

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЧЕРНІГІВСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ

ПРОГНОЗУВАННЯ ТА АНАЛІЗ ЧАСОВИХ РЯДІВ

МЕТОДИЧНІ ВКАЗІВКИ

**до практичних занять та самостійної роботи
студентів спеціальності 051 «Економіка»
освітня програма
«Економічна кібернетика», «Економічна аналітика»**

Обговорено і рекомендовано
на засіданні кафедри
інформаційних систем в економіці
Протокол №1
від 27 серпня 2018 р.

ЧЕРНІГІВ ЧНТУ 2018

Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика» / Укл.: Юрченко М. Є. – Чернігів: ЧНТУ, 2018. – 88 с.

Укладач: Юрченко Марина Євгеніївна, кандидат фізико – математичних наук, доцент кафедри інформаційних систем в економіці ЧНТУ

Відповідальний за випуск: Акименко Андрій Миколайович, завідувач кафедри інформаційних систем в економіці ЧНТУ, кандидат фізико-математичних наук, доцент

Рецензент: Маргасова Вікторія Геннадіївна, д.е.н., професор кафедри бухгалтерського обліку, оподаткування та аудиту ННІ економіки Чернігівського національного технологічного університету

ЗМІСТ

ВСТУП	5
1. МЕТА ТА ЗАВДАННЯ ДИСЦИПЛІНИ	5
2. ОСНОВНІ ПОНЯТТЯ	6
2.1. ОСНОВНІ ЕЛЕМЕНТИ ЧАСОВОГО РЯДУ	6
2.2 ОСНОВНІ МЕТОДИ ДОСЛІДЖЕННЯ ЧАСОВИХ РЯДІВ	9
2.3 ІНФОРМАЦІЙНЕ ПРЕДСТАВЛЕННЯ ДИНАМІКИ РОЗВИТКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ	12
3. ОСНОВНІ ХАРАКТЕРИСТИКИ ДИНАМІКИ РОЗВИТКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ.....	14
3.1 ХАРАКТЕРИСТИКИ ДИНАМІКИ ЧАСОВОГО РЯДУ	14
3.2 КОРИГУВАННЯ РІВНІВ ЧАСОВОГО РЯДУ	18
3.3 МЕТОД ІРВІНА	20
4. ВИПАДКОВІ ПРОЦЕСИ ТА ЧАСОВІ РЯДИ.....	23
4.1 ОСНОВНІ ЕЛЕМЕНТИ ТЕОРІЇ ВИПАДКОВИХ ПРОЦЕСІВ	23
4.2 ОПТИМАЛЬНИЙ ПРЕДИКТОР І ЙОГО ВЛАСТИВОСТІ.....	26
4.3 БІЛИЙ ШУМ.....	27
4.4 МАРКІВСЬКИЙ ПРОЦЕС.....	30
5. РОЗКЛАДЕННЯ (ДЕКОМПОЗИЦІЯ) ЧАСОВОГО РЯДУ	31
6. НЕСТАЦІОНАРНІ ЧАСОВІ РЯДИ	35
6.1 ТИПИ НЕСТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ.....	35
6.2 НЕЛІНІЙНІ ДИНАМІЧНІ ПРОЦЕСИ.....	37
7. ІДЕНТИФІКАЦІЯ ЧАСОВИХ РЯДІВ.....	38
8. ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ З ВИКОРИСТАННЯМ АRІМА- МОДЕЛЕЙ.....	41
8.1 ОСНОВНІ ПОНЯТТЯ ПРО ЛІНІЙНІ ПАРАМЕТРИЧНІ МОДЕЛІ ЧАСОВИХ РЯДІВ	41
8.2 АНАЛІЗ ЧАСОВИХ РЯДІВ БОКСА-ДЖЕНКІНСА	41
8.3 ДІАГНОСТИКА МОДЕЛІ	46
8.4 КРИТЕРІЇ ВИБОРУ КРАЩОЇ МОДЕЛІ	47
8.5 ПРОГНОЗУВАННЯ ЗА ДОПОМОГОЮ АRІМА МОДЕЛЕЙ.....	49
9 ПРОГНОЗУВАННЯ ТЕНДЕНЦІЇ НА ОСНОВІ ЗГЛАЖУВАННЯ ЧАСОВИХ РЯДІВ	53

9.1 ПРОГНОЗУВАННЯ ТЕНДЕНЦІЇ ЧАСОВОГО РЯДУ ЗА СЕРЕДНІМИ ХАРАКТЕРИСТИКАМИ	53
9.2 ПРОГНОЗУВАННЯ ТЕНДЕНЦІЇ ЧАСОВОГО РЯДУ ЗА АНАЛІТИЧНИМИ МЕТОДАМИ ЗГЛАДЖУВАННЯ.....	55
9.3 ВИДИ КРИВИХ ЗРОСТАННЯ	59
9.4 ВИБІР КРИВОЇ ЗРОСТАННЯ.....	66
ЗАВДАННЯ ДЛЯ САМОКОНТРОЛЮ	76

ВСТУП

Сучасні дослідження макроекономічної динаміки, процесів перехідної економіки, фінансових ринків спираються на аналіз взаємозв'язків соціально-економічних даних, що має вигляд часових рядів. Урахування часової структури даних щодо реальних економічних процесів дозволяє адекватно відобразити їх в економіко-математичних моделях. Усвідомлення цього факту зумовило бурхливий розвиток специфічних методів аналізу таких даних. Знання цих методів і способів застосування їх до прогнозування соціально-економічних процесів є необхідною складовою підготовки економістів-дослідників (аналітиків).

1. МЕТА ТА ЗАВДАННЯ ДИСЦИПЛІНИ

Метою вивчення навчальної дисципліни "Прогнозування та аналіз часових рядів" є надання студентам знань і навичок, необхідних для практичного аналізу часових рядів.

Об'єктом вивчення навчальної дисципліни є економічні процеси розширеного відтворення.

Предметом вивчення навчальної дисципліни є сукупність методів і моделей розробки соціально – економічних прогнозів

У результаті вивчення навчальної дисципліни студент повинен

знати:

- предмет та мету курсу;
- основні підходи до аналізу часових рядів;
- сучасні методи аналізу часових рядів;
- особливості прогнозування часових рядів.

вміти:

- провести попередній та статистичний аналіз часового ряду;
- зробити прогноз досліджуваного процесу на певний час у майбутнє;
- визначити, як потрібно впливати на об'єкт, щоб одержати задані параметри його функціонування;
- вибрати та застосувати сучасні методи аналізу часового ряду;
- визначити, чи існують взаємозв'язки між обраними параметрами процесу, і який їх вид;
- визначити, чи існує взаємозв'язок між деякими процесами;
- зробити змістовні висновки на основі проведеного аналізу.

При вивченні навчальної дисципліни необхідно звернути увагу на:

- ознайомлення студентів з основами математичного апарату, необхідного для розв'язання теоретичних і практичних економічних задач;

- розвиток логічного мислення та підвищення загального рівня математичної культури;
- здобуття навичок дослідження прикладних питань та умінь перевести задачу на математичну мову;
- формування навичок самостійного вивчення навчальної літератури з теорії часових рядів;
- застосування отриманих знань до аналізу економічних процесів.

Зв'язок з іншими дисциплінами. Навчальна дисципліна «Прогнозування та аналіз часових рядів» базується на циклі дисциплін професійної та практичної підготовки, зокрема «Математичний аналіз», «Теорія ймовірностей», «Математична статистика», «Економетрика».

2. ОСНОВНІ ПОНЯТТЯ

Динаміка темпів інфляції, курсів акцій, валют або темпів зростання ВВП є предметом детального вивчення економістів. Основною задачею таких досліджень постає прогнозування з метою прийняття економічних рішень, вживання заходів економічної політики, що спрямована на антикризове (антициклічне) регулювання тощо. Для складання прогнозів зазвичай використовують економетричні моделі динаміки, які дозволяють дослідити тенденцію процесу, періодичність його коливань та можливі непередбачувані впливи екзогенних факторів.

Динамічні процеси, що відбуваються в економічних системах, найчастіше проявляються у вигляді ряду послідовно розташованих у хронологічному порядку значень того чи іншого показника, який у своїх змінах відображає хід розвитку досліджуваного явища в економіці. Ці значення, зокрема, можуть служити для обґрунтування (або заперечення) різних моделей соціально-економічних систем. Вони служать також основою для розробки прикладних моделей особливого виду, званих *трендовими моделями*.

2.1. ОСНОВНІ ЕЛЕМЕНТИ ЧАСОВОГО РЯДУ

Економетричну модель можна побудувати, використавши два типи вихідних даних:

дані, що характеризують сукупність різних об'єктів у певний момент (період) часу;

дані, що характеризують один об'єкт за ряд послідовних моментів (періодів) часу.

Моделі, побудовані за даними першого типу, називаються *просторовими моделями*.

Моделі, побудовані за даними другого типу, називають **моделями часових рядів**.

Часовим рядом називають послідовність значень статистичного показника (ознаки), впорядковану у хронологічному порядку. Застосовують також терміни "**ряд динаміки**", "**динамічний ряд**".

В англomовній літературі використовують термін "**time series**". Окремі спостереження часового ряду називають його **рівнями**, або **елементами**. Кожний рівень ряду відповідає певному моменту часу. Рівні ряду можуть набувати як детермінованих, так і випадкових значень. Порядок розташування рівнів є істотною характеристикою ряду і не може змінюватися довільно. Іноді кожному моменту часу приводять у відповідність декілька значень різних показників досліджуваного об'єкта. Тоді отримують **багатовимірний часовий ряд**.

На рис. 1.1–1.8 наведено деякі приклади часових рядів. Основними **завданнями** дослідження часових рядів є:

- видокремлення та опис основних характерних особливостей ряду; підбір статистичної моделі, що найкращим у певному розумінні способом відображає ряд;
- прогнозування майбутніх значень показників, що утворюють ряд, за попередніми спостереженнями;
- підготовка рекомендацій з управління процесом, що породжує досліджуваний часовий ряд.
- аналіз часових рядів, як правило, передбачає проведення таких основних етапів:
 - графічне подання й попередній аналіз поведінки часового ряду;
 - видокремлення і видалення закономірних складових ряду (тренду, сезонних та циклічних компонент);
 - видокремлення і видалення низько- та високочастотних складових (фільтрація);
 - дослідження випадкової складової часового ряду, що залишилася після видалення вищезазначених компонент;
 - побудова і перевірка адекватності моделі випадкової складової;
 - побудова загальної моделі досліджуваного ряду;
 - дослідження отриманої моделі і прогнозування майбутньої поведінки об'єкта що вивчається;
 - вивчення взаємодії між різними часовими рядами, що характеризують певну систему або процес.

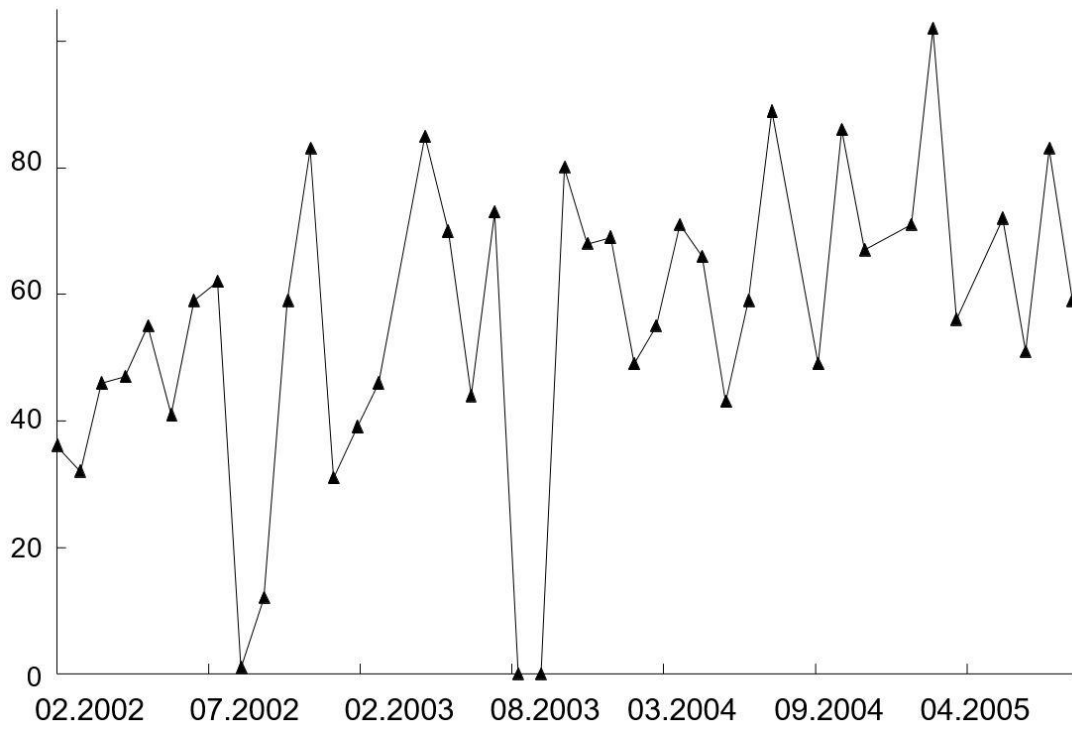


Рис. 2.1. Динамічний ряд кількості затверджених докторських дисертацій (за даними, опублікованими у бюлетені ВАК України у 2002–2005 рр.)

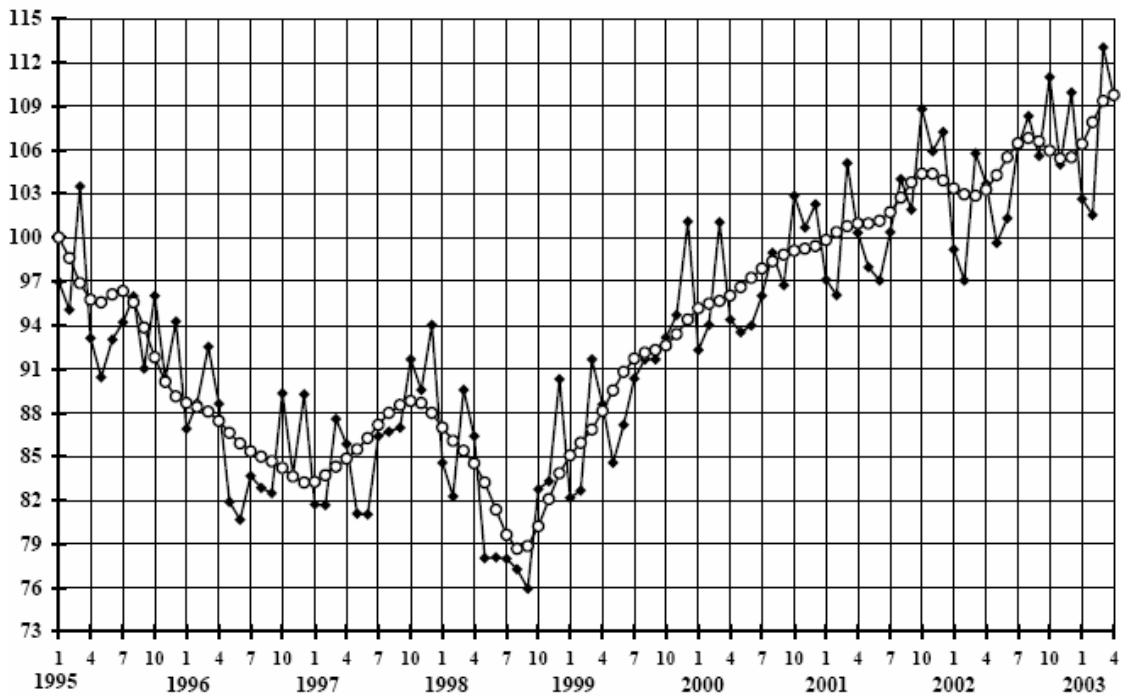


Рис. 2.2. Динаміка індексу промислового виробництва у Польщі у відсотках до рівня 1995 р.

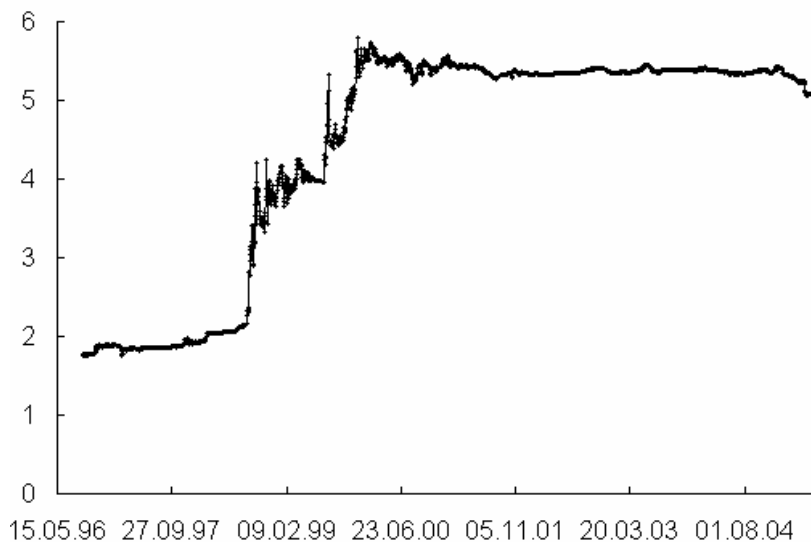


Рис. 2.3. Динаміка курсу "гривня-долар США" у 1996–2005 рр. за даними Національного банку України

2.2 ОСНОВНІ МЕТОДИ ДОСЛІДЖЕННЯ ЧАСОВИХ РЯДІВ

При дослідженні часових рядів застосовують основні методи:

- 1. Кореляційний аналіз**, який дає змогу виявляти істотні періодичні залежності та їх лаги (затримки) всередині певного процесу (автокореляція) або між декількома процесами (кроскореляція).
- 2. Спектральний аналіз**, що застосовують для визначення періодичних та квазіперіодичних компонент часового ряду.

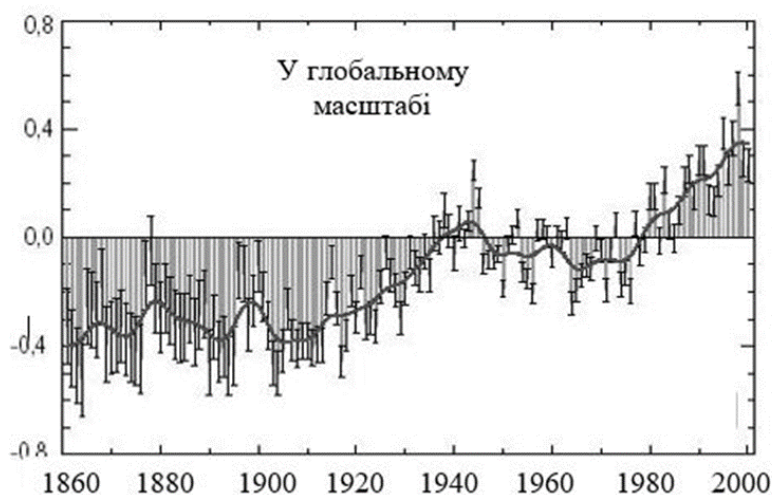


Рис. 2.4. Коливання температури на поверхні Землі у 1860–2000 рр. за даними

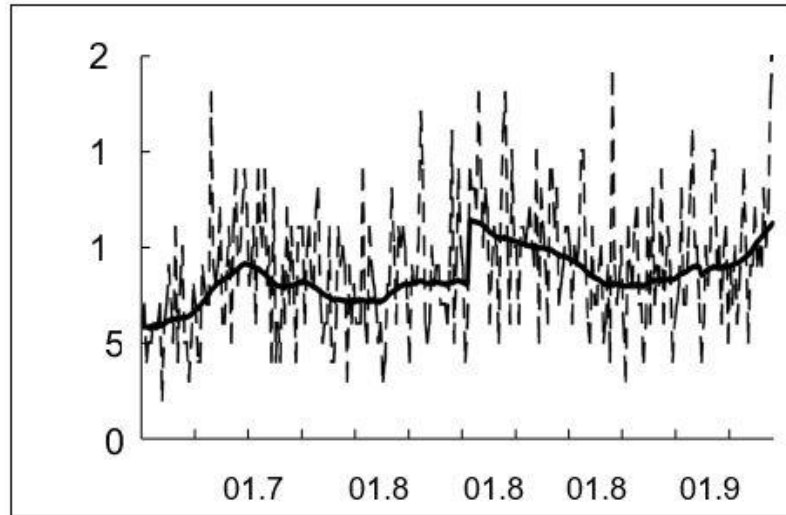


Рис. 2.5. Щомісячна кількість автокатастроф на дорогах штату Аріizona

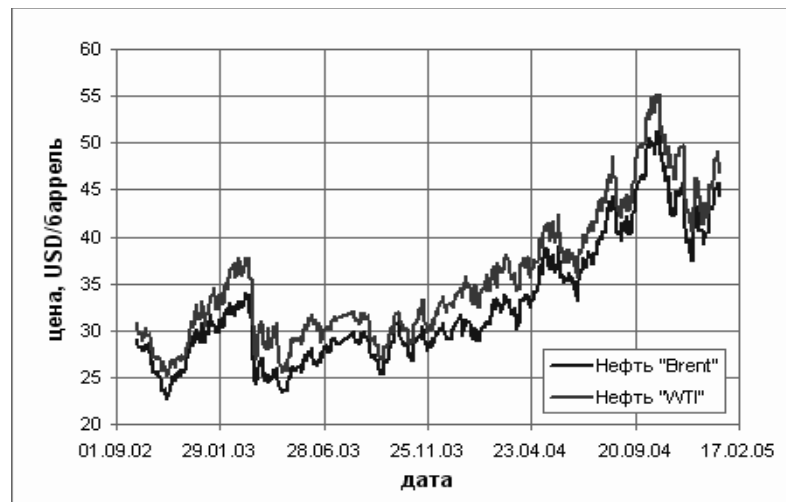


Рис. 2.6. Динаміка світових цін на нафту

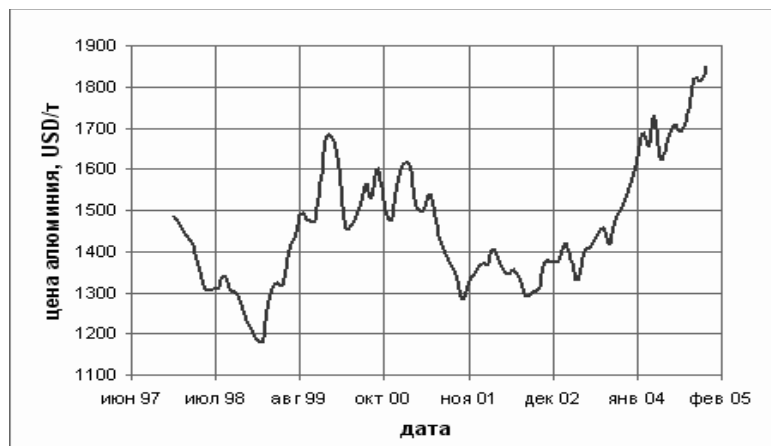


Рис. 2.7. Динаміка світових цін на алюміній

3. Методи згладжування та фільтрації, призначені для перетворення часових рядів з метою видалення з них високочастотних та сезонних коливань.

4. Методи авторегресії та ковзних середніх, які використовують для опису і прогнозування процесів, що здійснюють випадкові коливання навколо певного середнього значення.

5. Методи прогнозування, що дають можливість на основі обраної моделі часового ряду оцінювати його найбільш імовірні значення в майбутньому.

Залежно від характеру часового параметра розрізняють **моментні** та **інтервальні** часові ряди. У моментних рядах рівні характеризують значення показника станом на певний момент часу. Їм відповідають змінні типу запасу (stock variables). Зокрема, моментними є ряди цін на різноманітні товари станом на певні моменти часу, курсів валют, кількості населення, маси об'єкта, електричного напруження у мережі, температури хворого тощо.

В інтервальних рядах рівні характеризують значення показника за певні періоди (інтервали) часу. Їм відповідають змінні типу потоку (flow variables). Прикладами є ряди динаміки виробництва продукції (річного, квартального, місячного, добового), міграції населення, прибутку підприємства, валового національного доходу, розсіюваної потужності, витрат реагентів у хімічному реакторі тощо. Зокрема, ряди, наведені на рис. 1.2–1.4, 1.6, 1.7 є моментними, а наведені на рис. 1.1, 1.5 – інтервальними.

Часовий ряд, сформований із перших різниць рівнів моментного ряду, є інтервальним. Часовий ряд, рівні якого отримані підсумовуванням наростаючим підсумком всіх рівнів інтервального ряду, починаючи з певного визначеного рівня, є моментним. Співвідношення між моментними та інтервальними часовими рядами аналогічно співвідношенню між функціями та їх похідними.

Розрізняють **повні** й **неповні** часові ряди. У неповних часових рядів відсутні рівні, що відповідають певним моментам часу.

Рівні рядів динаміки можуть бути абсолютними, відносними або середніми значеннями певних показників. Якщо вони є не величинами, вимірюваними безпосередньо, а похідними від них – середніми, відносними й т.д., то відповідні ряди називають **похідними**. Зокрема, похідними є ряди середньодобового або середньомісячного виробництва певної продукції. Похідним є ряд, показаний на рис. 1.2.

Особливістю інтервальних рядів є можливість підсумовування їх рівнів. Результатом цього є так звані **нагромаджені підсумки**, які є значеннями певного показника за відповідний період часу. Наприклад, підсумовуючи добові обсяги виробництва продукції протягом місяця, отримаємо загальний обсяг її виробництва за відповідний місяць.

Суми рівнів моментного ряду не мають змісту. Зокрема, якщо ці рівні є значеннями середньодобової температури, то, підсумовуючи їх протягом певного періоду, ми не отримаємо величини, яка б мала

фізичний зміст. Для цього нам потрібно поділити результат на кількість днів у досліджуваному періоді. Тоді ми отримаємо величину, що є середнім значенням температури за відповідний період часу. Різниця рівнів моментного ряду є зміною показника за відповідний період.

Часовим рядам притаманні такі властивості:

- рівні часового ряду, як правило, не є статистично незалежними;
- члени часового ряду зазвичай не є однаково розподіленими.

Успішність аналізу часового ряду істотно залежить від правильності вибору інтервалу між його сусідніми рівнями. Зручніше вибирати рівновіддалені рівні ряду, але це не завжди є можливим. Зокрема, офіційні курси валют і фондові індекси не визначаються у вихідні та святкові дні, тому рівні відповідних рядів не є рівновіддаленими. Вибір надмірно великого інтервалу може призвести до втрати інформації про суттєві особливості ряду. Крім того, кількість членів ряду при цьому може виявитися недостатньою для застосування деяких методів аналізу. З іншого боку, надмірно малі інтервали збільшують обсяг розрахунків і можуть затінювати особливості динаміки показника випадковими флуктуаціями.

Важливою умовою аналізу є порівнянність рівнів ряду. Їх непорівнянність може бути зумовлена змінами законодавства та нормативної бази, методик розрахунку показників, меж адміністративно-територіальних одиниць і держав, технологій виробництва, термінологій, природними, екологічними та техногенними катастрофами й іншими факторами. Інформація про ряд має бути достатньо повною. Зокрема, для рядів, що містять циклічну складову, залежно від застосовуваних методів математичної обробки, потрібна інформація щодо проміжку, який перевищує 3–6 повних циклів. При побудові регресійних моделей необхідно мати ряди, довжина яких у кілька разів перевищує кількість параметрів, що визначаються.

Рівні часових рядів можуть включати аномальні значення (викиди). Причинами їх появи найчастіше є помилки під час збору та обробки інформації. Але викиди також можуть бути зумовлені реальними стрибками в динаміці досліджуваних показників.

2.3 ІНФОРМАЦІЙНЕ ПРЕДСТАВЛЕННЯ ДИНАМІКИ РОЗВИТКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ

Соціально-економічні процеси найчастіше спостерігаються у вигляді ряду послідовних, розташованих у хронологічному порядку значень того чи того показника.

Теоретично вимірювання можна реєструвати безперервно, але зазвичай їх здійснюють через однакові проміжки часу, тобто дискретно, і нумерують за елементами вибірки. Складовими ряду спостережень є

числові значення показника, які називають **рівнями** ряду, та **моменти** або **інтервали часу**, до яких належать рівні. Часовий ряд (ЧР) можна записати у стислому вигляді:

$$y_t, t = 1, 2, \dots, n,$$

де t — рівновіддалені моменти спостережень (година, доба, місяць, квартал, рік тощо)¹. Під **довжиною** часового ряду розуміють час, що минув від першого до останнього моменту спостереження. Часто довжиною ряду називають кількість рівнів n , які утворюють часовий ряд.

Залежно від характеру досліджуваних соціально-економічних показників часові ряди поділяють на моментальні, інтервальні та похідні.

Часові ряди, утворені показниками, що характеризують економічне явище на певні моменти часу, називають **моментальними**; прикладом такого ряду є дані щодо розміру виданих позичок відділенням Ощадбанку (табл. 2.1.).

Таблиця 2.1

МОМЕНТАЛЬНИЙ ЧАСОВИЙ РЯД

Дата надання позички	01.10.	05.10.	12.10.	23.10.	03.11.	07.11.
Розмір наданої позички, тис. грош. од.	3747	3710	3839	3783	3747	3710

Якщо рівні часового ряду утворюються шляхом агрегування за певний проміжок (інтервал) часу, такі ряди називають інтервальними часовими рядами; приклад наведено в табл. 1.1.2.

Таблиця 2.2

ІНТЕРВАЛЬНИЙ ЧАСОВИЙ РЯД

Місяць	Січень	Лютий	Березень	Квітень	Травень
Валовий внутрішній продукт, млн грн	6578	7016	7353	7353	7941

Часові ряди можуть бути створені як із абсолютних значень економічних показників, так і з середніх або відносних величин — це **похідні** ряди; приклад такого ряду наведено в табл. 1.1.3.

Таблиця 2.3

ЧАСОВИЙ РЯД, УТВОРЕНИЙ ІЗ СЕРЕДНІХ ЗНАЧЕНЬ ПОКАЗНИКА

Місяць	Січень	Лютий	Березень	Квітень	Травень
Середня зарплата	152,2	153,7	165,8	161,6	163,71

¹ Разом із часовим рядом іноді досліджують варіаційний ряд, який одержують із вхідного завдяки впорядкуванню за величиною рівня ряду. Отже, у варіаційному ряду на першому місці стоїть не перше за часом спостереження y_1 , а перше за величиною, тобто останнім буде мінімальне значення ряду.

загалом, грн/міс.					
-------------------	--	--	--	--	--

3. ОСНОВНІ ХАРАКТЕРИСТИКИ ДИНАМІКИ РОЗВИТКУ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ

3.1 ХАРАКТЕРИСТИКИ ДИНАМІКИ ЧАСОВОГО РЯДУ

Для аналізу соціально-економічних показників абсолютні рівні моментальних або інтервальних часових рядів, а також рівні середніх величин часто доводиться перетворювати на відносні величини. Найпоширеніші характеристики динаміки розвитку соціально-економічних процесів та їхні розрахунки наведено в табл. 3.1.

Таблиця 3.1

ХАРАКТЕРИСТИКИ ДИНАМІКИ ЧАСОВОГО РЯДУ

Характеристики	Розрахункові формули
1	2
1. Абсолютний приріст	$\Delta y_i = y_i - y_{i-k}$
2. Коефіцієнт зростання	$K_{i(зр)} = \frac{y_i}{y_{i-k}}$
3. Коефіцієнт приросту	$K_{i(нр)} = \frac{y_i - y_{i-k}}{y_{i-k}}$
4. Темп зростання	$T_{i(зр)} = \frac{y_i}{y_{i-k}} \cdot 100\% = K_{i(зр)} \cdot 100\%$
5. Темп приросту	$T_{i(нр)} = T_{i(зр)} - 100\%$, або $T_{i(нр)} = \frac{y_i - y_{i-k}}{y_{i-k}} \cdot 100\%$
6. Середня арифметична	$\bar{y} = \frac{\sum y_i}{n}$
7. Середня хронологічна	$\bar{y} = \frac{\sum y_i t}{\sum t}$
8. Середній абсолютний приріст	$\bar{\Delta y}_k = \frac{y_i - y_{i-k}}{k}$
9. Середній темп зростання	$\bar{T}_{(зр)} = n-1 \sqrt[n]{\frac{y_n}{y_1}}$
10. Середній темп приросту	$\bar{T}_{(нр)} = \bar{T}_{(зр)} - 100\%$

Для визначення змін, що відбуваються з досліджуваним явищем, передусім обчислюють швидкість розвитку цього явища за часом. Показником швидкості слугує *абсолютний приріст*, який характеризує

величину зміни показника за інтервал часу між порівнюваними періодами й обчислюється за формулою:

$$\Delta y_i = y_i - y_{i-k}, \quad (3.1)$$

де y_i — i -й рівень часового ряду ($i = 2, 3, \dots, n$);

k — індекс початкового рівня; $k = 1, 2, \dots, n - 1$ і може бути обраний будь-яким залежно від мети дослідження: за $k = 1$ отримують *ланцюгові* показники, за $k = i - 1$ отримують *базові* показники із базовим початковим рівнем ряду тощо.

Точніше, швидкість зміни показника характеризує приріст за одиницю часу; ця величина має назву *середнього абсолютного приросту*:

$$\overline{\Delta y}_k = \frac{y_i - y_{i-k}}{k}. \quad (3.2)$$

Зокрема, середній абсолютний приріст за весь період спостереження для заданого часового ряду дорівнює:

$$\overline{\Delta y} = \frac{y_n - y_1}{n - 1} \quad (3.3)$$

і характеризує середню швидкість зміни часового ряду, де n — індекс останнього спостереження.

Для визначення відносної швидкості зміни економічного явища як одиницю часу використовують відносні показники: коефіцієнти зростання й приросту (якщо ці показники виражені у відсотках, їх називають відповідно темпами зростання й приросту). Зазначимо, що в усіх наступних формулах індекс початкового рівня, стосовно якого здійснюють порівняння, також визначають за допомогою індексу k , як і раніше для показника абсолютного приросту.

Коефіцієнт зростання для i -го періоду обчислюють за формулою:

$$K_{i(\text{зр})} = \frac{y_i}{y_{i-k}}, \quad (3.4)$$

$K_{i(\text{зр})} > 1$, якщо рівень підвищується; $K_{i(\text{зр})} < 1$, якщо рівень зменшується; за $K_{i(\text{зр})} = 1$ рівень не змінюється.

Коефіцієнт приросту дорівнює:

$$K_{i(\text{пр})} = K_{i(\text{зр})} - 1 \text{ або } K_{i(\text{ппр})} = \frac{y_i - y_{i-k}}{y_{i-k}} \quad (3.5)$$

На практиці часто застосовують показники *темпу зростання* й *темпу приросту*:

$$T_{i(\text{зр})} = \frac{y_i}{y_{i-k}} \cdot 100 \% , \quad (3.6)$$

де $T_{i(\text{зр})}$ — темп зростання для i -го періоду;

$$T_{i(\text{нр})} = T_{i(\text{зр})} - 100 \% \text{ або } T_{i(\text{нр})} = \frac{y_i - y_{i-k}}{y_{i-k}} \cdot 100 \% , \quad (3.7)$$

де $T_{i(\text{нр})}$ — темп приросту для i -го періоду. Темп приросту показує, на скільки відсотків рівень одного періоду збільшився стосовно рівня іншого періоду, тобто цей показник характеризує відносну величину приросту у відсотках.

Порівняння абсолютного приросту та темпу приросту за той самий інтервал часу показує, що в реальних економічних процесах уповільнення темпу приросту часто не супроводжується зменшенням абсолютних приростів.

Абсолютне значення одного відсотка приросту визначають як відношення абсолютного приросту Δy_i до темпу приросту у відсотках $T_{i(\text{нр})}$

Середню швидкість зміни показника, що вивчається, за певний період характеризує також *середній темп зростання*. Його розраховують за формулою середньої геометричної:

$$\bar{T}_{(\text{зр})} = \sqrt[n-1]{T_{1(\text{зр})} \cdot T_{2(\text{зр})} \cdot \dots \cdot T_{n(\text{зр})}} = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100 \% , \quad (3.8)$$

де $T_{1(\text{зр})}, T_{2(\text{зр})}, \dots, T_{n(\text{зр})}$ — середні темпи зростання за окремі інтервали часу.

Відповідно *середній темп приросту* визначають як:

$$\bar{T}_{(\text{нр})} = \bar{T}_{(\text{зр})} - 100 \% . \quad (3.9)$$

Показник середнього темпу зростання, обчислюваний за формулою середньої геометричної (3.8), має суттєві недоліки, оскільки ґрунтується на зіставленні останнього та початкового рівнів часового ряду, проміжні рівні до уваги не беруться. У разі суттєвого коливання рівнів використання середнього геометричного темпу зростання для статистичного аналізу

може призвести до серйозних помилок, внаслідок чого реальна тенденція часового ряду буде викривлена.

Сучасні способи розрахунків середнього темпу зростання певною мірою позбавлені недоліків середньої геометричної.

Наприклад, для розрахунків середнього темпу зростання пропонується використовувати формулу:

$$\bar{T}_{(зр)} = \sqrt[n-1]{\frac{\hat{y}_n}{\hat{y}_1}}, \quad (3.10)$$

де \hat{y}_1, \hat{y}_n — згладжені за рівнянням тренду (рівнянням кривої зростання) перший та останній рівні часового ряду. Порядок побудови моделі тренду розглянуто в розділі 2.2 частини II. У моделі тренду враховано коливання проміжних рівнів часового ряду, тому обчислені за нею значення \hat{y}_1 та \hat{y}_n та середній темп зростання (3.10) точніше характеризуватимуть зміну економічного явища впродовж інтервалу дослідження.

Якщо тенденція часового ряду не змінюється, використовують характеристику *середнього рівня ряду*. В інтервальному ряду динаміки з однаково розташованими в часі рівнями середній рівень ряду обчислюють за формулою простої середньої арифметичної (тут і далі додавання ведеться за всіма періодами спостережень):

$$\bar{y} = \frac{\sum y_i}{n}. \quad (3.11)$$

Якщо інтервальный ряд має неоднаково розташовані в часі рівні, тоді середній рівень ряду (так звану *середню хронологічну*) обчислюють за формулою зваженої арифметичної середньої, де вагою є тривалість часу (наприклад, кількість років), упродовж якого рівень постійний:

$$\bar{y} = \frac{\sum y_i t}{\sum t}, \quad (3.12)$$

де t — кількість періодів часу, для яких значення рівня y_i не змінюється.

Для моментального ряду з однаково розташованими в часі рівнями середню хронологічну розраховують за формулою:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + y_3 + \dots + y_{n-1} + \frac{1}{2}y_n}{n-1}, \quad (3.13)$$

де n — кількість рівнів ряду.

Середню хронологічну для моментального часового ряду з неоднаково розташованими в часі рівнями розраховують за формулою:

$$\bar{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \dots + (y_{n-1} + y_n)t_{n-1}}{2\sum t}. \quad (3.14)$$

Тут n — кількість рівнів ряду, а t — період часу, що відокремлює 1-й рівень ряду від $(t + 1)$ -го рівня.

3.2 КОРИГУВАННЯ РІВНІВ ЧАСОВОГО РЯДУ

Часовий ряд правильно відображає об'єктивний закон зміни економічного показника, коли рівні цього ряду є порівнянними, однорідними, сталими та мають достатню сукупність спостережень. Невиконання однієї із цих умов робить некоректним застосування математичного апарату для аналізу часового ряду.

Порівнянність означає, що рівні часових рядів повинні мати однакові одиниці вимірювання, однакову періодичність обліку окремих спостережень, однаковий ступінь агрегування, обчислюватися за тією самою методикою. В економіці й соціології найпоширенішими є такі причини непорівнянності:

- за територією, внаслідок зміни кордонів регіону, за яким збирають статистичні дані;
- за колом охоплення об'єктів і підпорядкуванням або формою власності. Наприклад, унаслідок переходу частини підприємств конкретного об'єднання до іншого;
- за часовим періодом, коли дані кількох років наведено за станом на різні дати, або місяці мають різну тривалість, на порівнянність економічних і соціологічних даних впливають свята;
- через розбіжність у структурі одиниць сукупності, для якої їх обчислено. Наприклад, дані стосовно кількості населення залежать не лише від зміни кількості народжених і померлих, а й від зміни вікового складу населення впродовж періоду спостереження;
- за вартісними показниками. Навіть у тих випадках, коли значення цих показників фіксуються в незмінних цінах, їх часто важко зіставити.

Існують й інші причини. Непорівнянність часових рядів неможливо усунути лише формальними методами, тому на неї зважають у процесі

змістовного тлумачення рядів спостережень і результатів їхнього статистичного аналізу.

Однорідність означає відсутність нетипових, аномальних спостережень, а також викривлень тенденції. Під **аномальним рівнем** розуміють окреме значення рівня часового ряду, яке не відповідає потенційним можливостям економічної системи, що вивчається, і яке, залишаючись рівнем ряду, чинить суттєвий вплив на значення основних характеристик часового ряду. *Формально аномальність виявляється як несподіваний стрибок (або спад) із подальшим поступовим встановленням попереднього рівня.* Аномальність призводить до зміщення оцінок і, отже, до спотворення результатів аналізу. Причинами аномальних спостережень можуть бути помилки технічного порядку, або *помилки першого роду*: агрегування та дезагрегування показників, під час передання інформації та з інших технічних причин. Помилки першого роду слід виявляти й виправляти. Крім того, аномальні рівні в часових рядах можуть виникати через *помилки другого роду*: значення відображають об'єктивний розвиток процесу, але істотно відхиляються від загальної тенденції розвитку процесу; значення, що виникають через зміну методики обчислення, тощо. Ці помилки трапляються епізодично, тобто дуже рідко, і не підлягають усуненню. Для виявлення аномальних рівнів часових рядів використовують методи, призначені для статистичних сукупностей (метод Ірвіна тощо). Засоби описової статистики та обчислення їх за даними вибіркового спостережень наведено в табл. 3.2.

Таблиця 3.2

ОСНОВНІ СТАТИСТИЧНІ ХАРАКТЕРИСТИКИ ВИПАДКОВОЇ ВИБІРКИ

Характеристики	Оцінки вибіркового значень
1	2
1. Середні значення:	
арифметичне	$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$
геометричне	$y_G = \sqrt[n]{y_1 \cdot y_2 \cdot \dots \cdot y_n} = (y_1 \cdot y_2 \cdot \dots \cdot y_n)^{1/n}$
гармонійне	$\frac{1}{y_H} = \frac{1}{n} \sum_{i=1}^n \frac{1}{y_i}$
2. Дисперсія	$\hat{\sigma}^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1) \text{ (незміщена оцінка)}$
Середньоквадратичне відхилення (СКВ)	$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$
3. Середнє абсолютне лінійне відхилення $\left[\text{MAD} \right]$	$\text{MAD} = \left(\frac{\sum_{i=1}^n y_i - \bar{y} }{n} \right)$
4. Початкові моменти: другого, третього, четвертого порядку	$n_2 = \left(\frac{\sum_{i=1}^n y_i^2}{n} \right); n_3 = \left(\frac{\sum_{i=1}^n y_i^3}{n} \right); n_4 = \left(\frac{\sum_{i=1}^n y_i^4}{n} \right)$
5. Моменти центральні:	

<p>другого, третього, четвертого порядку</p>	$m_2 = \left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) / n$ $m_3 = \left(\sum_{i=1}^n (y_i - \bar{y})^3 \right) / n$ $m_4 = \left(\sum_{i=1}^n (y_i - \bar{y})^4 \right) / n$
<p>9. Коefіцієнт асиметрії його незміщена оцінка СКВ</p>	$A = m_3 / (m_2)^{3/2}$ $A_H = \left[\frac{(n-1)^{3/2}}{(n-2)} \right] A$ $S_H = \left[\frac{n(n-1)(n-2)(n+1)(n+3)}{n^2} \right]^{1/2}$
<p>10. Показник ексцесу його незміщена оцінка СКВ</p>	$E = \frac{m_4}{(m_2)^2} - 3$ $E_H = \left[\frac{(n-1)(n-2)(n-3)}{n^2} \right] \left[\frac{(n+1)E + 6}{n} \right]$ $S_E = \left[\frac{4n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)} \right]^{1/2}$
<p>11. Коefіцієнти варіації: за розмахом за середнім абсолютним лінійним відхиленням за СКВ медіана мода мінімальне значення ряду максимальне значення ряду розмах</p>	$\frac{R}{\bar{y}}$ $\frac{MAD}{\bar{y}}$ $\frac{\hat{\sigma}_y}{y}$ $m_e = y_n / 2$ <p>m_0 — характеризує величину, яка найчастіше спостерігається</p> y_{\min} y_{\max} $R = y_{\max} - y_{\min}$

3.3 МЕТОД ІРВІНА

Метод ґрунтується на порівнянні сусідніх значень ряду та розрахунку характеристики λ_t , яка дорівнює:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\hat{\sigma}_y}; t = 2, 3, \dots, n, \quad (3.15)$$

де $\hat{\sigma}_y$ — оцінка середньоквадратичного відхилення вибіркового ряду y_t , яка розраховується з використанням формул:

$$\hat{\sigma}_y = \sqrt{\frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n}}, \quad \bar{y} = \frac{\sum_{t=1}^n y_t}{n}.$$

Розрахункові значення λ_2, λ_3 тощо порівнюють із критичним значенням λ_α , і якщо вони не перевищують критичне, то відповідні рівні y_t вважаються нормальними. Критичні значення для рівня значущості $\alpha = 0,05$ (помилка 5 %) наведено в табл. 3.3.

Таблиця 3.3

n	2	3	10	20	30	50	100
λ_α	2,8	2,3	1,6	1,3	1,2	1,1	1,0

Критерій Ірвіна не «сприймає» аномальність, якщо вона виявляється в середині ряду зі стрімкою динамікою, тобто коли стрибок великий, але не перевищує рівнів наприкінці періоду спостережень, оскільки величина $\hat{\sigma}_y$ характеризує відхилення значень показника від середнього рівня за всією сукупністю спостережень.

Модифікація цього методу пов'язана із послідовним розрахунком $\hat{\sigma}_y$ не за всією сукупністю, а за трьома спостереженнями. Так, для всіх або лише для підозрюваних в аномальності рівнів розраховують оцінки середнього і середньоквадратичного відхилення для двох сусідніх із ними значень:

$$\bar{y}_t = \frac{y_{t-1} + y_{t+1}}{2}, \quad t = 2, 3, \dots, n-1, \quad (3.16)$$

$$\hat{\sigma}_y = \sqrt{\frac{\sum_{t=1}^n (y_{t-1} - \bar{y}_t)^2 + (y_{t+1} - \bar{y}_t)^2}{2}}. \quad (3.17)$$

Обчислюють величину

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\hat{\sigma}_y}, \quad t = 2, 3, \dots, n. \quad (3.18)$$

Розраховані ковзні значення λ_t порівнюють із критичними значеннями λ_α (див. табл. 3.3.) для $n = 3$.

Викривлення тенденції свідчить про зміну закономірності розвитку процесу або про зміну методики обчислення значень показника. Якщо точно встановлено, що причиною аномальності є помилки першого роду, то аномальні спостереження замінюють або простою середньою арифметичною двох сусідніх рівнів ряду, або відповідними значеннями за кривою, що згладжує цей часовий ряд. Не перевіряють часові ряди з періодом сезонності, більшим за одиницю, а також кінцеві рівні періоду спостережень.

Якщо значення наприкінці часового ряду «випадає» із загальної тенденції, то без додаткової інформації стосовно причин «випадіння» в кінці ряду неможливо визначити, чи це спостереження аномальне, чи відбувається зміна тенденції. У цьому разі важливо провести якісний аналіз змін, що відбуваються, або дочекатися надходження результатів нового спостереження. Якщо викривлення тенденції пояснюється зміною методики обчислення показника, то рівні, що передують викривленню тенденції, можуть бути використані для оцінювання характеристик динаміки і побудови моделі за умови, що вони будуть обчислені за новою методикою. Якщо таке обчислення неможливе, ці рівні ряду треба виключити з розгляду. Якщо викривлення тенденції відображає зміну закономірності розвитку процесу, то за інформаційну базу для статистичного аналізу можна взяти лише значення, що відповідають останнім змінам.

Стійкість часового ряду відбиває перевагу закономірності над випадковістю у зміні рівнів ряду. На графіках стійких часових рядів унаочнюється закономірність, а на графіках несталих рядів зміни послідовних рівнів постають хаотичними, тож пошук закономірностей формування значень рівнів таких рядів марний.

Достатня сукупність спостережень насамперед характеризує повноту даних. Достатня кількість спостережень визначається залежно від мети дослідження динаміки. Якщо метою є описовий статистичний аналіз, то період дослідження можна обрати будь-який, на власний розсуд. Якщо мета дослідження — побудова прогнозу моделі, тоді для статистичного аналізу, який розглядає незалежні спостереження з однаковою розподілом, кількість рівнів динамічного ряду має бути якомога більшою і, як правило, не менш як утричі має перевищувати період упередження прогнозу й становити більше 7. У разі використання квартальних або місячних даних для дослідження сезонності й прогнозування сезонних процесів часовий ряд має містити квартальні або місячні дані не менш як за чотири роки, навіть якщо складають прогноз на 1—2 квартали (місяці).

У методах нелінійної динаміки підхід до формування достатньої кількості даних відрізняється від прийнятого більшістю статистиків. У стандартній статистичній теорії чим більше даних точок спостережень, тим краще, бо спостереження передбачаються як незалежні. Нелінійні динамічні системи характеризуються процесами із довготривалою

пам'яттю. Тому для них охоплення більшого періоду часу є важливішим, ніж збільшення кількості точок спостережень. Наприклад, щоденна вибірка за чотири роки, або 1040 спостережень, не дадуть такого результату, як щомісячні дані за сорок років, або загалом 480 спостережень. Причина полягає в тому, що щоденні дані утворюють лише один чотирирічний цикл, а щомісячні — десять циклів. Нелінійні процеси мають так звану «стрілу часу». Збільшення «частоти» даних часто навіть ускладнює аналіз і не поліпшує значущості результату.

4. ВИПАДКОВІ ПРОЦЕСИ ТА ЧАСОВІ РЯДИ

4.1 ОСНОВНІ ЕЛЕМЕНТИ ТЕОРІЇ ВИПАДКОВИХ ПРОЦЕСІВ

Для аналізу часового ряду $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ порядок у послідовності t_1, t_2, \dots, t_n є суттєвим, тобто час виступає одним із визначальних чинників. Це відрізняє часовий ряд від звичайної випадкової вибірки, де індекси вводять лише для зручності ідентифікації. Принциповою відмінністю часового ряду від простих статистичних сукупностей є:

- по-перше, рівні часового ряду не є незалежними. Інакше кажучи, якщо майбутні значення змінної можна визначити, то вони є функцією від минулих значень цієї змінної;
- по-друге, рівні часового ряду неоднаково розподілені. Закон розподілу ймовірностей цих випадкових величин і, зокрема, їхні математичні сподівання та дисперсії можуть залежати від часу.

Отже, не можна поширювати властивості та правила статистичного аналізу випадкових вибірових спостережень на часові ряди. Порушення умови незалежності між спостереженнями призводить до негативних наслідків застосування цих методів. Наприкінці 1980-х — на початку 1990-х років дослідники остаточно переконалися, що лише врахування часової структури даних про реальні економічні процеси дають змогу адекватно відобразити їх в економіко-математичних моделях. Усвідомлення цього факту зумовило перегляд багатьох макроекономічних теорій і побудов та бурхливий розвиток специфічних методів аналізу таких даних, що дістали назву *аналіз часових рядів*.

Потужним математичним апаратом дослідження зміни соціально-економічних показників у їхній динаміці нині є теорія випадкових (стохастичних) процесів. Випадковий процес описують деякою функцією від часу, значення якої в будь-які моменти часу є випадковими величинами. Наведемо основні поняття та визначення теорії випадкових процесів, необхідні для подальшого аналізу часових рядів.

Реалізацією випадкового процесу ξ_t називають послідовність n результатів спостережень $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ певного економічного процесу в моменти часу t_1, t_2, \dots, t_n .

Динамічним або **часовим рядом** (*time series*) будемо називати послідовність спостережень y_{t_i} , отриманих у рівновіддалені моменти часу, а відповідну йому ймовірнісну модель — **дискретним випадковим** або **одnofакторним стохастичним процесом**.

Оскільки випадковий дискретний процес являє собою сукупність випадкових величин, то його найповнішою статистичною характеристикою є сумісна функція розподілу, або функція щільності розподілу. Щоб задати всі ймовірнісні властивості часового ряду, потрібна сукупність функцій розподілу, а саме одновимірна, двовимірна, тривимірна функції розподілу тощо: $F_1(\xi_{t_1}); F_2(\xi_{t_1}, \xi_{t_2}); F_3(\xi_{t_1}, \xi_{t_2}, \xi_{t_3}); \dots$. Індокси у величин $\xi_{t_1}, \xi_{t_2}, \dots, \xi_{t_n}$ означають, що випадкові величини розглядаються в моменти часу t_1, t_2, \dots, t_n і вони мають сумісну функцію розподілу. Якщо взяти інші моменти часу, то функція розподілу буде іншою. Така сукупність функцій розподілу цілковито характеризує випадковий процес.

Стаціонарні процеси. Економетричне моделювання відбувається, як правило, на підставі лише однієї реалізації випадкового процесу, тож ясно, що про оцінювання сукупності всіх функцій розподілу взагалі годі казати. Окрім того, якщо процес поводить ся так, що його основні статистичні характеристики з часом змінюються, то за короткий проміжок часу спостережень про нього взагалі нічого не можна сказати. Проблема втрачає гостроту, якщо розглядати вузький клас випадкових процесів, який дістав назву **стаціонарних випадкових процесів**. Під стаціонарністю розуміють такі випадкові процеси, деякі властивості яких не змінюються з часом.

Однією з важливіших властивостей стаціонарного випадкового процесу є **ергодичність**. Вона полягає в тому, що кожна окрема реалізація випадкового процесу є так би мовити «повноважним» представником усієї сукупності можливих реалізацій. Звідси для ергодичних процесів основні характеристики можна приблизно розраховувати не за кількома реалізаціями, як це робиться в загальному випадку, а за будь-якою однією реалізацією за доволі тривалий проміжок часу. В практичних розрахунках розглядають стаціонарний процес у широкому сенсі.

Стаціонарний часовий ряд у широкому сенсі — це процес, для якого математичне сподівання та дисперсія існують і є сталими величинами, що не змінюються в часі, а автокореляційна (автоковаріаційна) функція залежить лише від різниці між двома моментами часу $t_1 - t_2 = \tau$ і не залежить від конкретного періоду часу. Тобто для реалізації випадкового

процесу $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ основні моменти залишаються постійними й обмеженими у разі зміни часу t , для якого вони розраховуються, а саме:

математичне сподівання: $M(y_t) = \mu_y < \infty$, для всіх t ;

дисперсія: $Var(y_t) = D(y_t) = M(y_t - M(y_t))^2 = \sigma^2 < \infty$, для всіх t ;

автоковаріація порядку τ :

$$Cov(y_t, y_{t+\tau}) = M((y_t - M(y_t))(y_{t+\tau} - M(y_{t+\tau}))) = \gamma_\tau < \infty, \tau = 1, 2, 3, \dots$$

для всіх t .

Для отримання практичних оцінок часових рядів користуються такими формулами:

математичне сподівання:

$$\hat{\mu}_y = \frac{\sum_{t=1}^n y_t}{n};$$

дисперсія:

$$\hat{\sigma}_y^2 = \hat{\gamma}_0 = \frac{\sum_{t=1}^n (y_t - \hat{\mu}_y)^2}{n}; \quad (4.1)$$

автоковаріація порядку τ :

$$\hat{\gamma}_\tau = \frac{\sum_{t=1}^{n-\tau} (y_t - \hat{\mu}_y)(y_{t+\tau} - \hat{\mu}_y)}{n - \tau}.$$

Зрушення в часі τ називають **часовим лагом**. Зауважимо, що $Cov(y_t, y_{t+\tau})$, коли $\tau = 0$, дорівнює дисперсії:

$$\begin{aligned} Cov(y_t, y_{t+0}) &= \\ &= Cov(y_t^2) = D(y_t) = \gamma_0. \end{aligned}$$

При цьому $|\gamma_k| \leq \gamma_0$. Можна розглядати функцію γ_k як усі можливі значення автоковаріацій, де τ перебирає цілочисельні значення від $-\infty$ до ∞ . Сукупність значень автоковаріацій за всіх можливих значень τ називають **автоковаріаційною функцією** випадкового процесу. Автоковаріаційна функція стаціонарного часового ряду залежить лише від

різниць моментів часу ($t_1 - t_2 = \tau$). Ця функція парна, і досить розглядати невід'ємні τ .

Коефіцієнт автокореляції між зрушеними на τ рівнями часового ряду - це автоковаріація, розділена на корінь із добутку двох дисперсій, та оскільки дисперсія стала, отримуємо просто σ^2 або γ_0 . Розраховують коефіцієнт автокореляції за формулою:

$$\rho_\tau = \frac{\gamma_k}{\gamma_0}. \quad (4.2)$$

Вираз (4.2) визначає **автокореляційну функцію (АКФ)** часового ряду, яка показує наскільки статистично залежними є значення часового ряду для різних зрушень τ у часі (наприклад, для річних спостережень рік чи два роки тощо). Автокореляційна функція стаціонарного часового ряду залежить лише від різниці між двома моментами часу $t_1 - t_2 = \tau$, і є парною функцією, тобто $\rho(\tau) = \rho(-\tau)$. Задаючи різні значення $\tau = 1, 2, 3, \dots$, отримують послідовність значень $\rho_1, \rho_2, \rho_3, \dots$

Графік автокореляційної функції називають **корелограмою**. За корелограмою можна визначити запізнення, із яким зміна показника y_t позначається на його наступних значеннях.

У широкому сенсі оцінки наведених статистик є консистентними, тобто для них існує межа щодо ймовірності, яка збігається з їхніми справжніми значеннями для генеральної сукупності. Далі замість стаціонарності в широкому сенсі будемо просто говорити стаціонарність, оскільки інші різновиди стаціонарності не розглядатимуться.

Інтуїтивно можна очікувати, що небагато (якщо взагалі знайдуться) часових рядів соціально-економічних показників будуть стаціонарними, оскільки зростаючі й спадні значення є головною рисою соціально-економічних показників.

4.2 ОПТИМАЛЬНИЙ ПРЕДИКТОР І ЙОГО ВЛАСТИВОСТІ

Практика розроблення різноманітних прогнозів спирається на цілу систему методів, які стосуються оцінювання (прогнозування) величин, недоступних для безпосереднього спостереження в конкретний момент, і їх потрібно знайти за допомогою доступних для вимірювання (спостереження) супровідних величин.

Завдання стохастичного прогнозування полягає в тому, щоб за відомою сукупністю спостережень $\vec{Y} = (y_1, y_2, \dots, y_k)$ за випадковими величинами побудувати таку функцію від цих величин, яку можна було б

використати для оцінювання прогнозованої величини $\hat{y}_p : f(\vec{Y}) = y_p$. При цьому y_p та \vec{Y} стохастично пов'язані, тобто мають сумісну щільність розподілу $F(y_p, \vec{Y})$. Наприклад, y_p належить до майбутнього, а \vec{Y} — до теперішнього. Функцію $f(\vec{Y})$ називають **предиктором** величини y_p за \vec{Y} .

Для вимірювання точності предиктора, як правило, використовують середньоквадратичну похибку $MSE = M(f(\vec{Y}) - y_p)^2$. Предиктор, який мінімізує MSE в заданому класі предикторів, називають **оптимальним предиктором**, або прогнозом.

Розроблення методів побудови оптимальних предикторів становить зміст стохастичного прогнозування. На підставі предиктора, можна одержати варіанти прогнозу, що відповідатимуть сформульованим гіпотезам та умовам, урахованим під час його побудови.

Теорія стохастичного прогнозування величини y_p за супровідними величинами \vec{Y} передбачає, що сумісний закон розподілу $F(\vec{Y}, y_p)$ — відомий. У практичному застосуванні точний вид залежності між y_p та \vec{Y} найчастіше невідомий і пошук найкращого предиктора обмежується лінійними прогнозами, тобто коли $f(\vec{Y}) = \beta_0 + \beta_1 y_1 + \dots + \beta_k y_k$. Тоді оптимальний лінійний предиктор існує та збігається із функцією регресії y_p на \vec{Y} , тобто задається як умовне математичне сподівання¹:

$$f^*(\vec{Y}) = M(y_p | \vec{Y}) = \hat{\beta}_0 + \hat{\beta}_1 y_1 + \dots + \hat{\beta}_k y_k.$$

Ця функція має максимальну кореляцію з y_p серед усіх лінійних предикторів. Для побудови оптимального предиктора досить знати перші та другі моменти початкового розподілу $F(y_p, \vec{Y})$, які знаходять шляхом оброблення результатів відомих спостережень. Підставляючи ці оцінки замість теоретичних характеристик, отримують **емпіричний предиктор**, який використовують для прогнозування майбутніх значень y_p .

Отже, головну роль у статистичному підході до прогнозування відіграє вибір відповідної моделі. В разі наповнення її числовими параметрами вона стає безпосереднім інструментом прогнозування — предиктором.

4.3 БІЛИЙ ШУМ

¹ Доведення цього твердження можна знайти у [22].

Білим шумом (White noise) називають часові ряди, рівні яких мають середню, що дорівнює нулю, сталу дисперсію та нульову коваріацію послідовних спостережень, тобто нульову автокореляцію. Наприклад, залишки регресії, що задовольняють умовам теореми Гауса-Маркова, є «білим шумом»: $M(\varepsilon_t) = 0$; $D(\varepsilon_t) = \sigma^2$; $Cov(\varepsilon_t, \varepsilon_{t+\tau}) = 0$ за $\tau \neq 0$.

Наведене визначення білого шуму характеризує його як стаціонарний ряд. Хоча стаціонарний ряд необов'язково буде білим шумом, оскільки може мати середню або коваріацію, відмінні від нуля.

Якщо $\varepsilon_t \sim N(0, \sigma^2)$, то йдеться про гаусівський білий шум, хоча змінна білого шуму не обов'язково підпорядковується закону нормального розподілу. Найкращим передбаченням або прогнозом білого шуму є його нульове середнє значення. Білий шум відіграє важливу роль в аналізі часових рядів. На практиці білий шум трапляється не надто часто, але він утворює складніші процеси. Прикладом цього є процес випадкового блукання.

Випадкове блукання (Random walk). Іноді його називають броунівським рухом. Це стохастичний процес, де зміна рівня ряду, скажімо, рівня інфляції, досягається додаванням до нього випадкової змінної ε_t із постійною дисперсією та середнім, що дорівнює нулю. Випадкове блукання задається так:

$$y_t = y_{t-1} + \varepsilon_t, \quad (4.3)$$

де ε_t — білий шум. Цей процес можна розглядати як авторегресію із коефіцієнтом 1. Зазначимо, що лише ε_t має нульову середню та постійну дисперсію.

За умови наявності певної початкової точки y_0 підстановка у (4.3) значень змінної за попередні моменти часу дає вираз

$$y_t = y_0 + \varepsilon_t + \varepsilon_{t-1} + \dots = y_0 + \sum_{k=0}^{t-1} \varepsilon_{t-k},$$

який за t , що прямує до нескінченності, включатиме необмежену кількість доданків $\sum_{k=0}^{t-1} \varepsilon_{t-k}$, кожен із яких має нульове математичне сподівання та ненульову дисперсію σ_ε^2 .

Обрахуємо математичне сподівання процесу випадкового блукання:

$$M(y_t) = M(y_0) + M\left(\sum_{\tau=0}^{t-1} \varepsilon_{t-\tau}\right) = y_0 + \sum_{\tau=0}^{t-1} M(\varepsilon_{t-\tau}) = y_0 + 0 = const,$$

тобто математичне сподівання задовольняє умові стаціонарності.

Дисперсія процесу випадкового блукання дорівнює $D(y_t) = : = M\{(\sum_{\tau=0}^t \varepsilon_{t-\tau})^2\}$. Після розкриття дужок подвоєні добутки після взяття математичного сподівання будуть дорівнювати нулю, і залишиться математичне сподівання суми квадратів. Ураховуючи властивості дисперсії білого шуму, одержимо $D(y_t) = t \cdot \sigma_\varepsilon^2$. Отже, процес випадкового блукання не стаціонарний, оскільки дисперсія y_t зростає з часом.

Прогноз такого процесу на 1 крок уперед дорівнює

$$\hat{y}_t(1) = M(y_{t+1}|y_t, \dots, y_1) = M(y_t + \varepsilon_{t+1}).$$

Але $y_{t+1} = y_t + \varepsilon_{t+1}$ незалежно від y_{t-1}, \dots, y_1 . Отже прогноз на крок уперед становить: $\hat{y}_t(1) = y_t + M(\varepsilon_{t+1}) = y_t$. На два кроки уперед: $\hat{y}_t(2) = M(y_{t+2}|y_t, \dots, y_1) = M(y_t + \varepsilon_{t+1} + \varepsilon_{t+2}) = y_t$ тощо. Прогноз на k кроків уперед становитиме:

$$\hat{y}_t(k) = M(y_{t+k}|y_t, \dots, y_1) = M(y_t + \varepsilon_{t+1} + \dots + \varepsilon_{t+k}) = y_t.$$

Хоча величина прогнозової оцінки зі зростанням періоду випередження прогнозу k залишається постійною, дисперсія помилки прогнозу зростає. Так, помилка прогнозу на один крок вперед дорівнює $e_1 = y_{t+1} - \hat{y}_{t+1} = y_t + \varepsilon_{t+1} - y_t = \varepsilon_{t+1}$ та її дисперсія дорівнює $M(\varepsilon_{t+1}^2) = \sigma_\varepsilon^2$.

Для прогнозу на два кроки вперед

$$e_2 = y_{t+2} - \hat{y}_{t+2} = y_t + \varepsilon_{t+1} + \varepsilon_{t+2} - y_t = \varepsilon_{t+1} + \varepsilon_{t+2},$$

а дисперсія

$$M(\varepsilon_{t+1} + \varepsilon_{t+2})^2 = 2\sigma_\varepsilon^2,$$

оскільки ε_{t+1} та ε_{t+2} — незалежні. Аналогічно для прогнозу на k кроків уперед дисперсія помилки прогнозу становитиме $k \cdot \sigma_\varepsilon^2$. Середньоквадратичне відхилення прогнозу зростає пропорційно \sqrt{k} , і можна оцінити інтервал надійності прогнозу.

Якщо рівняння $y_t = y_{t-1} + \varepsilon_t$, де збурення ε_t є білим шумом, переписати як $\Delta y = y_t - y_{t-1} = \varepsilon_t$, отримаємо процес білого шуму. Приріст, або першу різницю (*first difference*), можна розглядати як інший часовий ряд Δy_t , який є стаціонарним. Перехід до перших різниць є розповсюдженим засобом зведення нестационарного часового ряду до стаціонарного.

Іноді випадкове блукання може передбачати елемент зсунення. Зсунення означає тенденцію (дрейф). Отже, випадкове блукання зі зсуненням — це випадкове блукання із дрейфом. Наприклад:

$$y_t = y_{t-1} + \alpha + \varepsilon_t, \quad (4.4)$$

де α — стала величина.

Тепер

$$\hat{y}_t(1) = M(y_{t+1} | y_t, \dots, y_1) = y_t + \alpha$$

та

$$\hat{y}_t(k) = y_t + k \cdot \alpha$$

Середньоквадратичне відхилення прогнозу в цьому разі не зміниться, оскільки: $e_1 = y_{t+1} - \hat{y}_{t+1} = y_t + \alpha + \varepsilon_{t+1} - y_t - \alpha = \varepsilon_{t+1}$.

Прогноз зростає лінійно за k , а інтервал надійності прогнозу розширюється пропорційно \sqrt{k} .

Середнє значення перших різниць становить швидкість зростання фактичного ряду спостережень, при цьому кожна зміна $\Delta y = y_t - y_{t-1}$ не залежить від усіх попередніх змін і має ідентичний розподіл імовірностей.

4.4 МАРКІВСЬКИЙ ПРОЦЕС

Марківськими називають процеси, в яких стан об'єкта в кожен наступний момент часу визначається станом поточного моменту і не залежить від того, яким шляхом об'єкт досяг поточного стану. Це стаціонарна послідовність незалежних, однаково розподілених випадкових величин. У термінах кореляційного аналізу часових рядів марківський процес можна описати таким чином: існує статистично значущий кореляційний зв'язок початкового ряду із рядом, зрушеним на один часовий інтервал, і цей зв'язок відсутній із рядами, зрушеними на два, три

тощо часові інтервали. В ідеальному випадку ці коефіцієнти кореляції дорівнюють нулю.

За допомогою рівняння авторегресії такий ряд можна представити як:

$$y_t = \alpha_t y_{t-1} + \varepsilon_t \text{ або } y_t = \rho y_{t-1} + \varepsilon_t, \quad (4.5)$$

розкладаючи y_{t-1} , отримуємо: $y_t = \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 y_{t-2} = \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 \varepsilon_{t-2} + \rho^3 y_{t-3}$ тощо. Очевидно, що y_t залежить від усіх минулих (але не майбутніх) ε_t . Якщо $M(\varepsilon_t) = 0$, то й $M(y_t) = 0$. Знайдемо добуток (4.5) на y_{t-1} і визначимо математичне сподівання:

$$M(y_t, y_{t-1}) = \rho M(y_{t-1}^2) \text{ або } Cov(y_t, y_{t-1}) = \rho \cdot D(y_t),$$

остаточно $\rho = \frac{Cov(y_t, y_{t-1})}{D(y_t)}$, тобто ρ є першою автокореляцією процесу.

$$Cov(y_t, y_{t-k}) = \rho \cdot Cov(y_{t-1}, y_{t-k}) \text{ розділимо на } D(y_t) \quad \rho_k = \rho \rho_{k-1} =: \\ = \rho^2 \rho_{k-2} \dots = \rho^k.$$

Отже, всі кореляції марківського процесу можна виразити через першу автокореляцію.

Окрім марківських стаціонарних процесів авторегресії часто трапляються процеси Юла, в яких ураховано авторегресію не лише першого, а й другого порядку, тобто

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \varepsilon_t.$$

5. РОЗКЛАДЕННЯ (ДЕКОМПОЗИЦІЯ) ЧАСОВОГО РЯДУ

Реальні часові ряди в економіці, як правило, є динамічно нестабільними, отже — не стаціонарними, і поняття стаціонарності процесу часто є лише зручною абстракцією для застосування статистичних методів. Кожен рівень часового ряду формується під впливом великої кількості чинників, які відображають закономірність і випадковість його формування. В аналізі часових рядів прийнято представляти часовий ряд y_t у вигляді суми систематичної складової (середньої) та випадкового відхилення від неї:

$$y_t = f(t) + \varepsilon_t, \quad (5.1)$$

де $f(t)$ — не випадкова функція часу (детермінована частина);

ε_t — випадкова, недетермінована частина.

Завдання розкладення часового ряду полягає в аналізі чинників, що впливають на значення його рівнів, у вирізненні серед них головних і другорядних (випадкових), а потім серед головних — еволюційних та періодичних (сезонних тощо).

Еволюційні чинники визначають загальний напрям розвитку економічного показника, провідну його тенденцію. Тенденція — це не випадкова складова часового ряду, яка змінюється повільно, і описується за допомогою певної функції n_t , яку називають **функцією тренду** або просто **трендом**. Тренд відображає вплив на економічний показник деяких постійних чинників, дія яких акумулюється в часі. У широкому сенсі під трендом розуміють будь-який упорядкований процес, що відрізняється від випадкового, тобто функцію $f(t)$ у (5.1). Іноді під трендом розуміють також зміщення у часі математичного сподівання. Відносно v_t припускається, що це певна гладка функція, ступінь гладкості якої заздалегідь не відомий. Під ступенем гладкості розуміють мінімальний ступінь поліному, що найкраще згладжує компоненту n_t . На рис. 5.1 а) зображено умовний часовий ряд із тенденцією, що лінійно зростає.

Серед чинників, що визначають регулярні коливання ряду, розрізняють такі:

Сезонні, що відповідають коливанням, які мають періодичний або близький до нього характер упродовж одного року. Наприклад, ціни на сільгосппродукцію взимку вищі, ніж улітку; рівень безробіття в курортних містах у зимовий період зростає відносно до літнього. Сезонні чинники можуть охоплювати причини, пов'язані з діяльністю людини (свята, відпустки, релігійні традиції тощо). Так, у ряду щомісячних даних слід очікувати наявності сезонних коливань із періодом 12, у квартальних рядах — із періодом 4. На рис. 5.2 б) зображено умовний часовий ряд, який містить лише сезонну компоненту. Результат дії сезонних чинників моделюють за допомогою функції s_t .

Циклічні (кон'юнктурні) коливання схожі на сезонні, але виявляються на триваліших інтервалах часу. Циклічні коливання пояснюються дією довготермінових циклів економічної, демографічної або астрофізичної природи. Наприклад, за багаторічними спостереженнями активність сонця має циклічність у 10,5—11 років, причому сплески сонячної радіації впливають на врожайність зернових культур, репродуктивну властивість тварин тощо. Отже динаміка показника містить характерні зміни, що повторюються з однаковою циклічністю. Результат дії циклічних чинників моделюють за допомогою функції c_t .

Тренд, сезонна й циклічна компоненти не є випадковими, тому їх називають **систематичними компонентами часового ряду**.

Випадкові чинники не підлягають вимірюванню, але неминуче супроводжують будь-який економічний процес і визначають стохастичний характер його елементів. До випадкових чинників можна віднести помилки вимірювання, випадкові збурення тощо. Деякі часові ряди, наприклад стаціонарні, не мають тенденції та сезонної складової, кожен наступний рівень їх утворюється як сума середнього рівня ряду і випадкової (додатної або від'ємної) компоненти. Приклад такого ряду демонструє рис. 5.2 в). Результат впливу випадкових чинників позначається **випадковою компонентою** ε_t , яку обчислюють як залишок або похибку, що залишається після вилучення з часового ряду систематичних компонент. Це не означає, що така складова не підлягає подальшому аналізу, оскільки містить лише хаос.

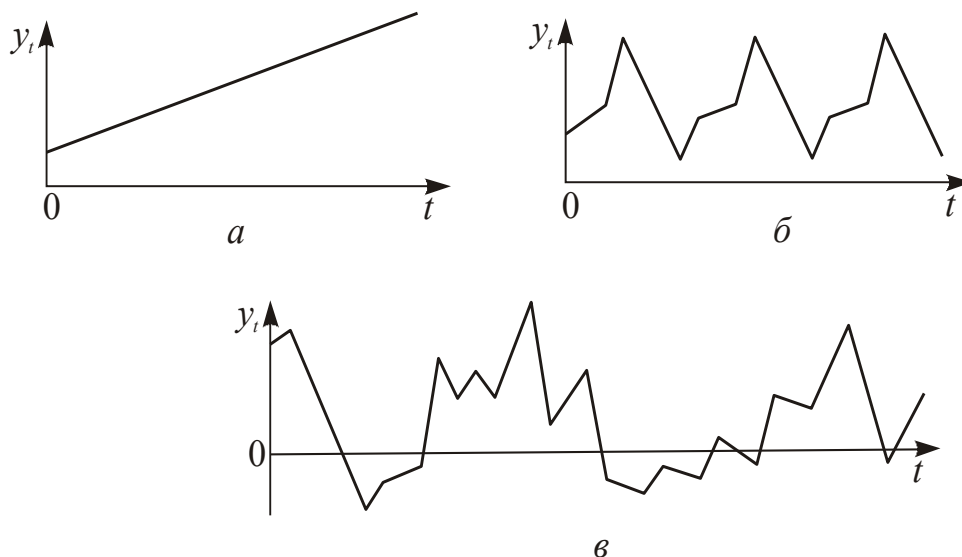


Рис. 5.2. Головні компоненти часового ряду:
 а — тренд, що зростає; б — сезонна компонента;
 в — випадкова компонента

За **декомпозицією Вольда** суто недетермінований стаціонарний у широкому сенсі випадковий процес можна записати у вигляді:

$$y_t - \mu = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \sum_{\tau=0}^{\infty} \psi_{\tau} \cdot \varepsilon_{t-\tau} \quad (5.2)$$

де μ — детермінована складова або математичне сподівання цього процесу, ε_t — білий шум з обмеженими математичним сподіванням та дисперсією. Розкладення Вольда (5.2) ще називають лінійним фільтром, начебто білий шум пропустили крізь лінійний фільтр. Це означає, що, не

втрачаючи цілого, обмежуються зручним лінійним представленням і переходять до вивчення стаціонарних процесів.

Щоб вираз (5.2) мав сенс, повинна виконуватися умова збіжності за ймовірністю, оскільки підсумовуються випадкові величини. Ця умова

записується, як $\sum_{\tau=0}^{\infty} \psi_{\tau}^2 < \infty$. Припускається, що $\psi_0 = 1$. Чим більший

ваговий коефіцієнт ψ_{τ} , тим більший вплив випадкового збурення в момент $t - \tau$ на поточний момент t .

Аналіз випадкової компоненти є важливою інформативною частиною дослідження часових рядів. Пояснюється це тим, що в короткотерміновому та певною мірою середньотерміновому прогнозуванні результати прогнозу тісно пов'язані із випадковою компонентою, тоді як у довготерміновому прогнозуванні головну увагу приділяють визначенню тенденції й взаємозв'язків між чинниками.

Очевидно, реальні дані цілковито не відповідають лише одній із наведених функцій, тож часовий ряд $y_t, t = 1, 2, \dots, n$ можна уявити у вигляді розкладення:

$$y_t = h_t + s_t + c_t + e_t, t = 1, 2, \dots, n \quad (5.3)$$

або різноманітних поєднань окремих функцій. Однак завжди припускають обов'язкову наявність випадкової складової. Розкладення (декомпозиція) часового ряду відбувається за такими варіантами моделей:

модель тренду

$$y_t = h_t + e_t, t = 1, 2, \dots, n; \quad (5.4)$$

модель сезонності

$$y_t = s_t + \varepsilon_t, t = 1, 2, \dots, n; \quad (5.5)$$

тренд-сезонна модель

$$y_t = h_t + s_t + e_t, t = 1, 2, \dots, n. \quad (5.6)$$

Моделі тренду й сезонності (тренд-сезонні) можуть відобразити як відносно постійну сезонну хвилю (цикл), так і динамічно змінювану залежно від тренду. Перша форма — (5.3—5.6) належить до **адитивних**, друга

$$y_t = h_t \cdot s_t \cdot c_t \cdot e_t, t = 1, 2, \dots, n, \quad (5.7)$$

до *мультиплікативних* моделей.

Моделі для врахування циклічних чинників будують аналогічно до тренд-сезонних, тільки замість сезонної складової вводять циклічну.

Процес окремого обчислення функцій n_t, c_t, s_t і ε_t називають *фільтрацією компонент* часового ряду y_t . Процедура оцінювання детермінованої частини разом з усіма невинпадковими компонентами має назву *згладжування* часового ряду.

Успішне розв'язання завдань виявлення й моделювання дії розглянутих складових чинників є підґрунтям, відправним пунктом для зрозуміння механізму формування соціально-економічного процесу та його прогнозування.

Утім, слід пам'ятати, що операція розкладення часового ряду, яка є допустимою з математичної точки зору й корисною для моделювання динаміки зміни показників у часі, подеколи може ввести в оману. Зокрема, за такого підходу дуже спрощеним може виявитися припущення стосовно незалежного впливу названих компонент, їхньої чіткої структури.

6. НЕСТАЦІОНАРНІ ЧАСОВІ РЯДИ

6.1 ТИПИ НЕСТАЦІОНАРНИХ ЧАСОВИХ РЯДІВ

За видом нестационарності часові ряди, що застосовують в економічній практиці, розподіляють на ряди типу: *TS*, *DS*, тренд-сезонні, нелінійні.

Часовий ряд типу TS (trend stationary process). До цього типу відносять нестационарні часові ряди із *детермінованим поліноміальним трендом* $y_t = P_k(t) + \varepsilon_t$, де $P_k(t)$ поліном ступеня k від t , а ε_t — стаціонарний процес, який не обов'язково є білим шумом. Наприклад, простий лінійний тренд $y_t = a + bt + \varepsilon_t$. Тут нестационарна змінна y_t виражена через детермінований, тобто невинпадковий тренд. Попри те, що додавання стаціонарної змінної призводить до коливань навколо тренду й робить y_t випадковою, ми, власне, маємо інформацію тільки про середнє значення $y_t = a + bt$, тобто часовий ряд характеризується наявністю тренду в середньому значенні. Отже, ані поточна, ані минулі події не змінюють довготермінових прогнозів цього процесу. Вплив випадкового збурення ε_t (поточний шок) забувається одразу на наступному кроці ($t + 1$). Похибка довготермінового прогнозу буде мати обмежену дисперсію σ_ε^2 , тому невизначеність є обмеженою навіть у далекому майбутньому.

Нестационарний процес типу *TS* зводять до стаціонарного за допомогою кількох методів. Наприклад, для лінійного тренду $y_t = a + bt + \varepsilon_t$ перехід до стаціонарності може відбуватися:

шляхом виділення лінійного тренду. Наприклад, будують лінійну регресію за часом і розглядають стаціонарний залишок $y_t - \hat{y}_t$;

узяттям перших різниць: різниці двох суміжних рівнів часового ряду

$$\Delta^1 y_t = y_t - y_{t-1} \quad (6.1)$$

є першими різницями ряду y_t , або $\Delta^1 y_t = a + bt - [a + b(t-1)] + (\varepsilon_t - \varepsilon_{t-1})$. Звідси $\Delta^1 y_t = b + u_t$, де $u_t = \varepsilon_t - \varepsilon_{t-1}$ — випадкова величина, розподіл якої цілком визначається розподілом величини ε_t . Окрім того, перші різниці часового ряду з лінійною тенденцією мають постійне математичне сподівання, що дорівнює певній константі b , не залежній від t .

Загалом якщо часовий ряд має тенденцію, що виражається через поліном ступеня d , то різниці порядку d

$$\Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1} \quad (6.2)$$

є випадковими величинами з постійним математичним сподіванням. У цьому разі $M(\Delta^{d+1} y_t) = 0$.

Якщо тенденція часового ряду відповідає експоненціальному або степеневому тренду, то метод послідовних різниць слід застосовувати не до початкового ряду, а до його логарифмів. Наприклад, процес у своєму розвитку наближається до певної величини a та може бути представлений у вигляді

$$Y_t = ae^{bt} + \varepsilon_t, \quad \lambda > 0, \quad (6.3)$$

де ε_t — аналогічна (6.1). Тоді процес, утворений із величин $z_t = \ln y_t - \ln y_{t-1}$, має постійне середнє (b) і може бути зведений до стаціонарного процесу авторегресії.

Часовий ряд типу *DS* (differencing stationary process). Це ряди без періодичної складової та тенденції зростання, але наявність тренду в дисперсії засвідчує їхню нестационарність. Прикладом таких рядів є процес випадкового блукання $y_t = y_{t-1} + \varepsilon_t$. Як уже зазначалося, цей процес накопичує випадкові збурення від усіх попередніх шоків, тобто має нескінченну пам'ять. Такий процес описують стохастичним трендом і

зводять до стаціонарного шляхом узяття першої різниці, звідси й відповідна назва.

Тренд-сезонні часові ряди окрім тренду містять чітко виражені сезонні коливання, які, своєю чергою, спричинюють нестационарність. Якщо процес включає періодичні (сезонні) коливання навколо середнього значення з періодом m , тобто

$$y_{t+m} \approx y_t \quad (6.4)$$

із точністю до випадкової складової, то d цьому разі різниці через m часових інтервалів представляють стаціонарний процес

$$\Delta_m y_t = y_{t+m} - y_t, \quad t = 0, 1, 2, \dots \text{ де } m \text{ — const,} \quad (6.5)$$

середнє значення якого збігається із середнім значенням початкового ряду.

Амплітуда сезонних коливань може зростати з часом і не обов'язково лінійно. Ці ряди характеризуються наявністю тренду в середньому значенні й дисперсії.

6.2 НЕЛІНІЙНІ ДИНАМІЧНІ ПРОЦЕСИ

До цього типу відносять часові ряди зі складною структурою, вони мають тренд і містять різні види коливань, зокрема сезонні та циклічні. Структуру таких рядів узагалі не можна описати за допомогою відомих функцій, оскільки для різних ділянок часового ряду набір цих функцій буде різним, тобто в цьому разі можна говорити про ряди зі змінною структурою, які характерні для нелінійних динамічних процесів. Вони спостерігаються в динаміці цін на ринках капіталу тощо.

Лише в останні роки завдяки розвитку математичних методів нелінійної динаміки та комп'ютерних технологій з'явилася можливість досліджувати такі процеси. У певному аспекті [29] будь-який динамічний процес зрештою є детермінованим, і моделювання його як реалізації випадкового процесу є зручним спрощенням. Невипадковий часовий ряд відображає невинуваткову природу впливів. Стрибки даних відповідають стрибкам впливових чинників і відбивають властиву їм кореляцію. Детерміновані процеси, що виглядають як випадкові, у теорії нелінійностей називають **детермінованим хаосом**. Добре відомо, що просте детерміноване нелінійне різницеве рівняння може породжувати надзвичайно складні часові траєкторії, які видаються випадковими. Наприклад, рівняння, яке трапляється в аналізі фінансових ринків $y_{t+1} = \alpha y_t (1 - y_t)$, де y_t є ціною облігацій. У багатьох економічних застосуваннях значення параметру α лежить між 1 та 4, таким чином виключають від'ємні значення рівноваги для y_t і уникають прямування

процесу до нескінченності. За зміни α від 1 до 4 динаміка системи зазнає суттєвих змін. Наприклад, для $1 < \alpha < 3$, за будь-яким відхиленням від $y_t = 0$, динаміка процесу прямує до рівноваги $y_t = 1 - \frac{1}{\alpha}$. Разом із тим для $3,75 < \alpha < 4$ спостерігатиметься нескінченна кількість циклів із різною періодичністю і нескінченне число положень рівноваги з еволюцією процесу залежно від початкового його стану. Такий тип поведінки називають «хаосом». Властивістю цього процесу є те, що хоча він детермінований, випадкове блукання є задовільною моделлю для описання механізму породження даних. У цьому випадку зміни y_t неможливо передбачити, хоча всю траєкторію перебігу процесу цілком можна передбачити.

7. ІДЕНТИФІКАЦІЯ ЧАСОВИХ РЯДІВ

Структуру часового ряду в деяких випадках можна визначити графічно. Це стосується, наприклад, таких компонент ряду, як тренд і сезонні коливання. Однак чисту випадковість інколи помилково сприймають як наявність певної структури, і, навпаки, за шумом можна не розгледіти існування структури. Тому потрібні методи або інструменти, за допомогою яких можна було б звести нанівець ефект впливу шуму, після чого з'ясувати характеристики ряду, необхідні для побудови відповідної прогнозової моделі. Як правило, спочатку з'ясовують, із яким процесом доведеться працювати — стаціонарним чи нестаціонарним. Для будь-якого нестаціонарного ряду важливо визначити ознаку його нестаціонарності: чи описується він детермінованим трендом, чи є інтегрованим процесом і описується стохастичним трендом (лінійним або нелінійним), визначити наявність періодичної складової.

Перевірка стаціонарності часового ряду. Стаціонарні часові ряди передбачають, що процес породження наявних даних є лінійним. Вони не мають тренду або періодичної зміни середнього та дисперсії.

Перевірку гіпотез стосовно сталості середнього значення та дисперсії часового ряду можна здійснити кількома способами. Найпростішими з них є перевірка значущої відмінності двох середніх значень для деяких підмножин вибірки (наприклад, для першої та останньої третин усього обсягу даних) за z — критерієм (критерій перевірки гіпотези про рівність середніх двох нормально розподілених вибірок) і для дисперсії, якщо справедливе припущення про нормальний розподіл, можна використати F-критерій. Розглянемо два поширені методи: метод перевірки різниць середніх рівнів і метод Форстера-Стьюарта.

Метод перевірки різниць середніх рівнів. Реалізація цього методу передбачає такі чотири кроки.

Крок перший. Вхідний часовий ряд $y_1, y_2, y_3, \dots, y_n$ розподіляють на дві приблизно однакові за кількістю спостережень частини: в першій частині n_1 першої половини рівнів вхідного ряду, у другій — решта рівнів n_2 ($n_1 + n_2 = n$).

Крок другий. Для кожної з цих частин розраховують середні значення й дисперсії:

$$\bar{y}_1 = \frac{\sum_{t=1}^{n_1} y_t}{n_1}; \quad \hat{\sigma}_1^2 = \frac{\sum_{t=1}^{n_1} (y_t - \bar{y}_1)^2}{n_1 - 1}; \quad \bar{y}_2 = \frac{\sum_{t=1}^{n_2} y_t}{n_2};$$

$$\hat{\sigma}_2^2 = \frac{\sum_{t=1}^{n_2} (y_t - \bar{y}_2)^2}{n_2 - 1}.$$

Крок третій. Перевірка рівності (однорідності) дисперсій обох частин ряду за допомогою F -критерію, що порівнює розрахункове значення цього критерію:

$$F = \begin{cases} \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2}, & \text{якщо } \hat{\sigma}_1^2 > \hat{\sigma}_2^2 \\ \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2}, & \text{якщо } \hat{\sigma}_1^2 < \hat{\sigma}_2^2 \end{cases} \quad (7.1)$$

із табличним (критичним) значенням критерію Фішера F_α із заданим рівнем значущості α . Якщо розрахункове значення F менше за табличне F_α , то гіпотезу про рівність дисперсій приймають, і можна переходити до четвертого кроку. Якщо F більше або дорівнює F_α , гіпотезу про рівність дисперсій відхиляють і доходять висновку, що цей метод не дає відповіді щодо наявності тренду.

На **четвертому кроці** перевіряють гіпотезу про відсутність тренду за допомогою t -критерію Стьюдента. Для цього визначають розрахункове значення критерію Стьюдента за формулою:

$$t = \frac{|\bar{y}_1 - \bar{y}_2|}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad (7.2)$$

де $\hat{\sigma}$ — оцінка середньоквадратичного відхилення різниць середніх:

$$\hat{\sigma} = \sqrt{\frac{(n_1 - 1)\hat{\sigma}_1^2 + (n_2 - 1)\hat{\sigma}_2^2}{n_1 + n_2 - 2}}.$$

Якщо розрахункове значення t менше за табличне t_{α} , то нульову гіпотезу не відхиляють, тобто тренд відсутній, інакше — тренд є. Зазначимо, що в цьому разі табличне значення t_{α} приймають для числа ступенів вільності, яке дорівнює $n_1 + n_2 - 2$, до того ж цей метод застосовують суто для рядів із монотонною тенденцією. Недолік методу полягає у неможливості правильно визначити існування тренду в тому разі, коли часовий ряд містить точку зміни тенденції у середині ряду.

8. ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ З ВИКОРИСТАННЯМ ARIMA-МОДЕЛЕЙ

8.1 ОСНОВНІ ПОНЯТТЯ ПРО ЛІНІЙНІ ПАРАМЕТРИЧНІ МОДЕЛІ ЧАСОВИХ РЯДІВ

Стаціонарні часові ряди можна представити широким класом лінійних параметричних моделей. Найпоширенішими є моделі авторегресії $\langle AR \rangle$, ковзної середньої $\langle MA \rangle$ та змішані $\langle ARMA \rangle$. Ця класифікація застосування цих моделей не обмежується стаціонарними процесами. Так, ряди зі специфічною однорідною нестационарністю можна звести до стаціонарних і описувати модифікованою формою моделі $ARMA$, відомої як модель Бокса-Дженкінса.

Лінійні параметричні моделі дістали загальну назву *авторегресійні інтегровані моделі ковзної середньої (ARIMA)*. Вони ґрунтуються на припущенні лінійності процесу породження даних і описують стаціонарний процес, який має три ознаки: p — порядок авторегресії, d — необхідний порядок інтегрування, тобто кількість разів взяття різниць для зведення початкового часового ряду до стаціонарного, q — порядок ковзної середньої в моделі.

8.2 АНАЛІЗ ЧАСОВИХ РЯДІВ БОКСА-ДЖЕНКІНСА

Практичне використання $ARMA$ -моделей пов'язують із появою методики їхньої побудови, розробленої Г. Боксом та Г. Дженкінсом [30]. Методика передбачає такі послідовні процедури:

1. Ідентифікація моделі часового ряду.
2. Оцінювання параметрів моделі.
3. Діагностика побудованої моделі.
4. Використання моделі для прогнозування майбутніх значень часового ряду.

Ці процедури можуть неодноразово повторюватися в процесі уточнення моделі. Розглянемо кожен етап алгоритму детальніше.

Ідентифікація моделі. Під час побудови моделі аналізу часових рядів виникає проблема визначення її із мінімальною кількістю параметрів. Ця проблема має назву *ідентифікація*.

Визначення порядку $ARMA$ -моделі на етапі ідентифікації складається із розв'язання двох відносно незалежних проблем:

- 1) аналізу стаціонарності процесу та визначення порядку d оператора переходу до послідовних різниць: $z_t = (1 - L)^d y_t$;

2) вибору параметрів p і q в моделі $ARMA$, яка описує стаціонарний ряд як процес авторегресії та ковзної середньої.

З'ясування стаціонарності часового ряду здійснюють за допомогою методів, розглянутих у розділі 2 частини 2. У разі нестаціонарності ряду для визначення порядку різницевого оператора можна скористатися емпіричним критерієм, сутність якого полягає у тому, що знаходять такі значення d , за якими вираз

$$\sqrt{\frac{1}{n-d} \sum_{t=1}^{n-d} (z_t - \bar{z}_t)^2}, \quad (8.1)$$

де \bar{z}_t — середнє значення стаціонарного процесу $z_t = \Delta^d y_t$, $t = 1, 2, \dots, n-d$, буде мінімальним. Величина критерію (8.1) зі збільшенням значення d зменшуватиметься доти, доки різницевий оператор не стане стаціонарним. Подальше підвищення порядку d різницевого оператора спричинить лише зростання дисперсії, а отже, збільшення її.

Систематичну складову можна також виключити з ряду, оцінивши її за методом найменших квадратів або будь-яким іншим методом згладжування часового ряду.

Коли стаціонарний ряд одержано, визначають порядок $ARMA$ -моделі. На цьому етапі вельми корисними є графічні методи, а також порівняння автокореляційної та часткової автокореляційної функції із відповідними функціями відомих $ARMA$ -процесів, наведених у табл. 8.1.

Таблиця 8.1

ХАРАКТЕРИСТИКИ $ARMA$ -МОДЕЛЕЙ

Модель	АКФ	ЧАКФ
Білий шум	усі нулі	усі нулі
$MA(1)$	нулі після ρ_1	спадна після ϕ_{11}
$MA(2)$	нулі після ρ_2	спадна після ϕ_{22}
$MA(q)$	нулі після ρ_q	спадна після ϕ_{qq}
$AR(1)$	геометрично спадна після ρ_1	нулі після ϕ_{11}
$AR(2)$	геометрично спадна після ρ_2	нулі після ϕ_{22}
$AR(p)$	геометрично спадна після ρ_p	нулі після ϕ_{pp}
$ARMA(1, 1)$	геометрично спадна після ρ_1	спадна після ϕ_{11}
$ARMA(p, q)$	геометрично спадна після ρ_p	спадна після ϕ_{qq}

У загальному випадку, якщо використовують вибірку спостережень, розмір якої часто є відносно малим, можна очікувати, що точної відповідності між даними й теоретичною моделлю не буде. Це може призвести до вибору на цьому кроці двох або трьох пробних моделей $ARMA(p, q)$ моделей, які мають кілька пар часових лагів p в авторегресійному процесі та лагових змінних q у моделі ковзної

середньої. Вибір із кількох моделей найдоцільнішої для подальшого аналізу й прогнозування здійснюється за допомогою методів діагностичної перевірки, що розглядатимуться далі.

Оцінювання параметрів моделі. Після того, як процес ідентифікації визначив початковий варіант стаціонарної моделі $ARMA$ -моделі, цю модель пристосовують до даних спостережень шляхом знаходження оцінок параметрів $a = (a_1, \dots, a_p)$ та $b = (b_1, \dots, b_q)$. Раніше було показано, що модель $ARMA$ -моделі порядку (p, d, q) , що враховує нестационарні процеси, зводиться за допомогою перших різниць до стаціонарної моделі порядку $(p, 0, q)$. Тому процедура обчислення коефіцієнтів розглядається тільки для стаціонарної моделі.

Параметри AR -моделі можуть бути оцінені за допомогою звичайного методу найменших квадратів (виходять зсунуті, але консистентні оцінки), та його не можна застосувати до MA або $ARMA$ моделей. Наприклад, для $MA(1)$ -моделі

$$y_t = \mu + \varepsilon_t - b_1 \varepsilon_{t-1}$$

неможливо оцінити параметри, користуючись лише спостереженнями y_t , оскільки невідомі значення параметрів для розрахунку.

Метод Бокса-Дженкінса [30] запропонували використовувати процедуру нелінійної оптимізації: процедуру пошуку на мережі (*grid-search procedure*). Це ітеративна процедура, в якій оцінки параметрів мінімізують суму квадратів залишків. Запишемо $MA(1)$ -моделі як $\varepsilon_t = y_t - \mu + b \cdot \varepsilon_{t-1}$. Аналізуючи оцінки АКФ та ЧАКФ, можна зробити попередні припущення відносно значень параметрів. Можна використовувати вибіркове середнє (для μ) та першу автокореляцію (для b). Припустімо, що вони дорівнюють 100 та 0,2. Тоді модель має вигляд $\varepsilon_t = y_t - 100 + 0,2\varepsilon_{t-1}$. Припускаючи, що ε_0 дорівнює нулю, можна отримати оцінки ε_t для t від 1 до n та розрахувати суму квадратів залишків s_1 . Вибір нових початкових значень для μ та b дає нове значення s_2 суми квадратів залишків. Потім перевіряються інші початкові дані, і остаточними оцінками стають значення коефіцієнтів моделі, за якими s є мінімальним.

За часів Бокса і Дженкінса, через значні обмеження на використання комп'ютерів, для оцінювання коефіцієнтів розроблялись окремі методи для кожної моделі. Зараз учені розробили загальний **метод максимальної правдоподібності**, який уможливорює отримання консистентних та асимптотично ефективних оцінок коефіцієнтів для будь-якої моделі [29].

Головна ідея застосування методу полягає у припущенні, що дані мають деякий імовірнісний розподіл та обчислюється ймовірність потрібної події. Це на загал залежить від деяких невідомих параметрів.

Використовуючи дані, можна максимізувати ймовірність цієї події. Коефіцієнти, за яких досягається максимум імовірності відповідної події, є необхідними оцінками параметрів. Іноді дуже важко знайти ці оцінки в аналітичному вигляді. В такому разі використовують числові методи оптимізації функції правдоподібності.

Будемо виходити зі спостережень $z = (z_1, z_2, \dots, z_n)$ (для цього знадобляться $N = n + d$ спостережень $y_{-d+1}, \dots, y_0, y_1, \dots, y_n$ над початковим рядом y_t). Запишемо модель $ARMA(p, q)$ -процес у вигляді

$$z_t = \sum_{i=1}^p a_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j} \quad (8.2)$$

Цей процес містить $p + q + 1$ невідомих параметрів: $a_i, b_j, \sigma_\varepsilon^2$.

Нехай $A = (a_1, a_2, \dots, a_p)^T$, $B = (b_1, b_2, \dots, b_q)^T$, $Z = (z_1, z_2, \dots, z_n)^T$, визначимо матрицю $V(A, B)$ таким чином, що $Dz_t = \sigma^2 V(A, B)$. Нехай також ε_t мають нормальний розподіл. Тоді логарифм функції правдоподібності має вигляд

$$\ln L(\sigma^2, A, B) = -\frac{1}{2} (n \ln \sigma^2 + \ln V(A, B) + \frac{Z^T (V(A, B))^{-1} Z}{\sigma^2}). \quad (8.3)$$

Оцінки $\hat{\sigma}^2, \hat{A}, \hat{B}$ отримують завдяки максимізації зазначеного логарифму функції правдоподібності. Існують також ефективніші методи обчислювання функції правдоподібності.

Функція правдоподібності. Розглянемо побудову функції правдоподібності більш детально. Нехай $y_t, t = 1, 2, \dots, n$ — вибірка, яка має імовірнісний розподіл $P(y_t|A)$, де A — набір невідомих параметрів. Припустімо, що y_t є незалежними, кожне із імовірнісним розподілом $P(y_t|A)$, а сумісний розподіл цілої сукупності $y_t, t = 1, 2, \dots, n$ подано формулою:

$$P(y_1, y_2, \dots, y_n | A) = P(y_1|A) \cdot P(y_2|A) \cdot \dots \cdot P(y_n|A) = \prod_{t=1}^n P(y_t|A). \quad (8.4)$$

Для відповіді на запитання, яке саме значення A максимізує ймовірність породження моделлю саме вибірки $y_t, t = 1, 2, \dots, n$, потрібно максимізувати функцію правдоподібності:

$$L(A) = \prod_{t=1}^n P(y_t | A) \quad (8.5)$$

Для подальшої оптимізації необхідно точно знати розподіл вибірки. Припустімо, що аналізується модель

$$y = f(X, a) + \varepsilon,$$

де y — часовий ряд,

X — матриця екзогенних змінних,

ε — вектор збурень, який має нормальний розподіл із нульовим вектором математичних сподівань та коваріаційною матрицею Θ . Тоді функція правдоподібності матиме вигляд:

$$L(a) = (2\pi)^{-\frac{n}{2}} \Theta^{-\frac{1}{2}} e^{-\frac{1}{2}(y-f(X,a))^t \Theta^{-1}(y-f(X,a))}. \quad (8.6)$$

У загальному випадку ані функція правдоподібності, ані її логарифм не є лінійними, тож знайти максимум функції правдоподібності в аналітичному вигляді дуже важко. Тому потрібно використовувати числові методи знаходження максимуму функції, наприклад метод Гауса, загальний алгоритм яких складається з таких кроків:

- 1) Покласти початкові значення для вектора A_i .
- 2) Визначити напрям руху для A_i , в якому значення $L(A_i)$ збільшується.
- 3) Визначити довжину кроку й обчислити нове значення A_{i+1} .
- 4) Перевірити критерій зупинки. Якщо алгоритм треба продовжити, то задаємо $i = i + 1$ й повертаємося до кроку 2.

Звичайним критерієм зупинки є $|L(A_{i+1}) - L(A_i)| < \delta$, де δ — наперед задане мале число.

За допомогою відповідних функцій правдоподібності відбувається тестування гіпотез. Розглянемо критерії перевірки гіпотези H^0 проти альтернативної H^1 у загальному випадку. Існує три основні класи тестових статистик: тест Вальда, тест за допомогою множників Лагранжа, тест на основі відношень значень функцій правдоподібностей. Усі ці критерії мають підґрунтям максимізацію функції правдоподібності. Вони є асимптотично еквівалентними. Ключовою різницею між цими трьома підходами є вибір оцінки для розрахунків. Метод відношень функцій правдоподібності є найстарішим з усіх цих тестів, він був розроблений Нейманом та Пірсоном у 1928 році. Сутність методу полягає в порівнянні значень функцій правдоподібності H^0 за умови (із обмеженнями) та без її

врахування (без обмежень). Наприклад, нехай без врахування H^0 оцінкою є \hat{A} , при врахуванні умови H^1 оцінкою буде \hat{A} . Тоді

$$LR = \frac{L(\hat{A})}{L(\hat{A})} > 1, \quad (8.7)$$

оскільки із визначення максимуму функції правдоподібності $L(\hat{A}) > L(\hat{A})$.

Потрібно визначити, якою може бути величина $L(\hat{A}) - L(\hat{A})$, щоб можна було прийняти гіпотезу H^0 , тобто чи є суттєвими обмеження, включені до H^0 . Відповідна статистика має вигляд:

$$LRT = 2(\log L(\hat{A}) - \log L(\hat{A})) \approx \chi_m^2. \quad (8.8)$$

Отже, для перевірки гіпотези необхідно підрахувати значення LRT та порівняти його з χ^2 — статистикою, де кількість ступенів свободи визначається кількістю обмежень у гіпотезі H^0 . Якщо $LRT \geq \chi_m^2$, то гіпотеза H^0 відхиляється.

8.3 ДІАГНОСТИКА МОДЕЛІ

Після знаходження оцінок параметрів треба перевірити, чи є побудована модель адекватною. Існує кілька різновидів критеріїв, що визначають значущість та стійкість параметрів, властивості залишків та придатність моделі для прогнозування. Розглянемо додаткові можливості діагностики, специфічні для *ARIMA*-моделей.

Перевірка залишків. Усі теоретичні моделі містять випадкову компоненту, тож, якщо оцінена модель коректна, залишки мають бути «білим шумом». Залишки моделі отримують відніманням від реальних спостережень значень, обчислених за моделлю, тобто $e_t = y_t - \hat{y}_t$.

Для моделі *AR(p)*-процесу $y_t = \sum_{j=1}^p a_j y_{t-j} + \varepsilon_t$ послідовність залишків будують за правилом:

$$e_t = y_t - \sum_{j=1}^p \hat{a}_j y_{t-j}, \quad t = p+1, \dots, n. \quad (8.9)$$

Зазначимо, що e_t не визначені для $t < p$.

Для $MA(q)$ -моделі $y_t = \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j}$ залишки обчислюють рекурентно:

$$e_j = 0, j \leq 0, e_1 = y_1, \quad (8.10)$$

$$e_2 = y_2 - \hat{b}_1 e_1 \text{ тощо, та для } t > q \text{ } e_t = y_t - \sum_{j=1}^q \hat{b}_j e_{t-j}.$$

Нарешті, для модель $ARMA(p, q)$ -процесу маємо:

$$e_t = 0, t \leq p$$

$$e_t = y_t - \sum_{j=1}^p \hat{a}_j y_{t-j} - \sum_{j=1}^q \hat{b}_j e_{t-j}, t \leq \max(p, q). \quad (8.11)$$

Отримані залишки треба перевірити на відповідність «білому шуму». Для цього обчислюють АКФ та ЧАКФ залишків і перевіряють їхню статистичну значущість за критеріями Бокса-Пірса, Льюнга-Бокса, стандартне відхилення. Інший критерій розглядає розподіл залишків, який вважають нормальним у разі малої вибірки.

8.4 КРИТЕРІЙ ВИБОРУ КРАЩОЇ МОДЕЛІ

Коли задовільними виявляються кілька моделей, потрібне правило вибору між ними. Бокс і Дженкінс запропонували принцип ощадливості, згідно з яким, маючи кілька адекватних моделей, треба обрати модель із найменшою кількістю параметрів. Для використання цього принципу треба формалізувати правило компромісу між точністю пристосування моделі та кількістю її параметрів. Існує кілька підходів до розв'язання цієї проблеми.

Порівняння моделей. Припустимо, що розраховано задовільну $ARMA(p, q)$ -модель часового ряду за методом максимальної правдоподібності, причому L_0 — максимальне значення функції правдоподібності. Тепер те саме розрахуємо для моделі $ARMA(p + 1, q)$ та моделі $ARMA(p, q + 1)$. Отримаємо значення L_1 та L_2 відповідно. Згідно зі стандартною теорією тестування функції правдоподібності (8.8), якщо початкова модель $ARMA(p, q)$ є коректною, то статистики $2(L_1 - L_0)$ та $2(L_2 - L_0)$ розподілені як χ^2 -розподіл. Така перевірка є дуже простою.

Але якщо дані сильно корелюють між собою, таке тестування може давати неправильні результати.

Числові критерії. На відміну від попередніх тестувань, числові критерії лише дають певне значення, за яким можна судити про адекватність моделі. Загальну характеристику критеріїв наведено в таблиці 8.2 [29].

Таблиця 8.2

ЧИСЛОВІ КРИТЕРІЇ		
Назва критерію	Формула підрахунку	Бажаний екстремум
Коефіцієнт детермінації	$R^2 = 1 - \frac{\sum_{t=1}^n e_t^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$	1
Скоригований коефіцієнт детермінації	$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-(p+q)}$	1
Інформаційний критерій Акаїке (AIC)	$AIC = \ln \left(\frac{\sum_{t=1}^n e_t^2}{n - (p+q)} \right)$	Min
Інформаційний критерій Шварца-Ріссанена (SIC)	$SIC = \ln \left(\frac{\sum_{t=1}^n e_t^2}{n - (p+q)} \right) + \frac{p+q}{n} \ln n$	Min
Критерій Ханнана-Квіна (HQ)	$HQ = \ln \left(\frac{\sum_{t=1}^n e_t^2}{n - (p+q)} \right) + \frac{c(p+q)}{n} \ln(\ln n), c \geq 2$	Min
Прогнозовий критерій (FC)	$FC = \frac{\sum_{t=1}^n e_t^2}{n - (p+q)} \left(1 + \frac{p+q}{n} \right)$	Min

Однією із пропозицій є обчислення **коефіцієнта детермінації** (R^2) та **зваженого коефіцієнта детермінації** (\bar{R}^2). Однак цей метод непридатний для різницевої змінної z у деяких моделях.

Інформаційні критерії ґрунтовані на мінімізації певних статистик, що мають стандартні розподіли.

Інформаційний критерій Акаїке (AIC) розглядає нелінійне компромісне співвідношення між дисперсією залишків і значенням загальної кількості оцінюваних параметрів ($p + q$), оскільки моделі з більшою кількістю оцінюваних параметрів можна віддати перевагу лише за пропорційно великого зменшення дисперсії залишків.

Інформаційний критерій Шварца-Ріссанена (SIC) надає більшій ваги ($p + q$) порівняно з AIC за $n > 7$, тобто зростання кількості оцінюваних параметрів потребує вагомішого зменшення дисперсії залишків для SIC, ніж для AIC.

Критерій Ханнана-Квіна (HQ). Тут вага при $(p + q)$ є більшою за 2, якщо $n > 15$.

Прогнозовий критерій (FC) використовує похибку передбачення.

Вибір між цими критеріями є довільним, оскільки всі статистики змінюються в одному напрямі в разі збільшення кількості оцінюваних параметрів. На практиці користуються одним із них.

8.5 ПРОГНОЗУВАННЯ ЗА ДОПОМОГОЮ ARIMA МОДЕЛЕЙ

В *ARIMA* моделях, під час прогнозування змінної для майбутнього моменту часу, лагові значення цієї змінної, які слугують пояснюючими змінними (регресорами) моделі, можна розглядати або фіксованими на вибіркових значеннях, або випадковими. Перша можливість призводить до умовного прогнозу, на кшталт множинної регресії, друга — до безумовного прогнозу. Отже, у прогнозуванні за моделлю типу *ARIMA* розглядають умовні та безумовні прогнози. Відомо, що умовна дисперсія випадкової величини не перевищує її безумовну дисперсію, тому точність умовного прогнозу завжди вища.

Якщо модель правильно специфіковано, то можливі два джерела помилок прогнозів: невизначеність майбутніх значень випадкової величини ε , відсутність точних значень коефіцієнтів моделі (наявні тільки їхні оцінки).

Під час прогнозування за моделлю *ARIMA* від наявної вибірки залежать як оцінки коефіцієнтів моделі, так і значення регресорів, тому важко аналітично виразити умовну дисперсію помилки прогнозу через спостереження часового ряду. Як правило, обмежуються припущенням про те, що коефіцієнти відомі точно. Зрозуміло, що таке припущення зменшує дисперсію помилки прогнозу й тим самим збільшує уявну точність як умовного, так і безумовного прогнозів.

Для досягнення мінімуму середньоквадратичної помилки (MSE) потрібно взяти умовне математичне сподівання: $M\{y_{t+\tau} | y_1, \dots, y_t\}$.

Прогноз за моделлю MA (q): $y_t = \theta + \varepsilon_t + b_1\varepsilon_{t-1} + b_2\varepsilon_{t-2} + \dots + b_q\varepsilon_{t-q}$.
Якщо коефіцієнти моделі точно відомі, і є значення y_t для $t \in [1, n]$, то безумовним точковим прогнозом для будь-якого моменту часу буде математичне сподівання процесу, тобто θ . Умовним прогнозом для моменту часу $t + 1$ буде умовне математичне сподівання:

$$\hat{y}_t(1) = M\{\theta + \varepsilon_{t+1} + b_1\varepsilon_t + \dots + b_q\varepsilon_{t-q+1} | y_1, \dots, y_t\}. \quad (8.12)$$

Серед випадкових величин ε , що знаходяться ліворуч, є такі, що пов'язані зі спостереженнями. Оскільки спостереження складаються із «модельного значення» й похибки, умовні математичні сподівання усіх складових, окрім ε_{t+1} , не дорівнюють нулю.

Наприклад, $M\{\varepsilon_t | y_1, \dots, y_t\}$ є залишком між спостереженням і розрахунком (прогнозом) за моделлю, тобто $e_t = y_t - \hat{y}_t$. Тому умовні математичні сподівання від усіх минулих значень випадкової складової треба замінити відповідними залишками. Так само будується прогноз на 2 й більше кроків уперед. Усі майбутні ε замінюються нулями, а минулі — залишками, які можна обчислити. Отже, для моделі $MA(q)$ прогноз залежить від того, які похибки були на попередніх кроках. Починаючи із кроку $(q + 1)$ умовний прогноз є математичним сподіванням θ , тобто умовний прогноз збігається з безумовним.

Умовна дисперсія помилки прогнозу на 1 крок випередження становить:

$$\begin{aligned} D(y_{t+1} - \hat{y}_t(1) | y_1, \dots, y_t) &= \\ &= M\{((\theta + \varepsilon_{t+1} + b_1\varepsilon_t + \dots + b_q\varepsilon_{t-q+1} - \\ &- \theta - b_1e_t - \dots - b_qe_{t-q+1}) | y_1, \dots, y_t)^2\} = \sigma_\varepsilon^2. \end{aligned} \quad (8.13)$$

Аналогічно дисперсія прогнозу на 2 кроки випередження дорівнює:

$$D(y_{t+2} - \hat{y}_{t+2} | y_1, \dots, y_t) = (1 + b_1^2)\sigma_\varepsilon^2, \quad (8.14)$$

а дисперсія на τ кроків становить

$$(1 + b_1^2 + \dots + b_\tau^2)\sigma_\varepsilon^2 \text{ для } \tau < q. \quad (8.15)$$

Якщо $\tau \geq q$, дисперсія помилки умовного прогнозу стає такою самою як і для безумовного прогнозу, тобто дорівнює дисперсії випадкового процесу y_t .

Прогноз за моделлю $AR(p)$: $y_t = \theta + a_1y_{t-1} + a_2y_{t-2} + \dots + a_py_{t-p} + \varepsilon_t$.
Для прогнозу на один крок уперед можна записати:

$$\begin{aligned}
\hat{y}_t(1) &= M\{y_{t+1}|y_1, \dots, y_t\} = \\
&= M\{\theta + a_1 y_t \dots + a_p y_{t-p+1} + \varepsilon_{t+1}|y_1, \dots, y_t\} = \\
&= \theta + a_1 y_t \dots + a_p y_{t-p+1}
\end{aligned} \tag{8.16}$$

Тобто у рівняння моделі підставляють p минулих значень реалізації часового ряду. Для прогнозу на два кроки вперед отримують:

$$\begin{aligned}
\hat{y}_t(2) &= M\{y_{t+2}|y_1, \dots, y_t\} = \\
&= M\{\theta + a_1 y_{t+1} \dots + a_p y_{t-p+2} + \varepsilon_{t+2}|y_1, \dots, y_t\}.
\end{aligned} \tag{8.17}$$

Математичне сподівання від випадкової похибки ε знов дасть 0, умовне математичне сподівання від y_1, y_2, \dots, y_t дорівнює цим самим значенням, але до цього виразу входить умовне математичне сподівання від y_{t+1} , отримане на попередньому кроці. Можна підставити його вираз і отримати розгорнуту формулу через значення реалізації. Насправді зручніше розглядати рекурентне співвідношення, яке пов'язує послідовні значення прогнозу. Це співвідношення є лінійним різницевим рівнянням порядку p , і його розв'язок прагне, якщо збільшується t , до величини $\frac{\theta}{1 - a_1 - \dots - a_p}$, тобто знов таки до безумовного прогнозу.

Умовну дисперсію помилки прогнозу розраховують аналогічно до випадку моделі ковзної середньої, але доведення стають досить громіздкими навіть для моделей невеликого порядку. Наприклад, для моделі $AR(2)$ без вільного члена прогноз на один крок випередження становить: $\hat{y}_t(1) = a_1 y_t + a_2 y_{t-1}$, та $y_{t+1} = a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}$. Очевидно, що дисперсія помилки прогнозу на 1 крок дорівнює:

$$\begin{aligned}
D(y_{t+1} - \hat{y}_t(1)|y_1, \dots, y_t) &= \\
&= M\{((\theta + \varepsilon_{t+1} + b_1 \varepsilon_t + \dots + b_q \varepsilon_{t-q+1} - \\
&\quad - \theta - b_1 \varepsilon_t - \dots - b_q \varepsilon_{t-q+1})|y_1, \dots, y_t)^2\} = \sigma_\varepsilon^2.
\end{aligned} \tag{8.18}$$

Для прогнозу на 2 кроки відповідно отримуємо:

$$\hat{y}_t(2) = a_1 \hat{y}_t(1) + a_2 y_t = a_1 (a_1 y_t + a_2 y_{t-1}) + a_2 y_t,$$

$$y_{t+2} = a_1 y_{t+1} + a_2 y_t + \varepsilon_{t+2} = a_1 (a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}) + a_2 y_t + \varepsilon_{t+2}$$

Дисперсія помилки прогнозу на 2 кроки випередження дорівнює:

$$(1 + b_1^2) \sigma_\varepsilon^2. \quad (8.19)$$

Для прогнозу на 3 кроки отримуємо:

$$\begin{aligned} \hat{y}_t(3) &= a_1 \hat{y}_t(2) + a_2 \hat{y}_t(1) = a_1 (a_1 (a_1 y_t + a_2 y_{t-1}) + a_2 y_t) + a_2 (a_1 y_t + a_2 y_{t-1}) \\ y_{t+3} &= a_1 y_{t+2} + a_2 y_{t+1} + \varepsilon_{t+3} = a_1 (a_1 (a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}) + a_2 y_t + \varepsilon_{t+2}) + \\ &+ a_2 (a_1 y_t + a_2 y_{t-1} + \varepsilon_{t+1}) + \varepsilon_{t+3}. \end{aligned}$$

Дисперсія помилки прогнозу на 3 кроки дорівнює:

$$(1 + b_1^2 + b_2^2 + 2b_1^2 b_2 + b_1^4) \sigma_\varepsilon^2. \quad (8.20)$$

Очевидно, що дисперсія помилки прогнозу збільшується з кожним кроком.

Значно простішими виходять вирази для дисперсії помилки прогнозу, якщо перейти від $AR(p)$ представлення до еквівалентного MA представлення: $y_t = \theta + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q} + \dots$ із необмеженою кількістю складових. Тоді дисперсію помилки прогнозу на τ кроків можна виразити формулою

$$\sigma_\varepsilon^2 \sum_{j=0}^{\tau-1} \psi_j^2 \quad (\psi_0 = 1). \quad (8.21)$$

Для загальної моделі $ARMA(p, q)$ потрібно об'єднати все те, про що говорилося вище. За моделлю, підставляючи туди для часу t спостереження y_t та розраховані значення залишків, обчислюють прогнозовані значення y_t , а для майбутніх моментів часу — замінюють залишки нулями і замість y_t підставляють їхні прогнозовані значення. Дисперсію помилки прогнозу обчислюють за формулою (8.21).

Наприклад, для моделі $ARMA(1, 1)$:

$$\begin{aligned} y_t &= (1 - aL)^{-1} (1 + bL) \varepsilon_t = (1 + aL + a^2 L^2 + \dots) (1 + bL) \varepsilon_t = \\ &= \varepsilon_t + \eta_1 \varepsilon_{t-1} + \eta_2 \varepsilon_{t-2} + \dots, \end{aligned}$$

де $\eta_1 = a + b$, $\eta_2 = a(a + b)$, ..., $\eta_k = a^{k-1}(a + b)$, починаючи з другого коефіцієнти спадають за геометричною прогресією. Звідси легко обчислити дисперсію помилки прогнозу на τ кроків:

$$\sigma_\varepsilon^2 \sum_{j=0}^{\tau-1} \eta_j^2 = \sigma_\varepsilon^2 \left(1 + \sum_{j=0}^{\tau-1} a^{2j} (a + b)^2 \right) = \sigma_\varepsilon^2 \left(1 + \frac{(a + b)(1 - a^{2\tau})}{1 - a^2} \right). \quad (8.22)$$

Для цієї моделі дисперсія помилки прогнозу асимптотично дорівнює дисперсії часового ряду.

В усіх розглянутих випадках умовний точковий прогноз асимптотично наближається до математичного сподівання ряду, а дисперсія помилки прогнозу — до дисперсії ряду. Це означає, що для стаціонарного процесу вплив наявної інформації на прогноз та його точність асимптотично спадає до нуля. До того ж за збільшення горизонту прогнозування дисперсія помилки не перевищує дисперсії часового ряду. Цей висновок, на жаль, є наслідком нереалістичного припущення про те, що коефіцієнти моделі відомі точно.

9 ПРОГНОЗУВАННЯ ТЕНДЕНЦІ НА ОСНОВІ ЗГЛАЖУВАННЯ ЧАСОВИХ РЯДІВ

9.1 ПРОГНОЗУВАННЯ ТЕНДЕНЦІ ЧАСОВОГО РЯДУ ЗА СЕРЕДНІМИ ХАРАКТЕРИСТИКАМИ

Задача згладжування часового ряду має таке формулювання. Для реалізації (часового ряду) y_{t_i} ($i = 1, 2, \dots, n$) деякого випадкового процесу ξ_t , потрібно найкращим чином визначити оцінку деякої невинуватої компоненти (тренду) \hat{y}_t , яка в кожен фіксований момент часу є середнім значенням випадкової величини ξ_t і відображає основні закономірності зміни досліджуваного показника в часі.

Найпростішим способом прогнозування вважається підхід, який визначає прогнозну оцінку від фактично досягнутого рівня за допомогою середнього рівня, середнього приросту, середнього темпу зростання.

Екстраполяція на основі середнього рівня ряду. Під час екстраполяції соціально-економічних процесів на основі середнього рівня ряду прогнозоване значення беруть як середнє арифметичне значення попередніх рівнів ряду, тобто точковий прогноз $\hat{y}_n(\tau)$, зроблений у момент часу $t = n$ на період упередження τ , розраховують за формулою:

$$\hat{y}_n(\tau) = \bar{y}. \quad (9.1)$$

Інтервал надійності для прогнозу середньої за невеликої кількості спостережень визначається як

$$\hat{y}_n(\tau) \pm t_\alpha \hat{\sigma}_{\bar{y}}, \quad (9.2)$$

де t_α — критичне значення t — критерію Стьюдента із $n-1$ ступенями свободи й рівнем значущості α ; $\hat{\sigma}_{\bar{y}}$ — оцінка середньої квадратичної похибки середнього ($\hat{\sigma}_{\bar{y}} = \frac{\hat{\sigma}}{\sqrt{n}}$, де $\hat{\sigma}$ — оцінка середньоквадратичного відхилення спостережень).

Отриманий інтервал надійності враховує невизначеність, приховану в оцінці середньої величини. Однак залишається припущення, що прогнозований показник дорівнює середньому вибірковому значенню, тобто за такого підходу не зважають на те, що окремі значення показника коливалися навкруги середнього в минулому, і це також відбуватиметься в майбутньому. Отже, загальна дисперсія включає коливання вибіркової середньої та коливання індивідуальних значень навколо середнього і

становить величину $\hat{\sigma}^2 + \frac{\hat{\sigma}^2}{n}$, а інтервал надійності для прогнозованої оцінки ряду дорівнює:

$$\hat{y}_n(\tau) \pm t_\alpha \hat{\sigma} \sqrt{1 + \frac{1}{n}}. \quad (9.3)$$

Екстраполяцію за середнім абсолютним приростом можна бути виконати в тому разі, коли загальна тенденція розвитку вважається лінійною.

Прогнозову оцінку $\hat{y}_n(\tau)$ одержують за формулою:

$$\hat{y}_n(\tau) = y_n + \tau \cdot \overline{\Delta y}, \quad (9.4)$$

де $\overline{\Delta y}$ — середній абсолютний приріст.

Екстраполяцію за середнім темпом зростання можна виконувати у разі, коли є підстави вважати, що загальна тенденція динамічного ряду характеризується експоненціальною кривою. Прогноз $\hat{y}_n(\tau)$, зроблений у

момент часу $t = n$ на період випередження τ , у цьому разі розраховують за формулою:

$$\hat{y}_n(\tau) = y_n \cdot \bar{T}_{зр}^\tau, \quad (9.5)$$

де $\bar{T}_{зр}$ — середній темп зростання, розрахований за середньою геометричною (3.8).

Інтервал надійності прогнозу за середнім абсолютним приростом і середнім темпом зростання можна одержати лише тоді, коли ці середні визначаються за допомогою статистичного оцінювання параметрів відповідно лінійної та експоненціальної кривої.

Усі три способи привертають увагу багатьох працівників статистичних органів завдяки своїй простоті та легкості реалізації. Однак, крім зазначених позитивних якостей, вони мають кілька суттєвих недоліків. По-перше, всі фактичні спостереження є результатом закономірності та випадковості, отже, виходити тільки з останнього спостереження неправильно. По-друге, немає можливості оцінити слушність використання середньої характеристики ряду в кожному конкретному випадку. По-третє, не завжди можна розрахувати інтервал надійності, до якого потрапляє прогнозована величина, і визначити його ймовірність. У зв'язку із цим екстраполяцію за середніми характеристиками ряду застосовують лише як орієнтир майбутнього розвитку або якщо неможливо використати інші статистичні методи (наприклад, за дуже малої кількості спостережень).

9.2 ПРОГНОЗУВАННЯ ТЕНДЕНЦІЇ ЧАСОВОГО РЯДУ ЗА АНАЛІТИЧНИМИ МЕТОДАМИ ЗГЛАДЖУВАННЯ

До методів аналітичного згладжування відносять регресійний аналіз разом із методом найменших квадратів та його модифікаціями. Виявити основну тенденцію аналітичним методом — означає надати досліджуваному процесу однакового розвитку впродовж усього часу спостереження. Тому для цих методів важливо обрати оптимальну функцію детермінованого тренду v_t (кривої зростання), яка згладжує ряд спостережень y_t .

Регресійний аналіз. Оцінювання параметрів кривих зростання здійснюють на підставі побудови моделі регресії, в якій пояснювальною змінною є час:

$$y_t = v_t + \varepsilon_t \quad t = 1, 2, \dots, n, \quad (9.6)$$

де v_t — функція тренду (крива зростання);

ε_t — невідомі випадкові похибки.

Виходячи з теоретичних міркувань крива зростання може описуватися будь-якою математичною функцією v_t . Оцінювання цієї функціональної залежності здійснюють за вибірковими спостереженнями y_t , $t = 1, 2, \dots, n$, а вибір методу оцінювання залежить від виду кривої й стохастичного походження випадкових похибок ε_t . Якщо функція v_t лінійна за параметрами, наприклад, має вигляд алгебраїчного полінома ступеня p :

$$v_t = a_0 + a_1 t + \dots + a_p t^p, \quad (9.7)$$

і при цьому довжина часового ряду n суттєво перевищує ступінь полінома p , а випадкові залишки ε_t мають властивості «білого шуму», тобто

$$M(\varepsilon_t) = 0, \\ M(\varepsilon_t \varepsilon_{t+k}) = \begin{cases} \sigma^2, & \text{якщо } k = 0 \\ 0, & \text{якщо } k \neq 0 \end{cases} \quad (9.8)$$

тоді оцінки \hat{a} параметрів a можна одержати методом найменших квадратів (МНК). МНК-оцінки параметрів лінійної регресії за умови мінімізації суми квадратів відхилень точок вхідного часового ряду y_t від їхніх згладжених значень \hat{y}_t :

$$S = \sum_{t=1}^n (y_t - \hat{a}_0 - \hat{a}_1 t - \dots - \hat{a}_p t^p)^2 \quad (9.9)$$

обчислюють за формулою:

$$\hat{A} = (X^T X)^{-1} X^T Y, \quad (9.10)$$

де \hat{A} — вектор оцінок параметрів моделі; $\hat{A}^T = (\hat{a}_0, \hat{a}_1, \dots, \hat{a}_p)$

X — матриця значень спостережень пояснювальних змінних розмірністю $n \times (p + 1)$, яка у разі (9.7) має вигляд

$$X = \begin{pmatrix} 1 & 1 & 1^2 & \dots & 1^p \\ 1 & 2 & 2^2 & \dots & 2^p \\ \dots & \dots & \dots & \dots & \dots \\ 1 & n & n^2 & \dots & n^p \end{pmatrix},$$

вектор-стовпчик спостережень залежної змінної $Y = (y_1, y_2, \dots, y_n)^T$.

Побудована модель прогнозу має супроводжуватися додатковою інформацією стосовно її точності та адекватності. Якщо умова сталості дисперсії та взаємної незалежності випадкових похибок моделі (9.8) не виконується, застосовують узагальнений МНК; модель, у якій функція v_t є нелінійною за параметрами, потребує техніки статистичного аналізу нелінійних моделей регресії тощо.

Для розрахунку в момент часу $t = n$ прогнозової оцінки $\hat{y}_n(\tau)$ на період випередження τ потрібно оцінити параметри лінійного тренду \hat{A} та підставити їх у рівняння тренду (наприклад, (9.7)), де $t = n + \tau$.

Методи, розроблені для статистичних сукупностей, уможливають визначення інтервалу надійності прогнозу, який залежить від стандартної похибки оцінки прогнозованого показника, від часу випередження прогнозу, від довжини прогнозової бази та обраного рівня значущості.

Наприклад, у разі прямолінійного тренду інтервал надійності прогнозу U_y має вигляд

$$U_y = \hat{y}_n(\tau) \pm t_\alpha \hat{\sigma}_y \sqrt{1 + \frac{1}{n} + \frac{3(n + 2\tau - 1)^2}{n(n^2 - 1)}}, \quad (9.11)$$

де τ — період випередження;

$\hat{y}_n(\tau)$ — точковий прогноз на момент часу $n + \tau$;

n — кількість спостережень у часовому ряді (довжина прогнозової бази);

$\hat{\sigma}_y$ — оцінка стандартної похибки (середньоквадратичною відхилення)

оцінки \hat{y}_t , $\hat{\sigma}_y = \sqrt{\frac{\sum (y_t - \hat{y}_t)^2}{n - 2}}$;

t_α — табличне значення критерію Стюдента для рівня значущості α і числа ступенів свободи $n - 2$.

Іноді для розрахунку інтервалів надійності прогнозу відносно лінійного тренду застосовують наведену вище формулу в дещо перетвореному вигляді:

$$U_y = \hat{y}_n(\tau) \pm t_\alpha \hat{\sigma}_y \sqrt{1 + \frac{1}{n} + \frac{(t_\tau - \bar{t})^2}{\sum (t - \bar{t})^2}},$$

або

$$U_y = \hat{y}_n(\tau) \pm t_\alpha \hat{\sigma}_y \sqrt{1 + \frac{1}{n} + \frac{t_\tau^2}{\sum t^2}}, \quad (9.12)$$

якщо перенести початок відліку часу на середину періоду спостережень ($\bar{t} = 0$), де t — порядковий номер рівня ряду ($t = 1, 2, \dots, n$); $n + \tau - t$ — час, для якого здійснюють прогноз; t — час, що відповідає середині періоду спостережень вхідного ряду; підсумок робиться за всіма спостереженнями.

Формула для розрахунку інтервалів надійності прогнозу відносно тренду, який має вид полінома другого або третього порядку, виглядає так:

$$U_y = \hat{y}_n(\tau) \pm t_\alpha \hat{\sigma}_y \sqrt{1 + \frac{1}{n} + \frac{t_\tau^2}{\sum t^2} + \frac{\sum t^4 - 2t_\tau^2 \sum t^2 + nt_\tau^4}{n \sum t^4 - (\sum t^2)^2}}. \quad (9.13)$$

Аналогічно розраховують інтервали надійності для кривих зростання, які можна звести до лінійної функції.

Розглянутий розрахунок інтервалів надійності прогнозів на підставі кривих зростання, що ґрунтується на висновках і формулах теорії регресійного аналізу, для часових рядів не зовсім правомірний, оскільки динамічні ряди, як уже зазначалося, відрізняються від статистичних сукупностей. Тому до оцінювання інтервалів надійності для кривих зростання слід підходити з певною обережністю. Якщо припустити, що випадкова змінна ε_t ($y_{t+\tau} - \hat{y}_t(\tau) = \varepsilon_t$) є стаціонарним часовим рядом, то похибка прогнозу становитиме

$$y_{t+\tau} - \hat{y}_t(\tau) = \varepsilon_{t+\tau} + b_1 \varepsilon_{t+\tau-1} + b_2 \varepsilon_{t+\tau-2} + \dots + b_{\tau-1} \varepsilon_{t+1}. \quad (9.14)$$

Звідси

$$MSE = M(y_{t+\tau} - \hat{y}_t(\tau))^2 = (1 + b_1^2 + b_2^2 + \dots + b_{\tau-1}^2)\sigma_\varepsilon^2 \quad (9.15)$$

Динамічним мультиплікатором збурення ε_t , тобто величиною, яка показує, на скільки зміниться значення часового ряду через τ періодів залежно від поточного збурення, є $\frac{\partial y_{t+\tau}}{\partial \varepsilon_t} = b_\tau$. Очевидно, що вплив збурення буде спадати з часом, тому $\lim_{\tau \rightarrow \infty} \frac{\partial y_{t+\tau}}{\partial \varepsilon_t} = 0$.

Методи прогнозування, ґрунтовані на методах регресії, використовують для короткотермінового та середньострокового прогнозування. Вони не допускають адаптації: з отриманням нових даних процедура побудови прогнозу має повторюватися спочатку. Оптимальна довжина періоду випередження визначається окремо для кожного економічного процесу з урахуванням його статистичної нестабільності. Ця довжина, як правило, не перевищує для рядів річних спостережень третини обсягу даних, а для квартальних і місячних рядів — двох років.

9.3 ВИДИ КРИВИХ ЗРОСТАННЯ

Для відображення економічних процесів існує велика кількість видів кривих зростання. Щоб правильно підібрати найдоцільнішу криву для моделювання й прогнозування економічного явища, необхідно знати особливості кожного виду кривих. Криві зростання описують різні тенденції економічних процесів, наприклад, життєвий цикл товару, процес нагромадження капіталу, маркетингові зусилля фірм тощо. В економічній практиці вже здобуто певний досвід і розроблено певні типи кривих, які найчастіше використовують у соціально-економічних дослідженнях. До таких кривих належать: поліноміальні, експоненціальні та S-подібні криві зростання.

Поліноміальні криві зростання можна використовувати для апроксимації (наближення) та прогнозування економічних процесів, у яких майбутній розвиток не залежить від досягнутого рівня. Простіші поліноміальні криві зростання мають вигляд:

$$\begin{aligned} \hat{y}_t &= a_0 + a_1 t \text{ (поліном першого ступеня),} \\ \hat{y}_t &= a_0 + a_1 t + a_2 t^2 \text{ (поліном другого ступеня),} \\ \hat{y}_t &= a_0 + a_1 t + a_2 t^2 + a_3 t^3 \text{ (поліном третього ступеня) тощо.} \end{aligned} \quad (9.16)$$

Поліноміальні моделі лінійні за параметрами. Параметри цих моделей (лінійної, квадратичної, полінома третього ступеня) мають такі економічні тлумачення: a_1 — лінійний приріст, a_2 — прискорення зростання, a_3 — характеризує динаміку прискорення зростання.

Для полінома першого ступеня характерний постійний приріст. Якщо обчислити перші прирости за формулою $\Delta_t^1 = y_t - y_{t-1}$, $t = 2, 3, \dots, n$, то вони будуть постійними величинами та дорівнюватимуть a_1 .

Якщо перші прирости обчислити для полінома другого ступеня, то вони матимуть лінійну залежність від часу і ряд із перших приростів $\Delta_1^1, \Delta_2^1, \dots$ на графіку буде представлений прямою лінією. Другі прирости $\Delta_t^2 = \Delta_t^1 - \Delta_{t-1}^1$ для полінома другого ступеня будуть постійними.

Для полінома третього ступеня перші прирости будуть поліномами другого ступеня, другі прирости будуть лінійною функцією часу, а треті прирости, які обчислюють за формулою $\Delta_t^3 = \Delta_t^2 - \Delta_{t-1}^2$, будуть постійними величинами.

Звідси можна відзначити такі властивості поліноміальних кривих зростання:

- від полінома високого ступеня шляхом розрахунку послідовних різниць (приростів) можна перейти до полінома нижчого порядку;
- значення приростів для поліномів будь-якого порядку є сталими величинами.

Експоненціальні криві використовують для зображення швидко зростаючих або спадних економічних процесів. Використання експоненціальних кривих зростання передбачає, що майбутній розвиток залежить від досягнутого рівня, тобто приріст залежить від значення функції.

В економіці використовують два різновиди експоненціальних кривих: просту експоненту та модифіковану експоненти.

Проста експонента може набувати різноманітних еквівалентних форм.

$$\hat{y}_t = ab^t, \text{ основна форма } b > 0. \quad (9.17)$$

$$\hat{y}_t = ae^{\beta t}, \text{ } b \text{ замінюємо на } e^\beta, \text{ де } \beta = \ln(b). \quad (9.18)$$

$$\hat{y}_t = a(1-r)^t, \text{ } b \text{ замінюємо на } (1-r), \text{ де } r = b-1. \quad (9.19)$$

$$\hat{y}_t = e^{\alpha + \beta t}, \text{ де } a \text{ замінюємо на } e^\alpha, \text{ і } b \text{ на } e^\beta. \quad (9.20)$$

$$\hat{y}_t = 10^{\alpha + \beta t}, \text{ де } a \text{ замінюємо на } 10^\alpha, \text{ і } b \text{ на } 10^\beta, \quad (9.21)$$

де a й b — додатні числа, при цьому якщо $b > 1$, то функція зростає, якщо $b < 1$ — спадає.

Усі ці форми використовують на практиці для описання різних економічних процесів, наприклад, форму (9.19) найчастіше використовують у фінансах, де r означає норму річного відсотка.

Логарифми ординат простої експоненти лінійно залежать від часу.

Наприклад, для функції (9.17) $\ln \hat{y}_t = \ln a + t \ln b$, тобто темп зростання постійний для будь-якого моменту часу. Якщо ця крива застосовується для зображення інфляції, то коефіцієнт b буде характеризувати темп інфляції. Можна помітити, що ордината цієї функції змінюється з постійним темпом приросту. Якщо взяти відношення приросту до самої ординати, то воно

буде сталою величиною: $\frac{\Delta_t^1}{y_t} = \frac{y_t - y_{t-1}}{y_t} = 1 - \frac{1}{b}$.

Модифікована експонента має вигляд:

$$\hat{y}_t = a + bc^t, \quad (9.22)$$

де постійні величини: $b < 0$, $0 < c < 1$, а константа a має назву асимптоти цієї функції, тобто значення функції необмежено наближаються (знизу) до величини a . Можуть бути й інші варіанти модифікованої експоненти, але на практиці найчастіше трапляється розглянута вище функція. Наприклад, якщо на ринку з'являється новий товар, який супроводжується широкою рекламою, то спочатку попит на цей товар буде досить великий і швидкість продажу товару буде значною. Із часом продаж буде стабілізуватися та дійде до певного рівня насичення. У таких випадках фаза уповільненого зростання відсутня, і найкраще згладжування дасть модифікована експонента.

Логарифми перших приростів цієї функції лінійно залежать від часу, а якщо взяти відношення двох послідовних приростів, то воно буде сталою величиною:

$$\frac{\Delta_t^1}{\Delta_{t-1}^1} = \frac{y_t - y_{t-1}}{y_{t-1} - y_{t-2}} = c.$$

Модифікована експонента слугує базовою кривою, на підставі якої за допомогою певних перетворень отримують криву Гомперця (9.28) і логістичну криву (9.29), які використовують частіше.

Степенева крива. Рівняння степеневі кривої має вигляд:

$$\hat{y}_t = at^b. \quad (9.23)$$

Степенева крива добре згладжує показники, які з часом монотонно зростають, якщо $b > 0$, або спадають, якщо $b < 0$. Зокрема, за $b = -1$, $\hat{y}_t = a/t$. Це рівняння задає гіперболу, асимптотами якої є вісі координат, а добуток змінних є сталою величиною ($\hat{y}_t \cdot t = a$). В економіці такій умові задовольняє крива попиту з одиничною еластичністю: відсоток збільшення одиниці часу t на такий самий відсоток зменшує залежну змінну \hat{y}_t . На практиці степеневі функції використовують для зображення різноманітних економічних процесів. Найвідомішою з них є виробнича функція Кобба-Дугласа. Крім того, вони застосовуються для зображення кривих байдужості, а також попиту на товари різних категорій (так звана крива Торнквіста) тощо.

Гіперболічна крива I типу. Звичайна гіпербола задається рівнянням:

$$\hat{y}_t = a + b/t. \quad (9.24)$$

Для цього типу гіперболи за $b > 0$ значення \hat{y}_t зменшується мірою зростання t і асимптотично наближається до a . Такого виду криву можна застосовувати для вирівнювання й прогнозування показника, який із часом спадає до певного відмінного від нуля рівня.

За $b < 0$ значення \hat{y}_t додатне, тільки якщо $t > b/a$; збільшення t приводить у цьому випадку і до збільшення \hat{y}_t з асимптотичною межею, що дорівнює a . Таким типом гіперболи доцільно зображувати зростаючі процеси з насиченням.

Гіперболічна крива II типу. Цей тип гіперболи задається рівнянням

$$\hat{y}_t = 1/(a + bt) \quad (9.25)$$

За $b > 0$ значення \hat{y}_t прагнуть до нуля у разі необмеженого збільшення часу t ; за $b < 0$ значення \hat{y}_t прагне до нескінченності, якщо t наближається до a/b . Остання ситуація на практиці мало ймовірна.

Гіперболічна крива III типу (проста раціональна залежність). Задається рівнянням

$$\hat{y}_t = t/(a + bt). \quad (9.26)$$

Для цього типу гіперболи незалежно від коефіцієнта b за $t=0$ $\hat{y}_t = 0$. Для додатних значень b значення \hat{y}_t зростає та асимптотично прагне до величини $1/b$ за необмеженого збільшення t . За від'ємного b ця крива, як і гіпербола другого типу, стає нестійкою за $t = a/b$.

S-подібна крива. В економіці поширені процеси, які спочатку поступово зростають, прискорюються, а потім знов уповільнюються свій розвиток, прагнучи певної межі. Наприклад, процес введення промислового об'єкта до експлуатації або зміна попиту на товари, що мають межу насичення тощо. Для моделювання таких процесів використовують так звані S-подібні криві зростання, які мають вигляд:

$$\hat{y}_t = e^{a+b/t} \quad (9.27)$$

Насправді ця крива має форму S тільки за від'ємного значення b та за умови, що його абсолютне значення більше за a . Якщо крива (3.2.22) справді має форму S, вона використовується для зображення повного циклу розвитку динамічних процесів. Повний цикл таких процесів починається з повільного зростання, потім настає фаза бурхливого розвитку і, нарешті, розвиток завершується періодом насичення (тобто асимптотичного наближення до величини e^a). Таке чергування фаз властиве багатьом соціально-економічним процесам. Для S-подібної кривої точку перегину, в якій швидкість зростання досягає максимального значення, знаходять розв'язок рівняння $f''(t) = 0$, де $f''(t)$ — друга похідна за t кривою $f(t)$. Для S-подібної кривої точкою перегину, тобто точкою, в якій зростання коефіцієнта нахилу дотичної змінюється спадом, буде точка $t = -b/2$. Утім, на практиці для опису таких процесів замість S-подібної кривої використовують більш гнучкі й адекватні криві: Гомперця та логістичну.

Крива Гомперця має такий аналітичний вираз:

$$\hat{y}_t = ab^{c^t}, \quad (9.28)$$

де c, b — додатні параметри, причому $c < 1$; параметр a — асимптота функції.

У кривій Гомперця виокремлюють чотири ділянки: на першій приріст функції незначний, на другій — збільшується, на третій ділянці приріст майже постійний, на четвертій — відбувається вповільнення темпів приросту, і функція необмежено наближається до значення a . В результаті конфігурація кривої нагадує латинську літеру S. Точкою перегину цієї

кривої буде $t_p = \frac{1}{\ln c} \ln\left(-\frac{1}{\ln b}\right)$ зі значенням функції \hat{y}_{t_p} , яке дорівнює $\hat{y}_{t_p} = \frac{a}{e}$, де $e = 2,71828$. Логарифм цієї функції ($\ln \hat{y}_t = \ln a + c^t \ln b$) є модифікованою експонентою; логарифм відношення першого приросту до самої ординати функції лінійною функцією часу.

На підставі кривої Гомперця будується, наприклад, динаміка показників рівня життя; модифікації цієї кривої використовують у демографії для моделювання показника смертності тощо.

Логістична крива, або крива Перла-Ріда — зростаюча функція, яку найчастіше записують у вигляді

$$\hat{y}_t = \frac{1}{a + bc^t} \quad (9.29)$$

У цьому виразі b і c — додатні параметри; a — граничне значення функції за нескінченного зростання часу.

Якщо взяти похідну від цієї функції, можна побачити, що швидкість зростання логістичної кривої у будь-який момент часу пропорційна досягнутому рівню функції й різниці між граничним значенням a і досягнутим рівнем. Логарифм відношення першого приросту функції до квадрата її значення (ординати) є лінійною функцією від часу.

Конфігурація графіка логістичної кривої близька до графіка кривої Гомперця, але, на відміну від останнього, логістична крива має точку симетрії, яка збігається із точкою перегину. Точка перегину дорівнює $t_p = \frac{1}{\ln c} \ln \frac{a}{b}$. Значення \hat{y}_{t_p} у точці перегину дорівнює $\hat{y}_{t_p} = \frac{1}{2}a$.

Метод найменших квадратів і процедури регресійного аналізу є доцільними для випадку, коли рівняння кривої зростання \hat{y}_t після деяких перетворень можна звести до лінійної регресії. У таблиці 9.1. наведено криві зростання, які найчастіше спостерігаються в соціально-економічних дослідженнях, їхні математичні функції \hat{y}_t та перетворення, необхідні для зведення функцій до лінійного вигляду.

Таблиця 9.1

ВИДИ КРИВИХ ЗРОСТАННЯ

Основні види кривих зростання	Математична функція	Лінеаризація функції
1	2	3
Лінійна (поліном першого ступеня)	$\hat{y}_t = a_0 + a_1 t$	Не потрібна
Квадратична (поліном другого ступеня)	$\hat{y}_t = a_0 + a_1 t + a_2 t^2$	$\hat{y}_t = a_0 + a_1 z_1 + a_2 z_2$
Поліном третього ступеня	$\hat{y}_t = a_0 + a_1 t + a_2 t^2 + a_3 t^3$	$\hat{y}_t = a_0 + a_1 z_1 + a_2 z_2 + a_3 z_3$

Експонента (проста)	$\hat{y}_t = ae^{bt}$	$\ln \hat{y}_t = \ln a + bt$
Логарифмічна крива	$\hat{y}_t = a + b \cdot \ln t$	$\hat{y}_t = a + bt, t = \ln t$
S-подібна крива	$\hat{y}_t = e^{a+b/t}$	$\ln \hat{y}_t = a + b/t$
Обернена логарифмічна крива	$\hat{y}_t = \frac{1}{a + b \ln t}$	$Y = \frac{1}{y}, \hat{Y} = a + bz, z = \ln t$
Степенева	$\hat{y} = at^b$	$\ln \hat{y} = \ln a + b \ln t$
Гіперболічна крива I типу	$\hat{y} = a + b/t$	$\hat{y} = a + bz, z = 1/t$
Гіперболічна крива II типу	$\hat{y} = \frac{1}{a + bt}$	$Y = \frac{1}{y}, \hat{Y} = a + bt$
Гіперболічна крива III типу	$\hat{y}_t = \frac{t}{a + bt}$	$Y = \frac{1}{y}, \hat{Y} = a'z + b',$ $a' = b, b' = a, z = \frac{1}{t}$
Модифікована експонента	$\hat{y} = a + bc^t$	$c = \frac{(n-1) \sum_{t=1}^{n-1} y_t y_{t+1} - \sum_{t=1}^{n-1} y_t \sum_{t=1}^{n-1} y_{t+1}}{(n-1) \sum_{t=1}^{n-1} y_t^2 - (\sum_{t=1}^{n-1} y_t)^2}$; ; $b = \frac{n \sum c^t y_t - \sum c^t \sum y_t}{n \sum c^{2t} - (\sum c^t)^2}$; $a = \frac{\sum y_t - b \sum c^t}{n}$
Крива Гомперця	$\hat{y} = ab^{c^t}$	$\ln \hat{y} = \ln a + c^t \ln b$
Логістична крива	$\hat{y} = \frac{1}{a + bc^t}$	$Y = \frac{1}{y}, \hat{Y} = a + bc^t$

Як бачимо з таблиці 9.1, у практиці криволінійного вирівнювання широко використовують два види перетворень: логарифмування ($\ln y$) і зворотне перетворення $\left(\frac{1}{t}\right)$. При цьому можливі перетворення як залежної змінної y , так і незалежної t або обох одночасно. Параметри S-подібних кривих (Гомперця та логістичної кривої) визначаються складнішим способом. Їх можна отримати із модифікованої експоненти, так само, як були отримані зі звичайної лінійної регресії криві, розглянуті раніше. Криві, побудовані за модифікованою експонентою, задаються трьома параметрами (замість двох параметрів у лінійній залежності). Спочатку визначають параметр c , а потім два інші параметри: a та b . У табл. 9.1 наведено відповідні перетворення функції Гомперця та логістичної кривої на модифіковану експоненту.

Апроксимація спостережень складними функціями дає задовільне наближення до фактичних спостережень, але зменшує сталість моделі на інтервалі упередження прогнозу. Тому використовувати для прогнозування такі моделі (наприклад, поліном вище другого ступеня) слід обережно. В комп'ютерних програмах використовують близько двох

десятків моделей. Зазначимо, що пошук параметрів функції Гомперця та логістичної кривої, через неможливість їхньої лінеаризації здійснюють методом багатовимірної числової оптимізації.

9.4 ВИБІР КРИВОЇ ЗРОСТАННЯ

Правильно встановити вид кривої, тобто вид аналітичної залежності значення показника від часу — одне з найважчих завдань. Обрана функція тренду має задовольняти такі умови: бути теоретично обґрунтованою; мати якнайменшу кількість параметрів; параметри функції повинні мати економічне тлумачення; оцінені значення тренду мають якомога менше відрізнятися від відповідних фактичних спостережень часового ряду.

Вибір форми кривої для згладжування певною мірою залежить від мети згладжування: інтерполяції або екстраполяції. У першому випадку метою є досягнення найбільшої близькості до фактичних рівнів часового ряду. У другому — виявлення основної закономірності розвитку явища, стосовно якої можна припустити, що в майбутньому вона збережеться.

В основі вибору кривої лежить теоретичний аналіз сутності економічного явища, зміни якого відображаються часовим рядом. Іноді до уваги беруть міркування стосовно характеру зростання рівнів ряду. Так, якщо зростання випуску продукції передбачається у вигляді арифметичної прогресії, то згладжування відбувається за прямою; якщо зростання йде в геометричній прогресії, то згладжування виконують за показниковою функцією.

На практиці під час попереднього аналізу часового ряду обирають, як правило, дві-три криві зростання для подальшого дослідження і побудови трендової моделі часового ряду. Розглянемо проблему вибору виду кривої зростання для конкретного часового ряду.

Метод послідовних різниць (Тінтнера). Цей метод може бути використаний для визначення порядку (ступеня) апроксимаційного полінома, якщо, по-перше, рівні часового ряду складаються лише із двох компонент: тренду та випадкової, і, по-друге, тренд є досить гладеньким, щоб його можна було згладити поліномом певного ступеня. Алгоритм застосування методу ідентичний алгоритму визначення порядку інтеграції нестационарного процесу і передбачає такі кроки.

1. Розраховують різниці (прирости) до d -го порядку включно:

$$\begin{aligned}\Delta_t^1 &= y_t - y_{t-1}; \\ \Delta_t^2 &= \Delta_t^1 - \Delta_{t-1}^1; \\ &\dots \\ \Delta_t^d &= \Delta_t^{d-1} - \Delta_{t-1}^{d-1}.\end{aligned}\tag{9.30}$$

Для апроксимації економічних процесів зазвичай розраховують різниці до четвертого порядку.

2. Для вхідного ряду та для кожного різницевого ряду обчислюють дисперсії за такими формулами:

для вхідного ряду —

$$s_0^2 = \frac{\sum_{t=1}^n y_t^2 - \frac{1}{n} \left(\sum_{t=1}^n y_t \right)^2}{n-1}; \quad (9.31)$$

для різницевого ряду d -го порядку ($d = 1, 2, \dots$) —

$$s_d^2 = \frac{\sum_{t=d+1}^n (\Delta_t^d)^2}{(n-d)c_{2d}^d}, \quad (9.32)$$

де c_{2d}^d — біноміальний коефіцієнт.

3. Порівнюють значення кожної наступної дисперсії із попередньою, тобто розраховують різниці $|s_d^2 - s_{d-1}^2|$, і якщо для будь-якого k ця величина не перевищує певної наперед заданої додатної величини, тобто порядок величин дисперсій однаковий, то ступінь апроксимаційного полінома має дорівнювати $d - 1$.

Необхідно зазначити, що для визначення тренду в економічних часових рядах не слід використовувати поліноми дуже великого порядку, оскільки отримані в такий спосіб функції згладжування відобразатимуть випадкові відхилення, а не детерміновану складову, що суперечить поняттю тенденції.

Метод характеристик приросту є універсальним методом попереднього вибору кривих зростання. Він ґрунтується на використанні окремих характерних властивостей кривих, розглянутих вище. За цього методу вхідний часовий ряд попередньо згладжують методом простої змінної середньої. Наприклад, для інтервалу згладжування $m = 3$ згладжені рівні розраховують за формулою:

$$\bar{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}, \quad (9.33)$$

причому щоб не втратити перший та останній рівні, їх згладжують за формулами:

$$\bar{y}_1 = \frac{5y_1 + 2y_2 - y_3}{6}, \quad \bar{y}_n = \frac{-y_{n-2} + 2y_{n-1} + 5y_n}{6}. \quad (9.34)$$

Далі обчислюють перші середні прирости

$$\bar{\Delta}_t = \frac{\bar{y}_{t+1} - \bar{y}_{t-1}}{2}, \quad t = 2, 3, \dots, l, n - 1; \quad (9.35)$$

другі середні прирости

$$\bar{\Delta}_t^{(2)} = \frac{\bar{\Delta}_{t+1} - \bar{\Delta}_{t-1}}{2}, \quad (9.36)$$

а також ряд похідних величин:

$$\frac{\bar{\Delta}_t}{\bar{y}_t}; \log \bar{\Delta}_t; \log \frac{\bar{\Delta}_t}{\bar{y}_t}; \log \frac{\bar{\Delta}_t}{\bar{y}_t^2}. \quad (9.37)$$

Відповідно до характеру зміни середніх приростів і похідних показників обирають вид кривої зростання для вхідного часового ряду [27], при цьому використовують відомості з табл. 9.2.

Таблиця 9.2.

ВИБІР КРИВОЇ ЗРОСТАННЯ ЗА ХАРАКТЕРОМ ЗМІНИ ПОКАЗНИКА

Показник	Характер зміни показника з часом	Вид кривої зростання
Перший середній приріст $\bar{\Delta}_t$	Майже однаковий	Поліном першого порядку (пряма)
$\bar{\Delta}_t$	Змінюється лінійно	Поліном другого порядку (парабола)
Другий середній приріст $\bar{\Delta}_t^{(2)}$	Змінюється лінійно	Поліном третього порядку (кубічна парабола)
$\frac{\bar{\Delta}_t}{\bar{y}_t}$	Майже однаковий	Проста експонента
$\log \bar{\Delta}_t$	Змінюється лінійно	Модифікована експонента
$\log \frac{\bar{\Delta}_t}{\bar{y}_t}$	Змінюється лінійно	Крива Гомперця
$\log \frac{\bar{\Delta}_t}{\bar{y}_t^2}$	Змінюється лінійно	Логістична крива

ПРИКЛАДИ РОЗВ'ЯЗУВАННЯ ЗАДАЧ

Задача 1. Є часовий ряд

$$t: 1 \ 2 \ 3 \dots 9$$

$$y_t: 25 \dots \dots \dots 10$$

Відомо, що, $\sum y_t = 130$; $\sum y_t^2 = 3100$; $\sum_{t=2}^n y_t y_{t-1} = 2552$

Визначити для цього часового ряду значення коефіцієнта автокореляції першого порядку.

Розв'язок. Значення коефіцієнта визначимо за формулою:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}$$

Розпишемо всі компоненти цієї формули. Чисельник перетворимо наступним шляхом:

$$\begin{aligned} \sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2) &= \sum_{t=2}^n (y_t y_{t-1} - \bar{y}_1 y_{t-1} - \bar{y}_2 y_t + \bar{y}_1 \bar{y}_2) \\ &= \sum_{t=2}^n y_t y_{t-1} - \bar{y}_1 \sum_{t=2}^n y_{t-1} - \bar{y}_2 \sum_{t=2}^n y_t + (n-1) \bar{y}_1 \cdot \bar{y}_2 \end{aligned}$$

Значення середніх обчислюємо за відповідними формулами; при цьому значення сум розраховуються з урахуванням крайніх значень часового ряду:

$$\sum_{t=2}^n y_{t-1} = \sum_{t=1}^n y_t - y_n = 130 - 10 = 120$$

$$\sum_{t=2}^n y_t = \sum_{t=1}^n y_t - y_1 = 130 - 25 = 105$$

$$\bar{y}_1 = \frac{105}{8} = 13,125; \quad \bar{y}_2 = \frac{120}{8} = 15$$

Звідки:

$$\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2) = 2552 - 13,125 \cdot 120 - 15 \cdot 105 + 8 \cdot 13,125 \cdot 15 = 977.$$

Аналогічно розраховуємо кожен член в знаменнику:

$$\begin{aligned} \sum_{t=2}^n (y_t - \bar{y}_1)^2 &= \sum_{t=2}^n (y_t^2 - 2\bar{y}_1 \cdot y_t + \bar{y}_1^2) = \\ &= \sum_{t=2}^n y_t^2 - 2\bar{y}_1 \cdot \sum_{t=2}^n y_t + (n-1)\bar{y}_1^2 = \\ &= \sum_{t=1}^n y_t^2 - y_1^2 - 2\bar{y}_1 \cdot \sum_{t=2}^n y_t + (n-1)\bar{y}_1^2 = \\ &= 3100 - 25^2 - 2 \cdot 13,125 \cdot 105 + 8 \cdot 13,125^2 = 1096,87. \\ \sum_{t=2}^n (y_t - \bar{y}_2)^2 &= \sum_{t=2}^n (y_{t-1}^2 - 2\bar{y}_2 \cdot y_{t-1} + \bar{y}_2^2) = \\ &= \sum_{t=2}^n y_{t-1}^2 - 2\bar{y}_2 \cdot \sum_{t=2}^n y_{t-1} + (n-1)\bar{y}_2^2 = \\ &= \sum_{t=1}^n y_t^2 - y_n^2 - 2\bar{y}_2 \cdot \sum_{t=2}^n y_{t-1} + (n-1)\bar{y}_2^2 = \\ &= 3100 - 10^2 - 2 \cdot 15 \cdot 120 + 8 \cdot 15^2 = 1200. \end{aligned}$$

Результат визначимо по вихідній розрахунковій формулі:

$$r_1 = \frac{977}{\sqrt{1096,875 \cdot 1200}} = \frac{977}{1147,28} = 0,852$$

Приклад 2. На основі квартальних даних обсягів продажів підприємства за 2010-2015 рр. була побудована аддитивна модель тимчасового ряду, трендова компонента якої має вигляд:

$$T = 200 + 3 \cdot t \quad (t = 1, 2, \dots)$$

Показатели за 2014 г. приведены в таблице:

Квартал	Фактичний обсяг продажів	Компонента адитивної моделі		
		трендова	сезонна	випадкова
1	2	3	4	5
1	200			-11
2			15	5
3	250		32	
4				

Визначити відсутні в таблиці дані, з огляду на що загальний обсяг продажів за 2014 р склав 1000 тис.грн.

Розв'язок. В першу чергу визначимо всі значення трендової компоненти. Щоб використовувати наявне рівняння тренда, треба визначити моменти часу, які відповідають 2014 р. Оскільки модель відноситься до періоду 2010 - 2015 рр., тобто охоплює 6 років, квартальні тимчасові позначки змінюються від 1 до 24. У цьому випадку 2014 г. (передостанній в досліджуваному періоді) відповідає моментам часу 17, 18, 19 і 20.

Після підстановки в рівняння тренда отримуємо

$$T_1 = 200 + 3 \cdot 17 = 251;$$

$$T_2 = 200 + 3 \cdot 18 = 254;$$

$$T_3 = 200 + 3 \cdot 19 = 257;$$

$$T_4 = 200 + 3 \cdot 20 = 260.$$

Далі відсутні величини для першого, другого і третього кварталів обчислюємо по балансу з рівняння для адитивної моделі часового ряду:

$$S_1 = y_1 - T_1 - E_1 = 200 - 251 - (-11) = -40;$$

$$y_2 = T_2 + S_2 + E_2 = 254 + 15 + 5 = 274;$$

$$E_3 = y_3 - T_3 - S_3 = 250 - 257 - 32 = -39.$$

Залишилося визначити тільки величини для четвертого кварталу, для якого відомими є тільки значення трендової компоненти. В умовах задачі задано загальний обсяг продажів за рік. Оскільки відомі є продажі за три перші квартали, четвертий визначається:

$$y_4 = 1000 - y_1 + y_2 + y_3 \approx 1000 - 200 + 274 + 250 \approx 276.$$

Для розрахунку сезонної компоненти за 4 - й квартал скористаємося тим, що в адитивній моделі сума сезонних компонент за один період повинна дорівнювати нулю:

$$S_4 = -S_1 + S_2 + S_3 \approx -40 + 15 + 32 \approx -7.$$

Останнє значення в таблиці - випадкову компоненту за 4 - й квартал - обчислюємо по балансу з рівняння адитивної моделі, оскільки всі інші компоненти вже відомі:

$$E_4 = y_4 - T_4 - S_4 = 276 - 260 + 7 = 23.$$

Квартал	Фактичний	Компонента адитивної моделі
---------	-----------	-----------------------------

		трендова	сезонна	випадкова
1	2	3	4	5
1	200	251	- 40	-11
2	274	254	15	5
3	250	257	32	- 39
4	276	260	- 7	23

Приклад 3. На основі поквартальних даних за 9 останніх років була побудована мультиплікативна модель деякого часового ряду. Рівняння тренду в цій моделі має вигляд:

$$T_1 = 10,8 + 0,1 \cdot t.$$

Скориговані значення сезонної компоненти наступні: в 1-му кварталі - 1,5; в 3-му кварталі - 0,6; в 4-му кварталі - 0,8. Визначити сезонну компоненту за 2 - й квартал і прогноз модельованого показника за 2 - й і 3 - й квартали наступного року

Розв'язок. У мультиплікативній моделі сума скоригованих сезонних компонент за один період повинні дорівнювати кількості цих коефіцієнтів, тобто чотирьом. Звідси знаходимо відсутню сезонну компоненту за 2-й квартал:

$$S_2 = 4 - (S_1 + S_3 + S_4) = 4 - (1,5 + 0,6 + 0,8) = 1,1.$$

Для прогнозування по мультиплікативній моделі скористаємося співвідношенням, в якому не будемо враховувати випадкову компоненту. При цьому слід мати на увазі, що 2-й і 3-й квартали наступного року будуть ставитися в рамках даної моделі відповідно до 38-ї та 39-ї позначок часу відповідно:

$$\begin{aligned} \hat{Y}_{38} &= (10,8 + 0,1 \cdot 38) \cdot 1,1 = 16,06; \\ \hat{Y}_{39} &= (10,8 + 0,1 \cdot 39) \cdot 0,6 = 8,82. \end{aligned}$$

Приклад 4. На основі помісячних даних за останні 5 років була побудована аддитивна тимчасова модель споживання тепла в районі. Скориговані значення сезонної компоненти наведені в таблиці.

Січень	+ 27	Травень	- 20	Вересень	- 10
Лютий	+ 22	Червень	- 34	Жовтень	+ 12
Березень	+ 15	Липень	- 42	Листопад	+20
Квітень	- 2	Серпень	- 18	Грудень	?

Рівняння тренду виглядає наступним чином:

$$T = 300 + 1,1 \cdot t.$$

Визначити значення сезонної компоненти за грудень, а також точковий прогноз споживання тепла на 2-й квартал наступного року.

Розв'язок. У адитивній моделі часового ряду сума скоригованих сезонних компонент за один період, в даному випадку за рік, повинна дорівнювати нулю. Звідси значення сезонної компоненти за грудень:

$$S_{12} = - \sum_{i=1, (i \neq 12)}^{12} S_i = -(7 + 22 + 15 - 2 - 20 - 34 - 42 - 18 - 10 + 12 + 20) = -30.$$

Прогноз споживання тепла розраховується за формулою для детермінованої складової ряду, в якій не враховується випадкова складова, оскільки вона не прогнозується. Для розрахунку трендової компоненти слід мати на увазі, що в другому кварталі наступного року (квітень, травень, червень) відповідають позначки часу 64, 65 і 66. Прогноз за другий квартал складається з прогнозів за квітень, травень і червень.

$$f(\text{квітень}) = 600 + 1,1 \cdot 64 - 2 = 368,4;$$

$$f(\text{травень}) = 600 + 1,1 \cdot 65 - 20 = 351,5;$$

$$f(\text{червень}) = 600 + 1,1 \cdot 66 - 34 = 338,6;$$

$$f(2\text{-й квартал}) = 368,4 + 351,5 + 338,6 = 1058,5.$$

Приклад 5. Дана таблиця:

Момент часу	$t-3$	$t-2$	$t-1$	t	$t+1$
S^*	130				
S	145	165	190	210	-

де S^* , S - очікуваний і дійсний обсяги пропозиції. Визначити значення відповідно до моделі адаптивних очікувань, прийнявши $\lambda = 0,55$

Розв'язок. Розрахунок очікуваних значень проводимо за формулою:

$$S_{t+1}^* = \lambda S_t + (1 - \lambda) \bar{S}_t^*,$$

яка модифікується для кожного моменту часу $(-2, t-1, t)$:

$$S_{t-2}^* = \lambda S_{t-3} + (1 - \lambda) \bar{S}_{t-3}^* = 0,55 \cdot 145 + (1 - 0,55) \cdot 130 = 138,25;$$

$$S_{t-1}^* = \lambda S_{t-2} + (1 - \lambda) \bar{S}_{t-2}^* = 0,55 \cdot 165 + (1 - 0,55) \cdot 138,25 = 152,96;$$

$$S_t^* = \lambda S_{t-1} + (1 - \lambda) \bar{S}_{t-1}^* = 0,55 \cdot 190 + (1 - 0,55) \cdot 152,96 = 173,33;$$

$$S_{t+1}^* = \lambda S_t + (1 - \lambda) \bar{S}_t^* = 0,55 \cdot 210 + (1 - 0,55) \cdot 173,33 = 193,50.$$

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Основні поняття і види часових рядів. Класифікація часових рядів. Вимоги порівнянності, однорідності, стійкості, достатньої сукупності спостережень. Коригування рівнів часових рядів. Метод Ірвіна.
2. Розрахунок характеристик динаміки розвитку економічних процесів. Статистичні характеристики часових рядів. Структурний аналіз часового ряду (тренд, циклічна, сезонна, випадкова складові).
3. Коригування рівнів часових рядів. Метод Ірвіна.
4. Розрахунок характеристик динаміки розвитку економічних процесів.
5. Статистичні характеристики часових рядів.
6. Структурний аналіз часового ряду (тренд, циклічна, сезонна, випадкова складові).
7. Стаціонарні та нестаціонарні процеси.
8. Білий шум.
9. Ідентифікація часового ряду.
10. Перевірка стаціонарності та визначення порядку інтегрування (метод Форстера-Стьюарта, метод Діккі-Фуллера).
11. Прогнозування за середніми характеристиками ряду: екстраполяція на основі середнього рівня ряду, екстраполяція за середнім абсолютним приростом, екстраполяція за середнім темпом зростання.
12. Аналітичні методи згладжування часових рядів.
13. Екстраполяція трендів на основі кривих зростання.
14. Вибір функції, що характеризує тенденцію ринкової кон'юнктури: лінійна, експоненціальна, степенева, гіперболічна, логістична, функція Гомперця тощо. Зведення кривої зростання до лінійної регресії.
15. Метод ковзної середньої.
16. Метод центрованих та зважених ковзних середніх.
17. Метод простого експоненціального згладжування. Вибір параметрів згладжування. Розрахунок прогнозу.
18. Адаптивні методи прогнозування: Брауна, Хольта.
19. Методи фільтрації сезонної компоненти.
20. Фільтрація сезонної компоненти за допомогою індексу сезонності.
21. Метод декомпозиції часового ряду.
22. Ітераційні методи фільтрації.
23. Моделі прогнозування сезонних процесів.
24. Множинна регресія.
25. Прогнозування на основі багатofакторних регресійних моделей.
26. Індивідуальні експертні методи: метод інтерв'ю, аналітичні експертні оцінки.

27. Колективні експертні методи: метод комісій, колективної генерації ідей («мозкової атаки»), дельфійський метод.
28. Методика розрахунку експертних оцінок.
29. Етапи проведення експертизи.
30. Вибір експертів та визначення їх чисельності.
31. Оцінка відносної важливості та методика її розробки.
32. Оцінка питомої ваги різних видів рішень.
33. Перевірка адекватності моделі прогнозування.
34. Перевірка випадковості коливань рівнів залишкової послідовності.
35. Перевірка відповідності випадкової компоненти нормальному закону розподілу.
36. Перевірка рівності математичного сподівання випадкової компоненти нулю.
37. Перевірка незалежності значень випадкової компоненти.
38. Перевірка точності моделі прогнозування.
39. Основні статистики: коефіцієнти детермінації та кореляції, F- критерій, побудова довірчих інтервалів параметрів за допомогою t- критерію.
40. Прогнозування за трендовою моделлю. Основні статистики міри точності прогнозів. Побудова інтервалів прогнозування.
41. Перевірка прогнозної якості моделі.
42. Критерії визначення якісного прогнозу.
43. Прогноз раціональних сподівань.
44. Стандартні критерії ефективності та незміщеності.
45. Параметричні та непараметричні показники точності прогнозу.

ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

ПЕРЕЛІК ПИТАНЬ

1. Експертні методи прогнозування.
2. Метод експертних оцінок. Його види, переваги і недоліки.
3. Методи колективної експертної оцінки.
4. Ранжирування.
5. Етапи колективної генерації ідей («мозкова атака»).
6. Особливості застосування методу «Делфі» в соціально-економічному прогнозуванні.
7. Побудова методу «сценаріїв» та прогнозних графів.
8. Оцінка ступеня узгодженості думок експертів.
9. Оцінка статистичної надійності моделі.
10. Надання моделям статистичної надійності.
11. Прогнозування окремих макроекономічних показників.
12. Моделі економічного зростання. Макроекономічна виробнича функція. Основні типи виробничих функцій
13. Регресійна модель інфляції.

14. Економетрична модель прогнозу темпу інфляції та обсягу виробництва. Середньострокова агрегована макромодель ціноутворення.
15. Моделі прогнозування структури національної економіки. Лінійна статична міжгалузева модель (МГБ).
16. Прогнозування валютних ринків. Прогнозування цін на ринку акцій.

ЗАВДАННЯ ДЛЯ САМОКОНТРОЛЮ

1. За даними за 30 місяців деякого часового ряду x_t були отримані коефіцієнти автокореляції рівнів:

$$1) \begin{matrix} r_1 = 0,63 & r_2 = 0,38 & r_3 = 30,72 & r_4 = 0,97 & r_5 = 0,55 & r_6 = 0,40 \\ r_7 = 0,65 \end{matrix}$$

$$2) \begin{matrix} r_1 = 0,62 & r_2 = 0,48 & r_3 = 30,62 & r_4 = 0,98 & r_5 = 0,65 & r_6 = 0,42 \\ r_7 = 0,75 \end{matrix}$$

$$3) \begin{matrix} r_1 = 0,73 & r_2 = 0,39 & r_3 = 30,71 & r_4 = 0,77 & r_5 = 0,38 & r_6 = 0,50 \\ r_7 = 0,71 \end{matrix}$$

$$4) \begin{matrix} r_1 = 0,53 & r_2 = 0,31 & r_3 = 31,22 & r_4 = 0,87 & r_5 = 0,54 & r_6 = 0,40 \\ r_7 = 0,55 \end{matrix}$$

$$5) \begin{matrix} r_1 = 0,73 & r_2 = 0,39 & r_3 = 31,12 & r_4 = 0,87 & r_5 = 0,51 & r_6 = 0,42 \\ r_7 = 0,75 \end{matrix}$$

$$6) \begin{matrix} r_1 = 0,68 & r_2 = 0,28 & r_3 = 30,52 & r_4 = 0,91 & r_5 = 0,45 & r_6 = 0,48 \\ r_7 = 0,55 \end{matrix}$$

$$7) \begin{matrix} r_1 = 0,58 & r_2 = 0,29 & r_3 = 31,01 & r_4 = 0,90 & r_5 = 0,48 & r_6 = 0,51 \\ r_7 = 0,59 \end{matrix}$$

$$8) \begin{matrix} r_1 = 0,61 & r_2 = 0,28 & r_3 = 30,62 & r_4 = 0,92 & r_5 = 0,45 & r_6 = 0,68 \\ r_7 = 0,73 \end{matrix}$$

$$9) \begin{matrix} r_1 = 0,59 & r_2 = 0,41 & r_3 = 29,99 & r_4 = 0,83 & r_5 = 0,49 & r_6 = 0,31 \\ r_7 = 0,81 \end{matrix}$$

$$10) \begin{matrix} r_1 = 0,72 & r_2 = 0,29 & r_3 = 32,01 & r_4 = 0,77 & r_5 = 0,51 & r_6 = 0,44 \\ r_7 = 0,67 \end{matrix}$$

де r_i – коефіцієнти автокореляції i -го порядку.

Побудуйте корелограму та охарактеризуйте структуру ряду. Визначте, яка можлива функціональна форма залежності для побудови рівняння регресії.

2. У таблиці наведено дані про розміри запасів компанії «Альфа» на початок кожного місяця. Дослідіть ряд на автокореляцію (до лагу $t = 8$). Побудуйте корелограму. Зробіть висновки щодо наявності тенденції та періодичних коливань.

Вихідні дані до задачі 2

1.

№ з/п	Спостереження	№ з/п	Спостереження
1.	231	21.	255
2.	322	22.	285
3.	118	23.	250
4.	351	24.	300
5.	191	25.	225
6.	320	26.	285
7.	275	27.	250
8.	205	28.	225
9.	295	29.	125
10.	240	30.	295
11.	355	31.	250
12.	175	32.	355
13.	285	33.	280
14.	200	34.	370
15.	290	35.	250
16.	220	36.	290
17.	400	37.	225
18.	275	38.	270
19.	185	39.	180
20.	370	40.	270

2.

№ з/п	Спостереження	№ з/п	Спостереження
1.	234	21.	256
2.	321	22.	286
3.	116	23.	252
4.	365	24.	301
5.	191	25.	226
6.	320	26.	286
7.	275	27.	251
8.	205	28.	226
9.	292	29.	126
10.	240	30.	296
11.	356	31.	251
12.	177	32.	356
13.	285	33.	281
14.	201	34.	371
15.	291	35.	251
16.	221	36.	291
17.	401	37.	226
18.	276	38.	271
19.	186	39.	181
20.	371	40.	271

3.

№ з/п	Спостереження	№ з/п	Спостереження
1.	240	21.	256
2.	325	22.	286
3.	122	23.	255
4.	361	24.	302
5.	192	25.	227

6.	327	26.	287
7.	280	27.	255
8.	211	28.	228
9.	300	29.	129
10.	245	30.	290
11.	361	31.	255
12.	180	32.	360
13.	290	33.	285
14.	205	34.	375
15.	295	35.	255
16.	225	36.	295
17.	405	37.	230
18.	280	38.	275
19.	190	39.	185
20.	372	40.	275

4.

№ з/п	Спостереження	№ з/п	Спостереження
1.	241	21.	256
2.	327	22.	288
3.	115	23.	252
4.	355	24.	301
5.	190	25.	226
6.	328	26.	290
7.	275	27.	255
8.	205	28.	227
9.	299	29.	128
10.	246	30.	295
11.	360	31.	256
12.	179	32.	358
13.	285	33.	283
14.	208	34.	370
15.	291	35.	255
16.	225	36.	292
17.	401	37.	225
18.	275	38.	278
19.	190	39.	180
20.	371	40.	271

5.

№ з/п	Спостереження	№ з/п	Спостереження
1.	244	21.	257
2.	324	22.	277
3.	119	23.	255
4.	355	24.	303
5.	191	25.	258
6.	326	26.	289
7.	275	27.	260
8.	209	28.	245
9.	292	29.	125
10.	245	30.	285
11.	358	31.	251
12.	179	32.	365
13.	285	33.	282
14.	206	34.	371
15.	291	35.	253
16.	222	36.	291
17.	405	37.	255
18.	277	38.	273
19.	189	39.	181
20.	372	40.	271

6.

№ з/п	Спостереження	№ з/п	Спостереження
1.	245	21.	257
2.	322	22.	295
3.	117	23.	250
4.	345	24.	310
5.	192	25.	235
6.	310	26.	295
7.	276	27.	260
8.	215	28.	235
9.	296	29.	165
10.	243	30.	285

11.	357	31.	250
12.	185	32.	365
13.	281	33.	290
14.	203	34.	380
15.	292	35.	260
16.	222	36.	291
17.	403	37.	235
18.	277	38.	272
19.	186	39.	181
20.	371	40.	270

7.

№ з/п	Спостереження	№ з/п	Спостереження
1.	245	21.	265
2.	322	22.	286
3.	125	23.	254
4.	355	24.	310
5.	190	25.	226
6.	326	26.	285
7.	276	27.	270
8.	206	28.	225
9.	285	29.	135
10.	242	30.	295
11.	365	31.	260
12.	176	32.	365
13.	295	33.	281
14.	202	34.	371
15.	291	35.	260
16.	230	36.	291
17.	405	37.	235
18.	285	38.	271
19.	186	39.	180
20.	371	40.	271

8.

№ з/п	Спостереження	№ з/п	Спостереження
1.	233	21.	254
2.	321	22.	285
3.	125	23.	254
4.	353	24.	300
5.	196	25.	225
6.	320	26.	288
7.	275	27.	258
8.	206	28.	225
9.	295	29.	125
10.	250	30.	291
11.	356	31.	250
12.	175	32.	357
13.	285	33.	280
14.	211	34.	372
15.	291	35.	251
16.	220	36.	290
17.	409	37.	225
18.	275	38.	274
19.	185	39.	181
20.	370	40.	270

9.

№ з/п	Спостереження	№ з/п	Спостереження
1.	215	21.	265
2.	320	22.	295
3.	119	23.	250
4.	355	24.	301
5.	190	25.	225
6.	327	26.	285
7.	275	27.	251
8.	206	28.	225
9.	295	29.	126
10.	240	30.	295
11.	356	31.	251
12.	185	32.	355
13.	284	33.	280
14.	200	34.	376
15.	291	35.	250

16.	220	36.	291
17.	401	37.	225
18.	275	38.	272
19.	186	39.	188
20.	370	40.	260

10.

№ з/п	Спостереження	№ з/п	Спостереження
1.	215	21.	265
2.	320	22.	284
3.	115	23.	250
4.	355	24.	309
5.	199	25.	215
6.	320	26.	285
7.	285	27.	251
8.	205	28.	225
9.	295	29.	135
10.	243	30.	294
11.	355	31.	259
12.	175	32.	355
13.	288	33.	283
14.	200	34.	373
15.	280	35.	251
16.	220	36.	296
17.	408	37.	225
18.	275	38.	275
19.	184	39.	184
20.	371	40.	271

3. Побудуйте адитивну модель на підставі наведених квартальних даних про споживання безалкогольних напоїв у 2013-2016 роках, у млн дкл.

Вихідні дані до задачі 3

1.

Рік	Квартал	Млн дкл, y_t
2003	1	24,4
	2	52,6
	3	60,4
	4	34,0
2004	1	32,7
	2	56,2
	3	67,3
	4	36,2
2005	1	37,8
	2	65,3
	3	73,1
	4	38,4
Разом		578,4

2.

Рік	Квартал	Млн дкл, y_t
2003	1	24,6
	2	52,7
	3	61,4
	4	34,8
2004	1	32,9
	2	56,4
	3	67,5
	4	36,7
2005	1	37,8
	2	66,3
	3	73,2
	4	38,5

Разом	582,8
-------	-------

3.

Рік	Квартал	Млн дкл, y_t
2003	1	24,7
	2	51,6
	3	61,4
	4	35,0
2004	1	32,9
	2	56,1
	3	68,3
	4	36,6
2005	1	34,8
	2	65,8
	3	74,1
	4	38,8
Разом		494

4.

Рік	Квартал	Млн дкл, y_t
2003	1	21,4
	2	55,6
	3	69,4
	4	34,0
2004	1	32,9
	2	52,2
	3	67,8
	4	31,2
2005	1	37,8
	2	68,3
	3	73,1
	4	38,9
Разом		582,6

5.

Рік	Квартал	Млн дкл, y_t
2003	1	25,0
	2	52,6
	3	66,4
	4	37,0
2004	1	35,7
	2	56,9
	3	61,3
	4	46,2
2005	1	47,8
	2	75,3
	3	63,1
	4	48,4
Разом		615,7

6.

Рік	Квартал	Млн дкл, y_t
2003	1	44,4
	2	51,6
	3	60,5
	4	34,7
2004	1	32,9
	2	56,8
	3	67,3
	4	46,2
2005	1	37,9

	2	66,3
	3	73,3
	4	48,4
Разом		620,3

7.

Рік	Квартал	Млн дкл, у _t
2003	1	54,4
	2	52,9
	3	62,4
	4	54,0
2004	1	32,9
	2	66,2
	3	67,9
	4	31,2
2005	1	38,8
	2	65,7
	3	73,9
	4	28,4
Разом		628,7

8.

Рік	Квартал	Млн дкл, у _t
2003	1	54,4
	2	52,8
	3	63,4
	4	34,9
2004	1	42,7
	2	57,2
	3	67,9
	4	32,2
2005	1	38,8
	2	65,8
	3	71,1
	4	39,4
Разом		620,6

9.

Рік	Квартал	Млн дкл, у _t
2003	1	54,4
	2	82,6
	3	61,5
	4	35,1
2004	1	42,7
	2	46,2
	3	67,9
	4	46,2
2005	1	47,8
	2	61,3
	3	71,1
	4	28,4
Разом		645,2

10.

Рік	Квартал	Млн дкл, y_t
2003	1	34,4
	2	52,6
	3	66,4
	4	54,0
2004	1	32,7
	2	56,5
	3	67,5
	4	36,2
2005	1	47,8
	2	75,3
	3	73,1
	4	38,9
Разом		635,4

4. У таблиці наведено динаміку курсів валют на Форексі *євро/долар* за 28 торгових днів: Зобразіть часовий ряд графічно. Визначте, чи містить ряд детерміністичний тренд. У разі його наявності звести ряд із нестационарного в стационарний. Перевірити отриманий ряд на стаціонарність за допомогою критерію, заснованого на медіані.

Вихідні дані до задачі 4

1.

№ з/п	Курс	№ з/п	Курс
1	1,2681	15	1,3453
2	1,2721	16	1,3579
3	1,2693	17	1,3623
4	1,2720	18	1,3629
5	1,2877	19	1,3801
6	1,2992	20	1,3692
7	1,3006	21	1,3834
8	1,3075	22	1,3933
9	1,3058	23	1,3925
10	1,3064	24	1,3994
11	1,3261	25	1,3873
12	1,3398	26	1,3919
13	1,3312	27	1,3960
14	1,3486	28	1,4078

2.

№ з/п	Курс	№ з/п	Курс
1	1,2682	15	1,3753
2	1,2724	16	1,3579
3	1,2683	17	1,3623
4	1,2720	18	1,3829
5	1,2878	19	1,3801
6	1,2982	20	1,3892
7	1,3006	21	1,3834
8	1,3075	22	1,3938
9	1,3058	23	1,3825
10	1,3068	24	1,3994
11	1,3261	25	1,3973
12	1,3398	26	1,3919
13	1,3312	27	1,3960
14	1,3486	28	1,4078

3.

№ з/п	Курс	№ з/п	Курс
1	1,3681	15	1,3453
2	1,2721	16	1,3577
3	1,2593	17	1,3623
4	1,2720	18	1,3639
5	1,2877	19	1,3801
6	1,2996	20	1,3672
7	1,3006	21	1,3834
8	1,3175	22	1,3933
9	1,3058	23	1,3825
10	1,3074	24	1,3994
11	1,3261	25	1,3877
12	1,3395	26	1,3910
13	1,2312	27	1,3960
14	1,3488	28	1,4078

4.

№ з/п	Курс	№ з/п	Курс
1	1,2781	15	1,3457
2	1,2721	16	1,3679
3	1,2696	17	1,3623
4	1,2730	18	1,3620
5	1,2877	19	1,3801
6	1,2992	20	1,3792
7	1,3007	21	1,3834
8	1,3075	22	1,3937
9	1,3068	23	1,3925
10	1,3064	24	1,3094
11	1,3261	25	1,3873
12	1,3388	26	1,3910
13	1,3312	27	1,3960
14	1,3486	28	1,4078

5.

№ з/п	Курс	№ з/п	Курс
1	1,2611	15	1,3553
2	1,2621	16	1,3579
3	1,2698	17	1,3623
4	1,2720	18	1,3626
5	1,2879	19	1,3821
6	1,2992	20	1,3692
7	1,3016	21	1,3834
8	1,3075	22	1,3953
9	1,3068	23	1,3925
10	1,3064	24	1,3894
11	1,3268	25	1,3873
12	1,3398	26	1,3910
13	1,3412	27	1,3960
14	1,3486	28	1,4178

6.

№ з/п	Курс	№ з/п	Курс
1	1,2581	15	1,3753

2	1,2721	16	1,3979
3	1,2698	17	1,3623
4	1,2720	18	1,3621
5	1,2377	19	1,3801
6	1,2999	20	1,3092
7	1,3006	21	1,3834
8	1,3065	22	1,3133
9	1,3058	23	1,3925
10	1,3084	24	1,3904
11	1,3261	25	1,3873
12	1,3390	26	1,3919
13	1,3312	27	1,3860
14	1,3476	28	1,4070

7.

№ з/п	Курс	№ з/п	Курс
1	1,3681	15	1,3053
2	1,3721	16	1,3570
3	1,2993	17	1,3223
4	1,2720	18	1,3629
5	1,2777	19	1,3841
6	1,2992	20	1,3692
7	1,3106	21	1,3854
8	1,3075	22	1,3033
9	1,3054	23	1,3925
10	1,3064	24	1,3904
11	1,3061	25	1,3873
12	1,3390	26	1,3419
13	1,3312	27	1,3960
14	1,3586	28	1,4071

8.

№ з/п	Курс	№ з/п	Курс
1	1,2688	15	1,3653
2	1,2791	16	1,3979
3	1,2693	17	1,3623
4	1,2729	18	1,3529
5	1,2867	19	1,3809
6	1,2997	20	1,3692
7	1,3006	21	1,3534
8	1,3095	22	1,3933
9	1,3058	23	1,3924
10	1,3084	24	1,3984
11	1,3261	25	1,3873
12	1,3394	26	1,3019
13	1,3312	27	1,3960
14	1,3488	28	1,4077

9.

№ з/п	Курс	№ з/п	Курс
1	1,2781	15	1,3453
2	1,2729	16	1,3509
3	1,2693	17	1,3623
4	1,2120	18	1,3029
5	1,2877	19	1,3801
6	1,2092	20	1,3690
7	1,3006	21	1,3834
8	1,3005	22	1,3233
9	1,3058	23	1,3925
10	1,3004	24	1,3904
11	1,3261	25	1,3873
12	1,3398	26	1,3919
13	1,3012	27	1,3060
14	1,3486	28	1,4171

10.

№ з/п	Курс	№ з/п	Курс
1	1,2981	15	1,3653
2	1,2721	16	1,3579
3	1,2690	17	1,3693
4	1,2720	18	1,3629

5	1,2807	19	1,3811
6	1,2999	20	1,3692
7	1,3106	21	1,3634
8	1,3075	22	1,3939
9	1,3059	23	1,3925
10	1,3164	24	1,3904
11	1,3261	25	1,3879
12	1,3399	26	1,3919
13	1,3322	27	1,3066
14	1,3480	28	1,4015

ТЕМАТИКА ЛАБОРАТОРНИХ ЗАНЯТЬ

На лабораторні заняття виносяться не всі, а найважливіші та найскладніші питання дисципліни. Решту питань студенти опрацьовують самостійно. При підготовці до семінарських і практичних занять потрібно передусім прочитати конспект лекцій з відповідних тем, вивчити понятійний апарат і осягнути логіку навчального матеріалу. Потім доцільно скористатися методичними порадами до відповідної теми. Нарешті, якщо деякі питання залишилися незрозумілими, звернутися до літературних джерел, рекомендованих у цьому розділі.

ЛАБОРАТОРНЕ ЗАНЯТТЯ 1. Попередній аналіз рядів динаміки та вибір методу прогнозування.

1. Розрахунок характеристик динаміки економічних процесів.
2. Декомпозиція часового ряду.
3. Ідентифікація моделі часового ряду.

Література: 1, 2, 5, 7.

Контрольні питання:

1. Форма надання статистичної інформації.
2. Багатовимірні часові ряди. Однорідність, стійкість і достатня сукупність спостережень часових рядів.
3. Показники динаміки часового ряду.
4. Статистичні характеристики часових рядів.

ЛАБОРАТОРНЕ ЗАНЯТТЯ 2. Прогнозування на основі одновимірних часових рядів.

1. Основні поняття і види часових рядів.
2. Складові часового ряду.
3. Основні показники часового ряду.
4. Метод прогнозування – екстраполяція
5. Методи аналізу часових рядів.
6. Метод аналізу часових рядів - ковзного середнього.
7. Метод аналізу часових рядів - зваженого ковзного середнього.

Література: 1,2,7,8.

Контрольні питання:

1. Систематичні компоненти часового ряду. Випадкові компоненти часового ряду.

2. Аналіз динаміки часового ряду. Стаціонарні та нестаціонарні часові ряди.
3. Авторегресійні процеси. Коефіцієнти автокореляції.
4. Механічні методи згладжування часових рядів (згладжування по двох точках; метод простої ковзкої середньої; метод зваженої ковзкої середньої; метод простого експоненційного згладжування).

ЛАБОРАТОРНЕ ЗАНЯТТЯ 3. Аналітичні методи згладжування часових рядів

1. Метод прогнозування – екстраполяція
2. Метод екстраполяції на основі кривих зростання
3. Використання множинної регресії для одночасної оцінки тренду та сезонного чинника.
4. Адаптивні методи прогнозування.
5. Метод адаптивного згладжування Брауна;
6. Метод Хольта—Уінтерса; метод Тейла—Вейджа.

Література: 1,2,7,8.

Контрольні питання:

1. Метод Хольта
2. Метод адаптивного згладжування Брауна;
3. Метод Хольта—Уінтерса;
4. Метод Тейла—Вейджа; метод Харрісона.

ЛАБОРАТОРНЕ ЗАНЯТТЯ 4. Економетричні методи прогнозування

1. Кореляційний аналіз. Перевірка значущості коефіцієнтів кореляції. Довірчі інтервали прогнозу.
2. Регресійний аналіз. Визначення функції регресії. Оцінка невідомих значень залежної змінної.
3. Аналіз залишків.
4. Прогнозування на основі регресійної моделі. Незміщена оцінка точкового прогнозу. Лінеаризація моделі. Порухення допущень регресійного аналізу (мультиколінеарність, гетероскедастичність, автокореляція).

Література: 1, 2, 3, 7, 8, 12, 13, 15.

Контрольні питання:

1. Розрахунок стандартної помилки оцінки залишків σ_ε
2. Аналіз дисперсії
3. Перевірка значущості коефіцієнта кореляції між Y та ε .
4. Алгоритм методу найменших квадратів.

ЛАБОРАТОРНЕ ЗАНЯТТЯ 5. Прогнозування на основі тренд-сезонних моделей.

1. Типи трендових моделей.
2. Способи вибору кривої динаміки часового ряду.
3. Методи прикладної статистики для вибору кривої динаміки часового ряду

3. Методи вибору кривих зростання.
4. Адиитвні тренд-сезонні моделі.
5. Мультиплікативні тренд-сезонні моделі.
6. Моделювання на основі рядів Фур'є.

Література: 2,3,6,7,8,12.

Контрольні питання:

1. Методи фільтрації сезонної компоненти s_t .
2. Проблема аналізу сезонності (та/або циклічності).
3. Ітераційні методи фільтрації. Розрахунок сезонної хвилі.
4. Авторегресійний процес (AR). Інтегрування (I). Процес ковзної середньої(MA). ARIMA-моделі

ЛАБОРАТОРНЕ ЗАНЯТТЯ 6. Прогнозування основних макроекономічних показників

1. Проста макроекономічна модель.
2. Моделі зростання.
3. Виробничі функції.
4. Комплексні макроекономічні моделі.
5. Моделі світової економіки.
6. Гіпотеза ефективного ринку.
7. Прогнозування валютних ринків.
8. Прогнозування цін на ринку акцій.

Література: 2,3,6,7,8,12.

Контрольні питання:

1. Загальна характеристика комплексних макроеконометричних моделей.
2. Види макроеконометричних моделей.
3. Економічна сутність комплексних економетричних моделей.
4. Процес побудови комплексної економетричної моделі.
5. Прості макроеконометричні моделі прогнозування.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

ОСНОВНА

1. Закон України «Про державне прогнозування та розроблення програм економічного і соціального розвитку України». — К., 2000.
2. Присенко Г.В., Равікович Є.І. Прогнозування соціально-економічних процесів. — К.: КНЕУ, 2005.
3. Саркісян С. А., Голованов Л. В. Прогнозирование развития больших систем. — М.: Статистика, 1975.
4. Красс И. А. Математические модели экономической динамики. — М.: Советское радио, 1976.
5. Класс А. и др. Введение в эконометрическое моделирование. — М.: Финансы и статистика, 1978.

6. *Винн Р., Холден К.* Введение в прикладной эконометрический анализ. — М.: Финансы и статистика, 1981.
7. Рабочая книга по прогнозированию. — М.: Мысль, 1982.
8. *Горелова В. Л., Мельникова Е. Н.* Основы прогнозирования систем. — М.: Высшая школа, 1986.
9. *Льюис К. Д.* Методы прогнозирования экономических показателей. — М.: Финансы и статистика, 1986.
10. Статистическое моделирование и прогнозирование / Под ред. А. Г. Гранберга. — М.: Финансы и статистика, 1990.
11. *Гальперин В. М.* и др. Макроэкономика. — СПб.: Экономическая школа, 1994.
12. *Уотшем Т. Дж., Паррамоу К.* Количественные методы в финансах. — М.: Финансы, 1999.
13. *Науменко В., Панасюк Б.* Впровадження методів прогнозування і планування в умовах ринкової економіки. — К.: Глобус, 1995.
14. *Наконечний С. І., Терещенко Т. О., Романюк Т. П.* Економетрія: Навч. посібник. — К.: КНЕУ, 1997.
15. *Холден К., Піл Д. А., Томпсон Дж. Л.* Економічне прогнозування: Вступ. — К.: Інформтехніка, ЕМЦ, 1996.
16. *Белецький В. М., Бакуменко В. Д.* Моделювання макроекономічних процесів. — К.: УАДУ при Президентіві України, 1998.

ДОДАТКОВА

17. *Лук'яненко І., Краснікова Л.* Економетрика: Практикум з використанням комп'ютера. — К.: Знання, КОО, 1998.
18. *Костіна Н. І., Алексеев А. А., Василик О. Д.* Фінанси: система моделей і прогнозів: Навч. посібник. — К.: Четверта хвиля, 1998.
19. *Айвазян С. А., Мхитарян В. С.* Прикладная статистика и основы эконометрики: Учебник для вузов. — М.: ЮНИТИ, 1998.
20. Секторальні моделі прогнозування економіки України / За ред. В. М. Гейця. — К.: Фенікс, 1999.
21. *Боровиков В. П., Ивченко Г. И.* Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере: Учеб. пособие. — М.: Финансы и статистика, 1999.
22. Экономико-математические методы и прикладные модели: Учеб. пособие для вузов / Под ред. В. В. Федосеева. — М.: ЮНИТИ, 2000.
23. *Дубров А. М.* Многомерные статистические методы: Учебник. — М.: Финансы и статистика, 2000.
24. *Шелобаев С. И.* Математические методы и модели в экономике, финансах, бизнесе: Учебное пособие для вузов. — М.: ЮНИТИ-ДАНА, 2000.
25. *Кобелев Н. Б.* Практика применения экономико-математических методов и моделей: Учебно-практическое пособие. — М.: ЗАО Финстатинформ, 2000.

26. *Петерс Е.* Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка: Пер. с англ. — М.: Мир, 2000.
27. *Лук'яненко І. Г., Городніченко Ю. О.* Сучасні економетричні методи у фінансах. — К.: Літера ЛТД, 2002.
28. *Черняк О. І., Ставицький А. В.* Динамічна економетрика: Навч. посібник. — К., 2000.
29. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. — М.: Мир, 1974.
30. *Горчаков А. А., Орлова И. В.* Компьютерные экономико-математические модели. — М.: ЮНИТИ, 1995.
31. *Парсаданов Г. А.* Планирование и прогнозирование социально-экономической системы: Учебное пособие для вузов. — М.: ЮНИТИ-ДАНА, 2001.
32. *Колемаев В. А.* Математическая экономика. — М.: ЮТИТИ, 2002.
33. Прогнозування і розробка програм. (Методичні рекомендації) / За ред. В. Ф.Беседіна. — К.: Науковий світ, 2000.