побудовано залежності загальної внутрішньокластерної відстані (4) та міжкластерної відстані (5) від кількості кластерів  $K$  (див. графік 2). Цілком очевидно, що характеристики виявлених кластерів відповідають теоретично очікуваним: спадною є загальна внутрішньокластерна відстань  $w(K)$ , натомість міжкластерна відстань  $b(K)$  – зростає.

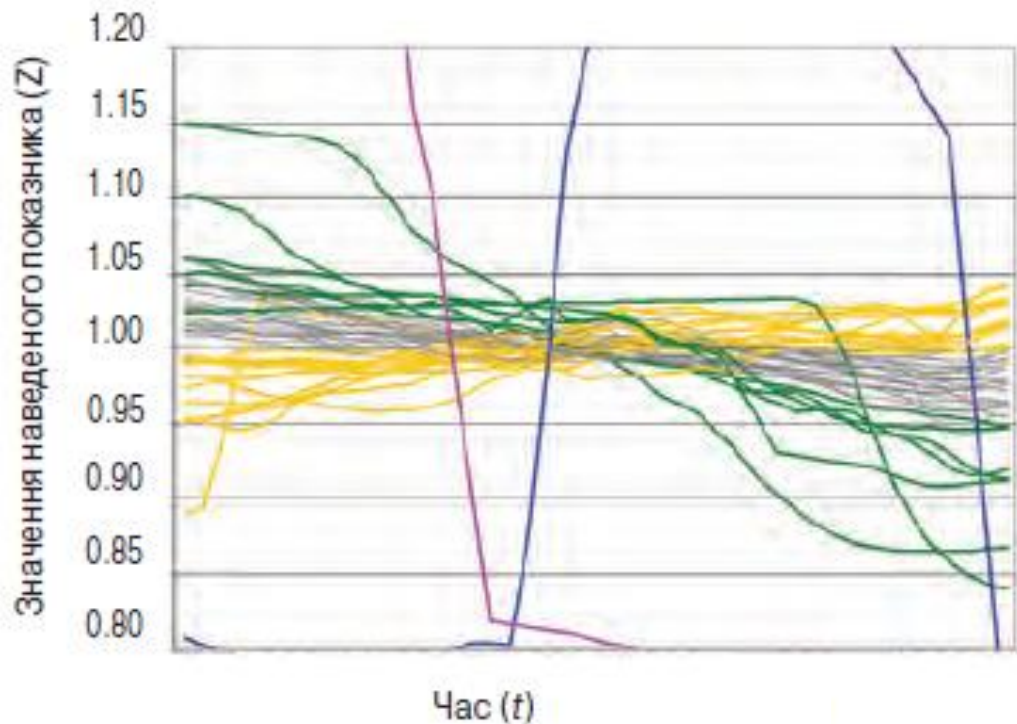
Також заслуговує на увагу те, що швидкість спадання загальної внутрішньокластерної відстані зменшується. Оптимальна кількість кластерів відповідно до мінімуму функції (6) становить  $K = 4$ . Водночас із міркувань дотримання симетрії та ізоляції разюче відмінних від загальносистемного трендів кількість кластерів було збільшено до  $K = 5$ .

Результат кластеризації з кількістю кластерів  $K = 5$  відображено на графіку, з якого видно, що існують такі кластери:

- два ізольовані кластери, кожен містить по одному банку, які мають значні відхилення від загальносистемного тренду (блакитне коло вгорі та бузкове внизу);
- група банків зі значними низхідними трендами (позначені зеленим знизу);
- група банків із висхідною динамікою залишків (позначені жовтим);
- найбільша група банків, динаміка залишків яких близька до середньої системної (позначені сірими колами).

Остаточна перевірка кластеризації здійснюється шляхом позначення окремих груп трендів на графіках часової динаміки кольорами відповідних кластерів (див. графік 3).

Аналогічні результати отримано для іншого показника – залишків строкових гривневих коштів фізичних осіб, їх наведено на графіку 4. Порівняння з попереднім результатом свідчить про те, що для фізичних осіб низхідні тренди є значно характернішими. А загалом тренди є вельми схожими між собою. Іншими словами, дії фізичних осіб з депозитними вкладками в різних банках значно однорідніші, ніж дії юридичних осіб.



**Графік 4. Кластерний аналіз часової динаміки залишків гривневих строкових коштів фізичних осіб 46 найбільших банків України**

Порівняння між собою властивостей кластеризації інших, досліджених у роботі показників, наведено в таблиці.

У таблиці наведено загальну внутрішньокластерну відстань  $w(K = 1)$ ,  $w(K = 5)$ , її відносний спад  $\delta = w(K=1) - w(K=5)w(K=1)$  та оптимальну кількість кластерів  $K_{opt}$ , визначені відповідно до мінімуму функції (6). Таблиця відсортована у зворотному порядку за відстанню  $w(K = 5)$ , яка відображає ступінь несхожості між собою трендів показників різних банків – нестабільність показників. Напівжирним виділено рядки показників, результати аналізу яких наводилися вище.

За даними таблиці бачимо, що найнестабільнішими впродовж досліджуваного періоду були залишки коштів юридичних осіб, адже їх динаміка для різних банків є найбільш різноманітною. Водночас більшу нестабільність мали кошти в доларах США. Динаміка строкових коштів,

залучених від юридичних осіб, суттєво не відрізняється від динаміки відповідних коштів на вимогу. Для фізичних осіб, на відміну від юридичних осіб, динаміка строкових коштів є вельми однорідною, серед яких найбільш однорідною була динаміка залишків коштів у доларах США. Цей факт можемо пояснити адміністративними заходами Національного банку, що діяли в цей період. Відносна змінна  $\delta$  відображає те, як швидко зменшується загальна внутрішньокластерна відстань при збільшенні кількості кластерів та свідчить про схильність до кластеризації трендів різних показників або про наявність помітно відмінних від загалу трендів. Як результат для великих значень  $\delta$  відповідними є малі значення оптимальної кількості кластерів Корт. Зокрема, за даними таблиці бачимо, що найбільшу схильність до групування мають залишки гривневих строкових коштів.

Додаткового дослідження потребує обґрунтування вибору тривалості досліджуваного періоду ( $T$ ). Спостереження за зміною параметрів кластеризації при зміні досліджуваного періоду в майбутньому допоможе виявити зміну динаміки показників – появу нових чи зникнення існуючих кластерних груп.

Формування системи показників для оцінки фінансового стану повинне ґрунтуватися на економіко-статистичному дослідженні тенденцій розвитку банківської системи. Потрібно досліджувати не тільки тенденції розвитку банківської системи, але й тенденції розвитку окремих банківських груп.

Банки, що діють на одному локально обмеженому територіальному ринку, досить залежні для того, щоб поведінка одних банків впливала на інші. Банківський ринок України має неоднорідну структуру, разом із великими, досить потужними банками існує і велика кількість малих «кишенькових» банків.

Особливість банківської системи України полягає ще і в тому, що існують великі державні банки, які оперують великим обсягом державних коштів, і їхній вплив на банківський ринок не можна не помітити.

Банківська система України є досить спотвореною версією ідеальної банківської системи: банкрутство одних банків не веде до колапсу банківської чи фінансових систем. Протягом періоду існування України спостерігається постійне виведення коштів із банків-банкротів та перелив їх в інші фінансові фонди, у результаті ми маємо вичерпання фінансових фондів Фонду гарантування вкладів України. Таким чином, система резервування грошових коштів, величина регулятивного капіталу не гарантують безпеку вкладників, банки не відповідають за своїми зобов'язаннями.

Повертаючись до теоретичних питань дослідження, потрібно зазначити, що для підтвердження гіпотези про схожість поведінки банків-банкротів, банків – успішних підприємців потрібно проводити ще додаткові дослідження для виокремлення взаємозв'язків між банками за іншими показниками, крім показників власного капіталу. Відповідне майбутнє розроблення економетричної моделі, що буде розраховуватися в режимі реального часу за отриманими результатами, повинне стати пріоритетом для того, щоб дані результати були актуальними.

Через відносну складність розрахунків та тривалість обробки даних автори даної статті обрали тільки обмежену кількість елементів балансу банку за статтею «власний капітал банку», але, як видно з результатів дослідження, множини показників власного капіталу банку не вистачає для визначення якісного групування банків. Це зумовлено багатьма чинниками. Наприклад, якісна характеристика фінансового стану банку зумовлена дотриманням збалансованості, взаємозв'язку і взаємоузгодженості складових елементів: власного капіталу, активів, зобов'язань, ліквідності, платоспроможності, прибутковості та ризиків.

Водночас ідея кластерного аналізу банків залишиться актуальною ще до того, коли науковці розроблять інструменти для вирішення даного питання.

## ВИСНОВКИ

Запропонована методика застосування кластерного аналізу для групування часової динаміки показників діяльності банків дає змогу:

- аналізувати спосіб утворення загальносистемної тенденції;
- виявляти характерні кластерні групи за змінами показників та позиціонувати банки в ці групи.

Зокрема, особливої уваги потребують банки зі значними відхиленнями від загальносистемної тенденції динаміки показників. Застосування запропонованої методики в практиці банківського нагляду в поєднанні з традиційними інтерполяційними методами аналізу часових рядів доповнить показники раннього реагування та забезпечить якісне посилення ефективності нагляду.

Отже, кластерний аналіз динаміки показників діяльності сукупності банків у певному часовому інтервалі дає змогу виокремлювати характерні групи трендів – кластери трендів, на підставі приналежності банку до яких можна дійти висновку щодо характерних особливостей діяльності банку. У подальшому отримані результати можуть бути використані для побудови системи індикаторних показників для раннього реагування та оцінювання ризиків банку.

**ПЕРЕЛІК ВИКОРИСТАНИХ ЛІТЕРАТУРНИХ ДЖЕРЕЛ**

1. Харари Ф. Теория графов. – М.:Мир, 1973. – 311 с. – (Russian source).
2. Кристофидес Н. Теория графов. Алгоритмический подход. – М.: Мир, 1978.– 429 с. – (Russian source).
3. David J. Hand, ISBN-13: 978-0262082907, Principles of Data Mining(Adaptive Computation and Machine Learning),A Bradford Book, 2001. – 584 p.
4. Graphviz – Graph Visualization Software.– [Електроннийресурс].– Режимдоступу: [http:// www.graphviz.org](http://www.graphviz.org).
5. Капулло М. Кластеризація часових рядів показників банків / М. Капулло // Вісник Національного банку України. – 2015. –№ 2. – С. 44–47. – [Електронний ресурс]. – Режим доступу: [http://nbuv.gov.ua/UJRN/Vnbu\\_2015\\_2\\_19](http://nbuv.gov.ua/UJRN/Vnbu_2015_2_19)