

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

АНАЛІЗ НЕСТАЦІОНАРНИХ І НЕЛІНІЙНИХ ПРОЦЕСІВ В ЕКОНОМІЦІ ТА ФІНАНСАХ

Навчальний посібник

Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського
як навчальний посібник для здобувачів ступеня бакалавра
за освітньою програмою «Системний аналіз і управління»,
спеціальності F4 Системний аналіз та наука про дані

Електронне мережеве навчальне видання

Київ
КПІ ім. ІГОРЯ СІКОРСЬКОГО
2025

УДК 519.2+519.8(330.43)

Укладачі: *Бідюк Петро Іванович*, д-р техн. наук, проф.
Тимошук Оксана Леонідівна, канд. техн. наук, доц.
Гуськова Віра Геннадіївна, д-р філос.
Просянкін-Жарова Тетяна Іванівна, канд. техн. наук, доц.
Левенчук Людмила Борисівна, д.ф.

Рецензент *Данилов В.Я.*, д-р техн. наук, проф.
Національний технічний університет України «КПІ імені Ігоря Сікорського»

Відповідальний редактор *Дмитрієва О.А.*, д-р техн. наук, проф.

*Гриф надано Методичною радою КПІ ім. Ігоря Сікорського
(протокол № 1 від 02.10.2025 р.)
за поданням вченої ради навчально-наукового інституту
(протокол № 8 від 29.09.2025 р.)*

Аналіз нестационарних і нелінійних процесів в економіці та фінансах [Електронний ресурс] : навч. посіб. для здобувачів ступеня бакалавра за спец. F4 Системний аналіз та наука про дані / уклад.: П. І. Бідюк та ін. – Електрон. текст. дані (1 файл). – Київ : КПІ ім. Ігоря Сікорського, 2025. – 385 с.

У посібнику викладено методи математичного моделювання та прогнозування нестационарних і нелінійних процесів в економіці, фінансах та екології. Видання містить опис методики моделювання, тестів на нелінійність і нестационарність процесів. Особлива увага приділена аналізу процесів із трендами, моделюванню коінтегрованих і гетероскедастичних процесів, побудові моделей у просторі станів.

У посібнику наведено сучасні методи прогнозування, зокрема на основі байєсових мереж, та запропоновано підходи до оцінювання параметрів із використанням рекурсивного методу найменших квадратів і Монте-Карло.

Посібник призначений для здобувачів ступеня бакалавра і магістра за спеціальністю «Фінанси, банківська справа та страхування», а також буде корисним для аспірантів, викладачів і практиків у сфері моделювання економічних і фінансових процесів.

УДК 519.2+519.8(330.43)

Реєстр. № НП 25/26-015. Обсяг 17,5 авт. арк.
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
проспект Берестейський, 37, м. Київ, 03056
<https://kpi.ua>
Свідцтво про внесення до Державного реєстру видавців, виготовлювачів
і розповсюджувачів видавничої продукції ДК № 5354 від 25.05.2017 р.

П. І. Бідюк, О. Л. Тимошук, В. Г. Гуськова, Т.І. Просянкін-Жарова, Л. Б. Левенчук, 2025

♥ КПІ ім. Ігоря Сікорського, 2025

ВСТУП

Метою посібника є: 1) аналіз та систематизація деяких існуючих методів моделювання і прогнозування стаціонарних і нестаціонарних нелінійних процесів, поданих часовими рядами і часовими перерізами даних; 2) подання методики моделювання і прогнозування лінійних та нелінійних нестаціонарних процесів; 3) вироблення рекомендацій щодо комплексного застосування методів прогнозування. Розглянуто спеціальні структури моделей гетероскедастичних та коінтегрованих процесів, які забезпечують отримання високоякісних оцінок прогнозів нестаціонарних процесів зі змінною у часі дисперсією і трендами різних порядків.

Структура книги ґрунтується на системному підході до аналізу фінансових, економічних та деяких інших процесів, який передбачає врахування в моделях можливих невизначеностей при створенні формалізованого опису процесів, зокрема це невизначеності структурного і статистичного характеру. Структурні невизначеності зумовлені значною нестаціонарністю багатьох реальних процесів, яка призводить до швидкої зміни характеристик структури процесу (наприклад, порядку моделі, часу затримки по входу, типу розподілу даних і параметрів випадкового збурення), що потребує адаптації формалізованого опису структури моделі до фактичних характеристик процесу.

Невизначеності статистичного характеру зумовлені недостатньою інформацією стосовно випадкових збурень (збурення в тій чи іншій мірі існують завжди), наявністю похибок вимірів даних та оцінювання структури і параметрів моделі, можливою наявністю екстремальних значень. Як правило, при формалізованому описанні таких невизначеностей слід застосувати розподіли ймовірностей різних типів, хоча здебільшого збурення досить обґрунтовано вважається нормальним процесом, оскільки випадкове збурення – це, як правило, комбінований вплив кількох випадкових процесів. Необхідно зазначити, що на сьогодні існує достатньо розвинений інструментарій для врахування і зменшення впливу невизначеностей різної природи на остаточні результати моделювання, прогнозування і керування. Основна умова – ці інструменти потрібно коректно вибирати і застосовувати.

Також відомо, що у більшості випадків моделювання та прогнозування ми будуємо оптимальні за параметрами математичні моделі та оцінюємо характеристики моделей і прогнозів за множиною деяких статистичних критеріїв. При цьому оптимізація стосується не тільки безпосередньо

параметрів (коефіцієнтів) моделі, але й статистичних параметрів якості моделі. Тобто, у загальному випадку, фактично виникає необхідність багатокритеріальної оптимізації всього процесу побудови моделі та оцінювання прогнозу. Ця задача виходить за межі змісту цього посібника, але про неї необхідно пам'ятати і, за можливості, розв'язувати.

Основне завдання посібника – надати студентам, аспірантам і зацікавленим фахівцям відповідного профілю інформацію стосовно деяких поширених методів моделювання і прогнозування нестационарних та нелінійних процесів в економіці і фінансах на основі статистичних даних. Незважаючи на стрімкий розвиток і широке практичне застосування методів штучного інтелекту для аналізу даних, відомі методи аналізу часових рядів та їх модифікації продовжують займати чільне місце в переліку інструментів аналізу розвитку процесів різної природи. Існуючі методи аналізу статистичних даних дають можливість дослідникам досягти необхідного поглибленого розуміння характеристик досліджуваних процесів, зрозуміти тонкощі розвитку процесів, та отримати бажані результати моделювання і прогнозування, виявити переваги і недоліки обраного для використання інструментарію.

Виникнення нестационарності і нелінійності фінансово-економічних процесів зумовлено багатьма факторами, наприклад, перехідними процесами в економіці (приклад: економіки країн, що утворились в результаті розпаду СРСР, економіки країн латинської Америки), нестабільністю законодавства, недостатньо високою якістю управління на різних рівнях ієрархії, швидким розвитком технологій, високою конкуренцією на внутрішньому та міжнародному ринках, військовими конфліктами і т. п. Як правило, економетричні і ймовірно-статистичні моделі дають можливість враховувати більшість можливих ефектів впливу і досягти високого ступеня адекватності моделей, а також бажаної якості прогнозів та керування. Це моделі процесів з детермінованими і випадковими трендами, моделі коінтегрованих та гетероскедастичних процесів, множина моделей, що дають можливість враховувати нелінійні ефекти, наприклад, білінійні моделі, поліноми різних типів, узагальнені лінійні моделі, радіальні базисні функції, байєсівські мережі і т. ін. Перевагою таких моделей є відносно невисокі обчислювальні витрати при їх побудові, зручний і прискорений аналіз адекватності, можливості структурно-параметричної адаптації та комбінування кількох моделей у більш складні структури з метою опису нелінійної поведінки вибраних процесів.

В процесі побудови моделей на основі статистичних даних важливо коректно вибирати і застосовувати методи оцінювання параметрів.

Застосування методу найменших квадратів вимагає нормального розподілу даних; оцінювання параметрів моделей, нелінійних стосовно змінних і параметрів виконується за допомогою різних методів. Аналіз процесів із нестационарною дисперсією виконується з використанням лінійних та нелінійних моделей, що потребує особливої уваги до коректного застосування методів оцінювання параметрів. Сучасна теорія оцінювання надає дослідникам широкий спектр методів оцінювання параметрів і змінних, але для досягнення високоякісних результатів моделювання цими методами потрібно правильно користуватись. Деякі методи і рекомендації подані в цьому посібнику.

Ймовірно-статистичні методи моделювання нестационарних та нелінійних процесів, розглянуті в посібнику, будуть корисними для студентів, аспірантів, викладачів відповідних курсів, а також для інженерів-практиків і будуть сприяти поглибленому аналізу поведінки сучасних фінансово-економічних процесів та використанню отриманого досвіду при прийнятті рішень.

ЗМІСТ

Вступ	3
Розділ 1. Мета побудови і деякі типи математичних моделей лінійних та нелінійних процесів	9
1.1. Мета побудови математичних моделей	9
1.2. Лінійні та нелінійні процеси і моделі	12
1.3. Поняття структури нелінійної математичної моделі	19
1.4. Два основних методи побудови математичних моделей	23
1.5. Узагальнений алгоритм побудови моделі	26
1.6. Вимоги до експериментальних даних, оцінок параметрів та	30
1.6.1. Вимоги до експериментальних даних	31
1.6.2. Вимоги до оцінок параметрів моделі	33
1.6.3. Вимоги до математичної моделі	35
1.7. Спрощена класифікація математичних моделей	38
1.8. Деякі типи регресійних та різницевих рівнянь	41
1.9. Запитання і вправи	46
Розділ 2. Методика побудови моделей нелінійних та	51
2.1. Побудова регресійних моделей	52
2.2. Методи оцінювання параметрів	59
2.3. Нестационарність і нелінійність у даних	65
2.4. Методи попередньої обробки даних	70
2.5. Запитання і вправи	75
Розділ 3. Аналіз стаціонарності процесів авторегресії з ковзним	81
3.1. Поняття стаціонарності	82
3.2. Умови стаціонарності для AR і MA процесів	87
3.3. Стаціонарність ARMA-процесів	93
3.4. Методи перевірки стаціонарності	98
3.5. Запитання і вправи	104

Розділ 4. Тести на нелінійність і нестационарність процесів	111
4.1. Тести на одиничний корінь	112
4.2. Тести на структурні злами	118
4.3. Тести на нелінійність часових рядів	124
4.4. Запитання і вправи	129
Розділ 5. Представлення дискретних моделей у просторі станів	135
5.1. Матрична форма представлення моделей	136
5.2. Канонічні форми представлення	141
5.3. Системи з випадковими збуреннями	147
5.4. Запитання і вправи	153
Розділ 6. Детерміновані і стохастичні тренди	159
6.1. Поняття тренду	160
6.2. Методи оцінювання трендів	165
6.3. Стационарність після видалення тренду	171
6.4. Запитання і вправи	176
Розділ 7. Аналіз коінтегрованих процесів	183
7.1. Поняття коінтеграції	184
7.2. Тест Енгла–Грейнджера	189
7.3. Тест Йогансена	195
7.4. Моделі корегування похибок (ECM)	201
7.5. Запитання і вправи	206
Розділ 8. Аналіз гетероскедастичних процесів	213
8.1. ARCH-моделі	214
8.2. GARCH-моделі	219
8.3. Узагальнені нелінійні гетероскедастичні моделі	225
8.4. Запитання і вправи	230
Розділ 9. Порогова модель стохастичної волатильності	237
9.1. Порогові процеси	238
9.2. Моделі стохастичної волатильності	243
9.3. Приклади застосування	249

9.4. Запитання і вправи	254
Література	327

РОЗДІЛ 1. МЕТА ПОБУДОВИ І ДЕЯКІ ТИПИ МАТЕМАТИЧНИХ МОДЕЛЕЙ ЛІНІЙНИХ ТА НЕЛІНІЙНИХ ПРОЦЕСІВ

Розділ присвячено аналізу етапів побудови математичних моделей процесів різної природи, представлених експериментальними або статистичними даними. Розглядається мета побудови моделі та елементи структури моделі, які необхідно визначити в результаті аналізу процесу або об'єкта, що описується математичною моделлю. Подано аналіз структурного і функціонального методів побудови моделей досліджуваних процесів; перший представляє собою теоретичний метод на основі аналізу законів та закономірностей функціонування процесу, а другий орієнтований на збір та використання експериментальних/статистичних даних.

Розглянуто узагальнений алгоритм побудови моделі за експериментальними/статистичними даними, який дає уявлення про загальні етапи моделювання, які можна застосовувати до процесів різної природи. Також сформульовано вимоги до експериментальних даних, оцінок параметрів та моделі в цілому. Дотримання цих вимог дає можливість будувати моделі високого ступеня адекватності і скорочувати витрати часу на побудову та аналіз якості (адекватності) моделі.

Розділ завершується переліком деяких типів математичних моделей, що описують стаціонарні і нестаціонарні процеси (ННП), а також процеси з нелінійністю стосовно змінних і параметрів.

1. Мета побудови математичних моделей

Можна вказати на три цілі побудови математичних моделей досліджуваних процесів та об'єктів, що часто зустрічаються на практиці:

- *поглиблене вивчення процесів з використанням імітаційного моделювання;*
- *прогнозування майбутніх значень змінних та параметрів за допомогою створеної моделі;*
- *використання моделі для створення (синтезу) системи керування.*

Поглиблене вивчення процесів (об'єктів) за допомогою математичних моделей з використанням імітаційного моделювання надає можливість проаналізувати такі характеристики досліджуваних процесів [1, 2]:

- ідентифікувати та оцінити кількісно причинно-наслідкові зв'язки між вхідними та вихідними змінними;
- дослідити яким чином ведуть себе змінні на виході при варіюванні вхідних у допустимому діапазоні їх значень;
- розглянути поведінку процесів на будь-яких часових інтервалах у прийнятному масштабі часу;
- виявити приховані режими функціонування об'єкта та причини їх виникнення (прикладом можуть бути Чорнобильська катастрофа; деякі, недостатньо глибоко досліджені хімічні реакції; взаємодія об'єктів у космосі і під водою; складні типи взаємодії нелінійних процесів і т. ін.);
- суттєво знизити вартість дослідження фактичної поведінки процесів та об'єктів за рахунок того, що не будується фізична модель, яка може бути набагато дорожчою ніж математична; при цьому, як правило, існує можливість (поступового) підвищення точності отриманих результатів моделювання за рахунок використання можливих методів оптимізації та адаптації моделей і оновленого математичного інструментарію (інструментальної платформи) з підвищеними характеристиками якості;
- виявити можливі/необхідні витрати на побудову та реалізацію системи керування або моніторингу досліджуваного об'єкта; як приклад, можна згадати військову операцію США «Буря у пустелі» у 1992 – 1993 рр., спрямовану на звільнення Кувейту від Іраку; основний критерій, який було закладено у імітаційну модель, такий: мінімізація втрат живої сили при реалізації запланованої операції, тобто американських солдат.

Математична модель, що будується для цієї мети, може бути дуже складною і трудомісткою, оскільки вона повинна враховувати тонкощі взаємодії кількісних і якісних змінних із можливим врахуванням протікання подій у реальному часі; тобто в такому випадку часто використовують імітаційне моделювання стосовно генерування множини випадкових процесів із заданими характеристиками. За допомогою імітаційних моделей можна виявити ефекти і явища, які недоступні безпосереднім спостереженням за допомогою приладів. Крім того, при проектуванні нових систем в різних галузях можна швидко змінювати варіанти реалізації та модифікації системи завдяки можливості її прискореного і прийнятного за вартістю дослідження за імітаційною моделлю, виявити вплив початкових умов та обмежень на ключові змінні і параметри.

Прогнозування значень змінних виконується, як правило, на основі набагато простіших моделей ніж поглиблене вивчення процесів. Для цієї мети

дуже зручними є відносно прості дискретні моделі у вигляді авторегресійних рівнянь (АР) та авторегресії з ковзним середнім (АРКС) [1, 2]. Часто для прогнозування та керування використовують також зручне уніфіковане представлення у просторі станів неперервних чи дискретних моделей.

Тип моделі залежить також від того, який прогноз необхідний – короткостроковий, середньостроковий чи довгостроковий. При виборі моделі необхідно враховувати, також, вимоги до типу оцінки прогнозу, яка необхідна за конкретною постановкою задачі: точкова, інтервальна або у формі розподілу ймовірностей. Задача прогнозування може бути досить складною, якщо стохастичний процес *нелінійний* та *нестационарний*. Загалом, оцінки прогнозів необхідні для розв’язання багатьох задач, у тому числі таких:

- керування та моніторинг на різних рівнях (малий, середній, великий);
- автоматичне (та автоматизоване) керування об’єктами у різних галузях економіки, військовій справі, космічних дослідженнях, технічних системах і технологічних процесах і т. ін.; (неможливо спроектувати і реалізувати систему автоматичного керування без розв’язання задачі прогнозування – це пов’язано із наявністю часу запізнення реакції на виході об’єкта стосовно вхідних впливів);
- автоматичне (та автоматизоване) діагностування і прогнозування станів в економіці, фінансах, у медицині й техніці;
- прогнозування розвитку процесів в екології;
- прогнозування погоди та можливих кліматичних змін;
- прогнозування різноманітних процесів при розв’язуванні завдань особистого характеру.

Для розв’язання задачі *синтезу систем керування вибраними процесами* використовують різні класи моделей у неперервному та дискретному часі, які представляють, як правило, в уніфікованій формі простору станів. У такому випадку моделі не повинні мати дуже високу точність (адекватність), тому що вони використовуються у замкненому контурі керування із від’ємним зворотним зв’язком. Це сприяє асимптотичному зменшенню похибок керування до мінімально можливих значень навіть при використанні досить грубих моделей.

Комп’ютерне моделювання системи керування надає можливість за мінімальний час дослідити ефективність застосування різних варіантів попередньої обробки даних (їх підготовки до побудови моделей та генерування керуючих впливів), законів керування, комбінацій керуючих дій та визначити ефективність системи керування при зміні вхідних керуючих сигналів в

широкому діапазоні їх значень. Комп'ютерна система керування може функціонувати як в автоматичному режимі (при керуванні технічними або технологічними процесами в реальному часі), так і в режимі порадики при керуванні економічними, екологічними, соціальними чи іншими складними системами і процесами.

При використанні комп'ютерної системи як порадики, керуючі дії, згенеровані комп'ютером (наприклад, у межах створеної системи підтримки прийняття рішень), порівнюються з експертними оцінками керуючих входних впливів і на основі прийнятного критерію якості вибираються найбільш прийнятні варіанти керуючих впливів. При цьому бажано спрогнозувати, також, наслідки прийняття того чи іншого рішення з використанням відповідних аналітичних критеріїв якості.

1.2 Лінійні та нелінійні процеси і моделі

Поширеним типом даних є часові ряди, які описують розвиток у часі відповідних процесів, що відбуваються в економіці, фінансах, технічних системах і технологічних процесах, живих організмах, екології, природоохоронних системах і т. ін. Досліджувані процеси можуть розглядатись на коротких і довгих проміжках часу і при різних обмеженнях, залежно від постановки задачі дослідження. Відповідно, вибірки даних мають різну довжину (потужність), інформативність, містять різні ефекти, які є специфічними для конкретної галузі.

Деякі типи процесів в економіці та фінансах, що описуються лінійними та нелінійними часовими рядами даних, подані на рис. 1.1.



Рис. 1.1 – Процеси в економіці та фінансах

Класифікація процесів, подана на рис. 1, є досить спрощеною і призначена тільки для того щоб підкреслити існування двох широких класів: лінійні та нелінійні. Кожний клас потребує використання для їх формального опису різних типів математичних моделей, у багатьох випадках різних методів оцінювання параметрів, різних підходів до оцінювання прогнозів.

Практично майже всі процеси, що відбуваються у природі та різних галузях наукової і практичної діяльності є нелінійними та нестаціонарними. Якщо ми описуємо вибрані процеси лінійними моделями, то це, як правило, означає, що ці процеси є кусково-лінійними, тобто лінійними на деяких часових інтервалах та в деякому обмеженому діапазоні зміни параметрів моделі, а загалом вони є нелінійними.

Інтегровані процеси – це поширені процеси з трендами, які можуть бути лінійними або нелінійними. Гетероскедастичні – процеси із змінною у часі дисперсією. Вони також містять нелінійності оскільки моделі дисперсії включають нелінійні компоненти. Системи (множини) коінтегрованих процесів включають процеси з трендами однакового порядку. При моделюванні коінтегровані процеси аналізуються в одному рівнянні.

Чи можна взагалі на практиці обійтись тільки лінійними моделями?

Однозначна відповідь – ні! Тому що існують процеси, які неможливо лінеаризувати, а якщо ми все-таки стараємось це зробити, то в результаті отримуємо неадекватні моделі, непридатні для подальшого практичного використання. Наведемо деякі приклади природних явищ, які свідчать про наявність нелінійності у їх розвитку.

Причини появи нелінійності у процесах:

Фазовий перехід. Вода, яка з'являється у результаті танення льодовика, призводить до значного швидкого (ступінчатого) підвищення рівня води у річці, яка протікає рядом з цим льодовиком. Тут спостерігається перетворення твердої фази води (льоду) у рідку, що є індикатором наявності нелінійності. Так само, штучний лід може швидко переходити з твердої фази у газоподібну.

Насичення. Якщо деяка система переходить у режим функціонування, який характеризується насиченням, то збільшення вхідного впливу, наприклад, на 20%, не приведе до збільшення реакції на виході також на 20%.

Наприклад, якщо ринок потребує деякого нового товару, то на перших порах прибуток від його продажу буде пропорційним об'єму товару, принесеного на ринок. Однак, з часом відбувається насичення ринку цим

товаром і прибуток вже не буде прямо пропорційним об'єму товару, поданому на ринок.

Так само нелінійним (тобто процесом з насиченням) є розвиток багатьох процесів у живих організмах, наприклад, насичення ліками, мікроелементами або їжею.

Синхронізація розвитку процесів. Данський вчений Хрістіан Гюйгенс відкрив явище синхронізації механічних годинників, якщо їх поставити рядом на дерев'яну основу. Синхронізація функціонування різних пристроїв – це нелінійний процес.

Інший приклад: сильні магнітні поля впливають на живі організми і примушують змінювати свої звичні ритми на ті, що несуть у собі магнітні поля.

Також може наступати синхронізація у біологічних системах. Люди, які проживають разом, з часом можуть набувати однакових (або подібних) звичок.

Нелінійні тренди. Нелінійні тренди (другого порядку і вище) зустрічаються у фінансово-економічних та деяких технологічних процесах, біології, кліматології і багатьох інших галузях. Так, у процесах ціноутворення на біржові активи зустрічаються складні нелінійні тренди 5-го та 7-го порядків при їх формальному описі поліномами від часу. Тренд, характерний для ВВП України у 90-х роках XX-го століття, можна описати поліномом 3-го порядку (цей тренд змінює свій напрям розвитку із зростаючого на спадаючий і навпаки) [25].

Як правило, нелінійні тренди виникають внаслідок одночасного впливу на досліджуваний процес багатьох інших випадкових процесів, які виконують роль збурень. Прикладом може бути вироблення валового внутрішнього продукту (ВВП) в слабких економіках, до яких відносяться, в першу чергу, сучасні перехідні економіки країн пострадянського простору. На них впливають випадкові збурення у вигляді низької організації праці, недостатньої кваліфікації менеджменту, коливання цін на енергоносії, застарілих технологій, локальні військові конфлікти і т. ін.

Приклад формального представлення впливу випадкового процесу – випадкове блукання [2, 23]:

$$y(k) = a_0 + y(k-1) + \varepsilon(k) \quad (1.1)$$

Еволюція процесу $\{y(k)\}$ визначається впливом випадкового процесу

$\{\varepsilon(k)\}$. $\Delta y(k) = y(k) - y(k-1) = a_0 + \varepsilon(k)$; k – дискретний час. Розв’язок цього простого рівняння можна знайти ітераційним способом; для початкової умови $y(0) = y_0$, можна записати:

$$\begin{aligned} k=1: \quad y(1) &= a_0 + y_0 + \varepsilon(1), \\ k=2: \quad y(2) &= a_0 + y(1) + \varepsilon(2) = a_0 + [a_0 + y_0 + \varepsilon(1)] + \varepsilon(2) = \\ &= y_0 + 2a_0 + \varepsilon(1) + \varepsilon(2) \end{aligned} \quad (1.2)$$

а для довільного моменту часу k можна записати по індукції:

$$y(k) = y_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) \quad (1.3)$$

У знайденому розв’язку міститься лінійний тренд $a_0 k$ і випадковий (нерегулярний) тренд у вигляді суми незважених випадкових величин $\sum \varepsilon(i)$. Тобто на детермінований тренд накладається випадкова складова, яка визначається незваженою сумою випадкових величин. Напрямок руху нерегулярного тренду визначається конкретною реалізацією випадкового процесу $\{\varepsilon(k)\}$.

Хаос. Підкидання монети для прийняття деякого рішення вносить випадковість у процес його прийняття. Ця випадковість носить нелінійний стрибкоподібний характер: 0 або 1, тобто варіанти рішення можуть відрізнитись радикально. Фактично, такий підхід до прийняття рішення має хаотичний характер, хоча напрям руху обмежений двома допустимими варіантами.

Динамічні системи, які характеризуються хаотичною поведінкою, часто описують системою з трьох нелінійних диференціальних рівнянь, розв’язок яких є дуже чутливим до початкових умов. Говорять, що такі системи описують хаос у технічному сенсі [37].

З іншого боку, генерування псевдовипадкових чисел – це також хаотичний процес. Генерування відбувається з використанням кусково-лінійних функцій, наприклад, це стосується лінійного конгруентного генератора.

Переключення режимів функціонування. Переключення режимів функціонування відбувається практично у двигунах всіх типів. На першому етапі після включення електродвигуна або двигуна внутрішнього згоряння спостерігається перехідний режим, який поступово переходить в усталений (номінальний) режим.

В економіці пострадянських країн на початку 1990-х років відбувся перехід економіки з режиму функціонування із спільною власністю на засоби виробництва (основні фонди) до режиму функціонування на основі приватної власності. Цей перехід продовжується десятки років з надзвичайно великими економічними втратами, а *час переключення* взагалі став повною *невизначеністю* для суспільства. Стрибкоподібні зміни дуже шкідливі для економіки, особливо якщо вони повторюються у часі. На жаль, ми маємо можливість спостерігати це явище сьогодні (на початку 21-го століття) [29, 44].

Люфт, тертя і гістерезис у технічних системах. Для технічних систем характерні нелінійні ефекти, пов'язані з виникненням люфту, тертя і гістерезису. Гістерезис (петля гістерезису) характерна для поведінки електромеханічних та електромагнітних систем.

Ми будемо вважати процес *нелінійним*, якщо його неможливо адекватно описати лінійною моделлю на інтервалі дослідження. При цьому довільна функція $f(\alpha \mathbf{x} + \mathbf{y})$ вважається лінійною, якщо для векторів $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ і будь-якого дійсного скаляра α можна записати:

$$f(\alpha \mathbf{x} + \mathbf{y}) = \alpha f(\mathbf{x}) + f(\mathbf{y}) \quad (1.4)$$

тобто для цієї функції виконується принцип суперпозиції для лінійних систем. Якщо функція не задовольняє такій умові, то її вважають нелінійною.

Приклади нелінійних процесів

Приклад 1.1. Система рівнянь Лоренца (з фізики рідини):

$$\begin{aligned} \dot{x}(t) &= a(y(t) - x(t)); \\ \dot{y}(t) &= r x(t) - y(t) - x(t)z(t); \\ \dot{z}(t) &= x(t)y(t) - b z(t) \end{aligned} \quad (1.5)$$

де $x(t)$ – конвективна інтенсивність (convective intensity) рідини; $y(t)$ – коливання температури рідини; $z(t)$ – відхилення від лінійності у вертикальному профілі конвекції (deviation from linearity in the vertical convection profile); a – число Прандтля; r – число Рейлі; b – коефіцієнт динаміки відхилення від лінійності у вертикальному профілі конвекції. Подана вище система нелінійна стосовно змінних і описує нелінійні коливання вказаних змінних.

Приклад 1.2. Псевдолінійна модель. Псевдолінійні моделі, тобто моделі, нелінійні стосовно змінних, містять лінійні комбінації змінних та вибраних (довільних) типів базисних функцій, наприклад, у такій формі [13, 14]:

$$y(t) = \sum_{i=1}^m \lambda_i f_i(\mathbf{x}(t)) + \varepsilon(t) \quad (1.6)$$

де $\mathbf{x}(t) = [y(k-l_1), y(k-l_2), \dots, y(k-l_d)] \in \mathbb{R}^d$ – вектор затриманих у часі вимірів основної змінної; $[l_1, l_2, \dots, l_d]$ – вектор значень лагів (цілі додатні значення, $l_i = 0, 1, 2, \dots$, $i = 1, 2, \dots, d$); λ_i – невідомі параметри моделі, які оцінюють на основі наявних даних; d – розмірність вектора лагових (затриманих у часі) змінних; $\varepsilon(k)$ – вектор випадкових збурень для досліджуваного процесу, який часто формально визначають так: $\{\varepsilon(k)\} \sim i.i.d.(\mu, V)$, де параметри розподілу $\mu = const$ і $V = const$ для стаціонарного процесу; *i.i.d.* = *independent and identically distributed* (незалежні та однаково розподілені випадкові величини). Функції $f_i(\mathbf{x}(t))$ можуть містити квадратичні і члени вищого порядку, наприклад, такі: $[y(k-i) y(k-l)]$.

Незважаючи на просту структуру, моделі такого типу знаходять широке застосування у нелінійному моделюванні завдяки можливості адаптації структури і параметрів до конкретного режиму функціонування та можливості спрощеного обчислення оцінок прогнозів на основі побудованої моделі.

Два типи нелінійності:

Досліджувані процеси (часові ряди) в економіці та фінансах можна віднести до одного з таких типів (класів) [1, 2]:

- лінійні (стосовно змінних і параметрів);
- нелінійні стосовно змінних (або квазілінійні);
- нелінійні стосовно параметрів;
- змішані нелінійності: одночасно стосовно змінних і параметрів.

Самими поширеними у застосуванні до розв'язання задач прогнозування та керування є лінійні моделі, широке використання яких зумовлене такими практичними причинами [5]:

- наявність розвиненої теорії побудови лінійних моделей на основі статистичних/експериментальних даних та аналізу їх характеристик;
- простота імітації лінійних процесів у випадку необхідності генерування даних для побудови та апробування моделей, виконання імітаційного моделювання і т. ін.;
- наявність досить широкого спектру методів оцінювання параметрів моделей, які можна успішно застосовувати до згенерованих та фактичних даних за виконання деяких умов;
- простота аналізу адекватності лінійної моделі досліджуваному процесу (об'єкту) за допомогою множини статистичних критеріїв якості;
- можливість побудови та використання прогнозуючих функцій на основі побудованих моделей у випадку необхідності (наприклад, з метою спрощення обчислювальних процедур);
- відносна простота побудови систем підтримки прийняття рішень для розв'язання задач адаптивного (структурно-параметричного) моделювання і прогнозування;
- простота реалізації функцій комбінування оцінок прогнозів, обчислених за альтернативними методами; у багатьох випадках це сприяє підвищенню якості оцінок прогнозів.

Основним недоліком лінійних моделей є те, що вони описують функціонування процесів у деякому обмеженому, вибраному проектувальником системи керування/прогнозування, лінійному режимі. При виході процесу за межі цього режиму модель перестає бути адекватною процесу, що призводить до суттєвого погіршення якості прогнозування та керування. Або ж така лінійна модель взагалі перестає бути придатною для її подальшого використання. Її потрібно адаптувати до нового режиму, якщо це допускає структура моделі, або ж для нового режиму функціонування об'єкта необхідно будувати нову лінійну модель. Така альтернатива досить часто використовується на практиці [2, 8].

Відповідно до типів процесів отримали назви і математичні моделі, які описують ці процеси. Параметри (коефіцієнти) лінійних і квазілінійних моделей оцінюють за допомогою звичайного методу найменших квадратів (МНК), а параметри моделей, нелінійних стосовно параметрів, оцінюють за допомогою методів нелінійного оцінювання, серед яких популярними (широко вживаними) є такі [9, 51]:

- нелінійний метод найменших квадратів (НМНК);
- метод максимальної правдоподібності (ММП);
- метод Монте-Карло для марковських ланцюгів (МКМЛ).
- оптимізаційні ітераційні процедури на основі відомих методів штучного інтелекту: генетичних та імунних алгоритмів і т. ін.

У спеціалізованих системах підтримки прийняття рішень використовують, як правило, різні методи оцінювання параметрів з метою порівняння та отримання можливості одержання альтернативних оцінок з подальшим вибором кращого варіанту для конкретної постановки задачі.

1.3 Поняття структури нелінійної математичної моделі

Введемо поняття структури математичної моделі, яке будемо використовувати надалі. Воно необхідне для створення методики побудови моделі за статистичними даними. Поняття структури моделі містить у собі такі елементи (складові):

1. Порядок моделі, тобто порядок диференціального, різницевого чи іншого рівняння, що використовується для опису динаміки досліджуваного процесу чи об'єкта. Наприклад, широко вживане стохастичне різницеве авторегресійне (АР) рівняння другого порядку має вигляд:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) \quad (1.7)$$

Тобто порядок цього різницевого рівняння визначається кількістю затриманих у часі значень основної змінної, що використовуються у правій частині рівняння для опису поведінки змінної в лівій частині. Стохастичним воно називається тому, що у правій частині наявна випадкова змінна $\varepsilon(k)$, призначення якої ми пояснимо трохи нижче.

Нелінійне різницеве рівняння може мати, наприклад, такий вигляд:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-4) + a_3 y(k-2) + \varepsilon(k) \quad (1.8)$$

У цьому рівнянні враховується сезонний ефект з періодичністю 4 у мультиплікативній формі (другий член у правій частині рівняння). Це рівняння можна вважати рівнянням 4-го порядку, оскільки у ньому використовується основна змінна з лагом 4.

2. Розмірність моделі. Вона визначається кількістю рівнянь, що використовуються для математичного опису поведінки об'єкта чи процесу. Процес, який описують одним рівнянням, називають одновимірним або скалярним. Процес, котрий описують двома і більше рівняннями, називають багатовимірним. Хоча більшість процесів у природі є багатовимірними, часто обмежуються одновимірними моделями, виходячи з простоти їх структури та зручності застосування.

Якщо модель, яка складається з кількох рівнянь, містить хоча б одне нелінійне, то вона вважається нелінійною у цілому.

3. Нелінійність та її тип. Визначити наявність нелінійності – не завжди проста задача. Так, для механічних і деяких інших систем наявність нелінійності можна визначити шляхом вивчення законів, закономірностей і особливостей їхнього функціонування. Наприклад, відомо, що для механічних систем характерною є наявність нелінійностей типу «люфт», «тертя» та гістерезис.

При побудові регресійних моделей частіше зустрічаються нелінійності стосовно змінних ніж нелінійності стосовно параметрів. Прикладами нелінійності стосовно змінних можуть бути поліноміальна регресія виду:

$$y(k) = a_0 + a_1 x(k) + a_2 x^2(k) + a_3 x^3(k) + \varepsilon(k) \quad (1.9)$$

або білінійна модель:

$$y(k) = a_0 + a_1 x_1(k)x_2(k) + a_2 x_1(k)x_3(k) + a_3 x_2(k)x_3(k) + \varepsilon(k) \quad (1.10)$$

Параметри таких рівнянь можна оцінювати звичайним методом

найменших квадратів (МНК) за належної побудови матриці вимірів (її вигляд буде розглянуто нижче). Нелінійність стосовно параметрів зумовлена наявністю у моделі добутків (або відношення) коефіцієнтів, наприклад так, як це має місце у логістичній регресії:

$$g_{\max}(x) = \frac{e^{x(k)}}{1 + e^{x(k)}} \quad (1.11)$$

$$x(k) = a_0 + a_1 \cdot \hat{S}1(k) + a_2 \cdot \hat{S}2(k) + a_3 \cdot \hat{S}3(k) +$$

$$+ a_4 \cdot \hat{P}(k) + a_5 \cdot \hat{R}1(k) + a_6 \cdot \hat{R}2(k) + a_7 \cdot \hat{R}3(k)$$

де $g_{\max}(k)$ – може представляти максимальну ціну активу на біржі;
 $x(k)$ – процес ціноутворення активу; $S1, S2, S3, R1, R2, R3$ – біржові індикатори. Модель використана для прогнозування максимальної ціни на напрямку руху процесу.

Коефіцієнти (параметри) таких моделей неможливо оцінити за допомогою звичайного МНК, а тому для розв'язання цієї задачі використовують нелінійний МНК, метод максимальної правдоподібності чи інші методи нелінійного оцінювання.

4. Час запізнення реакції на виході об'єкта відносно вхідного впливу. Цей параметр необхідно враховувати у лінійних та нелінійних моделях. Запізнення по входу (лаг) досить легко враховується як у неперервних, так і в дискретних моделях. Для моделі з дискретними змінними у вигляді різницевого рівняння

$$y(k) = a_0 + a_1 y(k-1) + a_2 x^2(k-d) + \varepsilon(k) \quad (1.12)$$

тут час запізнення d представляє собою ціле число, що дорівнює кількості періодів дискретизації вимірів, на яке запізнюється вихідний сигнал щодо моменту подачі вхідного. Тривалість періоду дискретизації вимірів залежить від динаміки конкретного процесу і може змінюватися в межах від декількох мікросекунд (або ще менше) у фізико-технічних системах і, наприклад, до одного року в макроекономіці.

Розглянемо модель неперервного процесу у вигляді передавальної функції із запізненням:

$$W(p) = \frac{K e^{-p\tau}}{1 + T p} \quad (1.13)$$

де K – статичний коефіцієнт передачі об’єкта; p – змінна Лапласа; T – постійна часу; τ – час запізнення по входу. Запізнення у дискретній формі d і запізнення у неперервній формі τ зв’язані між собою таким чином:

$$\hat{d} = \text{int}(\tau/T_s) \quad (1.14)$$

тут $\hat{d}$ – оцінка часу запізнення в дискретній формі ($\hat{d} = 0, 1, 2, \dots$).

5. Тип випадкового збурення, що діє на процес, його параметри і спосіб їх врахування. Під збуреннями розуміють вхідні впливи процесу, котрі створюють, як правило, негативні умови для його протікання і не використовуються з тих чи інших причин як керуючі. Збурення поділяють на детерміновані і стохастичні, а враховуються вони у моделі в адитивній чи мультиплікативній формах. У більшості випадків збурення має стохастичний характер. Вище ми подали різницеві рівняння, у яких збурення $\varepsilon(k)$ входить в адитивній формі. Приклад нелінійної мультиплікативної форми врахування збурення у моделі може бути таким:

$$h(k) = v(k)[\alpha_0 + \alpha_1 h(k-1)] \quad (1.15)$$

де $v(k)$ – мультиплікативне збурення.

Частіше всього збурення описують розподілами випадкових величин (статистичні моделі), але в окремих випадках детерміноване збурення можна виміряти і описати функціонально (математичні моделі збурень). Наприклад, можна виміряти температуру навколишнього середовища, яка впливає на протікання реакції у хімічному реакторі, і побудувати відповідну функціональну залежність температури від часу.

Вибір (оцінювання) структури моделі, адекватної процесу, задача не проста і вона розв'язується, як правило, ітераційно. Спочатку структуру моделі оцінюють наближено на підставі дослідження закономірностей протікання процесу, аналізу кореляційних функцій, візуального аналізу статистичних/експериментальних даних. При цьому вибирають кілька найбільш ймовірних структур-кандидатів за допомогою відповідних критеріїв. Потім обчислюють оцінки параметрів моделей-кандидатів і вибирають кращу з них, використовуючи відповідні статистичні характеристики (статистики або критерії) якості моделей [3, 51].

Якщо жодна з оцінених моделей-кандидатів не може вважатися адекватною для конкретного застосування, то необхідно досліджувати на інформативність статистичні/експериментальні дані, які можуть бути недостатньо інформативними для оцінювання моделі. У такому випадку може знадобитися повторний чи додатковий збір експериментальних/ статистичних даних і корегування структури моделі.

Зазначимо, що структуру і параметри моделі краще оцінювати за допомогою адаптивних алгоритмів, для створення яких необхідно вибрати (або створити) належний критерій адекватності та відповідні обчислювальні процедури.

4. Два основних методи побудови математичних моделей

Основними відомими методами побудови математичних моделей є такі:

- *структурний*;
- *функціональний*.

Структурний метод передбачає моделювання внутрішнього механізму взаємодії змінних, відображає їх фактичні взаємозв'язки.

Прикладом структурної моделі може бути модель коливань твердого тіла (конструкції) у матричній формі:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{F}(l, t) \quad (1.16)$$

де $\mathbf{M}$, $\mathbf{C}$, $\mathbf{K}$ – матриці мас, демпфування і жорсткості конструкції;
 $\mathbf{F}(l, t)$ – сила, що діє на конструкцію; $\mathbf{x}$ – переміщення елементів конструкції у

просторі; l – просторова координата; t – неперервний час. Досить часто функція $F(l, t)$ має нелінійну форму.

Критерієм правильності структурної моделі є однаковий характер поведінки основних змінних фактичного процесу і моделі.

Моделі з макроекономічної теорії

Розглянемо, наприклад, зростання інфляції внаслідок випуску додаткової грошової маси. Оскільки логіка цього процесу досить проста і існують статистичні дані, які ілюструють зростання інфляції, то можна постулювати, що інфляція описується диференціальним або різницеvim рівнянням першого/другого порядку (рис. 1.2). Загалом, *інфляція – це нелінійний процес*, що пояснюється комбінованим впливом багатьох детермінованих та випадкових факторів [25, 29].

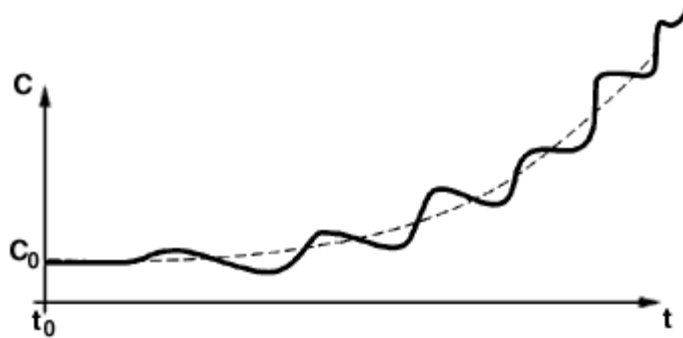


Рис.1.2 Можливий характер зростання інфляції

Структурним підходом можна скористатись, наприклад, для побудови математичної моделі процесу трансформування власності або макроекономіки в цілому. Для цього необхідно визначити вхідні керуючі впливи (змінні), збурення та вихідні змінні, а також визначити якого типу зв'язки існують між ними (рис. 1.3).



Рис. 1.3 Спрощене зображення макроекономічного процесу

Серед керуючих змінних макроекономічного процесу можна виділити внутрішні та зовнішні *інвестиційні потоки*, потоки *сировини, робочої сили, нові технології* та *структурні зміни* в промисловості в цілому, а також в окремих галузях. Метою використання *керуючих змінних* є досягнення заданих рівнів макроекономічних показників: рівня виробництва валового внутрішнього продукту (ВВП), індексу інфляції, індексу людського розвитку, середньої заробітної плати і т. ін.

Як правило, в такі моделі вводять у явному вигляді *збурення* – випадкові змінні (впливи), які негативно впливають на протікання процесу. Так, у випадку створення моделі макроекономічного процесу, збуреннями можуть бути, наприклад, такі фактори впливу [2]:

- помилкові рішення уряду країни і його низька кваліфікація;
- затримка платежів між підприємствами, галузями та державами;
- значні коливання цін на енергоносії;
- значний вплив економік високорозвинених країн на слабкі країни;
- хронічна технологічна і кваліфікаційна відсталість країни;
- швидкі (часто недостатньо обґрунтовані) зміни податкового законодавства;
- вплив значного об'єму капіталу за кордон, який часто є неконтрольованим;
- використання недостовірної статистичної інформації;
- локальні та глобальні війни.

Очевидно, що врахувати таку різноманітну інформацію в моделі надзвичайно складно, а тому випадкові змінні агрегують (об'єднують) і представляють в моделі однією чи двома випадковими змінними, які представляють можливі збурення. Велика кількість збурень, що діють на макроекономічні процеси, спричиняють виникнення нелінійних трендів, коливань та нелінійності інших типів.

На основі знання логіки взаємодії змінних процесу та використання відомих макроекономічних законів (наприклад, врівноваженого розвитку процесів) будується система рівнянь, які описують розвиток окремої галузі або макроекономіки в цілому.

Функціональний підхід використовують для формального опису процесу, не проникаючи глибоко у фактичну структуру цього процесу і взаємодію його змінних.

У даному випадку досліджуваний процес = “чорна скринька”.

(1.17)
25

Наприклад, для побудови моделі ціноутворення можна скористатись такими вхідними змінними: I – об’єм імпорту; D – об’єм грошової маси в обороті; P – місячний об’єм виробництва. За вихідну змінну можна взяти індекс споживчих цін C , тобто:

$$C = f(I, D, P) \quad (1.18)$$

Для побудови цієї моделі необхідно мати чотири часових ряди, які необхідні для обчислення оцінок коефіцієнтів моделі.

Очевидно, що функціональний підхід є простішим і гнучкішим від структурного і саме він найчастіше використовується на практиці. Гнучкість і відносна простота даного підходу дають можливість відносно швидко побудувати високоякісні (адекватні) моделі для прогнозування та синтезу систем керування.

1.5 Узагальнений алгоритм побудови моделі

Розглянемо узагальнений алгоритм побудови математичної моделі на основі експериментальних даних у вигляді кроків, поданих нижче. Він є узагальненим з точки зору можливості застосування до систем чи процесів практично будь-якого типу.

1. *Визначення мети побудови моделі, попереднє вивчення процесу (об’єкта).*

На цьому етапі визначається мета побудови моделі, тобто чи буде модель використовуватись для поглибленого вивчення процесу, прогнозування його стану чи керування. Виконується аналіз функціонування процесу на основі літературних джерел та (можливо) експериментальних даних при їх наявності з метою встановлення *числа входів і виходів, логіки взаємодії* складових частин процесу, визначення можливих *зовнішніх збурень* та їх типу: детерміновані чи випадкові. По можливості, необхідно встановити розподіл ймовірностей для випадкових збурень або функціональне описання для детермінованих збурень. Якщо існують моделі подібних процесів, їх також необхідно досконально вивчити та врахувати можливі переваги і недоліки.

2. *Попередня оцінка структури моделі.* На основі вивчення процесу, виконаного на першому етапі, необхідно попередньо встановити типи структур моделей-кандидатів. Їх може бути декілька в залежності від того наскільки невизначеною є інформація відносно процесу. Чим більшою є інформаційна невизначеність, тим більше структур моделей необхідно досліджувати в процесі побудови адекватної моделі.

3. *Планування експерименту та підготовка до його виконання.* На цьому етапі виконуються такі дії щодо планування виконання експерименту (у тих випадках, коли це можливо) з метою отримання експериментальних даних:

- визначаються діапазони зміни вхідних та вихідних величин, збурень;
- встановлюється дискретність зміни вхідних величин, період дискретизації вимірів (якщо змінні неперервні), визначаються типи вимірювальних та реєструючих приладів;
- режими функціонування процесу, для яких необхідно зібрати експериментальні/статистичні дані;
- якщо дані носять статистичний характер, то визначається періодичність їх збору та занесення в базу даних;
- тип, об'єм і якість продукції, що буде вироблена протягом експерименту, а також визначаються необхідні об'єми сировини та енергії.

4. *Виконання експерименту та формування бази даних.* На цьому етапі реалізується розроблений на третьому етапі план експерименту і формуються часові ряди з вимірів (чи статистичних даних), які будуть використані для обчислювання оцінок параметрів математичних моделей.

5. *Обчислення оцінок параметрів (коефіцієнтів) математичних моделей на основі експериментальних даних.* При цьому оцінюють параметри для всіх моделей-кандидатів, вибраних на другому етапі. Для виконання цієї задачі необхідно:

- вибрати метод оцінювання параметрів моделі в залежності від її структури;
- зробити попередню обробку даних; в залежності від конкретної задачі це може бути масштабування, логарифмування, цифрова та/або оптимальна фільтрація, видалення недостовірних даних, обробка екстремальних значень і т. ін.;
- обчислити оцінки (векторів) параметрів моделей з використанням альтернативних методів.

6. *Визначити ступінь адекватності кожної моделі-кандидата* процесу за допомогою належних статистичних критеріїв. Вибрати кращу модель з множини кандидатів.

7. Якщо побудована модель відповідає сформульованим вимогам (за точністю прогнозу чи якістю керування), то завершити процедуру; інакше, перейти на 8-й крок.

8. *Уточнити структуру моделі*, зібрати, за необхідністю, додаткові експериментальні дані, уточнити вибір методу оцінювання параметрів і перейти на крок 5.

Хоча планування та виконання експерименту для соціально-економічних та фінансових систем є досить складною задачею, в окремих випадках це цілком можливо, особливо, якщо підприємство має наміри впровадити нові інформаційні системи та технології обробки даних і методи прогнозування розвитку процесів на виробництві. Наприклад, цілком можливо спланувати та провести інвестиційний експеримент, експерименти з новими технологіями, новими типами продукції. У таких випадках доцільно спроектувати та реалізувати спеціалізовану систему підтримки прийняття рішень (СППР), яка надасть можливість досить швидко і з високою якістю розв'язувати поставлені задачі. Крім того, СППР надає можливість прискорено адаптуватись до нових умов, модифікувати інформаційні технології, методи моделювання, прогнозування та керування.

1.6 Вимоги до експериментальних даних, оцінок параметрів та моделі

1.6.1. Вимоги до експериментальних даних

1. Вимога *неперервності та синхронності даних*. Експериментальні дані повинні вимірюватись та реєструватись через однакові проміжки часу (період дискретизації вимірів, T_s). Цю вимогу необхідно виконувати для процесів будь-якого типу – технічних, економічних, екологічних і т. ін. Порухення цієї вимоги призводить до зміни спектрального складу вимірювального сигналу, що недопустимо, оскільки при цьому змінюється інформативність сигналу. Крім того, виміри входних та вихідних впливів (сигналів) необхідно робити синхронно, тобто в одні й ті ж моменти часу. Інакше, вони будуть непридатні для побудови передаточних функцій та інших моделей, оскільки порушуються причинно-наслідкові зв'язки між входами та виходами. Як правило, в системах моніторингу, діагностики та

керування реального часу задача збору вимірювальних даних має найвищий пріоритет [2, 10].

2. *Вибірка даних повинна бути представницькою.* Це означає, що вона повинна охоплювати досить довгий період часу, щоб включити в розгляд ті режими роботи, які передбачається описати математичною моделлю. Розрізняють два основних режими роботи процесів: *перехідний та усталений*. В перехідному режимі система керування переводить процес з деякого початкового стану в заданий. Перебування процесу в заданому (номінальному) стані протягом деякого відносно довгого проміжку часу називають усталеним режимом роботи.

Прикладом широко відомого перехідного процесу є процес нагрівання до кипіння вмісту каструлі на кухонній плиті. На початку цього процесу ми задаємо режим максимальних витрат енергії щоб скоротити тривалість процесу нагрівання. Після досягнення режиму кипіння витрати енергії можна суттєво скоротити і процес переходить в усталений режим “повільного” кипіння. Подачу великої кількості енергії в початковий момент часу можна порівняти з подачею на вхід об’єкта сигналу у вигляді сходинок, а зміну температури вмісту каструлі можна вважати за перехідну характеристику цього процесу. Очевидно, що безліч прикладів такого типу можна знайти в промисловості. На сьогодні в нашому суспільстві спостерігається перехідний процес від соціалістичного ладу, який ґрунтувався на суспільній власності на засоби виробництва, до капіталістичного з приватною власністю.

3. *Вибірка вимірювальних даних повинна бути інформативною.* Частіше всього інформативність пов’язують з *числом похідних*, що їх містять виміри. Чим більше число похідних можна отримати з вимірів, тим інформативнішим є сигнал (процес). Наприклад, припустимо, що досліджуваний процес описується диференціальним рівнянням другого порядку:

$$a_2 \frac{d^2 y}{dt^2} + a_1 \frac{dy}{dt} + a_0 = bu(t) \quad (1.19)$$

де $y(t)$ – сигнал на виході процесу; $u(t)$ – вхідний сигнал; $\theta^T = [a_0 \ a_1 \ a_2]$ – вектор коефіцієнтів (параметрів) рівняння, які необхідно оцінити за допомогою експериментальних даних. Очевидно, що оцінки коефіцієнтів a_2 і

a_1 можна обчислити тільки в тому випадку, якщо виміри $y(t)$ містять другу і першу похідну по часі.

Іноді інформативність формально визначають *величиною дисперсії* процесу, тобто чим більшою є дисперсія, тим вища його інформативність. Так, константа має нульову дисперсію і, відповідно, мінімальну інформативність.

Вимога інформативності виконується в тому випадку, коли вхідний сигнал задовольняє умові *достатнього збудження* процесу. Основна ідея достатнього рівня збудження полягає в тому, щоб смуга частот вхідного сигналу перекривала амплітудно-частотну характеристику процесу. Тобто вихідний сигнал $y(t)$ буде інформативним у тому випадку, коли достатньо інформативним буде вхідний сигнал $u(t)$. Умові інформативності (достатнього збудження) задовольняють такі основні вхідні сигнали: *білий шум*, *псевдовипадковий двійковий сигнал* та *одиничний імпульс*. Білий шум (гаусів процес) теоретично має нескінченний частотний спектр; достатньо широкі частотні спектри мають і два інші сигнали.

З одного боку, для збудження процесу на його вхід необхідно подавати інформативний сигнал типу білого шуму, а з іншого – такий сигнал може бути недопустимим з точки зору фізичних особливостей функціонування процесу (подача на вхід подібного сигналу може призвести до руйнування процесу чи до створення аварійної ситуації). Тому в таких системах часто використовують як збуджуючий сигнал завдання регулятору, якщо він має форму сходинок, тобто має фронт прямокутного імпульсу. В багатьох випадках можливе використання гармонічних сигналів, які сприймаються "легше" більшістю об'єктів ніж білий шум або одиничний імпульс. Так, при дослідженні механічних систем часто використовують одиничні імпульси, гармонічні збуджуючі сигнали та їх комбінації, а при дослідженні технологічних процесів до вхідного сигналу керування додають 10-15% білого шуму, який забезпечує достатній рівень „збудження” процесу.

1.6.2 Вимоги до оцінок параметрів моделі

Точність оцінок параметрів моделі залежить від якості вимірювальних даних, коректності попередньої обробки даних та від того, наскільки правильно обрано метод оцінювання. Так, для оцінювання параметрів лінійних і псевдолінійних (нелінійних відносно змінних) моделей можна застосовувати

звичайний МНК та його модифікації, а для оцінювання моделей, нелінійних стосовно параметрів, необхідно застосовувати нелінійний МНК, метод максимальної правдоподібності та інші методи, розроблені для оцінювання параметрів нелінійних моделей.

Існують такі стандартні вимоги до оцінок параметрів математичних моделей [8, 25]:

1. Оцінки повинні бути *незміщеними*. Це означає, що оцінки параметрів не повинні містити систематичної похибки, яка збільшує або зменшує оцінки параметрів на всіх вибірках даних, що використовуються, або на різних відрізках однієї вибірки. Формально незміщеність оцінок параметрів записують так:

$$E[\overset{\vee}{\theta}] = \theta \quad (1.20)$$

де E – символ математичного сподівання; $\overset{\vee}{\theta}$ – вектор оцінок параметрів; θ – істинне значення вектора параметрів.

2. Оцінки повинні бути *консистентними*, тобто оцінка $\overset{\vee}{\theta}$ вектора параметрів повинна наближатись до свого істинного значення θ по мірі збільшення об'єму вибірки даних. Оскільки оцінка $\overset{\vee}{\theta}$ – це випадкова величина, то наближення до істинного значення можливе тільки в імовірнісному сенсі. Консистентна оцінка повинна задовольняти такій умові:

$$P(|\overset{\vee}{\theta}_k - \theta| < \varepsilon) \rightarrow 1 \quad \text{при} \quad k \rightarrow \infty, \quad (1.21)$$

де $\varepsilon > 0$ – задане мале число; P – символ ймовірності; $\overset{\vee}{\theta}_k$ – оцінка вектора параметрів в момент k . Відомо, що довжина перехідного процесу при оцінюванні моделі залежить від кількості параметрів, що оцінюються, та вимірності моделі, а тому об'єми вибірок даних в усіх випадках повинні бути, по можливості, більшими. Проблеми з необхідними об'ємами даних виникають, як правило, при моделюванні економічних та соціальних систем; при побудові моделей технічних систем такі проблеми досить рідкісні.

3. Оцінки повинні бути *ефективними*, а це означає, що із множини допустимих, незміщених та консистентних оцінок необхідно вибрати ті, що є найближчими до оцінюваних параметрів, тобто ті, що мають найменші відхилення від середнього значення.

Іншими словами це вимога мінімальності дисперсії оцінки, яка формально записується так:

$$\text{Var}(\hat{\theta}) \rightarrow \min \quad (1.22)$$

Незміщені ефективні оцінки параметрів лінійної моделі можна отримати, наприклад, за допомогою методу найменших квадратів, якщо при оцінюванні виконуються такі умови:

- *похибка моделі* $e(k) = y(k) - \hat{y}(k)$ є центрованою величиною; де $y(k)$ – значення ряду, отримане експериментально (або статистичні дані); $\hat{y}(k)$ – оцінка змінної, отримана за допомогою побудованої моделі;
- *похибки моделі* – повинна бути некорельованим процесом, тобто у ньому має бути відсутня автокореляція похибок:

$$\text{cov}[e(k)] = E[e(k)e(k-l)] = \begin{cases} \sigma_e^2, & k = l, \\ 0, & k \neq l. \end{cases} \quad (1.23)$$

- *похибка моделі* – некорельована із залежною змінною $y(k)$.

Корельованість похибок означає, що вона містить інформацію про досліджуваний процес; модель не містить всю наявну інформацію про процес. Тобто необхідно коригувати структуру моделі таким чином, щоб похибки стали некорельованими.

1.6.3 Вимоги до математичної моделі у цілому

1. Модель повинна буди *адекватною* процесу чи об'єкту. Адекватність означає, що модель повинна мати такі характеристики [8]:

- а) відображати найбільш характерні зв'язки та взаємодію між змінними процесу;

- б) враховувати можливі керуючі дії (сигнали);
- в) враховувати вплив зовнішніх збурень та шуми (похибки) вимірів;
- г) враховувати початкові значення змінних та обмеження на них;
- д) модель повинна мати прийнятну якість за множиною відповідних статистичних критеріїв адекватності.

Формально адекватність визначають за допомогою ряду статистичних величин. Наприклад, часто використовують середньоквадратичну похибку моделі (СКП), яка обчислюється за формулою:

$$СКП(x_s, x_m) = \sqrt{\frac{1}{N} \sum_{k=1}^N [x_s(k) - x_m(k)]^2} \quad (1.24)$$

де $x_s(k)$ – вимір вихідної змінної об’єкта в момент k ; $x_m(k)$ – оцінка вихідного сигналу об’єкта, отримана за оціненою моделлю. Для лінійних моделей існує множина статистичних параметрів, що використовуються при оцінюванні адекватності, які будуть розглянуті нижче. Використання одного параметра для визначення ступеня адекватності моделі є некоректним підходом, оскільки *оцінки параметрів – це випадкові величини*, а тому збільшення кількості критеріїв адекватності сприяє підвищенню ймовірності вибору адекватної моделі.

2. Рівняння моделі повинні мати *розв’язок*, тобто бажано мати *аналітичний* або, якщо це неможливо, то *чисельний* розв’язок.

Одним із принципів, яких необхідно дотримуватись при побудові моделі є наступний: *“в моделі не повинно бути нічого зайвого крім необхідного для збереження основних статистичних характеристик досліджуваного процесу”*. Звичайно, що дотримуватись цього принципу досить непросто, і на практиці буває так, що модель дійсно має надзвичайно складну структуру, що також може бути оправдано необхідністю досягнення високого ступеня її адекватності процесу. Це особливо стосується нелінійних процесів. Але при побудові лінійних моделей у вигляді авторегресії чи авторегресії з ковзним середнім достатньо побудувати модель, статистичні характеристики якої співпадають з статистичними характеристиками часового ряду, на основі якого вона оцінюється. Такі спрощені моделі виявляються цілком придатними для прогнозування та керування процесами. Загалом питання складності моделі вирішується у кожному випадку окремо.

3. Модель повинна бути достатньо *універсальною*, щоб її можна було застосувати до описання класу однотипових процесів або до описання функціонування процесу в різних умовах.

Наприклад, для опису *моторної функції* людини (реакція на зовнішні збуджуючі сигнали) застосовують звичайне диференціальне рівняння другого порядку, яке представляють у вигляді функції передачі такого ж порядку:

$$W(s) = \frac{K e^{-\tau s}}{(1 - T_1 s)(1 - T_2 s)} \quad (1.25)$$

де K – статичний коефіцієнт передачі об'єкта; τ – час запізнення по входу, який в середньому дорівнює для людини 300-350 мс; T_1, T_2 – постійні часу. Така передаточна функція може використовуватись, наприклад, для опису реакції людини на зовнішні відео- або аудіосигнали, що поступають через систему візуального сприйняття чи аудіосистему (поширений приклад – водіння автомобіля чи іншої машини). Значення параметрів моделі можуть бути різними для різних людей, але структура моделі залишається незмінною. Таким чином, подана вище модель описує широкий клас біологічних систем і цілком відповідає умові універсальності.

При моделюванні технічних систем широко застосовують лінійні ланки першого і другого порядку, що відповідають звичайним диференціальним рівнянням таких же порядків. На основі таких простих ланок можна побудувати моделі високої складності. Дуже поширений в техніці та екології клас систем з розподіленими у просторі параметрами. Наприклад, процес розповсюдження домішок в атмосфері та водному середовищі, механічні коливання сонячних батарей та антен космічних супутників, коливання крила літака, локомотива з вагонами на залізниці, автомобіля з причепом, а також можна навести багато інших прикладів. Динаміку таких систем описують, як правило, диференціальними рівняннями з частинними похідними або використовують моделі, побудовані на основі методу скінченних елементів.

4. Вимога *робастності* (*robust* – сильний, міцний). Робастність означає, що модель повинна давати прийнятний прогноз вихідної змінної не тільки на тому відрізку часового ряду, на основі якого вона побудована, але і на будь-якому іншому відрізку, що відповідає вибраному режиму роботи. Робастність може розглядатись також як стійкість моделі по відношенню до

збурень, похибок та пропусків вимірів. Вимога робастності є особливо критичною для систем, що працюють в реальному часі, оскільки нестійка модель може стати причиною створення аварійної ситуації.

5. Вимога *адаптивності*. Ця вимога означає, що структуру і хоча б частину параметрів моделі можна уточнювати по мірі надходження нових вимірів від об'єкта. Ця вимога є обов'язковою при побудові моделей нестационарних систем, тобто систем, параметри (і структура) яких є функціями часу. Системи керування, що будуються для нестационарних об'єктів та процесів, називають адаптивними. Такі системи є досить складними з точки зору аналізу збіжності оцінок параметрів, поточного контролю адекватності моделей, похибок керування, а тому при проектуванні адаптивних систем необхідно особливу увагу приділяти питанням достатнього збудження об'єктів керування та коректного вибору методів оцінювання параметрів математичних моделей.

1.7 Загальна спрощена класифікація математичних моделей

В спеціальній літературі можна знайти класифікацію математичних моделей за різними критеріями, які приводять до визначення багатьох класів та підкласів моделей. Нижче розглянемо спрощену класифікацію, яка може бути корисною на практиці.

1. Описові (вербальні) моделі

– Технічне завдання, звіт, логічна схема взаємодії змінних, яка супроводжується словесним описом.

2. Математичні моделі

Вони будуються, наприклад, на основі рівнянь таких типів:

- дифференціальні рівняння;
- різницеві рівняння;
- алгебраїчні рівняння.

Математичні моделі розділяють на два широких класи:

а) *аналітичні* – моделі, які описують вибрані змінні процесу; як правило, такі моделі відтворюють один із аспектів функціонування процесу чи об'єкта, наприклад, динаміку валового внутрішнього продукту, моторну функцію людини чи рух автомобіля, потяга або літака;

б) *імітаційні* – моделі, які докладно відтворюють поточне функціонування процесу у вибраному масштабі часу; як правило, це досить складні моделі, які будують шляхом комбінування простіших компонент.

Імітаційні (ІМ) моделі нагадують активний фізичний експеримент з можливим використанням фактичних даних, отриманих безпосередньо з процесу.

Переваги імітаційних моделей:

- висока наочність результатів обчислень (проміжних та остаточних);
- динамічний характер відображення протікання процесу, що відповідає фактичному протіканню досліджуваних процесів;
- можливість врахування детермінованих і випадкових факторів (збурень), а також складних залежностей від них;
- простота корекції моделі (досить просто можна ввести у модель додаткове рівняння (чи систему рівнянь), правило і т. ін.);
- існує можливість дослідити процес на множині його реалізацій (тобто спланувати і виконати статистичний експеримент);
- практично необмежені можливості введення в модель таких, як
 - рівняння будь-якого типу;
 - логічні правила;
 - правила нечіткої логіки;
 - статистичні розрахунки;
 - оптимізаційні процедури;
- алгоритми адаптації;
- процедури прогнозування або підтримки прийняття рішень;
- об'єднання чітких алгоритмів з нечіткими (це не завжди просто, але корисно при дослідженні функціонування процесів з невизначеностями та при прийнятті рішень).

Недоліки імітаційних моделей:

- висока вартість побудови та подальшого практичного використання (супроводження функціонування) систем імітаційного моделювання;
- як правило, необхідно багато часу для створення таких систем;
- для створення таких моделей необхідно мати (організувати) групи висококваліфікованих виконавців з досвідом розв'язування подібних задач.

8. Деякі типи лінійних та нелінійних рівнянь, які використовують для опису часових рядів

Авторегресія: рівняння авторегресії описує пам'ять процесу, тобто вплив значень попередніх станів на його поточний стан [2, 50]:

$$y(k) = a_0 + a_1 y(k-1) + \dots + a_p y(k-p) = a_0 + \sum_{i=1}^p a_i y(k-i) + \varepsilon(k) \quad (1.26)$$

де $a_i, i=1, \dots, p$ – коефіцієнти моделі, які оцінюються на основі значень часового ряду; p – порядок авторегресії, який визначається числом затриманих в часі значень ряду, що використовуються у правій частині рівняння для опису динаміки змінної в момент k ; $k=1, 2, \dots$ – дискретний час; $\varepsilon(k)$ – випадкова величина, поява якої у моделі зумовлена такими причинами:

- вплив випадкових збурень на процес, що моделюється;
- похибки рівняння, зумовлені неточно вибраною структурою (можливо, що не враховано деякі регресори, введено непотрібні незалежні змінні або робиться спроба моделювати нелінійний процес за допомогою лінійного рівняння);
- методичні і обчислювальні похибки, які з'являються при обчисленні оцінок коефіцієнтів (параметрів) рівняння.

Розширена нелінійна авторегресія, яка містить нелінійність квадратичного типу:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + b_1 x_1^2(k) + b_2 x_2^2(k) + \varepsilon(k) \quad (1.27)$$

де x_1, x_2 – регресори (незалежні змінні).

Парна регресія – вона включає в правій частині незалежну змінну (регресор):

$$y(k) = a_0 + a_1 x(k) + \varepsilon(k) \quad (1.28)$$

де $x(k)$ – *регресор* (незалежна або екзогенна змінна), тобто $x(k)$ має три назви. Залежну змінну $y(k)$ називають ще *ендогенною* змінною.

Множинна регресія – множинна регресія відображає вплив декількох незалежних змінних на залежну:

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + \dots + a_p x_p(k) + \varepsilon(k) \quad (1.29)$$

де $x_1(k), \dots, x_p(k)$ – регресори рівняння. Таке рівняння може включати також авторегресійну частину.

Авторегресія + множинна регресія = змішана регресія:

$$y(k) = a_0 + \sum_{i=1}^l a_i y(k-i) + b_1 x_1(k) + b_2 x_2(k) + \dots + b_p x_p(k) + \varepsilon(k) \quad (1.30)$$

Авторегресія з ковзним середнім порядку (p, q) – (АРКС (p, q)):

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j \varepsilon(k-j) + \varepsilon(k) \quad (1.31)$$

Регресія, нелінійна відносно змінних (псевдолінійна регресія):

$$y(k) = a_0 + a_1 x(k) + a_2 x^2(k) + \dots + a_p x^p(k) + \varepsilon(k) \quad (1.32)$$

Тобто в даному випадку це поліноміальна регресія порядку P . Коефіцієнти псевдолінійної регресії оцінюються такими ж методами, що і лінійної, наприклад методом найменших квадратів (МНК) або максимальної правдоподібності (ММП).

Регресія, нелінійна стосовно параметрів. – Моделі, нелінійні стосовно параметрів, містять адитивні члени, які включають в себе добутки параметрів моделей або інші види зв'язку (крім адитивного) між параметрами [48]:

$$y(k) = a_0 + a_1 e^{bx(k)} + \varepsilon(k) \quad (1.33)$$

Для оцінювання параметрів таких моделей необхідно застосовувати методи нелінійного оцінювання – нелінійний метод найменших квадратів, метод максимальної правдоподібності, Монте Карло та інші.

Моделі гетероскедастичних процесів – тобто процесів, дисперсія яких змінюється в часі; гетероскедастичні процеси нестационарні та нелінійні за означенням, оскільки дисперсія описується квадратичним членом. Рівняння для умовної дисперсії (авторегресія першого порядку) [2, 51]:

$$h(k) = \beta_0 + \varepsilon^2(k-1) + \varepsilon_1(k) \quad (1.34)$$

де $h(k)$ – умовна дисперсія процесу в момент k ; $\varepsilon^2(k)$ – квадрат залишків; $\varepsilon_1(k)$ – похибка моделі в момент k . Докладно моделі гетероскедастичних процесів будуть розглянуті нижче.

Авторегресійна умовно гетероскедастична модель порядку p (АРУГ(p)):

$$h(k) = \beta_0 + \sum_{i=1}^p \beta_i \varepsilon^2(k-i) + \varepsilon_1(k) \quad (1.35)$$

Узагальнена авторегресійна умовно гетероскедастична модель (УАРУГ(p, q)):

$$h(k) = \beta_0 + \sum_{i=1}^p \beta_i \varepsilon^2(k-i) + \sum_{i=1}^q \alpha_i h(k-i) + \varepsilon_1(k) \quad (1.36)$$

де $\alpha, \beta, \gamma \geq 0$ (щоб уникнути появи від'ємних значень умовних дисперсій).

Експоненціальна модель УАРУГ (умовна дисперсія як асиметрична функція ε , тобто моделювання впливу попередніх значень $\varepsilon(k-i)$ на волатильність):

$$\log(h(k)) = \alpha_0 + \sum_{i=1}^p \alpha_i \frac{|\varepsilon(k-i)|}{h(k-i)} + \sum_{i=1}^p \gamma_i \frac{\varepsilon(k-i)}{h(k-i)} + \sum_{i=1}^q \beta_j \log(h(k-i)) \quad (1.37)$$

Модель УАРУГ-М (модифікована) – моделювання премії за ризик:

$$y(k) = \beta + \gamma h(k) + \varepsilon(k)$$

$$h(k) = a_0 + a \sum_{i=1}^p \varepsilon^2(k-i) + \sum_{i=1}^q h(k) + \varepsilon_2(k) \quad (1.38)$$

Модель для прогнозування волатильності за допомогою УАРУГ:

$$\begin{aligned} h(k+1) &= \beta_0 + \beta_1 \varepsilon^2(k) + \gamma h(k), \\ h(k+j) &= \beta_0 + (\beta_1 + \gamma_1)h(k+j-1) \end{aligned} \quad (1.39)$$

Рівняння для умовної дисперсії та коваріації:

$$\begin{aligned} h(s(k)) &= \alpha_0 + \alpha_1 \varepsilon^2(s(k-1)) + \alpha_2 h(s(k-1)), \\ h(f(k)) &= \beta_0 + \beta_1 \varepsilon^2(f(k-1)) + \beta_2 h(f(k-1)), \\ \text{cov}[s(k), f(k)] &= \gamma_0 + \gamma_1 \varepsilon(s(k-1)) + \gamma_2 \text{cov}[s(k-1), f(k-1)] \end{aligned} \quad (1.40)$$

Коефіцієнт хеджування при використанні двофакторної моделі:

$$b^*(k) = \frac{\text{cov}[s(k), f(k)]}{h^2(f)} \quad (1.41)$$

Моделі коінтегрованих процесів, тобто нестационарних процесів, які можна об'єднати в межах однієї моделі, що має стаціонарні статистичні характеристики.

Концепція коінтегрованості змінних передбачає існування довгострокового зв'язку між значеннями вибраних змінних. Тобто ми припускаємо існування спільної врівноваженої траєкторії руху цих змінних, від якої вони можуть відхилятися на коротких проміжках часу. Однак економічні механізми в цілому діють таким чином, що рівновага відновлюється і зберігається на довгих часових інтервалах шляхом корегування відповідних відхилень від врівноваженого стану [8, 51].

У випадку коінтегрованості змінних $x(k)$ і $y(k)$, для них може бути побудована модель корегування похибки, яка поєднує динаміку змінних на коротких проміжках часу з довгостроковим врівноваженим зв'язком та має такий вигляд:

$$\Delta x(k) = a_{10} + \sum_{i=1}^p b_{1i} \Delta x(k-i) + \sum_{i=1}^p c_{1i} \Delta y(k-i) + \lambda_1 e(k-1) + \varepsilon_1(k) \quad (1.42)$$

Коефіцієнти λ_1, λ_2 в наведених рівняннях називають швидкістю пристосування (корегування). Більш докладно побудова та характеристики моделей коінтегрованих процесів будуть розглянуті нижче.

Вище ми навели тільки деякі структури (типи) математичних моделей стаціонарних та нестаціонарних нелінійних процесів, які широко використовуються при формальному описі динаміки процесів різної природи. На сьогодні існує широка множина інших різновидів дискретних моделей, які описують різні складові фактичних процесів. В подальшому ми будемо намагатися виконати аналіз по можливості більшої кількості типів існуючих моделей, але необхідно зазначити, що зробити це в межах однієї, навіть досить об'ємної, публікації досить складно.

Запитання і вправи до першого розділу

1. Дайте визначення стаціонарного процесу у вузькому (слабкому) та широкому (сильному) сенсі.
2. Чим відрізняється коваріаційна стаціонарність від строгої стаціонарності?
3. Чому умова на суму коренів характеристичного рівняння AR-моделі є критичною для стаціонарності?
4. Яке значення має властивість ергодичності для стаціонарних процесів?
5. Чому випадкове блукання не є стаціонарним процесом?
6. Як пов'язане поняття стаціонарності з властивостями спектральної густини процесу?
7. Які обмеження накладаються на коефіцієнти MA-моделей для забезпечення інвертованості?
8. Чим відрізняється нестаціонарний процес із трендом від нестаціонарного з одиничними коренями?
9. Як використовується різницювання для перетворення нестаціонарного процесу у стаціонарний?
10. Чому стаціонарність є необхідною умовою для застосування багатьох методів прогнозування?
11. Як перевірити стаціонарність часового ряду візуально? Які ознаки можна помітити на графіку?

12. Опишіть покрокову процедуру тесту Дікі–Фуллера. Які є його обмеження?
13. У чому полягає суть тесту KPSS і чим він відрізняється від ADF?
14. Чи можна вважати стаціонарним ряд, у якого середнє значення зростає, але дисперсія залишається сталою?
15. Наведіть приклади економічних процесів, які є стаціонарними, і тих, що зазвичай нестаціонарні.
16. Які наслідки має неправильне трактування нестаціонарного процесу як стаціонарного у прогнозуванні?
17. Що таке сезонна нестаціонарність і як її можна виявити?
18. Як ARIMA-моделі враховують нестаціонарність у часових рядах?
19. Як можна використати автокореляційну та часткову автокореляційну функції (ACF і PACF) для перевірки стаціонарності?

РОЗДІЛ 2. МЕТОДИКА ПОБУДОВИ МАТЕМАТИЧНОЇ МОДЕЛІ НЕЛІНІЙНОГО ПРОЦЕСУ (ЧАСОВОГО РЯДУ)

Розглянемо методику моделювання, яка більше орієнтована на побудову моделей нелінійних фінансово-економічних, соціальних та екологічних процесів, для яких, як правило, набагато складніше поставити експеримент та отримати інформативні експериментальні дані у достатньому об'ємі. Хоча у поданому вигляді ця методика може бути успішно застосована також до побудови моделей динаміки технічних систем, технологічних процесів, деяких рухомих об'єктів і т. ін.

Узагальнена методика побудови математичної моделі процесу на основі експериментальних/статистичних даних, в основу якої покладено відомі ідеї Бокса і Дженкінса, складається з поданих нижче кроків.

– Виконання аналізу особливостей функціонування процесу, для якого будується модель, на основі літературних джерел, експертних оцінок фактичного протікання/розвитку процесу, візуального дослідження вимірів вхідних і вихідних змінних, представлених часовими рядами, та іншої доступної інформації про вибраний для дослідження процес (об'єкт). Попередній аналіз можливості використання лінійних та/або нелінійних моделей для формального опису досліджуваного об'єкта (процесу).

– Попередня обробка статистичних/експериментальних даних, тобто їх підготовка до оцінювання структури і параметрів математичної моделі. Сюди можуть входити такі операції: нормування, заповнення пропусків, цифрова та/або оптимальна фільтрація, обробка екстремальних значень, бутстреп (розмноження вибірки), структурування даних з метою підвищення їх інформативності.

– Аналіз часових рядів на можливу наявність нестационарності та нелінійності за допомогою множини відповідних статистичних критеріїв. При цьому бажано застосовувати більше одного критерію з метою підвищення ймовірності отримання коректного результату. Це пояснюється тим, що на кожному етапі моделювання обчислюються та використовуються ті чи інші типи оцінок елементів моделі (структури, параметрів, можливих оцінок прогнозів), які тільки наближено відповідають невідомим істинним значенням.

– Оцінити структури моделей-кандидатів, для чого необхідно виконати такі операції: обчислити та виконати аналіз кореляційної матриці для часових рядів залежної та незалежних змінних з метою визначення тих екзогенних змінних, які необхідно включити в модель. Обчислити автокореляційну (АКФ)

та часткову автокореляційну функції (ЧАКФ) залежної змінної з метою оцінювання порядку авторегресійної частини моделі; оцінити функції взаємної кореляції між залежною та незалежними змінними, визначити тип розподілу даних, оцінити лаги для регресорів і характеристики інших елементів структури математичної моделі.

– Вибрати і застосувати до наявних даних метод (методи) оцінювання параметрів математичних моделей-кандидатів вибраних структур. Найчастіше це модифікації звичайного методу найменших квадратів (МНК); нелінійний МНК (НМНК); метод максимальної правдоподібності (ММП); метод допоміжної змінної (МДЗ); метод Монте-Карло для марковських ланцюгів (МКМЛ) та інші.

– Вибрати кращу (кращі) з оцінених моделей-кандидатів за допомогою множини статистичних критеріїв адекватності. На цьому етапі також не варто зупинятись на одній моделі, а необхідно скористатись кількома кандидатами.

– Перевірити вибрані моделі на можливість їх використання для обчислення оцінок прогнозів – це задача, яка не потребує значних витрат часу, але дає можливість досить швидко оцінити якість побудованої моделі. Виконати аналіз якості оцінок прогнозів за множиною статистичних критеріїв якості і зробити остаточний вибір моделі для подальшого використання.

Нижче розглянемо докладніше кожний етап побудови моделі за статистичними/експериментальними даними.

Використання елементів методики побудови лінійних моделей

Незважаючи на те, що відома методика Бокса-Дженкінса та її модифікації запропоновані для побудови моделей лінійних процесів, деякі елементи цієї методики можна використати при побудові моделей нелінійних часових рядів. Наприклад, можна скористатись такими корисними процедурами аналізу даних:

- встановлення факту наявності залежності між змінними можна наближено оцінити за допомогою лінійних та нелінійних кореляційних функцій;
- оцінити значення лагів для основної (залежної) змінної можна за допомогою автокореляційних функцій;
- кореляційний аналіз залишків лінійних моделей дає можливість встановити наявність нелінійності та гетероскедастичності (дисперсія процесу змінюється у часі) у даних;
- деякі лінійні моделі досить легко модифікуються в нелінійні; наприклад, лінійну множинну регресію можна трансформувати у

нелінійну поліноміальну регресійну модель, а авторегресію з ковзним середнім у нелінійну розширену АРКС (НАРКС), наприклад, таким чином [15, 29]:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j v(k-i) + c x^m(k) + \varepsilon(k) \quad (2.1)$$

або у складнішій формі (структурі):

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) y(k-j) + \sum_{j=1}^q b_j v(k-i) + c x^m(k) + \varepsilon(k), \quad j=1,2,\dots \quad (2.2)$$

де m – ціле число, яке визначає ступінь нелінійності пояснюючої змінної; так само достатньо просто будується модель нелінійного тренду за допомогою поліноміальної або нелінійної функції іншого типу;

- оцінювання параметрів моделей, нелінійних стосовно змінних (*квазілінійні моделі*), можна виконувати за допомогою звичайного методу найменших квадратів (*ordinary least squares (OLS)*) або інших методів лінійного оцінювання;

- для побудови квазілінійних моделей можна застосовувати такі ж схеми адаптивного оцінювання структури і параметрів, як і до лінійних, тобто застосовувати відомі методи структурно-параметричної адаптації з використанням аналогічних або подібних статистичних критеріїв адекватності моделей і якості оцінок прогнозів;

- більшість статистичних критеріїв адекватності лінійних моделей можна використовувати для аналізу адекватності нелінійних моделей, наприклад, коефіцієнт множинної детермінації, суму квадратів похибок моделі, інформаційний критерій Акайке, F – статистику Фішера та деякі інші;

- в цілому, для побудови нелінійних моделей можна застосувати ті ж принципи системного аналізу, які використовуються при побудові лінійних моделей, а саме такі: – принцип ієрархічності створення систем підтримки прийняття рішень (СППР) для побудови математичних

моделей на основі статистичних даних; – принципи адаптивності, оптимальності, багатомодельності, інформативності і т. ін.;

– перевірка адекватності нелінійної моделі виконується на навчальній і перевіірочній вибірках, як це має місце при аналізі адекватності лінійних моделей; також рекомендується застосувати побудовану модель до розв’язання основної задачі – прогнозування, імітаційного моделювання та/або керування (це може бути система підтримки прийняття рішень).

1. Аналіз процесу

Аналіз процесу – це важливий етап побудови моделі, коректне виконання якого потребує досвіду дослідження фактичних процесів різної природи. Ігнорування цього етапу призводить до неможливості побудови моделі заданого ступеня адекватності досліджуваного процесу або об’єкту. Аналіз процесу (або досліджуваного об’єкту) спрямовується на виконання завдань, поданих нижче.

– Встановлення кількості входів і виходів – тобто визначення розмірності об’єкта (процесу). При цьому особлива увага приділяється аналізу тих входів і виходів, які необхідні для подальшого формалізованого опису вибраних режимів функціонування об’єкта.

– Встановлення логічних зв’язків між змінними та аналіз можливостей їх математичного опису за допомогою рівнянь вибраного типу (тобто коректного об’єднання в одному математичному виразі). При цьому розглядаються можливості опису об’єкта за допомогою лінійних та нелінійних моделей.

– Попереднє встановлення можливості існування нелінійних зв’язків між змінними на основі такої інформації: – наявних статистичних/експериментальних даних стосовно фактичного функціонування досліджуваного процесу (графіки поведінки змінних); – дослідження функціонування процесу у різних режимах (наприклад, у перехідному та усталеному); – застосування спеціальних тестів на лінійність/нелінійність; – використання інформації із спеціальної літератури щодо особливостей функціонування вибраного процесу.

– Визначення *кількості зовнішніх збурень* та їх типу (детерміноване чи стохастичне), а також встановлення типу розподілу ймовірностей для стохастичного збурення включно із можливістю існування багатовимірною випадкового впливу. Наприклад, на протікання макроекономічних процесів (особливо у слабких економіках) впливає множина суттєвих випадкових впливів, які сумарно спричиняють

нелінійність протікання багатьох процесів та призводять до появи випадкових трендів [23, 51]. Випадкові тренди цікаві з точки зору моделювання, прогнозування, а також їх використання у системах імітаційного моделювання.

– Встановлення можливості декомпозиції процесу на окремі підпроцеси, які є простішими як з точки зору їх функціонування, так і з точки зору формального математичного опису; при цьому *деякі процеси можуть бути лінійними, а інші – нелінійними*. Загалом декомпозиція – це досить складний процес, який виконується за допомогою спеціальних математичних методів.

– Якщо процес має ієрархічну структуру (наприклад, у системі можна виділити верхній та нижній рівні функціонування або більше рівнів), то необхідно чітко розмежувати ці рівні, визначити функції кожного з них і встановити які типи зв'язків існують між ними. Так, технологічні процеси часто можна досить просто розмежувати на два та більше рівнів ієрархії. Відповідно до ієрархії функціонування процесів проєктується та реалізується ієрархічна система керування або система підтримки прийняття рішень.

– Використання знань із спеціальної літератури стосовно особливостей функціонування конкретного типу процесу, відомих законів та закономірностей його протікання, виявлення вже існуючих типів математичних моделей процесу та досвіду їх теоретичного чи експериментального дослідження.

– За наявності в спеціальній літературі вже розроблених моделей досліджуваного процесу необхідно встановити їх переваги та можливі недоліки, а також встановити можливість їх подальшого практичного використання (можливо після необхідної модифікації) для розв'язання задач прогнозування та/або керування. Аналіз і використання існуючих моделей надає можливість суттєво скоротити час та інші витрати на побудову і використання необхідної моделі.

Отримана в результаті аналізу процесу інформація максимально використовується для попереднього оцінювання структури моделі або кількох моделей-кандидатів, параметри яких оцінюють надалі за допомогою наявних статистичних/експериментальних даних. В процесі виконання аналізу функціонування досліджуваного процесу (об'єкта) доцільно використовувати та порівнювати інформацію, що надходить з різних джерел, оскільки попередні дослідження могли бути спрямовані на розв'язання різних задач

(прогнозування, керування або поглиблений аналіз функціонування за імітаційними моделями у заданих режимах). Як правило, інформаційні потоки, що надходять з різних джерел, є взаємно доповнюючими і їх потрібно коректно об'єднати та використати.

2. Попередня обробка даних

Процес попередньої обробки експериментальних (статистичних) даних, як правило, включає такі операції:

- нормування та візуальну перевірку даних і, при необхідності, їх корегування; нормування даних означає їх логарифмування або приведення до зручного діапазону їх зміни, наприклад, від 0 до 1; від -1 до $+1$; від $+10$ до -10 і т. ін.;
- корегування даних полягає у заповненні пропусків та зменшенні можливих викидів (екстремальних значень; outliers), що виходять за встановлений основний діапазон значень змінних;
- цифрова фільтрація вимірів з метою відсіювання непотрібних спектральних складових;
- оптимальна фільтрація вимірів з метою зменшення впливу похибок вимірів та випадкових зовнішніх збурень на значення змінних стану, оцінювання невимірюваних компонент вектора стану, заповнення пропусків, обробка екстремальних значень;
- формування перших або різниць вищих порядків, які необхідні для аналізу відповідних складових часового ряду.

Поширеним методом нормування даних є їх логарифмування з подальшим формуванням додаткових часових рядів (процесів) з перших чи других різниць. Нагадаємо, що перші різниці представляють собою наближений дискретний аналог першої похідної (швидкості), а другі різниці – другої похідної (прискорення). Часто із значень ряду вилучають його середнє для того щоб отримати можливість працювати з відхиленнями, а не повними значеннями змінних (при цьому для формалізованого опису досліджуваних процесів використовуються, як правило, моделі у просторі станів). Застосування того чи іншого методу нормування даних визначається в кожному випадку окремо, залежно від постановки задачі дослідження.

Обробка екстремальних (аномальних) значень. Хоча виявлення та обробка екстремальних значень – це велика окрема тема для дослідження, розглянемо деякі можливості щодо розв’язання цієї проблеми. В подальшому будемо вважати дані *аномальними*, якщо вони виникли внаслідок впливу значних похибок вимірів або похибок, пов’язаних з некоректним збором статистичних даних. Якщо можна встановити факт наявності аномальних даних, то їх просто видаляють з розподілу.

Екстремальні значення – це правильно виміряні (зібрані) дані, які характеризують фактичні раптові (стрибкоподібні або outliers – англ.) зміни процесу. Підхід до розв’язання задачі дослідження (корегування) екстремальних значень спостережень залежить від поставленої мети. Якщо дослідника цікавить тільки факт наявності таких значень (наприклад, з метою виявлення умов, що призводять до появи екстремальних значень), то досить мати надійний критерій для виявлення таких спостережень.

Якщо ж ставиться задача виявлення і виключення екстремальних значень (наприклад, з метою покращення статистичних оцінок параметрів і моделей), то виникає задача – як коректно виконати обробку таких даних. Спираючись на критерій для визначення наявності екстремальних значень, можна визначити величину зміщення оцінок параметрів, зумовлену відхиленням процесу від стаціонарної поведінки.

Критерії аналізу екстремальних значень застосовують з такою метою:

- вирівняти спостереження перед аналізом (як правило, суттєво зменшити) ті великі значення, які суттєво перевищують більшість вимірів;
- переконатись, що дані містять аномальні значення, що свідчить про необхідність перегляду процедури отримання даних;
- виділити спостереження, які є цікавими з точки зору їх аномальності та, по можливості, описати встановлений ефект математично.

Відомий підхід до виявлення аномальних спостережень полягає у тому, що вибіркові спостереження розглядають як випадкові, нормально розподілені величини. При цьому для аналізу (виявлення екстремальних значень) створюється статистика (статистичний тест), яка є чутливою до різких відхилень такого типу. Необхідно встановити розподіл цієї статистики за нульовою гіпотезою, що всі спостереження належать нормальній сукупності, а потім відхилити цю гіпотезу, якщо виявиться, що значення обчисленої статистики їй протирічить [50, 51].

Розглянемо можливий критерій відкидання екстремальних значень. Нехай дана деяка вибірка $\{x_1, x_2, \dots, x_N\}$, $N \geq 3$, яка, за припущенням, є випадковою для випадкової змінної X з нормальним розподілом: $\{X\} \sim (\mu_x, \sigma_x^2)$. Позначимо відхилення від середнього через

$$\tilde{x}_i = x_i - \bar{x}, \quad i = 1, 2, \dots, N, \quad \text{де} \quad \bar{x} = \frac{1}{N} \sum_{k=1}^N x(k). \quad (2.3)$$

Якщо виділити одне значення із наявних спостережень, то вибіркове середнє для спостережень, що залишились, можна визначити так:

$$\sum_{\substack{k=1 \\ k \neq i}}^N \frac{x_i}{N-1} = \bar{x} - \frac{\tilde{x}}{N-1} \quad (2.4)$$

Якщо виділити декілька значень $x_1, x_2, \dots, x_r$, то вибіркове середнє дорівнює:

$$\bar{x} - \frac{\tilde{x}_1 + \tilde{x}_2 + \dots + \tilde{x}_r}{N-r} \quad (2.5)$$

Позначимо максимальне відхилення через $\tilde{x}_m = x_m - \bar{x}$. Тепер правило визначення екстремального значення можна сформулювати так: *при заданому значенні c спостереження x_m відкидається, якщо $|\tilde{x}_m| > c S_x$, де S_x – вибіркове стандартне відхилення змінної X .*

Якщо вибірка має досить великий об'єм, то значення x_m видаляється і аналіз продовжується. Величина константи c може змінюватись із зміною довжини вибірки; вона зв'язана неявно з t – статистикою:

де $\nu = N - 1$; α – рівень значимості; ν_0 – будь-яке інше число додаткових ступенів свободи, яке зв’язане з оцінюванням σ_x^2 за вибіркою, об’єм якої не дорівнює N ($\nu_0 = 0$, якщо такої інформації немає). Також існує наближений вираз для оцінювання значення c через розподіл F :

$$c \approx \left(\frac{\nu}{N} \right)^{1/2} \left(\frac{3F_{1-q}}{1 + [3F_{1-q} - 1]/(\nu + \nu_0)} \right)^{1/2} \quad (2.7)$$

де $q = \Delta \hat{\sigma}_x^2 \frac{\nu}{N}$; $\Delta \hat{\sigma}_x^2$ – очікуваний приріст дисперсії внаслідок появи екстремальних значень. При використанні виразу (2.2.4) значення c визначається додатнім значенням квадратного кореня.

Виразом (2.2.4) можна скористатись таким чином: якщо з ряду значення не видалялись, то допустимий (очікуваний) відносний приріст дисперсії (“премію”) $\Delta \hat{\sigma}_x^2$ необхідно помножити на ν/N і, таким чином, отримаємо величину q . За його допомогою можна знайти відповідну верхню процентну точку для відношення дисперсій F_{1-q} при трьох і $\nu + \nu_0 - 1$ ступенях свободи. За виразом (2.2.4) обчислимо значення c і застосуємо критерій до x_m . Очікуваний відносний приріст дисперсії залежить від того, наскільки ймовірною є поява екстремальних значень, наприклад, можна прийняти невеликий відносний приріст $\Delta \hat{\sigma}_x^2 = 0,01 \div 0,03$.

Наприклад, якщо $N = 4$, $\nu = 3$ і $\nu/N = 0,75$, то при $\Delta \hat{\sigma}_x^2 = 0,02$ маємо: $q = 0,02 \cdot 0,75 = 0,05$. При 3-х степенях свободи $F_{1-q} = F_{1-0,05} = 9,28$. Тепер знайдемо значення c :

Спостереження x_m необхідно видалити, якщо $|\tilde{x}_m| > 0,831 S_x$. Можливі інші підходи до аналізу екстремальних значень [9].

Приклад 2.1. Обчислення критерію для виявлення екстремального значення. Нехай є ряд значень: $X = \{23,2 \ 23,4 \ 23,5 \ 24,1 \ 25,5\}$. Встановити, чи можна вважати значення 25,5 екстремальним і чи необхідно його видалити з вибірки?

Розв'язок. Обчислимо: $\bar{x} = 23,9$; $\tilde{x}_5 = 25,5 - 23,9 = 1,6$; $S_x = 0,77$. Для $\alpha = 0,05$; $\nu = 4$ і $N = 5$ за виразом (2.2.3) маємо:

$$\left(\frac{15c^2}{9 - 4c^2} \right)^{1/2} = 2,776^3 \quad (2.9)$$

Звідси, за методом проб та помилок знайдемо, що $c = 1,49$. Згідно з критерієм

$$|1,6| > 1,49 \cdot 0,77 = 1,05 \quad (2.10)$$

тобто спостереження x_5 видаляється як аномальне.

2.3 Деякі типи моделей нелінійних процесів

Моделі нелінійних процесів з теорії часових рядів

До самих відомих типів моделей нелінійних процесів в економіці та фінансах відносяться подані нижче.

– Нелінійна авторегресія, яка містить тільки функції залежної змінної і випадковий процес, який апроксимує вплив невимірюваних зовнішніх збурень, похибок вимірів та оцінок структури і параметрів моделі. Типовими функціональними елементами таких моделей є такі: синус або косинус, знакова функція, цілостепеневі функції, логарифми модулів та відношення поліномів низького порядку.

- Нелінійні передавальні функції, у яких використовуються функції затриманих у часі значень залежної (основної) змінної, поточні та затримані у часі значення незалежних (пояснюючих) змінних.
- Білінійні моделі типу $y(k) = \sum_{j,l} \beta_{jl} y(k-j)\varepsilon(k-l) +$ інші подібні члени, які містять добутки компонент $x(k)$ та затриманих залишків $\varepsilon(k-i)$, де i – допустимі значення лагів. Такі рівняння є частиною багатовимірної білінійної системи [44].
- Моделі з нелінійними ковзними середніми, які містять функції затриманих у часі залишків $\varepsilon(k), e(k)$ та інших можливих випадкових процесів.
- Подвійні стохастичні моделі (стохастичні змінні і стохастичні параметри), які містять перехресні добутки затриманих у часі змінних $y(k)$, а також поточними і затриманими у часі компонентами $x_i(k)$, або випадковими параметричними процесами [47, 48].

Більшість нелінійних моделей розширюються членами лінійної авторегресії. Однак аналіз таких змішаних моделей пов'язаний із значними труднощами. Внаслідок наявності цих труднощів значення лагів, як правило, вибираються невеликими. Загалом, задача оцінювання значень лагів для нелінійних моделей потребує окремого розгляду.

Приклади моделей, нелінійних стосовно параметрів:

- степенева функція:

$$y(k) = a [x(k)]^b \text{ і перетворена: } \log[y(k)] = \log a + b \log[x(k)]; \quad (2.11)$$

- показникова функція:

$$y(k) = a \cdot b^{x(k)} \text{ і перетворена: } \log[y(k)] = \log a + x(k) \cdot \log b; \quad (2.12)$$

- показниково-степенева функція:

$$y(k) = a \cdot [x(k)]^b \cdot c^{x(k)} \text{ і перетворена: } \log[y(k)] = \log a + b \log[x(k)] + x(k) \log c; \quad (2.13)$$

– екологічна функція:

$$y(k) = a \cdot \exp(-b^2 (x-c))^2 \text{ і перетворена:} \quad (2.14)$$

$$\log y = \log a - b^2 c^2 \log e + 2b^2 c \log(e)x - b^2 (\log e)x^2.$$

– логістична функція:

$$y = \frac{a}{1 + b e^{-cx}} \text{ і перетворена: } \log\left(\frac{a}{y} - 1\right) = \log b - cx \log e; \quad (2.15)$$

$$y = \frac{a}{1 + e^{b-cx}} \text{ і перетворена: } \log\left(\frac{a}{y} - 1\right) = b \log e - cx \log e;$$

– окремий випадок логістичної функції:

$$y = \frac{a}{1 + \left(\frac{b}{x}\right)^c} \text{ і перетворена: } \log\left(\frac{a}{y} - 1\right) = c \log b - c \log x; \quad (2.16)$$

– ірраціональна функція:

$$y = \sqrt{a + bx + cx^2} \text{ і перетворена: } y^2 = a + bx + cx^2; \quad (2.17)$$

– гіперболічна функція:

$$y = \frac{1}{a + bx} \text{ і перетворена: } \frac{1}{y} = a + bx; \quad (2.18)$$

– дробово-раціональна функція:

$$y = \frac{x}{a + bx + cx^2} \text{ і перетворена: } \frac{x}{y} = a + bx + cx^2; \quad (2.19)$$

– функція Джонсона:

$$\log y = -\frac{a}{b+x} + c \text{ і перетворена: } \frac{1}{(\log y) - c} = -\frac{b}{a} - \frac{1}{a}x; \quad (2.20)$$

- модифікована експоненціальна:

$$y = a e^{bx} \text{ і перетворена: } \log y = \log a + bx \log e; \quad (2.21)$$

- функція Торнквіста 1-го типу:

$$y = \frac{ax}{b+x} \text{ і перетворена: } \frac{1}{y} = \frac{b}{a} \frac{1}{x} + \frac{1}{a}; \quad (2.22)$$

- функція Торнквіста 2-го типу:

$$y = \frac{a(x-b)}{x+c} \text{ і перетворена: } \frac{x-b}{y} = \frac{1}{a} x + \frac{c}{a}; \quad (2.23)$$

- функція Торнквіста 3-го типу:

$$y = \frac{ax(x-b)}{x+c} \text{ і перетворена: } \frac{x-b}{y} = \frac{1}{a} + \frac{c}{a} \frac{1}{x}. \quad (2.24)$$

Таким чином, деякі функції можна лінеаризувати стосовно їх параметрів шляхом перетворення змінних та логарифмування. Параметри вихідних (початкових) функцій можна визначити шляхом зворотних перетворень. Наприклад, якщо вихідна функція показникова або степенева з дробовим показником, то оцінки параметрів цих рівнянь можна знайти шляхом потенціювання параметрів лінеаризованих залежностей. Така лінеаризація зв'язків дає можливість застосувати для оцінювання параметрів метод найменших квадратів. Однак, отримані таким чином оцінки параметрів можуть не відповідати статистичним властивостям оцінок МНК. Зазначимо, що згідно теорії оцінювання існують методи уточнення цих оцінок.

Наведені вище функції відіграють особливо значну роль в економетрії при вивченні попиту та розв'язуванні інших задач. З поданих вище функцій найбільші труднощі виникають при оцінювання параметра a логістичної функції, параметра c функції Джонсона і параметра b функції Торнквіста 2-го і 3-го типів. Оскільки параметр a вказує на рівень насичення,

то його можна визначити наперед, виходячи з логіко-економічних міркувань. Його можна визначити, також, за допомогою чисельних методів. Але у будь-якому випадку необхідно виконувати докладний економічний аналіз досліджуваного явища.

Моделі, створені за економічною теорією

Моделі такого типу використовуються, як правило, для опису конкретних типів фінансових та/або економічних процесів. Вони можуть враховувати можливу оптимізацію розвитку досліджуваних систем на основі вибраних функціоналів вартості або функцій корисності. Прикладом такої узагальненої моделі може бути, наприклад, така [28]:

$$y(k) = \min(\beta^T \mathbf{w}(k), \theta^T \mathbf{w}(k)) + \varepsilon(k) \quad (2.25)$$

де за оцінку основної змінної $\hat{y}(k)$ береться менше з двох значень, обчислених за альтернативними функціями: $\beta^T \mathbf{w}(k)$ або $\theta^T \mathbf{w}(k)$. Модель такого типу виникає з аналізу розбалансування деяких простих ринків.

Якщо у моделі (2.3.1) зробити заміну “min” на ще одну змінну $z(k-d)$, яка може входити у вектор $\mathbf{w}(k)$, але не може дорівнювати одиниці, то отримаємо ще одну форму моделі:

$$y(k) = \beta^T \mathbf{w}(k) + \theta^T \mathbf{w}(k) F(z(k-d)) + \varepsilon(k), \quad (2.26)$$

$$F(z(k-d)) = \begin{cases} 0, & \text{якщо } z(k-d) \leq c; \\ 1, & \text{якщо } z(k-d) > c; \end{cases}$$

де

c – деяке порогове значення, яке використовується для переключення з одного типу моделі на інший; $d = 0, 1, 2, \dots$ – дискретний час затримки по входу. Загалом, регресію (2.3.2) називають моделлю з переключенням.

Скалярну версію цієї моделі (для одновимірного ряду) називають пороговою авторегресійною моделлю з двома режимами. Її можна узагальнити (розширити) на множину можливих режимів функціонування досліджуваного процесу, шляхом введення до розгляду функції такого вигляду:

$$F(z(k-d)) = \frac{1}{1 + \exp[-\gamma(z(k-d)-c)]}, \quad \gamma > 0 \quad (2.27)$$

Така модель отримала назву логістичної регресії з гладким переходом (ЛРГП) або logistic smooth transition regression (LSTR) у англomовному варіанті.

Функція F може бути також використана у формі функції щільності ймовірностей, а не кумулятивної функції розподілу (КФР). В одновимірному випадку це буде відповідати експоненційній авторегресійній моделі з гладким переходом (ЕАРГП) або ж exponential smooth transition autoregressive (ESTAR) model у англomовному варіанті.

Перехідна змінна може представляти зміну поточної політики або політичного режиму, значне швидке зростання інфляції, зміну напряму руху бізнесового циклу, різку зміну зовнішніх впливів (наприклад, виникнення локальної війни) і т. ін. Теоретично, моделі з переключенням режимів такого типу або ж їх згадані версії з гладкими переходами можуть використовуватись досить часто, наприклад, при переході від сповільненого до прискореного темпів виробництва. При цьому модель з гладким переходом може виявитись чутливішою до агрегування ефектів ніж модель з різким переходом типу [46].

Деякі відомі нелінійні моделі запропоновані в результаті досліджень поведінки економетричних часових рядів. Прикладом є нелінійна авторегресія для валового внутрішнього продукту (ВВП) і податкових надходжень у вигляді:

$$\begin{aligned} y_1(k) &= a_0 + a_1 y_1(k-1) + b_{12} \exp(y_2(k)) + a_2 x_1(k)x_2(k) + \varepsilon_1(k), \\ y_2(k) &= c_0 + c_1 y_2(k-1) + b_{21} \exp(y_1(k)) + c_2 x_1(k)x_2(k) + \varepsilon_2(k) \end{aligned} \quad (2.28)$$

де $y_1(k)$ – логарифм валового внутрішнього продукту; $y_2(k)$ – логарифм податкових надходжень; $x_1(k)$ – внутрішні інвестиції; $x_2(k)$ – зовнішні інвестиції. Моделі такого типу, що виникають з економічної теорії, не обов'язково кращі ніж ті, що будуються повністю на основі статистичних даних, але вони корисні для порівняльного аналізу остаточних результатів (після застосування до прогнозування та/або керування).

Деякі інші приклади моделей квазілінійного типу

Поліноміальна модель (поліном стосовно незалежної змінної) має такий вигляд:

$$y(k) = b_0 + b_1 x(k) + b_2 x^2(k) + \dots + b_m x^m(k) + \varepsilon(k) \quad (2.29)$$

де $\theta = [b_0 \ b_1 \ b_2 \ \dots \ b_m]^T$ – вектор параметрів моделі; m – порядок полінома; $\varepsilon(k)$ – випадковий процес, поява якого у правій частині рівняння зумовлена наявністю зовнішніх випадкових збурень (наявні завжди), похибками структури і параметрів моделі, похибками вимірів та обчислень; k – дискретний час, зв'язаний з неперервним реальним часом t періодом дискретизації вимірів: $t = k \cdot T_s$;

– поліноміальна модель з логарифмованими змінними:

$$\log[y(k)] = b_0 + b_1 \log[x(k)] + b_2 \log[x(k)]^2 + \dots + b_m \log[x(k)]^m + \varepsilon(k) \quad (2.30)$$

До квазілінійних відноситься також *білінійна модель* такого вигляду:

$$y(k) = b_0 + b_1 x_1(k)x_2(k) + b_2 x_1(k)x_3(k) + b_3 x_2(k)x_3(k) + \varepsilon(k) \quad (2.31)$$

а також узагальнена білінійна модель:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j v(k-j) + \sum_{i=1}^m \sum_{j=1}^s c_{i,j} y(k-i)v(k-j) + \varepsilon(k) \quad (2.32)$$

де p, q, m і s – додатні цілі числа. Умови стаціонарності для цієї моделі можна визначити після її представлення у просторі станів (перехідне рівняння у просторі станів) та подання станів у формі добутку минулих інновацій та векторів випадкових коефіцієнтів [50, 51].

Окремим узагальненням білінійної моделі є можливий варіант

рівняння для умовної гетероскедастичності, наприклад таке:

$$y(k) = \mu + \sum_{i=1}^s \beta_i \varepsilon^2(k) + v(k) \quad (2.33)$$

де $\varepsilon(k)$ – послідовність білого шуму (залишки авторегресійної моделі низького порядку, наприклад, першого або другого). Перші два умовні моменти для поданої моделі обчислюються за виразами:

$$E[y(k) | F(k-1)] = \mu;$$

$$Var[y(k) | F(k-1)] = \left(1 + \sum_{i=1}^s \beta_i \varepsilon(k-i) \right)^2 \cdot \sigma_v^2, \quad (2.34)$$

де $F(k-1)$ – інформація про процес на момент часу $(k-1)$, включно.

Приклад 2.2. Розглянемо задачу побудови моделі на основі щомісячних статистичних даних стосовно доходності $R(k)$ деяких акцій на біржі, всього 300 спостережень. Вибіркова ЧАКФ цього ряду даних свідчить про наявність значущих автокореляцій для лагів 1 і 3, а тому будуємо модель АР(3). Для квадратів залишків цього рівняння існують значущі автокореляції для лагів 1, 3. Таким чином, необхідно побудувати спеціальну білінійну модель такого виду:

$$R(k) = \mu + a_1 R(k-1) + a_3 R(k-3) + (1 + \beta_1 v(k-1) + \beta_3 v(k-3)) \cdot \varepsilon(k) \quad (2.35)$$

де $\{v(k)\}$ – процес ковзного середнього; $\{\varepsilon(k)\}$ – нормально розподілений процес.

Припускаючи, що умовний розподіл випадкового процесу $\varepsilon(k)$ нормальний, для оцінювання параметрів поданої вище моделі скористаємось умовним методом максимальної правдоподібності:

Знайдені параметри адекватності цієї моделі такі: $R^2=0,85$; $DW = 1,92$, а середня абсолютна похибка однокрокових прогнозів у процентах на навчальній вибірці склала: $MAPE = 5,2\%$.

Перевірка оцінок параметрів моделі на статистичну значущість свідчить, що всі оцінки є значущими на рівні значущості 5%. З останнього рівняння можна знайти вираз (модель) для випадкового процесу:

$$\hat{\varepsilon}(k) = \frac{R(k) - 0,0117 - 0,173R(k-1) + 0,089R(k-3)}{0,077 \cdot (1,0 + 0,383v(k-1) - 0,651v(k-3))} \quad (2.37)$$

де $\hat{\varepsilon}(k) = 0$, при $k \leq 3$, як для ряду значень залишків цієї конкретної моделі. За цим рівнянням розраховано значення ряду $\hat{\varepsilon}(k)$. Вибіркова автокореляційна функція для ряду значень $\hat{\varepsilon}(k)$ свідчить про відсутність значущої автокореляції, але цей ряд не можна вважати неінформативним, оскільки АКФ для ряду квадратів залишків $\hat{\varepsilon}^2(k)$ містить статистично значущі величини.

Для порівняння побудовано другий варіант моделі: модель AP(3) для самого процесу:

$$R(k) = 0,0119 + 0,231R(k-1) + 0,0073R(k-2) - 0,0194R(k-3) + \varepsilon(k) \quad (2.38)$$

яка має такі характеристики: $R^2=0,74$; $DW = 1,83$, а середня абсолютна похибка у процентах на навчальній вибірці склала: $MAPE = 6,1\%$.

Узагальнена авторегресія з умовною гетероскедастичністю УАРУГ(1, 3) для динаміки умовної дисперсії цього процесу має вигляд:

$$h^2(k) = 0,0025 + 0,297h^2(k-1) + 0,193\hat{\varepsilon}^2(k-1) + 0,315\hat{\varepsilon}^2(k-2) + 0,204\hat{\varepsilon}^2(k-3) \quad (2.39)$$

де $h^2(k)$ – умовна дисперсія ряду значень процесу $R(k)$, обчислена за ковзним вікном. Модель УАРУГ має такі характеристики адекватності: $R^2=0,56$; $DW=1,19$, а середня абсолютна похибка у процентах для одно-крокових прогнозів на навчальній вибірці склала: $MAPE=12,7\%$. Остання модель описує процес $R(k)$ повніше, оскільки модель динаміки дисперсії доповнює модель динаміки самого процесу.

Гнучкі статистичні параметричні моделі

Існують досить ефективні процедури (які дають можливість будувати адекватні моделі) моделювання нелінійних процесів, які ґрунтуються на математичних моделях, що містять лінійну та нелінійну складові:

$$y(k) = \beta^T \mathbf{z}(k) + \sum_{i=1}^p \alpha_i \varphi_i(\theta_i^T \mathbf{z}(k)) + \varepsilon(k) \quad (2.40)$$

де $\mathbf{z}(k)$ – вектор минулих значень основної змінної $y(k)$, а також минулих і поточних значень вектора пояснюючих змінних $\mathbf{x}(k)$ плюс стала величина. Перша складова цієї моделі лінійна, а $\varphi_i(x)$ – множина функцій, прикладами яких можуть бути такі:

- степенева функція $\varphi_i(x) \equiv x^i$, де змінна x – може бути затриманим у часі значенням y або іншою змінною;
- тригонометрична функція $\varphi_i(x) = \sin x$ або $\varphi_i(x) = \cos x$; рівняння (2.3.1) можна розширити квадратичною функцією $\mathbf{z}^T(k) \mathbf{A} \mathbf{z}(k)$, що приводить до гнучкої функціональної форми [25];
- $\varphi_i(x) = \varphi(x)$, $\forall i$, де $\varphi(x)$ – функція зв'язку, наприклад, функція щільності ймовірності або логістична функція вигляду:

— $\varphi(x)$ – також може бути непараметричною функцією.

Для аналізу перших трьох типів моделей існують теореми, у яких доводиться, що будь-яка гладка функція може бути апроксимована з достатньою точністю шляхом вибору належного значення параметра P у рівнянні (2.3.1). Цей параметр можна оцінити за допомогою оптимізаційних процедур з використанням належним чином сформованого критерію. Моделі такого типу часто використовують на практиці, оскільки вони забезпечують прийнятну адекватність досліджуваним процесам при невисокій складності. Вони також утворюють природну альтернативу непараметричним та напівпараметричним моделям.

Непараметричні моделі надають можливість створення функціонального опису процесів у кожній точці простору визначення пояснюючих змінних шляхом вибору деякої згладжуючої, а не явно заданої параметричної функції. Разом з тим, варто підкреслити, що відмінність між параметричними і непараметричними описами не є занадто помітною, про що свідчить застосування сплайнів або нейронних мереж з невизначеним порогом відсікання [46]. Загалом, це спрощує процес математичного моделювання і прогнозування.

Моделі із змінними у часі параметрами та залежністю від стану

Існує досить загальний клас нелінійних моделей у такій формі:

$$y(k) = \sum_{j=1}^p \varphi_j(\mathbf{x}(k-1)) y(k-j) + \mu(\mathbf{x}(k-1)) + \varepsilon(k) \quad (2.42)$$

де $y(k)$ – стохастичний вектор залежних величин розмірності $[n \times 1]$;
 $\mathbf{x}(k) = [y(k), y(k-1), \dots, y(k-n+1)]$ – вектор змінних стану, динаміка якого описується моделлю:

$$\mathbf{x}(k) = h(\mathbf{x}(k-1)) + \mathbf{F}(\mathbf{x}(k-1)) \mathbf{x}(k-1) + v(k) \quad (2.43)$$

Рівняння (2.42) може включати, також, члени з ковзним середнім. Тобто

у цьому випадку для опису досліджуваного процесу фактично використовуються дві моделі одночасно, що ускладнює процедуру формування структури моделі (2.42).

Рівняння (2.43) – це модель марковського типу у просторі станів, яку можна доповнити рівнянням вимірів, тобто загальна модель буде складатись з двох векторно-матричних рівнянь. Елементи матриці $\mathbf{F}(\cdot)$ – це можуть бути лінійні функції або поліноми низького порядку. Модель (2.43) може містити члени, які свідчать про наявність довгої пам'яті досліджуваного процесу, що досить часто має місце при дослідженні фінансово-економічних та екологічних процесів.

Непараметричне моделювання не накладає обмежень на використання моделей конкретного класу, тобто воно надає більше свободи при виборі складових (функціональних членів) для формування моделі досліджуваного процесу у цілому. Це дає можливість створювати моделі з елементів, які найкраще придатні для опису лінійних та нелінійних складових і таким чином приводити моделі до необхідного ступеня адекватності [33, 34].

Також непараметричне моделювання часових рядів *не вимагає подання моделей у явному вигляді*, хоча загалом модель такого класу може бути подана так:

$$y(k) = f(y(k-1), x(k-1)) + g(y(k-1), x(k-1)) \varepsilon(k) \quad (2.44)$$

Де

$$\begin{aligned} \mathbf{y}(k-1) &= [y(k-i_1), \dots, y(k-i_p)]^T \text{ та} \\ \mathbf{x}(k-1) &= [x(k-j_1), \dots, x(k-j_p)]^T \end{aligned} \quad (2.45)$$

вектори затриманих у часі значень відповідних змінних; $\{\varepsilon(k)\}$ – послідовність мартингальних різниць по відношенню до наявної інформації $\mathbf{I}(k) = \{y(k-i), i>0; x(k-i), i>0\}$. Як правило, припускається, що спільний процес $\{y(k), x(k)\}$ стаціонарний та існує залежність типу (2.3.4).

Структура такої моделі може бути узагальнена на багатовимірний випадок та

належне трансформування незалежних змінних.

Представником непараметричної моделі такого типу може бути *авторегресія з функціональними коефіцієнтами*:

$$y(k) = a_1(y(k-d))y(k-1) + \dots + a_p(y(k-d))y(k-p) + \sigma(y(k-d))\varepsilon(k) \quad (2.46)$$

де $a_1(\cdot), \dots, a_p(\cdot)$ – функції, що визначають коефіцієнти; $y(k-d)$ – змінна, яка визначає функціональну залежність коефіцієнта моделі (від її змінних); $\sigma(y(k-d))$ – ваговий коефіцієнт для випадкової змінної; $\{\varepsilon(k)\}$ – випадковий процес, що впливає на поведінку основної змінної. Модель у цілому називають функціональною авторегресією (ФАР).

Модель ФАР, яка включає експоненту в опис функціонального коефіцієнта, називають ЕКСАР (EXPAR) [24, 40]:

$$y(k) = \sum_{i=1}^p \{ \alpha_i + (\beta_i + \gamma_i y(k-d)) \exp(-\theta_i y^2(k-d)) \} + \varepsilon(k) \quad (2.47)$$

де $\theta_i \geq 0, i=1, \dots, p$. Очевидно, що коефіцієнти моделі ФАР можуть бути описані функціями інших видів.

В аналізі економетричних часових рядів як *непараметричну модель* часто використовують умовне математичне сподівання

$$\begin{aligned} f(\mathbf{y}, \mathbf{x}) &= \mathbf{M}(\mathbf{y}; \mathbf{x}) = \mathbf{M}(y_1, \dots, y_p; x_1, \dots, x_q) = \\ &= E[y(k) | y(k-i_1)=y_1, \dots, y(k-i_p)=y_p; x(k-i_1)=x_1, \dots, x(k-i_q)=x_q] \end{aligned} \quad (2.48)$$

і умовну дисперсію

$$\begin{aligned} \mathbf{V}(\mathbf{y}; \mathbf{x}) &= \mathbf{V}(y_1, \dots, y_p; x_1, \dots, x_q) = \\ &= Var[y(k) | y(k-i_1)=y_1, \dots, y(k-i_p)=y_p; x(k-i_1)=x_1, \dots, x(k-i_q)=x_q] \end{aligned} \quad (2.49)$$

Умовне середнє для $y(k)$ можна обчислити за допомогою за допомогою відповідного варіанту методу найменших квадратів за умови, що наявні виміри $y(k - i_1), \dots, y(k - i_p); x(k - i_1), \dots, x(k - i_q)$. Похідні від $M(y, x)$ також можна оцінити за непараметричним методом і надати їм економічну інтерпретацію. Умовну дисперсію можна використати для дослідження волатильності процесу:

$$V(y, x) = \sigma^2 g(y, x) \quad (2.50)$$

де $\sigma^2 = E[\varepsilon^2(k)]$; часто припускають, що $g(y, x) \equiv 1$.

Перевагою застосування непараметричного підходу до моделювання є його висока гнучкість, тобто немає обмеження на використання конкретного класу моделей, що має місце у випадку параметричних моделей. Однак, для отримання оцінок коефіцієнтів з необхідною ступінню варіабельності необхідно мати достатню кількість спостережень.

Труднощі із застосуванням непараметричного моделювання виникають у багатовимірному випадку, вони пов'язані із «прокляттям» розмірності. Збільшення кількості регресорів та значень лагів призводить до зменшення кількості спостережень в одиниці об'єму у просторі регресорів, що утруднює обчислення значущих непараметричних оцінок. Для подолання цих труднощів існують спеціальні методи.

Іншим методом подолання труднощів, пов'язаних із високою розмірністю, є використання напівпараметричних моделей. У таких моделях стосовно частини змінних припускається лінійна параметрична залежність, а до опису решти змінних застосовується непараметрична функціональна залежність.

2.4 Аналіз наявності нелінійності

Однією з проблем при оцінюванні структури моделі є встановлення факту наявності нелінійностей в досліджуваному процесі та їх типу. Встановлення факту наявності нелінійності досліджуваного процесу вимагає застосування до його формального опису нелінійної моделі. Для розв'язання цієї задачі обов'язково використовують *візуальний аналіз* даних та *формальні тести* на наявність нелінійностей (або тести на лінійність).

Досвідченому фахівцю з моделювання візуальний аналіз дає можливість оперативно виявити наявність участків з лінійним або нелінійним трендом, в якійсь мірі ідентифікувати наявність гетероскедастичності та значних викидів (імпульсів або екстремальних значень), які можуть суттєво впливати на адекватність моделі.

Необхідно зазначити, що існують окремі навчальні курси з візуального аналізу даних. Це свідчить про те, що не варто нехтувати такою доступною, але достатньо ефективною можливістю прискореного попереднього дослідження даних.



Рис. 2.1 – Методи ідентифікації нелінійності в даних

Існує також ряд формальних тестів на наявність нелінійності. Деякі методи подані на рис. 2.1. Ідея тесту, як правило, полягає у тому, що на основі даних будується рівняння прямої і обчислюється сума квадратів відхилень фактичних даних від цієї прямої.

Далі ця сума квадратів відхилень нормується за допомогою відповідно вибраної величини і отримане значення використовується як тестова статистика з метою формального підтвердження/відхилення гіпотези стосовно наявності нелінійності.

F-тест:

Розглянемо *F*-тест для встановлення факту наявності нелінійності. Цей тест можна застосувати у випадку, коли вибірку спостережень для одного і того ж процесу можна розділити на кілька груп:

$$\bar{F} = \frac{\frac{1}{m-2} \sum_{i=1}^m n_i (\bar{y}_i - \bar{y}_i^{\text{лінійний}})^2}{\frac{1}{n-m} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2} \quad (2.51)$$

де $\bar{y}_i$ – середнє значення для i –ї групи (вибірки або групи) даних; $\bar{y}_i^{\text{лінійний}}$ – середнє для лінійної апроксимації даних; m – кількість груп даних; n_i – кількість вимірів в i – й групі; n – загальна кількість наявних вимірів. Фактично, подана вище статистика представляє собою таке відношення:

$$\bar{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{СКВ фактичних значень } y(k) \text{ від групових середніх}} \quad (2.52)$$

де *СКВ* – сума квадратів відхилень. Пряма регресії будується на основі наявних статистичних даних. Нульова гіпотеза формулюється так: H_0 : процес лінійний.

Якщо статистика $\bar{F}$ з $v_1 = m - 2$ та $v_2 = n - m$ степенями свободи досягає або перевищує рівень значимості, то гіпотезу щодо лінійності необхідно відхилити.

Тестування з використанням методу множників Лагранжа

Оскільки оцінювання структури і параметрів нелінійних моделей – це складніша задача ніж для лінійних моделей, то природно використовувати методи, які не потребують оцінювання нелінійних структур. У випадках, коли модель не ідентифікується при формулюванні нульової гіпотези стосовно лінійності, то тести, які ґрунтуються на оцінюванні деякої нелінійної альтернативи, як правило, не існують. Тому результат використання методу

множників Лагранжа виявляється корисним для створення тестів на лінійність. Але крім цього методу також застосовуються інші можливості.

Розглянемо нелінійну модель такого вигляду:

$$y(k) = \beta^T \mathbf{w}(k) + f(\theta, \mathbf{w}(k), \mathbf{v}(k)) + u(k) \quad (2.53)$$

де $\mathbf{w}(k) = [1, y(k-1), \dots, y(k-p); x_1(k), \dots, x_l(k)]^T$ – вектор вимірів основної змінної та регресорів; $\mathbf{v}(k) = [u(k-1), \dots, u(k-q)]^T$ – вектор випадкових величин; $u(k) = g(\beta, \theta, \mathbf{w}(k), \mathbf{v}(k))\varepsilon(k)$; $\varepsilon(k)$ – мартингальний різницевий процес з такими характеристиками: $E[\varepsilon(k) | \mathbf{I}(k)] = 0$, $\text{cov}[\varepsilon(k) | \mathbf{I}(k)] = \sigma_\varepsilon^2$; $\mathbf{I}(k) = \{y(k-j), j > 0; \mathbf{x}(k-i), i \geq 0\}$ – наявна інформація у формі вимірів. У процесі побудови моделі необхідно знайти таку апроксимуючу функцію $f(\mathbf{w}(k))$, що $E[y(k) | \mathbf{I}(k)] = \beta^T \mathbf{w}(k) + f(\theta, \mathbf{w}(k), \mathbf{v}(k))$.

Таким чином, на основі цих припущень можна записати, що

$$E[u(k) | \mathbf{I}(k)] = 0 \quad \text{і} \quad \text{cov}[u(k) | \mathbf{I}(k)] = \sigma_\varepsilon^2 g^2(\beta, \theta, \mathbf{w}(k), \mathbf{v}(k)). \quad (2.54)$$

Припустимо, що функція $f(\cdot)$ є щонайменше двічі диференційованою стосовно параметрів $\theta = [\theta_1, \dots, \theta_m]^T$ і що $f(0, \mathbf{w}(k), \mathbf{v}(k)) = 0$. Таким чином, гіпотезу стосовно лінійності можна сформулювати так:

$$H_0: \theta = 0 \quad (2.55)$$

Також розглянемо випадок, коли $g \equiv 1$, що приводить до рівності $\varepsilon(k) = u(k)$. Для того щоб перевірити гіпотезу стосовно лінійності, запишемо умовну (псевдо-) логарифмічну функцію правдоподібності такого вигляду:

Відповідний скоринговий вектор, нормований величиною $1/\sqrt{N}$ має такий вигляд [46]:

$$\frac{1}{\sqrt{N}} \frac{\partial L}{\partial \theta} = \frac{1}{\sqrt{N} \sigma_u^2} \sum_{k=1}^N u(k) \mathbf{h}(k); \quad \mathbf{h}(k) = \left(\frac{\partial u(1)}{\partial \theta_1}, \dots, \frac{\partial u(k)}{\partial \theta_m} \right) \quad (2.57)$$

Ця величина буде ненульовою при прийнятій нульовій гіпотезі. Інформаційна матриця має блок-діагональний вигляд, при цьому кожний діагональний елемент, що множиться на σ_u^2 , утворює окремий блок. Таким чином, обернена величина для кожного блока, пов'язана з θ і оцінена стосовно H_0 , має вигляд:

$$\mathbf{I}_\theta = (N \sigma_u^2)^{-1} \sum_{k=1}^N \tilde{\mathbf{h}}(k) \tilde{\mathbf{h}}^T(k) - \sum_{k=1}^N \tilde{\mathbf{h}}(k) \mathbf{w}^T(k) \left(\sum_{k=1}^N \mathbf{w}(k) \mathbf{w}^T(k) \right)^{-1} \sum_{k=1}^N \mathbf{w}(k) \tilde{\mathbf{h}}^T(k) \quad (2.58)$$

де $\tilde{\mathbf{h}}(k)$ – це $\mathbf{h}(k)$, оцінене при сформульованій H_0 . Покладаючи $\tilde{\mathbf{u}} = [\tilde{u}(1), \dots, \tilde{u}(N)]^T$, отримаємо вираз для тестової статистики у вигляді:

$$LM = \tilde{\sigma}^{-2} \tilde{\mathbf{u}}^T \mathbf{H} (\mathbf{H}^T \mathbf{M}_w \mathbf{H})^{-1} \mathbf{H}^T \tilde{\mathbf{u}} \quad (2.59)$$

де $\mathbf{M}_w = \mathbf{I} - \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T$; $\tilde{\sigma}^2 = N^{-1} \sum_{k=1}^N \tilde{u}^2(k)$; елементами вектора $\tilde{\mathbf{u}}$ є залишки рівняння (2.4.1), консистентно оціненого при сформульованій H_0 і при $g \equiv 1$. Величина, розрахована за виразом (2.4.2), асимптотично має розподіл $\chi^2(m)$ при виконанні гіпотези H_0 .

На практиці тестова статистика часто розраховується з використанням оцінок параметрів, отриманих за звичайним МНК, таким чином:

- побудувати лінійну регресію основної змінної $y(k)$ на $\mathbf{w}(k)$:

$$y(k) = \beta^T \mathbf{w}(k) + u(k),$$
- обчислити залишки цієї моделі $\tilde{u}(k)$ і суму квадратів залишків SSR_0 ;
- побудувати регресію $\tilde{u}(k)$ на $\mathbf{w}(k)$, обчислити суму квадратів залишків цієї моделі SSR_1 ;
- обчислити тестову статистику

$$F_S(m, N - n - m) = \frac{(SSR_0 - SSR_1)/m}{SSR_1/(N - n - m)} \quad (2.60)$$

де $n = l + p + 1$; m – розмірність вектора θ ; обчислена величина має F – розподіл при $\theta = 0$. Використання F_S – статистики замість тесту χ^2 рекомендується при коротких вибірках [29].

Нелінійні кореляційні функції

Наявність нелінійності можна встановити також за допомогою вибірових *нелінійних кореляційних функцій* (НКФ), тобто кореляційних функцій, розрахованих за вибірками експериментальних (статистичних) даних. Наприклад, якщо дискретна НКФ

$$r_{yx^2}(s) = r_{y(k)x^2(k-s)} = \frac{1}{N} \frac{\sum_{k=s+1}^N \{ [y(k) - \bar{y}] [x(k-s) - \bar{x}]^2 \}}{\sigma_y \sigma_x^2}, \quad s = 0, 1, 2, 3, \dots, \quad (2.61)$$

містить значення, які суттєво відрізняються від нуля в статистичному сенсі, то процес містить квадратичну нелінійність відносно регресора x .

У багатьох випадках основною ідеєю, яка використовується при перевірці даних на лінійність, є така: якщо за нульову гіпотезу приймається, що процес лінійний, то залишки коректно побудованої лінійної моделі повинні бути некорельованими (незалежними). Порушення умови незалежності залишків свідчить про неадекватність побудованої моделі та невиконання гіпотези стосовно лінійності.

Тестування бікореляцій та крос-бікореляцій

Розглянемо два часових ряди $\{x(k)\}$, $\{y(k)\}$ потужністю N значень і нормованих таким чином:

$$x_n(k) = \frac{x(k) - \bar{x}}{\sigma_x}, \quad (2.62)$$

$$\text{де } \bar{x} = \frac{1}{N} \sum_{k=1}^N x(k) \quad \text{— вибіркве середнє; } \sigma_x^2 = \frac{1}{N-1} \sum_{k=1}^N [x(k) - \bar{x}]^2 \quad \text{—}$$

вибіркова дисперсія. Отримані після нормування ряди $\{x_n(k)\}$, $\{y_n(k)\}$ будуть мати нульове середню та одиничну дисперсію. Якщо виконується аналіз високочастотних (зареєстрованих з коротким періодом дискретизації вимірів) даних на коротких вибірках, то припущення стосовно стаціонарності не відіграє суттєвої ролі [51].

Якщо при тестуванні одновимірних процесів формулюється нульова гіпотеза, що ряд $\{x(k)\}$ — чисто випадковий шумовий процес, то не повинні існувати значущі бікореляції, обчислені за виразом:

$$E [x(k) x(k+r) x(k+s)] = 0 \quad (2.63)$$

де r, s — значення додаткових часових зсувів. Узагальнення цього виразу на розрахунок крос-бікореляції відбувається в результаті заміни $x(k+r)$ або $x(k+s)$ на іншу змінну, наприклад, на $y(k+s)$:

Можна стверджувати (тут без доведення), що статистичні тести (статистики) на основі бікореляції та крос-бікореляції можна записати у вигляді:

$$\begin{aligned} H_{xxx}(N) &= \sum_{s=-L}^L * \sum_{r=1}^L (N-m) C_{xxx}^2(r, s), \quad (*-s \neq 0) \\ H_{xxy}(N) &= \sum_{s=-L}^L * \sum_{r=1}^L (N-m) C_{xxy}^2(r, s), \quad (*-s \neq 0) \end{aligned} \quad (2.65)$$

Де

$$\begin{aligned} C_{xxx}(r, s) &= \frac{1}{N-m} \sum_{k=1}^{N-m} x(k) x(k+r) x(k+s) \\ C_{xxy}(r, s) &= \frac{1}{N-m} \sum_{k=1}^{N-m} x(k) x(k+r) y(k+s) \\ m &= \max(r, s), \quad L = N^c, \quad 0 < c < 0,5 \end{aligned} \quad (2.66)$$

Тут c – параметр, який вибирається користувачем. У роботах [33, 34] на основі моделювання за методом Монте-Карло вибрано значення $c = 0,4$. Загалом, вибір значення цієї константи має максимізувати потужність тесту, але одночасно забезпечувати належні апроксимаційні характеристики тестової статистики.

У роботі [33] доведено, що величина H_{xxx} асимптотично має розподіл χ^2 з кількістю ступенів свободи, що дорівнює числу квадратичних членів у сумі. Аналогічні аргументи застосовуються до доведення асимптотичних властивостей тестової крос-бікореляційної статистики стосовно розподілу χ^2 . Число ступенів свободи залишається у цьому випадку таким же [16].

Бікореляцію можна трактувати як кореляцію між поточними значеннями досліджуваного процесу та попередніми значеннями коефіцієнтів автокореляції. Значення крос-кореляції можна інтерпретувати як кореляцію між значеннями

одного фінансового ряду даних (x або y) і часовою крос-кореляцією (із зсувом вперед) між двома рядами ($x(k+r)$ $y(k+s)$). У роботі [16] бікореляційний тест застосовується до залишків авторегресійних моделей, що будуються для опису досліджуваних фінансових процесів, а крос-бікореляційний тест застосовується до залишків моделей у формі векторної авторегресії. На етапі попередньої фільтрації даних забезпечується вилучення з часових рядів даних лінійної та взаємної залежності відповідно.

Прогнозування з використанням крос-бікореляцій

Основною перевагою тесту на наявність нелінійності, що ґрунтується на обчисленні бікореляцій та крос-бікореляцій, є те, що він достатньо загальний і може виявити різні типи нелінійності в умовному середньому (виявлення наявності залежності третього порядку). На основі цього тесту можна вибрати також належну функціональну форму для побудови нелінійного рівняння з метою обчислення оцінки прогнозу.

Перепишемо рівності (2.4.1) і (2.4.2) так, щоб вони містили тільки лаги спостережуваних змінних. В результаті отримаємо такі вирази для бікореляції та крос-бікореляції:

$$\begin{aligned} E [x(k) x(k-r) x(k-s)] &= 0, \\ E [x(k) x(k-r) y(k-s)] &= 0. \end{aligned} \quad (2.67)$$

Тепер, якщо значення $x(k)$, стосовно якого буде обчислюватись оцінка прогнозу, відоме, то будуть відомими також значення $x(k-r)$, $x(k-s)$ і $y(k-s)$. Тобто комбінації цих величин можна використовувати у лінійній регресії для майбутніх значень досліджуваного процесу $x(k)$. Якщо максимально допустиме значення лагу авторегресії позначити через P , то модель для обчислення прогнозів можна подати у вигляді:

$$x(k) = \alpha_0 + \sum_{i=1}^P \alpha_i x(k-i) + \sum_{r=1}^P \sum_{s=1}^P \beta_{rs} x(k-r)x(k-s) + \varepsilon(k) \quad (2.68)$$

для одновимірного випадку, а для багатовимірного випадку можна скористатись системою рівнянь у вигляді векторної авторегресії (VAR):

$$x(k) = \alpha_{10} + \sum_{i=1}^p \alpha_{1i} x(k-i) + \sum_{j=1}^p \gamma_{1j} y(k-j) + \sum_{r=1}^p \sum_{s=1}^p \beta_{1rs} x(k-r)x(k-s) + \varepsilon_1(k)$$

$$y(k) = \alpha_{20} + \sum_{i=1}^p \alpha_{2i} x(k-i) + \sum_{j=1}^p \gamma_{2j} y(k-j) + \sum_{r=1}^p \sum_{s=1}^p \beta_{2rs} x(k-r)x(k-s) + \varepsilon_2(k)$$

де $\varepsilon_1(k)$ і $\varepsilon_2(k)$ – незалежні, однаково розподілені випадкові величини; $\alpha_0, \alpha, \gamma, \beta$ – параметри регресії. Оскільки ці рівняння лінійні стосовно параметрів (але містять мультиплікативні комбінації змінних), їх можна оцінювати за допомогою звичайного МНК.

Значення параметра зсуву P підбираються експериментально у процесі аналізу конкретних даних. У роботі [16] використано значення $P = 2$ для досліджених процесів. За побудованими моделями оцінено прогнози на 6 кроків, але вже після третього кроку за всіма моделями отримано оцінки, які збігаються до прогнозу довгострокового середнього.

Тестування без задавання конкретної альтернативи

Вище розглянуто тести на лінійність процесу з формулюванням конкретної нелінійної альтернативи. Разом з тим існують загальні тести, які не передбачають конкретної альтернативи. Одним з них є тест на основі специфікації похибок регресії (regression error specification test (RESET)).

Спочатку для опису процесу будується лінійна модель

$$y(k) = \beta^T \mathbf{w}(k) + \varepsilon(k) \quad (2.70)$$

де $\mathbf{w}(k) = [1, y(k-1), \dots, y(k-p); x_1(k), \dots, x_l(k)]^T$ – вектор вимірів основної змінної та регресорів; $u(k) = g(\beta, \mathbf{w}(k) \cdot \varepsilon(k))$; $\varepsilon(k)$ – мартингальний різницевий

процес з такими характеристиками: $E[\varepsilon(k)|\mathbf{I}(k)] = 0$, $\text{cov}[\varepsilon(k)|\mathbf{I}(k)] = \sigma_\varepsilon^2$;
 $\mathbf{I}(k) = \{y(k-j), j>0; \mathbf{x}(k-i), i\geq 0\}$ – наявна інформація у формі вимірів.
 Параметри цієї моделі оцінюються за допомогою звичайного методу найменших квадратів.

Для моделі (2.71)

$$\hat{\varepsilon}(k) = e(k) = y(k) - \hat{y}(k) \quad (2.71)$$

будується допоміжна регресія такого вигляду:

$$\hat{\varepsilon}(k) = \alpha^T \mathbf{w}(k) + \sum_{j=2}^h c_j \hat{y}^j(k) + \varepsilon^*(k) \quad (2.72)$$

де α, c – вектори параметрів; j – визначає степінь, з якою основна змінна $y(k)$ входить у модель; $\varepsilon^*(k)$ – залишки другої моделі (2.4.8). Нульова гіпотеза для тесту RESET формулюється таким чином:

$$H_0 : c_j = 0, \quad j=2, \dots, h \quad (2.73)$$

Якщо $\mathbf{w}(k) = [1, y(k-1), \dots, y(k-p)]^T$ і $h=2$, то це вже буде тест Кіннана.

Існують також інші варіанти моделі (2.4.8):

$$\hat{\varepsilon}(k) = \alpha^T \mathbf{w}(k) + \sum_{i=1}^p \sum_{j=i}^p c_{ij} y(k-i) y(k-j) + \varepsilon^*(k) \quad (2.74)$$

з нульовою гіпотезою стосовно лінійності: $H_0 : c_{ij} = 0, \quad \forall i, j$.

з нульовою гіпотезою стосовно лінійності: $H_0 : c_{ijl} = 0, \forall i, j, l$.

Q-статистика Люнга-Бокса для квадратів залишків

Статистика Люнга-Бокса застосовується до квадратів залишків моделі АРКС(p, q); тестова статистика розраховується за виразом [2]:

$$\hat{Q}(m) = N(N+2) \sum_{i=1}^m \frac{\hat{\rho}_i^2[\hat{\varepsilon}^2(k)]}{N-i} \quad (2.76)$$

де N – потужність вибірки даних; m – належним чином вибрана кількість значень дискретної автокореляційної функції $\hat{\rho}_i[\hat{\varepsilon}^2(k)]$, обчисленої для квадратів залишків $\hat{\varepsilon}^2(k)$. Якщо побудована модель АРКС(p, q) адекватна, то величина $Q(m)$ має асимптотично розподіл χ^2 з $m-p-q$ ступенями свободи.

За нульову та альтернативну гіпотези цього статистичного тесту приймаються такі:

$$\hat{H}_0 : \alpha_1 = \dots = \alpha_m = 0 \quad \hat{H}_1 : \text{хоча б один коефіцієнт } \alpha_i \neq 0, i=1, \dots, m \quad (2.77)$$

де α_i – коефіцієнти регресійної моделі

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \dots + \alpha_m \hat{\varepsilon}^2(k-m) + v(k) \quad (2.78)$$

для $k=m+1, \dots, N$. Оскільки тестова статистика розраховується для залишків моделі АРКС(p, q), то кількість ступенів свободи дорівнює $m-p-q$.

Біспектральний тест

Цей тест можна використовувати для аналізу процесу на лінійність та гауссовість. Для лінійного часового ряду біспектр буде постійним на всіх

частотах, а для нормального розподілу ця константа дорівнює нулю. Біспектр часового ряду обчислюється як перетворення Фур'є його третіх моментів. Так, третій момент стаціонарного ряду

$$y(k) = \mu + \sum_{i=1}^{\infty} \alpha_i v(k-i) \quad (2.79)$$

де $\alpha_i, i = 1, \dots, \infty$ – послідовність дійсних чисел при $\alpha_0 = 1$, $\{v(k)\} \sim i.i.d.$ – випадковий процес незалежних однаково розподілених чисел з явно визначеною функцією розподілу; розраховується за формулою:

$$c(u, v) = E[v^3(k)] \sum_{k=-\infty}^{\infty} \alpha_k \alpha_{k+u} \alpha_{k+v} \quad (2.80)$$

де u і v – цілі числа; $\alpha_0 = 1$; $\alpha_k = 0, k < 0$. Перетворення Фур'є для (2.4.11) має вигляд:

$$b_3(\omega_1, \omega_2) = \frac{8}{4\pi^2} \Gamma[-(\omega_1 + \omega_2)] \Gamma(\omega_1) \Gamma(\omega_2) \quad (2.81)$$

де $\Gamma(\omega) = \sum_{k=0}^{\infty} \alpha_k \exp(-ik\omega)$; $i = \sqrt{-1}$; ω_i – частоти. Функція спектральної щільності процесу $y(k)$ обчислюється за виразом:

$$p(\omega) = \frac{\sigma_v^2}{2\pi} |\Gamma(\omega)|^2 \quad (2.82)$$

де ω – частота. Таким чином, отримуємо такий корисний вираз для біспектральної функції:

Біспектральний тест ґрунтується на обчисленні та аналізі виразу (2.4.13). Функція $S(\omega_1, \omega_2)$ оцінюється на належно вибраній сітці точок (значень частот), потім перевіряється сталість оціненої величини. Це відбувається за аналогією аналізу статистики T^2 Хотеллінга [51].

Зазначимо, що для лінійного гауссівського процесу виконується рівність: $E[v^3(k)] = 0$ і біспектральна функція дорівнює нулю (статистично незначуща) на всіх частотах (ω_1, ω_2) . Таким чином, при $S(\omega_i, \omega_j) = 0, \forall \omega_i, \omega_j$ процес буде лінійним і третій момент вхідної послідовності $c(u, v) = 0$.

Такий підхід до встановлення наявності нелінійності має два недоліки. По-перше, оцінювання спектральної щільності потужності потребує застосування спеціальної попередньої обробки достатньо великої кількості вимірів вимірів у вигляді застосування часових вікон, усереднення, цифрової фільтрації і т. ін. По-друге, він не завжди може бути використаний при розв'язуванні задач ідентифікації (моделювання) систем, оскільки він не дає можливості одержати оцінки необхідних параметрів моделі у явному вигляді. Крім того, при розв'язуванні цих же задач не завжди є можливість одержати виміри вхідної послідовності або ж інформативний вхідний сигнал одержують штучно у вигляді спеціально згенерованих послідовностей, які не завжди можна подавати на вхід об'єкта внаслідок фізичних особливостей його функціонування.

Що стосується економічних процесів, то в цьому випадку, як правило, не можна поставити експеримент із процесом. Тому використовують тільки ті статистичні дані, які фактично можна зібрати в процесі дослідження. У загальному випадку при ідентифікації систем використовують три типи сигналів: вхідні, вихідні і збурення. При цьому вхідний керуючий сигнал вважають незалежним від збурення. У результаті виявляється неможливим з'ясувати деякі типи зв'язків між змінними.

Нелінійний детермінований тренд

Наявність нелінійного детермінованого тренду у процесі можна спрощено встановити шляхом оцінювання рівняння:

$$y(k) = a_0 + c_1 k + c_2 k^2 + \dots + c_m k^m + \varepsilon(k) \quad (2.84)$$

яке представляє собою поліном порядку m стосовно часу. Якщо хоча б один із коефіцієнтів c_i , $i = 1, \dots, m$ є статистично значимим, то гіпотеза щодо відсутності тренду відхиляється. Якщо тренд відносно швидко змінює свій напрямок розвитку і для нього важко знайти адекватний функціональний опис, то застосовують моделі випадкових трендів, які ґрунтуються на комбінаціях випадкових величин.

Метод групового урахування аргументів

Автоматично оцінює структуру математичної моделі *метод групового врахування аргументів* (МГВА), запропонований академіком Івахненком О.Г. (Київський інститут кібернетики). Його використовують для побудови моделей досліджуваних процесів у формі полінома Колмогорова-Габора. Сьогодні цей метод успішно застосовують для опису процесів різних класів, у тому числі і до моделювання процесів різної природи з нелінійністю та нестационарністю. Подальшим розвитком цього методу є нечіткий МГВА, який ґрунтується на нечіткому представленні параметрів оцінюваної моделі, які оцінюють методами лінійного програмування. Необхідно зазначити, що поліноміальна модель, яка будується за МГВА, може бути досить громіздкою і незручною для практичного використання, але цей недолік нівелюється високою адекватністю цієї моделі досліджуваному процесу.

5. Формування інших елементів структури моделі

На третьому етапі необхідно вибрати структури моделей кандидатів. Поняття структури моделі було розглянуто вище, але нагадаємо, що воно включає *порядок моделі* (найвищий порядок рівнянь, що його утворюють); *розмірність* (кількість рівнянь моделі); *час запізнення* по входу (лаг) та його оцінка; *можливі нелінійності* та їх тип; *зовнішні збурення* та їх тип (детерміновані або випадкові; адитивні та мультиплікативні) [2, 51].

Для того щоб визначити які незалежні змінні (регресори) необхідно включити в праву частину рівняння, обчислюють коефіцієнт кореляції між залежною та відповідною незалежною змінною.

Коефіцієнт кореляції, а в загальному випадку кореляційна функція, дозволяє встановити факт існування зв'язку між змінними. Кореляція може бути лінійною або нелінійною в залежності від типу функціональної залежності,

яка фактично має місце між змінними. В більшості практичних випадків розглядають лінійну кореляцію (взаємозв'язок) між змінними, але більш глибокий аналіз потребує використання нелінійних залежностей. Складну нелінійну залежність можна спростити, але знати про її існування необхідно для того щоб побудувати, при необхідності, (складнішу за структурою) модель процесу з вищим ступенем адекватності.

Кореляційна матриця дає можливість встановити існування зв'язку між залежною (ендогенною) змінною та незалежними (екзогенними) змінними у правій частині. Розглянемо кореляційну матрицю $\mathbf{R}$ розмірності 3×3 , яка будується для трьох змінних x, y, z :

$$\mathbf{R} = \begin{bmatrix} r_{yy} & r_{xy} & r_{zy} \\ r_{yx} & r_{xx} & r_{zx} \\ r_{yz} & r_{xz} & r_{zz} \end{bmatrix}, \quad \text{де } r_{yx} = r_{xy}, \quad r_{yz} = r_{zy}, \quad r_{xz} = r_{zy}. \quad (2.85)$$

Нехай y – показник якості технологічного процесу; x, z – технологічні параметри, які, за припущенням, впливають на показник якості. Тобто, ставиться задача встановлення існування залежності вигляду:

$$y = f(x, z), \quad (2.86)$$

яка може бути представлена у формі регресії змінної y на незалежні змінні x, z :

$$y(k) = a_0 + a_1 x(k) + a_2 z(k) + \varepsilon(k) \quad (2.87)$$

де k – дискретний час (наприклад, в долях секунди, секундах, хвилинах, годинах, днях, тижнях, місяцях і т.і.); $\varepsilon(k)$ – випадкова змінна, введення якої в модель пояснюється наступними причинами:

– часто буває неможливо встановити всі незалежні змінні, які впливають на залежну змінну, а тому наведене рівняння описує процес з похибкою;

– можуть існувати такі незалежні змінні, які неможливо виміряти і включити в модель, а тому їх розглядають як збурення і вважають, що їх спільний вплив на залежну змінну описується випадковою змінною $\varepsilon(k)$;

– в наведене вище регресійне рівняння можуть бути введені пояснюючі змінні, які є формально корельованими із залежною змінною, але фактично вони не впливають на неї;

– для будь-якого методу оцінювання параметрів рівнянь властиві методичні похибки, які повинні бути, по можливості, враховані в моделі.

Вважається, що сукупний вплив всіх вказаних факторів можна описати в деякій мірі за допомогою випадкової змінної $\varepsilon(k)$. Оскільки вона не вимірюється, то оцінити її значення (*похибку моделі або залишок*) можна тільки після оцінювання коефіцієнтів моделі, тобто

$$\hat{\varepsilon}(k) = e(k) = y(k) - \hat{y}(k) \quad (2.88)$$

де $\hat{y}(k)$ – оцінка змінної $y(k)$, отримана за допомогою моделі; $y(k)$ – фактичний вимір.

Для обчислення елементів матриці **R** необхідно мати синхронізовані в часі вибірки значень всіх трьох змінних y, x, z . Формула для розрахунку коефіцієнтів кореляції має вигляд:

$$r_{yx} = \frac{\frac{1}{N-1} \sum_{k=1}^N \{[x(k) - \bar{x}][y(k) - \bar{y}]\}}{\sigma_x \sigma_y} \quad (2.89)$$

де $\bar{x}, \bar{y}$ – вибіркові середні значення змінних x, y ; σ_x, σ_y – стандартні відхилення цих змінних, тобто корені квадратні з їх дисперсії. Наприклад,

$$\sigma_y = \sqrt{\sigma_y^2} = \left[\frac{1}{N-1} \sum_{k=1}^N [y(k) - \bar{y}]^2 \right]^{1/2} \quad (2.90)$$

де N – кількість вимірів змінної Y ; $\bar{Y}$ – вибіркове середнє значення ряду $\{y(k)\}$, яке обчислюється за відомою формулою:

$$\bar{y} = \frac{1}{N} \sum_{k=1}^N y(k) \quad (2.91)$$

Очевидно, що перед формальним обчисленням коефіцієнтів кореляції, необхідно зробити аналіз процесу і визначити наявність (або відсутність) логічних зв'язків між змінними. Це дозволяє ввести до розгляду тільки ті змінні, які дійсно впливають на залежну змінну, наприклад, на показник якості. Очевидно, що для правильного вибору незалежних (екзогенних) змінних необхідно досить глибоко знати процес, що моделюється.

На основі значень коефіцієнтів кореляції приймається рішення про їх включення у рівняння регресії:

$$y(k) = a_0 + b_1 x(k) + b_2 z(k) + \varepsilon(k) \quad (2.92)$$

яке може бути представлено в загальному вигляді як

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + a_3 x_3(k) + \dots + a_{p-1} x_{p-1}(k) + \varepsilon(k) \quad (2.93)$$

Можна показати, що між коефіцієнтами регресії b_1, b_2 і коефіцієнтами кореляції r_{yx}, r_{yz} існує однозначний взаємозв'язок.

Останнє рівняння представляє собою *лінійну регресію* P -го порядку, але досить часто необхідно застосовувати більш складні нелінійні моделі. Характерним представником нелінійної відносно змінних регресії є поліноміальна регресія порядку $P-1$:

$$y(k) = a_0 + a_1 x(k) + a_2 x^2(k) + a_3 x^3(k) + \dots + a_{p-1} x^{p-1}(k) + \varepsilon(k) \quad (2.94)$$

Хоча в це рівняння включено тільки одну незалежну змінну $x(k)$, очевидно, що воно може бути розширене будь-якими іншими змінними.

Для визначення необхідності введення в рівняння регресії авторегресійної складової необхідно обчислити і дослідити вибірккову автокореляційну функцію змінної $y(k)$. Рівняння з авторегресійною складовою другого порядку має вигляд:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + b_1 x(k) + b_2 z(k) + \varepsilon(k) \quad (2.95)$$

тобто в рівняння регресії введено авторегресійну (АР) складову другого порядку. Порядок авторегресії визначається за допомогою автокореляційної функції. Число коефіцієнтів автокореляційної функції, які відмінні від нуля в статистичному сенсі, і буде складати оцінку порядку авторегресії.

Природа авторегресії пояснюється існуванням так званої “пам’яті” процесу, яка проявляється у тому, що його поточний стан в значній мірі визначається попередніми станами. Наприклад, стан людини вранці залежить від того, яким було самопочуття ввечері та в попередні дні. На формування ринкових цін суттєво впливають значення цін в попередні періоди часу. Поточний стан технічної системи або технологічного процесу також залежить від їх стану у попередні моменти часу.

Вибіркова АКФ обчислюється за виразом:

$$r_y(s) = r_{y(k)y(k-s)} = \frac{1}{N-1} \frac{\sum_{k=s+1}^N \{[y(k) - \bar{y}][y(k-s) - \bar{y}]\}}{\sigma_y^2}, \quad s = 1, 2, 3, \dots, \quad (2.96)$$

де σ_y^2 – вибірккова дисперсія змінної $y(k)$; $\bar{y}$ – середнє значення вибірки даних.

Кількість коефіцієнтів АКФ, відмінних від нуля в статистичному сенсі, вказує на порядок авторегресійної частини моделі. Для стаціонарного процесу (це процес із постійними середнім значенням, дисперсією та коваріацією)

коефіцієнти $r_y(s)$ мають нормальний розподіл та нульове середнє. Коли ми говоримо, що значення коефіцієнтів автокореляційної функції повинні бути відмінними від нуля в статистичному сенсі, це означає, що існує вираз (формула), який дозволяє підтвердити або спростувати цей факт. Одним із загальноприйнятих підходів до встановлення того факту, що коефіцієнти АКФ суттєво відмінні від нуля в статистичному сенсі, є обчислення та аналіз значимості статистичного параметра (або просто статистики) Льюнга-Бокса $Q(r_k)$ за виразом [51]:

$$Q(r_k) = N(N + 2) \sum_{k=1}^s r_k^2 / (N - k) \quad (2.98)$$

де N – довжина вибірки даних змінної, для якої обчислено значення автокореляційної функції r_k ; S – число коефіцієнтів АКФ, які досліджуються на суттєву відмінність від нуля. Якщо дані згенеровані процесом АР чи АРКС, то значення $Q(r_k)$ асимптотично мають розподіл χ^2 з s ступенями свободи, а тому для перевірки їх значимості необхідно користуватись відповідними статистичними таблицями. Очевидно, що більші значення вибіркової автокореляційної функції приводять до більших значень $Q(r_k)$.

Так, для ідеального процесу білого шуму $Q(r_k) = 0$. Якщо значення $Q(r_k)$, обчислене за наведеним виразом, перевищує критичне значення з розподілу χ^2 з s ступенями свободи, то існує щонайменше одне значення $r(k)$, яке є відмінним від нуля в статистичному сенсі. Статистику Льюнга-Бокса можна застосовувати також для встановлення близькості залишків моделі до білого шуму. Однак, необхідно пам'ятати, що при обчисленні s значень кореляційної функції число ступенів вободи зменшується на число коефіцієнтів моделі. Таким чином, при аналізі залишків моделі АРКС(p, q) статистика $Q(r_k)$ має

розподіл χ^2 з $s - p - q$ ступенями свободи, а із врахуванням константи $s - p - q - 1$.

Цей етап закінчується формуванням структур декількох моделей-кандидатів з векторами параметрів $\theta_1, \dots, \theta_m$, де m – кількість кандидатів. Кандидатів може бути декілька, оскільки встановити структуру точно за один раз, як правило, неможливо. Загалом побудова моделі високого ступеня адекватності – це досить трудомісткий ітераційний процес, який вимагає значних зусиль. На наступному етапі уточнюють структуру моделі і оцінюють параметри моделей-кандидатів.

6. Уточнення структури нелінійної моделі

Якщо в результаті застосування тестів на нелінійність встановлено, що досліджуваний процес нелінійний, то структуру моделі необхідно оцінити за допомогою наявних статистичних (експериментальних) даних. Для визначення структури моделей (лагів) лінійних процесів використовують використовують кореляційний аналіз, зокрема автокореляційні (АКФ) та часткові автокореляційні функції (ЧАКФ) як було показано вище. Ступінь корельованості залежної змінної з незалежними можна встановити за допомогою кореляційної матриці навіть у випадку наявності нелінійності. Надалі для оцінювання параметрів лінійних моделей необхідно застосовувати рекурсивні процедури, які відслідковують зміни параметрів моделей. Якщо вибір структури моделі виконується у класі моделей, залежних від стану, то такий вибір є складним, оскільки клас цих моделей досить широкий.

Якщо клас параметричних моделей більше обмежений, то структура моделі-кандидата може бути конкретнішою. Так, у випадку побудови лінійних моделей типу АРКС методика оцінювання їх структури є достатньо чіткою і забезпечує вибір кількох структур, з яких після оцінювання параметрів вибирається краща модель.

Використовуючи тести на лінійність і візуальний аналіз даних, модель нелінійного процесу можна вибирати у класі порогових авторегресійних моделей [50, 51]. У роботі [18] розглядається побудова авторегресійних моделей з функціональними коефіцієнтами, а у дослідженні [17] розв'язується задача оцінювання структури адитивних авторегресійних моделей з функціональними коефіцієнтами. Ключовим елементом запропонованої процедури є

використання множини локальних регресійних моделей.

Застосування багатовимірних адаптивних регресійних сплайнів (multivariate adaptive autoregressive splines (MARS)) до опису нелінійних процесів розглядається у роботі [40]; адаптивні порогові авторегресійні моделі на сплайнах аналізуються у дослідженні [24]. У роботі [46] для опису нелінійних процесів запропоновані авторегресійні моделі з гладкими перехідними характеристиками (АРГПХ).

Авторегресійна модель з гладкою перехідною характеристикою має вигляд:

$$y(k) = \beta^T \mathbf{w}(k) + (\theta^T \mathbf{w}(k)) \mathbf{F}(z(k)) + \mathbf{v}(k) \quad (2.99)$$

де $E[\mathbf{v}(k) | \mathbf{I}(k)] = 0$; $\text{cov}[\mathbf{v}(k) | \mathbf{I}(k)] = \sigma^2$; $\theta = [\theta_0, \theta_1, \dots, \theta_m]^T$;
 $\beta = [\beta_0, \beta_1, \dots, \beta_m]^T$; $\mathbf{I}(k) = [y(k-j), j > 0; x(k-j, i), i = 1, \dots, l, j \geq 0]$; $m = p + l + 1$;
 $\mathbf{w}(k) = [1, y(k-1), \dots, y(k-p); x_1(k), \dots, x_l(k)]^T$. Функція $\mathbf{F}(z(k))$ може бути задана так:

$$\mathbf{F}(z(k)) = \frac{1}{1 + \exp[-\gamma (z(k) - c)]}, \quad \gamma > 0, \quad (2.100)$$

що приводить до логістичної моделі, або

$$\mathbf{F}(z(k)) = 1 - \exp[-\gamma (z(k) - c)]^2, \quad \gamma > 0, \quad (2.101)$$

що відповідає експоненційній моделі АРГПХ. Перехідною змінною $z(k)$ може бути будь-який елемент вектора $\mathbf{w}(k)$ або інша змінна.

Оцінювання структури нелінійної моделі на основі даних виконується за кілька етапів, поданих нижче.

1. Вибрати (оцінити) структуру лінійної моделі на основі кореляційного аналізу даних, як правило, низького порядку, яка необхідна

для того, щоб тестувати лінійність (нелінійність) досліджуваного процесу; для розв'язання цієї задачі необхідно користуватись відомою методикою моделювання лінійних процесів.

2. Виконати тестування процесу на лінійність, використовуючи побудовану лінійну модель як нульову гіпотезу (проти гладкої перехідної регресії – smooth transition regression (STR)); якщо гіпотеза стосовно лінійності відхиляється, то необхідно визначити перехідну змінну на основі даних.

3. Вибрати для подальшого дослідження модель логістичної регресії з гладким переходом (logistic smooth transition regression (LSTR)) або експоненційну регресію з гладким переходом (exponential smooth transition regression (ESTR)) для подальшого тестування (залишків) на лінійність.

4. За необхідності доповнити структури моделей-кандидатів такими нелінійними членами: $x^2(k)$, $x^3(k)$, $x(k)s(k)$, $x^2(k)s(k)$, де $x(k)$, $s(k)$ – регресори.

5. Обчислити оцінки параметрів сформованих структур моделей-кандидатів за допомогою коректно вибраних методів.

6. Вибрати кращу модель з множини кандидатів за множиною статистичних критеріїв адекватності моделі.

Тестування за пунктом 3 ґрунтується на такій моделі:

$$\hat{v}(k) = \beta_0^T \mathbf{w}(k) + \beta_1^T \mathbf{w}(k) z_d(k) + \beta_2^T \mathbf{w}(k) z_d^2(k) + \beta_3^T \mathbf{w}(k) z_d^3(k) + \eta(k) \quad (2.102)$$

де $z_d(k)$ – перехідна змінна; $\hat{v}(k)$ – залишки моделі лінійної регресії, оціненої за допомогою звичайного методу найменших квадратів:

$$y(k) = \beta^T \mathbf{w}(k) + v(k) \quad (2.103)$$

Якщо $z_d(k)$ – елемент вектора $\mathbf{w}(k)$, то вектор $\mathbf{w}(k) = [1, \tilde{\mathbf{w}}^T k]^T$ в (2.6.2) необхідно замінити на $\tilde{\mathbf{w}}(k)$, за винятком першого члена у правій частині цього рівняння.

Нульова гіпотеза стосовно лінійності формулюється у цьому випадку так:

$$H_{0d}: \beta_1, \beta_2, \beta_3 = 0 \quad (2.104)$$

Рівняння (2.6.2) також використовується для вибору перехідної змінної $z_d(k)$. Тестування виконується для всіх можливих кандидатів на перехідну змінну $z_d(k)$ і вибирається та змінна, для якої P – value приймає найменше значення (воно також має бути достатньо малим). Якщо така умова не виконується, то модель (процес) вважається лінійною.

Коректність цієї процедури тестування мотивується таким чином. Припустимо, що існує така гладка перехідна регресія (STR-модель), з перехідною змінною $z_d(k)$, що характеризується відповідними статистичними даними (нульова шіпотеза). У такому випадку застосування тесту на основі множників Лагранжа (Lagrange multiplier test або LM-test) проти цієї гіпотези характеризується оптимальною потужністю [28].

Якщо перехідна змінна вибрана для тестування некоректно, то результат тестування буде мати меншу потужність проти істинної альтернативи ніж у випадку коректного вибору цієї змінної. Таким чином, саме обгрунтоване відхилення нульової гіпотези припускає, що необхідно коректно вибирати перехідну змінну.

Якщо гіпотеза стосовно лінійності відхиляється і перехідна змінна вибрана, то на третьому кроці необхідно розв'язати задачу вибору між моделлю логістичної регресії з гладким переходом (logistic smooth transition regression (LSTR)) та експоненційною регресією з гладким переходом (exponential smooth transition regression (ESTR)). Це можна робити шляхом тестування множини нульових гіпотез з використанням моделі (2.4.2) і F – тесту. Ці гіпотези формулюються так:

$$H_{03}^*: \beta_3 = 0; \quad H_{02}^*: \beta_2 = 0 | \beta_3 = 0; \quad H_{01}^*: \beta_1 = 0 | \beta_2 = \beta_3 = 0. \quad (2.105)$$

Якщо P – value приймає найменше значення при тестуванні H_{02}^* , то

вибирається модель ESTR, інакше – модель LSTR.

Оцінювання значень лагів для моделі (2.6.1) виконується з використанням моделі (2.6.2) та статистичних критеріїв адекватності результуючих моделей. У відомих прикладах побудови моделей за статистичними даними, як правило застосовувався підхід зверху-вниз (general-to-specific), який ґрунтується на оцінюванні нелінійних регресійних моделей з гладким переходом (STR) або на оцінюванні нелінійних авторегресійних моделей з гладким переходом (STAR).

Задача оцінювання структури моделі виникає також при непараметричному моделюванні часових рядів. За початку структуру можна вибрати, наприклад, таку:

$$y(k) = f(y(k-1), x(k-1)) + g(y(k-1), x(k-1)) \varepsilon(k),$$

$$\text{де } y(k-1) = [y(k-i_1), \dots, y(k-i_p)]^T \text{ та} \quad (2.106)$$

$$x(k-1) = [x(k-j_1), \dots, x(k-j_q)]^T$$

– вектори затриманих у часі значень відповідних змінних; $\{\varepsilon(k)\}$ – послідовність мартингальних різниць по відношенню до наявної інформації $I(k) = \{y(k-i), i > 0; x(k-i), i > 0\}$. Тут виникає питання – які лаги включати у модель:

$$y(k-1) = [y(k-i_1), \dots, y(k-i_p)]^T \text{ та} \quad (2.107)$$

$$x(k-1) = [x(k-j_1), \dots, x(k-j_q)]^T$$

Також необхідно встановити – будуть функції f і g лінійними чи нелінійними, і будуть вони адитивними чи ні. Більше того, якщо в модель включаються члени із взаємодіючими змінних, то яким чином можна встановити структуру цих членів? У більш загальному випадку виникає питання стосовно того, чи можна використовувати у непараметричному аналізі такі функціональні форми як порогові або функції з гладкими переходами?

В аналізі нелінійних часових рядів це задачі поглибленого аналізу даних,

які ще потребують виконання подальших досліджень.

7. Оцінювання параметрів (коефіцієнтів) моделей-кандидатів

На *четвертому етапі* оцінюють коефіцієнти (параметри) рівняння, використовуючи принцип економії або збереження. Цей принцип означає, що *кількість коефіцієнтів, що оцінюються, не повинна перевищувати їх необхідне число* (“необхідність” можна визначити, наприклад, як необхідність збереження в моделі основних статистичних характеристик процесу). При моделюванні процесів будь-якої природи необхідно пам’ятати, що поведінку процесу необхідно *апроксимувати* за допомогою рівнянь, а не старатися описати її до найменших дрібниць. Необхідно враховувати також, що різні за структурою моделі можуть мати однакові властивості.

Модель, що оцінюється, повинна задовольняти принципу інверсії, тобто щоб за допомогою отриманого рівняння можна було б згенерувати початковий ряд, на основі якого оцінювались коефіцієнти. Це означає, що хоча модель і спрощена, вона повинна узгоджуватись з досліджуваним процесом за такими основними характеристиками як середнє, дисперсія та коваріація.

В процедурі оцінювання часто використовують не абсолютні значення змінних, а їх відхилення від середнього, тобто

$$y(k) = Y(k) - \mu_y \quad (2.108)$$

μ_y – середнє значення ряду. Якщо для оцінювання параметрів використовується рекурсивна процедура, то поточне середнє можна обчислювати за формулою:

$$\mu_y(k) = \mu_y(k-1) + \frac{1}{k} [y(k) - \mu_y(k-1)] \quad (2.109)$$

Найбільш поширеними методами оцінювання параметрів моделі є такі:

- *метод найменших квадратів (МНК);*
- *метод максимальної правдоподібності (ММП);*
- *метод допоміжної (інструментальної) змінної (МДП);*

- *нелінійний метод найменших квадратів (НМНК);*
- *метод Монте-Карло для марковських ланцюгів.*

та їх рекурсивні версії (РМНК, РММП, РМДП). Деякі методи будуть розглянуті в розділі, присвяченому методам оцінювання. Оцінки (звичайного) МНК обчислюють за допомогою наступного виразу:

$$\hat{\theta} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y} \quad (2.110)$$

де $\theta [p]$ – вектор оцінок параметрів вимірності p ; $\mathbf{X}[N \times p]$ – матриця вимірів; $\mathbf{y}[N]$ – вектор вимірів залежної змінної. В квадратних дужках вказана вимірність векторів і матриці. Елементи матриці вимірів обчислюються по-своєму для кожної конкретної моделі. Так, для моделі

$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + a_3 x_3(k) + \varepsilon(k) \quad (2.111)$$

матриця вимірів має вигляд:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1(1) & x_2(1) & x_3(1) \\ 1 & x_1(2) & x_2(2) & x_3(2) \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_1(N) & x_2(N) & x_3(N) \end{bmatrix} \quad (2.112)$$

Одиниці в першому стовпчику матриці $\mathbf{X}$ означають, що вимір при коефіцієнті a_0 завжди дорівнює одиниці.

Елементи матриці вимірів дещо ускладнюються у випадку використання поліноміальної моделі, але її також можна оцінювати за допомогою лінійних методів. Безпосереднє застосування методу мінімізації суми квадратів похибок до поліноміальної моделі порядку P приводить до формування такої матриці вимірів:

При такому представленні векторно-матричне рівняння для N вимірів залежної та незалежної змінних можна записати так: $\mathbf{y}' = \mathbf{X}' \boldsymbol{\theta}$; звідси

$\hat{\boldsymbol{\theta}} = [\mathbf{X}']^{-1} \mathbf{y}'$, де $\mathbf{y}' = \left[\sum_1^N y(k) \quad \sum_1^N x(k) y(k) \quad \dots \quad \sum_1^N x^p(k) y(k) \right]^T$; $\hat{\boldsymbol{\theta}}$ – вектор оцінок параметрів моделі. Тобто оцінку вектора параметрів можна знайти шляхом знаходження розв'язку системи лінійних (нормальних) рівнянь.

Для отримання незміщених, консистентних та ефективних оцінок вектора параметрів $\boldsymbol{\theta}$ лінійної регресійної математичної моделі, наприклад, моделі змішаної регресії:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + b_1 x(k) + b_2 z(k) + \varepsilon(k) \quad (2.114)$$

за допомогою методу найменших квадратів необхідно задовольнити наступні умови:

а) $\varepsilon(k)$ – некорельована послідовність випадкових чисел з нульовим середнім, тобто, $E[\varepsilon(k)] = 0$, $\text{cov}[\varepsilon(k)] = E[\varepsilon(k)\varepsilon(j)] = \begin{cases} \sigma_\varepsilon^2, & k = j; \\ 0, & k \neq j. \end{cases}$

б) послідовності $\varepsilon(k)$ і $y(k)$ не повинні бути корельовані між собою.

Зазначимо, що перевірити виконання наведених умов ми можемо тільки після оцінювання коефіцієнтів моделі, а до оцінювання можна тільки постулювати їх виконання. Тобто, після оцінювання моделі оцінка значень випадкового процесу визначається похибками моделі:

$$\hat{\varepsilon}(k) = e(k) = y(k) - \hat{y}(k) \quad (2.115)$$

що дає можливість виконати аналіз характеристик випадкового процесу $\{\varepsilon(k)\}$.

8. Діагностика моделей – вибір кращої з множини оцінених кандидатів

На *п'ятому етапі* аналізується якість моделі, тобто виконується перевірка оцінених кандидатів на адекватність процесу. Діагностика складається з наступних кроків [2, 51].

а) *Візуальне дослідження графіка похибок* моделі отримане за допомогою побудованої моделі. На графіку не повинно бути значних викидів та довгих інтервалів, на яких похибка приймає великі значення (тобто довгих інтервалів суттєвої неадекватності моделі). У випадку застосування рекурсивних методів оцінювання найбільші похибки оцінок будуть в перехідному процесі, коли інформаційна матриця ще не містить достатньо інформації про процес.

б) *Похибки моделі не повинні бути корельовані* між собою. Для аналізу наявності кореляції між значеннями похибок необхідно обчислити АКФ та ЧАКФ для ряду i за допомогою статистики визначити ступінь корельованості (наприклад, статистика вважається несуттєвою до рівня 10%).

Крім того, корельованість похибок визначають за допомогою статистики Дарбіна-Уотсона (DW), яка розраховується за формулою:

$$DW = 2 - 2\rho \quad (2.116)$$

де $\rho = E[e(k)e(k-1)]/\sigma_e^2$ – коефіцієнт кореляції між сусідніми значеннями похибки; σ_e^2 – дисперсія послідовності похибок $\{e(k)\}$. Таким чином, при повній відсутності кореляції між похибками $DW = 2$ – це ідеальне значення. Граничними значеннями для DW є 0 (при $\rho = 1$) та +4 (при $\rho = -1$).

Отримати формулу $DW = 2 - 2\rho$ можна досить просто. Автори цієї статистики (Durbin, Watson) запропонували скористатись для перевірки корельованості похибок моделі таким виразом:

$$DW = \frac{\sum_{k=2}^N [e(k) - e(k-1)]^2}{\sum_{k=1}^N e^2(k)} = \frac{\sum_{k=2}^N [e(k) - e(k-1)][e(k) - e(k-1)]}{\sum_{k=1}^N e^2(k)} \quad (2.117)$$

тобто, DW можна, в деякій мірі, трактувати як коефіцієнт автокореляції для (перших різниць) приростів похибок.

Розкриваючи квадрат різниці в чисельнику, отримаємо:

$$DW = \frac{\sum_{k=2}^N e^2(k)}{\sum_{k=1}^N e^2(k)} + \frac{\sum_{k=2}^N e^2(k-1)}{\sum_{k=1}^N e^2(k)} - 2 \frac{\sum_{k=2}^N e(k)e(k-1)}{\sum_{k=1}^N e^2(k)}, \quad (2.118)$$

де $\frac{\sum_{k=2}^N e^2(k)}{\sum_{k=1}^N e^2(k)} \approx 1$; $\frac{\sum_{k=2}^N e^2(k-1)}{\sum_{k=1}^N e^2(k-1)} \approx 1$; а $\frac{\sum_{k=2}^N e(k)e(k-1)}{\sum_{k=1}^N e^2(k-1)} = \rho$

Тому можна записати, що $DW = 2 - 2\rho$.

в) Для лінійної моделі 2-3 порядку оцінки параметрів повинні збігатися до усталених значень після 30-40 ітерацій алгоритму оцінювання. Якщо кількість ітерацій набагато перевищує вказані числа, то це свідчить про те, що процес може бути нестационарним.

г) Перевірка значимості параметрів моделі. *Статистика Стьюдента* або *t-статистика* (випадкова величина, що має *t*-розподіл), яка використовується для визначення значимості оцінки кожного коефіцієнта в статистичному сенсі, визначається за виразом:

$$t = \frac{\hat{a} - a^0}{SE_{\hat{a}}} \quad (2.119)$$

де $\hat{a}$ – оцінка коефіцієнта моделі; a^0 – нуль-гіпотеза (початкова гіпотеза) щодо цієї оцінки; $SE_{\hat{a}}$ – стандартна похибка оцінки. За нуль-гіпотезу щодо значимості оцінки можна висувати будь-яку: що коефіцієнт значимий, тобто,

($H_0: a^0 \neq 0$) або незначимий ($H_0: a^0 = 0$). Статистична теорія перевірки гіпотез пропонує висувати нуль-гіпотезу, яка є протилежною бажаному результату. В даному випадку бажаним результатом є значимість коефіцієнтів математичної моделі. Таким чином, необхідно висувати нульову гіпотезу, що коефіцієнт незначимий. Це дає можливість коректно підійти до визначення значимості оцінок коефіцієнтів та дещо спростити розрахунки.

Для того щоб встановити, чи є оцінка коефіцієнта значимою, необхідно знати довжину вибірки даних N (потужність вибірки); число ступенів свободи $f = N - n$, де n – число коефіцієнтів моделі, які оцінюються на основі ряду даних, і вибрати рівень значимості $\alpha = 1\%$ або $\alpha = 5\%$ або $\alpha = 10\%$ (для цих значень існують розраховані таблиці для критичних значень t -статистики). Помилка першого роду означає відхилення вірної гіпотези.

Користуючись значеннями N, f і α , з таблиць для t -розподілу знаходять критичне значення t -статистики, тобто $t_{кр}$. Для перевірки правильності висунутої гіпотези розраховане значення t порівнюють з критичним $t_{кр}$. Якщо

$$-t_{кр} < t < t_{кр} \quad \text{або} \quad |t| < |t_{кр}|, \quad (2.120)$$

то нуль-гіпотеза щодо незначимості коефіцієнта приймається (його можна не враховувати в регресії). Звідси випливає, що чим більшим є значення t -статистики для оцінки коефіцієнта, тим імовірніше, що цей коефіцієнт є значимим.

Загалом послідовність дій при перевірці значимості оцінок коефіцієнтів побудованої моделі можна сформулювати так:

- сформулювати нуль-гіпотезу щодо значимості коефіцієнта;
- обчислити значення t -статистики для кожного коефіцієнта регресії (це робить кожний пакет для математичного моделювання);

- за допомогою значень N , f і α знайти із таблиць для t -статистики її критичне значення;
- перевірити нуль-гіпотезу за наведеним вище простим правилом (аналіз виконання нерівності $-t_{кр} < t < t_{кр}$).

д) Коефіцієнт множинної детермінації, який обчислюється так:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST} \quad (2.121)$$

де $\text{var}(\hat{y})$ – дисперсія залежної змінної, оціненої за допомогою побудованої моделі; $\text{var}(y)$ – дисперсія вимірів залежної змінної;

$SSE = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2$ – сума квадратів похибок (залишків) моделі (*sum of*

squared errors); $SST = \sum_{k=1}^N [y(k) - \bar{y}]^2$ – загальна сума квадратів (*total sum of*

squares); $\bar{y}$ – середнє значення; $SSR = \sum_{k=1}^N [\hat{y}(k) - \bar{y}]^2$ – загальна сума квадратів для регресії (*sum of squares for regression*).

Очевидно, що найкращим значенням є $R^2 = 1$, тобто, коли дисперсії вимірів змінної, та цієї ж змінної, оціненої за рівнянням, збігаються. Цей параметр можна трактувати, також, як міру інформативності моделі, якщо вибрати за міру інформативності дисперсію. Таким чином, R^2 показує рівень інформативності моделі по відношенню до інформативності вибірки даних, за допомогою якої вона була оцінена.

е) Сума квадратів похибок для вибраної моделі повинна бути мінімальною, тобто,

$$\sum_{k=1}^N e^2(k) = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \rightarrow \min_{\hat{\theta}} \quad (2.122)$$

у порівнянні з усіма іншими моделями.

є) Для оцінки адекватності моделі також використовують *інформаційний критерій Акайке*

$$AIC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + 2n \quad (2.123)$$

та *критерій Байєса-Шварца*

$$BSC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + n \ln(N) \quad (2.124)$$

число параметрів моделі, які оцінюються за допомогою статистичних даних (p - число параметрів авторегресійної частини моделі; q - число параметрів ковзного середнього; 1 з'являється тоді, коли оцінюється зміщення (або *перетин*)).

Критерії Акайке і Байєса-Шварца містять в правій частині суму квадратів похибок, а тому за цими критеріями вибирають ту модель, для якої критерії приймають найменші значення. Введення нового регресора приводить до збільшення критерію (при цьому збільшується n), але, разом з тим, зменшується сума квадратів похибок і критерій в цілому зменшується. Якщо регресор не покращує модель, то критерій збільшується. Необхідно також зазначити, що асимптотичні властивості для довгих виборок кращі у критерія Байєса-Шварца, тобто, його рекомендують застосовувати при відносно великих значеннях N ($N > 100$).

ж) Окрім згаданих параметрів, для визначення адекватності моделі в цілому використовують *статистику Фішера*, яка пропорціональна відношенню:

$$F \sim \frac{R^2}{1 - R^2} \quad (2.125)$$

а для множинної (багатофакторної) регресії вона визначається за виразом

$$F = \frac{R^2}{1 - R^2} \cdot \frac{(N - p - 1)}{p} \quad (2.126)$$

де, як і раніше, N – число значень ряду; p – число параметрів моделі без врахування перетину (константи).

Таким чином, якщо $R^2 \rightarrow 1$, то $F \rightarrow \infty$. Порядок застосування F – статистики такий же, як і t – статистики. Нуль-гіпотезою є в даному випадку припущення про те, що модель неадекватна в цілому, тобто,

$$H_0 : a_1 = a_2 = \dots = a_p = 0 \quad (2.127)$$

проти альтернативної гіпотези

$$H_1 : \text{хоча б одне значення } a_i \text{ відмінне від нуля} \\ \text{в статистичному смислі.} \quad (2.127)$$

Значення $F_{\text{крит}}$ знаходять із таблиць для F – розподілу. Послідовність застосування цієї статистики можна представити наступним чином:

1. Сформулювати нуль-гіпотезу щодо адекватності моделі в цілому.

Наприклад, $H_0 : \text{модель неадекватна в цілому (або}$

$$H_0 : a_1 = a_2 = \dots = a_p = 0).$$

2. Розрахувати значення F для оціненої моделі (як правило, воно розраховується всіма пакетами статистичної обробки даних).

3. Задати рівень значимості $\alpha = 1\%$ або $\alpha = 5\%$ або $\alpha = 10\%$.

4. Користуючись значеннями N , f і α , знайти критичне значення $F_{\text{крит}}$ знаходять із таблиць для F – розподілу при $(p, N - p - 1)$ степенях свободи.

5. Перевірити нуль-гіпотезу:

якщо $F > F_{\text{крит}}$, то нуль-гіпотеза щодо неадекватності моделі в цілому відкидається на вибраному рівні значимості.

Коректне застосування описаної методики забезпечує можливість побудови адекватної математичної моделі процесу, якщо експериментальні дані відповідають вимогам *представництва* та *інформативності*. Перша вимога означає, що вибірка даних повинна охоплювати досить довгий проміжок часу, щоб повністю відображати функціонування того режиму процесу, для яких будується модель. Вимога *інформативності* означає, що вибірка повинна містити в собі об'єм інформації, достатній для оцінювання коефіцієнтів моделі. Наприклад, якщо моделюється процес другого порядку, то вибірка повинна забезпечувати коректне обчислення першої та другої похідної за допомогою відповідного полінома, що формально описує статистичні дані.

Іноді інформативність формально оцінюють також за допомогою величини дисперсії процесу, або за кількістю значимих гармонічних складових. Очевидно, що чим більше гармонік (частотних складових) містить вибірка, тим інформативнішою вона є. Кількість основних частот, які несуть 90% енергії сигналу, фактично має дорівнювати кількості похідних, які можна обчислити за поліномом, що описує поведінку процесу.

Умову інформативності даних пов'язують з умовою *достатнього збудження* процесу. Достатнє збудження означає, що вхідний сигнал повинен охоплювати всю смугу частот, які може пропускати на вихід досліджуваній об'єкт. Тобто вхідний сигнал повинен охоплювати всю амплітудно-частотну характеристику процесу. Ця вимога залишається однаковою для об'єктів будь-якої природи.

Статистичні параметри якості нелінійних моделей (логіт, пробіт)

Вище описані відомі методи аналізу фінансових даних, викладені їх основні ідеї, переваги і недоліки. Для розробки інтегрованого підходу необхідно зрозуміти, з якими недоліками на практиці зустрічаються аналітики, які працюють з цими методами, та які критерії слід використовувати для порівняння і аналізу якостей моделей, побудованих на основі цих методів. Нижче будуть розглянуті критерії, які використовуються на практиці, та уніфікований показник GINI, який є найбільш адекватним показником оцінки якості моделі, оскільки не залежить від порогу відсікання [8].

Для аналізу якості моделей і встановлення найкращої моделі для розв'язання певної задачі використовують кілька критеріїв для оцінювання

адекватності моделей:

- 1) загальна точність моделі;
- 2) помилки I-го та II-го роду;
- 3) ROC-крива та індекс GINI.

Загальна точність моделі (CA – common accuracy) визначається так:

$$CA = \frac{Correct\_Forecast}{N}, \quad (2.128)$$

де *Correct_Forecast* – кількість вірно спрогнозованих випадків, а *N* – загальна кількість випадків. Загальна точність моделі є дещо суб'єктивною оцінкою, оскільки вона залежить від частки дефолтів в моделі, а також від порогу відсікання. Для різних значень порогу точність моделі також буде приймати різні значення.

ROC-крива (Receiver Operation Characteristic – робоча характеристика приймача) показує залежність кількості вірно класифікованих позитивних прикладів від кількості невірно класифікованих негативних прикладів. Перші називають істинно позитивними, а інші – негативними множинами. При цьому припускається, що у класифікатора існує певний параметр, варіюючи який, можна отримати певне розбиття на класи. Цей параметр часто називають порогом або точкою відсікання (cut-off), в залежності від якого будуть отримані різні величини помилок I-го та II-го роду (табл. 2. 1).

Таблиця 2.1. Помилки I-го та II-го роду

	Прогноз моделі: Повернення кредиту (0)	Прогноз моделі: Дефолт (1)
Фактично: Повернення кредиту(0)	Вірно класифіковані (TP)	Помилки II-го роду (FN)
Фактично: Дефолт (1)	Помилки I-го роду (FP)	Вірно класифіковані (TN)

Для аналізу якості моделі найчастіше використовують такі відносні показники у процентах:

- ☐ частка істинно позитивних прикладів (True Positives Rate):

$$TPR = \frac{TP}{TP + FN} \quad (2.129)$$

❓ частка хибно позитивних прикладів (False Positives Rate):

$$FPR = \frac{FP}{TN + FP} \quad (2.130)$$

Зазвичай для аналізу якості моделей використовують ще дві характеристики: чутливість та специфічність.

Чутливість моделі – це частка істинно позитивних випадків, тобто

$$Se = TPR = \frac{TP}{TP + FN} \quad (2.131)$$

Специфічність моделі – це частка істинно негативних випадків, які були

вірно класифіковані моделлю: $Sp = \frac{TN}{TN + FP}$. Очевидно, що

$$Sp = \frac{TN + FP - FP}{TN + FP} = 1 - \frac{FP}{TN + FP} = 1 - FPR \quad (2.132)$$

Модель з високою чутливістю надає істинний результат при наявності позитивних випадків (виявляє позитивні приклади). Навпаки, модель із високою специфічністю найчастіше дає істинний результат за наявності негативних випадків (виявляє негативні приклади). Для побудови графіка ROC-кривої по осі Y відкладаються значення чутливості Se , а по осі X – долю хибно позитивних випадків FPR або $1 - Sp$.

Графік ідеального класифікатора ROC-кривої проходить через верхній лівий кут, де доля істинно позитивних випадків становить 1 (ідеальна чутливість), а доля хибно позитивних прикладів дорівнює нулю. Тому чим ближче крива наближається до верхнього лівого кута, тим кращою є здатність моделі передбачувати. Діагональна лінія відповідає класифікатору, який не здатний розпізнати ці два класи.

Оскільки візуальне порівняння ROC-кривих не завжди дозволяє визначити ефективнішу модель, застосовують оцінку площі під кривими.

Чисельний показник площі під кривою AUC (Area Under Curve) і обчислюється, наприклад, за методом трапецій:

$$AUC = \int f(x)dx = \sum_i \left[\frac{X_{i+1} + X_i}{2} \right] \cdot (Y_{i+1} - Y_i) \quad (2.133)$$

Більш зрозумілим і частіше згадуваним в літературі параметром оцінки якості моделі є індекс GINI, розроблений італійським статистиком Коррадо Джіні, який тісно пов'язаний з чисельним показником площі під ROC-кривою. *Індекс GINI* – це площа області між діагоналлю і кривою Лоренца, поділена на площу усієї області під діагоналлю. Індекс GINI широко використовується для аналізу роздільної здатності системи оцінювання при управлінні кредитними ризиками, тобто оцінки здатності моделі розділяти клієнтів на схильних та несхильних до дефолту. Якщо модель здатна оцінити клієнтів за ймовірністю дефолту, то більшість клієнтів, схильних до дефолту, мають отримати більшу ймовірність дефолту. Відповідно, найменша ймовірність дефолту має бути для клієнтів, несхильних до дефолту. Індекс GINI проілюстровано графом, де сукупний процент дефолту для клієнтів показано поряд із сукупним процентом клієнтів, коли вони упорядковані за ймовірністю дефолту (менша ЙД – зліва, більша – справа). Цей графік відомий як крива Лоренца (рис. 2.2).

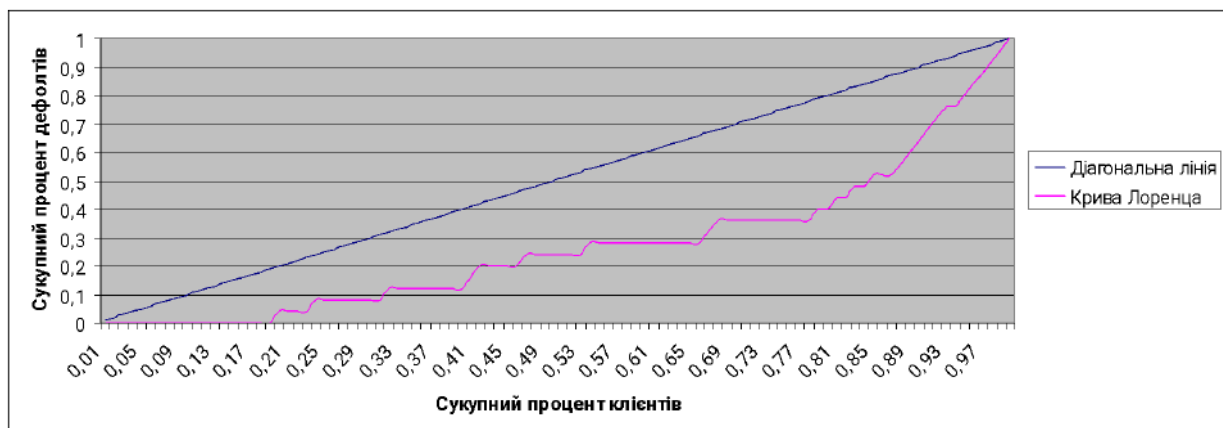


Рис. 2.2 Крива Лоренца для задачі кредитування

Діагональна лінія – це лінія «випадкової моделі». На осі X клієнти впорядковані за заданою ймовірністю дефолту. Очевидно, що 80% дефолтів зустрічаються саме серед останніх 20% клієнтів (20% клієнтів з найбільшою

ймовірністю дефолту). Очевидно, це означає, що модель виконує прийнятне розділення. Отже, чим більша область між діагоналлю та кривою, тим вища якість моделі. Дефолти серед клієнтів із рівною ймовірністю дефолту вважаються розподіленими випадково серед цих клієнтів. Індекс GINI можна визначити через площу фігури, що знаходиться під ROC-кривою, таким чином:

$$GINI = 2 \cdot AUC - 1.$$

Діапазон значень індексу GINI складає $0 \leq G \leq 1$, а моделі з найвищою розділювальною здатністю, тобто моделі, які роблять високоякісне сортування схильних до дефолту клієнтів і клієнтів, не схильних до дефолту, отримують найвищі коефіцієнти. На практиці оцінка якості моделі суттєво залежить від даних, за якими вона будується. Для застосування на практиці скорингу (оцінки фінансового стану нових клієнтів) індекс GINI на рівні 55% є вже дуже високим, в той час як для скорингу поведінки (оцінки фінансового стану існуючих клієнтів) індекс GINI зазвичай приймає значення вищі 70%. В літературі подано шкалу значень індексу GINI (табл. 2.2).

Таблиця 2.2. Оцінка якості моделі за площею AUC та індексом GINI

Інтервал AUC	Індекс GINI	Якість моделі
0,9 – 1,0	0,8 – 1,0	Відмінна
0,8 – 0,9	0,6 – 0,8	Дуже висока
0,7 – 0,8	0,4 – 0,6	Прийнятна
0,6 – 0,7	0,2 – 0,4	Середня
0,5 – 0,6	0 – 0,2	Незадовільна

Значення точок ROC-кривої можуть бути використані для знаходження оптимального порогу відсікання – компромісу між чутливістю та специфічністю моделі. Критеріями вибору порогу відсікання можуть бути:

- 1) вимога мінімальної величини чутливості (специфічності) моделі;
- 2) вимога максимальної сумарної чутливості і специфічності моделі, тобто $cut-off = \max_k (Se_k + Sp_k)$;
- 3) вимога балансу між чутливістю і специфічністю, тобто коли $Sp \approx Se$. $cut-off = \min_k |Se_k - Sp_k|$.

9. Резюме по методиці побудови моделей нелінійних та нестационарних процесів

Більшість процесів, що протікають у сучасній економіці, фінансах, екології і багатьох інших галузях діяльності в Україні та багатьох інших країнах світу, відносяться до класу нелінійних та нестационарних. Це досить просто доводиться за допомогою статистичних даних. Зазначимо, що нелінійні процеси, як правило, містять нестационарність того чи іншого типу. Кожний клас процесів потребує використання для його формального опису різних типів математичних моделей, у багатьох випадках різних методів оцінювання параметрів, різних алгоритмів оцінювання прогнозів.

Для процесів в економіці та фінансах характерні два типи нестационарності: *інтегрованість* та *гетероскедастичність*. Інтегрованість означає наявність тренду. Так, процес з лінійним трендом, позначають ще через $I(1)$, де символ $I(\cdot)$ означає інтегрованість, а цифра у дужках – порядок тренду. Якщо в економіці зустрічаються, як правило, процеси не вище другого порядку інтегрованості, то у фінансах і екології цей порядок може бути набагато вищим, наприклад, зустрічаються процеси $I(5)$ і вище [1, 2]. Для того щоб переконатись у цьому до даних застосовують відповідні статистичні тести на інтегрованість або будують поліноміальну модель тренду і аналізують статистичну значущість оцінених параметрів моделі.

Формально процес із трендом записують так: $E_k(y(k)) \neq const, k=0, \dots, N$, де $E_k(\cdot)$ – умовне математичне сподівання процесу на момент часу k . Умова полягає у тому, що для обчислення математичного сподівання у момент k ми повинні мати всю необхідну для цього інформацію про процес – значення вимірів, оцінки невимірюваних компонент. Необхідність побудови моделі тренду зумовлена тим, що необхідно прогнозувати довгострокові зміни процесів, тобто оцінювати довгострокові прогнози. Якщо одночасно аналізуються два або інтегрованих процеси з трендами однакового порядку, то їх називають коінтегрованими.

Наявність гетероскедастичності формально записують так: $Var(y(k)) \neq const, k=0, \dots, N$, тобто дисперсія процесу змінюється у часі на

інтервалі його дослідження. Тут k – дискретний час – послідовність цілих чисел, які зв'язані з реальним неперервним часом t періодом дискретизації вимірів T_s : $t = kT_s$. Побудова моделі динаміки дисперсії виникає у зв'язку з тим, що необхідно прогнозувати волатильність досліджуваних процесів (стандартне відхилення), яка використовується при побудові моделей ціноутворення, формуванні правил доходності, а також у системах економічної, технічної і медичної діагностики.

Загальна схема аналізу нелінійних нестационарних процесів подана на рис. 2.3. Підготовка даних до побудови математичної або статистичної моделі включає у себе такі обчислювальні операції: нормування у заданому діапазоні, заповнення пропусків з використанням альтернативних методів, обробку екстремальних значень, цифрову та оптимальну фільтрацію, розмноження вибірки і т. ін.

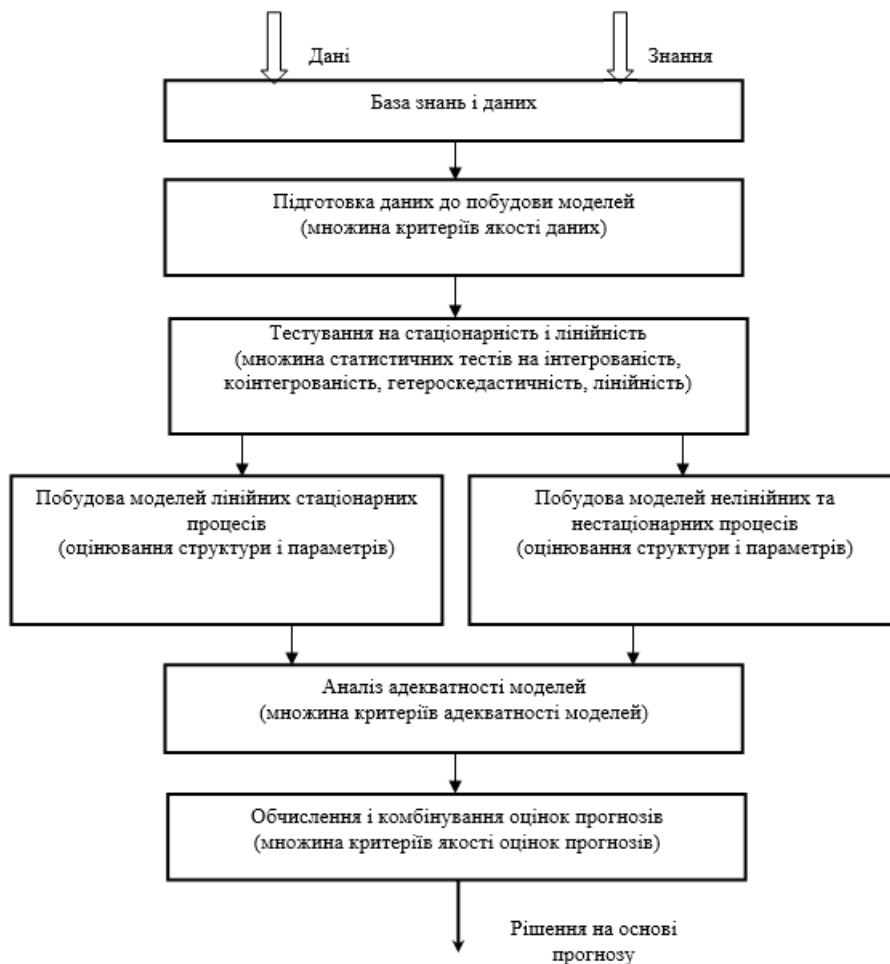


Рис. 2.3 Загальна схема аналізу нелінійних та нестационарних процесів

Для аналізу якості даних використовують множину статистичних критеріїв якості, які необхідні для перевірки інформативності вибірки, наявності екстремальних значень, оцінювання рівня шумових компонент, визначення кількості пропусків. Отримані результати стосовно якості даних використовують для покращення їх характеристик.

На цьому та інших етапах активно використовуються різні модифікації алгоритмів оптимальної фільтрації, які виконують такі функції: оптимальне оцінювання станів процесів; оцінювання невимірюваних компонент вектора стану досліджуваного процесу; заповнення пропусків вимірів; короткострокове прогнозування (як правило, на один крок, тобто один період дискретизації вимірів); адаптивне оцінювання коваріацій випадкових впливів (збурень) на протікання процесів, підготовка вимірів до їх подальшого використання іншими процедурами та моделями, наприклад, у процесі побудови ймовірнісно-статистичних моделей у формі статичних і динамічних байєсівських мереж, дерев рішень, звичайної і байєсівської регресії і т. ін.

Тестування на стаціонарність виконується за допомогою множини відповідних статистичних тестів, які дають можливість встановити наявність тренду і його порядок, а також наявність гетероскедастичності. Оскільки тренд може мати детермінований або стохастичний характер розвитку, то для його формального опису необхідно застосовувати різні типи моделей. Якщо досліджуються коінтегровані процеси, то додатково виконується тестування даних на наявність коінтегрованості. У подальшому позитивний результат стосовно наявності коінтегрованості використовується для побудови відповідної моделі коінтегрованих процесів, яку можна використати, наприклад, для коректного вибору керованих та керуючих змінних.

Після встановлення класу процесу – лінійний стаціонарний або нелінійний та/або нестаціонарний – переходимо до оцінювання структури і параметрів моделі у вибраному класі. Нагадаємо, ми оцінюємо структуру, що складається з 6-и елементів: розмірність моделі (кількість рівнянь, що утворюють модель); порядок моделі (максимальний порядок диференціального або різницевого рівняння); наявність нелінійності та її тип; час запізнення по входу (лаг) та його оцінка (тут можливий багатовимірний випадок); тип випадкового зовнішнього збурення і його параметри; можливі обмеження на ключові змінні моделі та її параметри. Структури моделей-кандидатів оцінюються на основі аналізу статистичних характеристик процесу, законів та закономірностей його протікання, а також накопичених знань стосовно аналізу

процесів конкретного класу. Для одного процесу доцільно оцінювати кілька можливих кандидатів, оскільки побудувати за одну спробу «найкращу» модель практично майже неможливо.

Успішне оцінювання структури дає можливість перейти до оцінювання параметрів (коефіцієнтів) моделей-кандидатів. Можливості стосовно вибору методу оцінювання досить широкі. Лінійні та псевдолінійні моделі можна оцінювати за звичайним методом найменших квадратів (МНК), а моделі, нелінійні стосовно параметрів, оцінюють за нелінійним МНК, методом максимальної правдоподібності (ММП), де це можливо, або за методом Монте-Карло для марковських ланцюгів (МКМЛ). Очевидно, що можна застосовувати інші методи. Останній метод – універсальний для різних типів розподілів даних і типів нелінійності. Загалом, для моделювання лінійних та нелінійних процесів необхідно проектувати та реалізовувати відповідні системи підтримки прийняття рішень на основі принципів системного аналізу, які забезпечать коректне використання методів підготовки даних, оцінювання структури і параметрів моделей, а також належний контроль обчислювальних процесів в системі.

10. Приклади побудови математичних моделей за допомогою статистичних даних

Приклад 2.3. Побудуємо математичні моделі динаміки для наступних макроекономічних процесів України: формування внутрішнього валового продукту (ВВП), індекс споживчих цін (ІСЦ) і грошовий агрегат М3. Для побудови використано фактичні щомісячні дані з січня 1996 по січень 2005 року, всього 109 значень. Кореляційна матриця для цих змінних має вигляд:

ІСЦ	ВВП	М3
1	-0.3043610	-0.2491123
-0.3043610	1	0.9317529
-0.2491123	0.9317529	1

Корельованість ІСЦ з ВВП та агрегатом М3 незначна; в подальшому ця інформація буде використана при побудові альтернативних варіантів математичних моделей процесів, що розглядаються. Корельованість між ВВП і

агрегатом МЗ складає 0,931 – це велике значення, яке може негативно вплинути на якість оцінок моделі при одночасному використанні цих змінних у правій частині рівняння.

10.1. Модель індексу споживчих цін

Авторегресійна модель індексу споживчих цін

Спочатку розглянемо можливість опису індексу споживчих цін за допомогою моделі авторегресії з ковзним середнім. Авторегресійні моделі самі прості за своєю структурою, але досить часто вони мають високу ступінь адекватності досліджуваному процесу, прийнятну для подальшого використання. Автокореляційна функція процесу подана в таблиці 2.3. При побудові моделі індекс споживчих цін будемо позначати як isc .

Таблиця 2.3. Автокореляційна функція процесу формування споживчих цін (ІСЦ)

Часові дані: 1996:01 2005:01						
Всього спостережень: 109						
АКФ	Часткова АКФ		АКФ	ЧАКФ	Q-стат.	Ймов
. ****	. ****	1	0.570	0.570	36.373	0.000
. **	. * .	2	0.230	-0.140	42.378	0.000
. *	. *	3	0.129	0.089	44.284	0.000
. *	. .	4	0.076	-0.022	44.957	0.000
. .	. .	5	0.064	0.043	45.433	0.000
. *	. *	6	0.120	0.101	47.137	0.000
. *	. *	7	0.180	0.088	50.960	0.000
. *	. * .	8	0.068	-0.136	51.515	0.000
. .	. *	9	0.052	0.105	51.840	0.000
. *	. .	10	0.072	-0.001	52.468	0.000

В таблиці 10.1 скорочення АКФ – автокореляційна функція, а ЧАКФ – часткова АКФ.

Як видно з АКФ, при побудові моделей необхідно починати з моделей нижчих порядків, які часто мають прийнятну адекватність процесу і забезпечують високу якість прогнозу. Результати оцінювання моделі авторегресії першого порядку за методом найменших квадратів подані в таблиці 2.4.

Таблиця 2.4. Результати оцінювання моделі AP(1) для ІСЦ

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1996:02 2005:01				
Використано спостережень: 108 після корегування				
Модель AP(1): $I=C(1)+C(2)*I(-1)$				
	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	43.29386	6.624742	6.535178	0.0000
C(2)	0.569826	0.065703	8.672790	0.0000
R-квадрат	0.415067	Середнє залежної змінної		100.7407
Скорегований R-квадрат	0.409549	Станд. відхил. зал. змінної		1.506189
Станд. похибка регресії	1.157368	Інформ. критерій Акайке		3.148519
Сума квадратів похибок	141.9871	Критерій Шварца		3.198188
		Стат. Дарбіна-Уотсона		1.931805

Отримана модель AP(1):

$$\begin{aligned}
 isc(k) &= a_0 + a_1 isc(k-1) + \varepsilon(k) = \\
 &= 43.294 + 0.57 isc(k-1) + e(k)
 \end{aligned}
 \tag{2.134}$$

де $e(k)$ – залишки (похибки) моделі, значення яких можна знайти у відповідному файлі комп'ютерної системи, що застосовується для побудови моделі. Статистичні характеристики адекватності моделі:

$$R^2 = 0,415; \quad J = CKП = 141,99; \quad DW = 1,931. \quad (2.135)$$

Коефіцієнт детермінації має низьке значення (0,415), сума квадратів похибок (СКП) досить високе (141,99), а статистика Дарбіна-Уотсона (1,931) наближається до найкращого значення. Таким чином, загалом адекватність моделі AP(1) досить низька, а тому структура моделі потребує уточнення. Характеристики якості (історичного – на навчальній вибірці) однокрокового прогнозу такі:

$$CeКП = 1,36; \quad CAП = 1,02; \quad CAПП = 1,008; \quad U = 0,0067, \quad (2.136)$$

тобто середньоквадратична похибка ($CeКП$), середня абсолютна похибка ($CAП$), середня абсолютна похибка в процентах ($CAПП$) і коефіцієнт Тейла, який свідчить про загальну придатність моделі для прогнозування (ідеальне значення – наближення до нуля). Критерій $CAПП$ свідчить про високу якість короткострокового прогнозу.

В таблиці 2.5 подані результати оцінювання авторегресійної моделі AP(3). Всі три коефіцієнти моделі статистично значимі.

Таблиця 2.5. Результати оцінювання моделі AP(3)

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1996:04 2005:01				
Всього спостережень після корегування крайніх значень: 106				
Моделі: $I = C(1) + C(2) * I(-1) + C(3) * I(-2) + C(4) * I(-3)$				
	Оцінки коэф-в	Станд. похиб.	t-статистика	Ймов.

C(1)	46.96556	9.330171	5.033729	0.0000
C(2)	0.613458	0.098563	6.223995	0.0000
C(3)	-0.157670	0.113613	-1.387787	0.1682
C(4)	0.077517	0.086493	0.896229	0.3722
R-квадрат	0.316655	Середнє залежної змінної		100.6604
Скорегований R-квадрат	0.296556	Станд. відхил. зал. змінної		1.372428
Станд. похибка регресії	1.151076	Інформ. критерій Акайке		3.156277
Сума квадратів похибок	135.1476	Критерій Шварца		3.256785
		Стат. Дарбіна-Уотсона		1.992389

Таким чином, можна записати таку лінійну модель:

$$\begin{aligned}
 isc(k) &= a_0 + a_1 isc(k-1) + a_2 isc(k-2) + a_3 isc(k-3) + \varepsilon(k) = \\
 &= 46,96 + 0,61isc(k-1) - 0,15isc(k-2) + 0,08isc(k-3) + e(k),
 \end{aligned}
 \tag{2.137}$$

Для цієї моделі спостерігається зменшення коефіцієнта детермінації від 0,415 до 0,317 (деяке погіршення); зменшення суми квадратів похибок від 141,99 до 135,148 і покращення статистики Дарбіна-Уотсона: від 1,931 до 1,992 (тобто похибки моделі можна вважати практично некорельованими). Отримані значення характеристик адекватності моделі:

$$R^2 = 0,317; \quad J = СКП = 135,148; \quad DW = 1,992. \tag{2.138}$$

Характеристики однокрокового прогнозу для даної моделі:

$$CeKP = 1,36; \quad CAП = 1,02; \quad CAПП = 1,01; \quad U = 0,0068, \tag{2.139}$$

тобто модель загалом придатна для прогнозування (за коефіцієнтом Тейла, який наближається до ідеального значення, 0,0068), а три інших показники свідчать про високу точність прогнозу. Необхідно зазначити, що статистичні показники якості прогнозу для моделей AP(1) і AP(3) є практично однаковими.

Розглянемо характеристики моделі вищого порядку. Згідно з аналізом автокореляційної функції, в таблиці 2.6 подані результати оцінювання моделі AP(7).

Таблиця 2.6. Результати оцінювання моделі AP(7) для ІСЦ

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1996:08 2005:01				
Всього використано спостережень після корегування крайніх значень: 102				
Модель: $I = C(1) + C(2) * I(-1) + C(3) * I(-2) + C(4) * I(-3) + C(5) * I(-5) + C(6) * I(-6) + C(7) * I(-7)$				
	Оцінки коэф-в	Станд. похибка	t- статистика	Ймов.
C(1)	33.13446	12.79240	2.590168	0.0111
C(2)	0.603211	0.101625	5.935625	0.0000
C(3)	-0.184907	0.117804	-1.569623	0.1198
C(4)	0.125878	0.101891	1.235415	0.2197
C(5)	-0.031530	0.102954	-0.306254	0.7601
C(6)	0.093246	0.117601	0.792904	0.4298
C(7)	0.064850	0.089527	0.724365	0.4706
R-квадрат	0.346348	Середнє залежної змінної		100.6667
Скорегований R- квадрат	0.305064	Станд. відхил. зал. змінної		1.388306
Станд. похибка регресії	1.157330	Інформ. критерій Акайке		3.196268

Сума квадратів похибок	127.2443	Критерій Шварца	3.376413
		Стат. Дарбіна-Уотсона	1.811821

Порівнюючи моделі $AP(3)$ і $AP(7)$, можна сказати, що коефіцієнт детермінації збільшився (від 0,317 до 0,346; сума квадратів похибок моделі зменшилась від 135,15 до 127,24, а статистика Дарбіна-Уотсона зменшилась від 1,992 до 1,811). Характеристики однокрокового прогнозу для моделі $AP(7)$: $SeKP = 1,36$; $CAП = 1,012$; $CAПП = 1,002$; $U = 0,0067$,

тобто на 0,008 зменшились середня абсолютна похибка і середня абсолютна похибка в процентах.

Результати оцінювання моделі 12-го порядку подані в таблиці 2.7.

Таблиця 2.7. Результати оцінювання моделі $AP(12)$ для ІСЦ

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1997:01 2005:01				
Всього використано спостережень після корегування крайніх значень: 97				
Модель: $I = C(1) + C(2) \cdot I(-1) + C(3) \cdot I(-2) + C(4) \cdot I(-3) + C(5) \cdot I(-6) + C(6) \cdot I(-7) + C(7) \cdot I(-8) + C(8) \cdot I(-9) + C(9) \cdot I(-11) + C(10) \cdot I(-12)$				
	Оцінки коеф-в	Станд. похибка	t-статистика	Ймов.
C(1)	26.23923	14.28763	1.836500	0.0697
C(2)	0.666410	0.102198	6.520741	0.0000
C(3)	-0.247996	0.122488	-2.024654	0.0460
C(4)	0.179378	0.103248	1.737345	0.0859
C(5)	-0.038240	0.096715	-0.395392	0.6935
C(6)	0.016650	0.112406	0.148120	0.8826
C(7)	-0.037042	0.111025	-0.333639	0.7395

C(8)	0.030055	0.095744	0.313915	0.7543
C(9)	0.020445	0.093392	0.218911	0.8272
C(10)	0.149194	0.081977	1.819964	0.0722
R-квадрат	Статистичн і 0.435002	<u>характеристики якості</u> <u>моделі</u> Середнє залежної змінної		100.6082
Скорегований R-квадрат	0.376553	Станд. відхил. зал. змінної		1.342857
Станд. похибка регресії	1.060301	Інформ. критерій Акайке		3.052366
Сума квадратів похибок	97.80880	Критерій Шварца		3.317800
		Статистика Дарбіна-Уотсона		1.940130

Модель AP(12) має кращі характеристики адекватності ніж попередні моделі, а саме: $R^2 = 0,435$; $J = СКП = 97,80$; $DW = 1,94$.

Значно зменшилась сума квадратів похибок, підвищилось значення R^2 , а значення статистики Дарбіна-Уотсона майже таке ж, як для моделі AP(3).
Характеристики однокрокового прогнозу для цієї моделі: $СеКП = 1,337$; $САП = 1,02$; $САПП = 1,013$; $U = 0,0066$.

Таким чином, характеристики однокрокового прогнозу також найкращі для моделі AP(7). Можна зробити висновок, що процес формування індексу оптових цін може бути описаний моделлю авторегресії AP(7) з високим ступенем адекватності. Ця модель забезпечує також отримання кращого однокрокового прогнозу.

Модель авторегресії для відхилень ІСЦ від середнього

Якщо із вихідних (фактичних) значень ряду ІСЦ відняти середнє ($isc(k) - 100,66$), $k = 1, \dots, 109$, то отримаємо ряд відхилень від середнього. Автокореляційна функція залишається фактично незмінною. Результати оцінювання моделі AP(1) подані в таблиці 2.8.

Три вибрані статистичні характеристики адекватності цієї моделі:

$$R^2 = 0,415; \quad J = CKП = 141,99; \quad DW = 1,931.$$

Характеристики якості однокрокового прогнозу такі:
 $SeCKП = 1,36; \quad CAП = 1,02; \quad CAПП = 100,40; \quad U = 0,662.$

Таблиця 2.8. Результати оцінювання моделі AP(1) для відхилень ІСЦ

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1996:02 2005:01				
Всього використано спостережень після корегування крайніх значень: 109				
Модель: $IV = C(1) + C(2) * IV(-1)$				
	Оцінки коеф-в	Станд. Похибка	t-статистика	Ймов.
C(1)	-0.007477	0.111831	-0.066857	0.9468
C(2)	0.569826	0.065703	8.672790	0.0000
R-квадрат	0.415067	Середнє залежної змінної		0.080741
Скорегований R-квадрат	0.409549	Станд. відхил. зал. змінної		1.506189
Станд. похибка регресії	1.157368	Інформ. критерій Акайке		3.148519
Скма квадратів похибок	141.9871	Критерій Шварца		3.198188
		Стат. Дарбіна-Уотсона		1.931805

Порівняно з моделлю AP(1) для повних значень вимірів (без віднімання середнього) середньоквадратична похибка і середня абсолютна похибка не змінились, але в 100 разів збільшилась $CAПП$ і коефіцієнт Тейла: від 0,0067 до 0,662. Таким чином, модель AP(1) для відхилень не придатна для прогнозування. Це можна пояснити тим, що відхилення мають різні знаки, що чисельно утруднює побудову моделі.

Авторегресія 13-го порядку для відхилень ІСЦ від середнього:

Таблиця 2.9. Результати оцінювання моделі AP(13) для відхилень ІСЦ.

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1997:02 2005:01				
Всього використано спостережень після корегування крайніх значень: 96				
Модель: $IV = C(1) + C(2) * IV(-1) + C(3) * IV(-2) + C(4) * IV(-3) + C(5) * IV(-5) + C(6) * IV(-6) + C(7) * IV(-7) + C(8) * IV(-8) + C(9) * IV(-9) + C(10) * IV(-11) + C(11) * IV(-12) + C(12) * IV(-13)$				
	Оцінки коэф-в	Станд. Похибка	t-статистика	Ймов.
C(1)	-0.045818	0.110205	-0.415756	0.6786
C(2)	0.706976	0.107583	6.571429	0.0000
C(3)	-0.261301	0.123875	-2.109395	0.0379
C(4)	0.194879	0.105046	1.855172	0.0671
C(5)	-0.110121	0.107924	-1.020352	0.3105
C(6)	0.043907	0.120182	0.365334	0.7158
C(7)	0.002351	0.115561	0.020347	0.9838
C(8)	-0.023657	0.114078	-0.207375	0.8362
C(9)	0.033030	0.096947	0.340705	0.7342
C(10)	0.002612	0.096331	0.027112	0.9784
C(11)	0.187169	0.109248	1.713242	0.0904
C(12)	-0.074932	0.085883	-0.872494	0.3854
R-квадрат	0.440922	Середнє зал. Змінної		-0.066250
Скорегований R-квадрат	0.367709	Станд. відхил. зал. змінної		1.342254
Станд. похибка регресії	1.067316	Інформ. крит. Акайке		3.084639
Сума квадратів похибок	95.68969	Критерій Шварца		3.405183

Логарифм правдоподібн.	-136.0627	Статистика Дарбіна-Уотсона	1.965193

Три вибрані статистичні характеристики адекватності цієї моделі мають такі значення: $R^2 = 0,440$; $J = СКП = 95,69$; $DW = 1,965$.

Характеристики якості однокрокового прогнозу: $СеКП = 1,346$; $САП = 1,02$; $САПП = 109,09$; $U = 0,848$.

Спостерігається погіршення характеристик прогнозу, особливо погіршилось значення $САПП$ і коефіцієнта Тейла: від 0,662 для моделі $AP(1)$ до 0,848 для моделі $AP(13)$. За цим параметром модель є непридатною для прогнозування. Коефіцієнти $C(6) - C(10)$ є незначущими, а тому їх можна видалити з моделі.

Після вилучення цих коефіцієнтів із моделі та відповідних їм складових процесу отримаємо модель, параметри якої подані в таблиці 2.10.

Таблиця 2.10. Результати оцінювання моделі $AP(13)$ для відхилень

Метод оцінювання: метод найменших квадратів (МНК)				
Скорегована часова вибірка даних: 1997:02 – 2005:01				
Всього використано спостережень після корегування крайніх значень: 96				
Модель: $IV = C(1) + C(2) * IV(-1) + C(3) * IV(-2) + C(4) * IV(-3) + C(5) * IV(-5) + C(6) * IV(-12) + C(7) * IV(-13)$				
	Оцінки коэф-в	Станд. похибка	t-статистика	Ймов.
C(1)	-0.046129	0.107257	-0.430077	0.6682
C(2)	0.700062	0.103640	6.754728	0.0000
C(3)	-0.256660	0.120078	-2.137440	0.0353
C(4)	0.195599	0.101703	1.923247	0.0576
C(5)	-0.081877	0.083177	-0.984368	0.3276

C(6)	0.190417	0.089745	2.121760	0.0366
C(7)	-0.067455	0.081058	-0.832180	0.4075
R-квадрат	0.438765	Середнє залеж. змінної		-0.066250
Скорегований R-квадрат	0.400929	Станд. відхил. залеж. змінної		1.342254
Станд. похибка регресії	1.038900	Інформ. критерій Акайке		2.984323
Сума квадратів похибок	96.05881	Критерій Шварца		3.171307
Логарифм правдоподібн.	-136.2475	Статистика Дарбіна-Уотсона		1.959533

Вибрані статистичні характеристики адекватності цієї моделі:
 $R^2 = 0,439$; $J = СКП = 96,05$; $DW = 1,96$.

Характеристики якості однокрокового прогнозу:
 $СеКП = 1,346$; $САП = 1,017$; $САПП = 108,2$; $U = 0,854$.

Таким чином, після видалення з моделі незначимих коефіцієнтів її характеристики залишились практично незмінними. В цілому можна зробити висновок, що побудовані авторегресійні моделі для відхилень ІСЦ від середнього не придатні для прогнозування. Це свідчить про те, що побудувати модель для процесу, який має різні знаки вимірів в різні моменти часу (тобто розвиток процесу відбувається в двох квадрантах), що є складнішим ніж для процесу, який змінюється в межах одного квадранту.

Авторегресія з ковзним середнім для ІСЦ

Розглянемо можливість формального опису ІСЦ за допомогою моделі АРКС. Характеристики моделі АРКС(1,1) подані в таблиці 2.11.

Таблиця 2.11. Результати оцінювання моделі АРКС(1, 1) для ІСЦ

Метод оцінювання: метод найменших квадратів
Скорегована часова вибірка даних: 1996:02 2005:01

Всього використано спостережень після корегування крайніх значень: 108				
Змінна	Оцінки коеф-в	Станд. похибка	t- статистика	Ймов.
C	100.6529	0.263645	381.7742	0.0000
AR(1)	0.546826	0.093191	5.867793	0.0000
KC(1)	0.060696	0.133153	0.455838	0.6494
R-квадрат	0.415932	Середнє залеж. змінної		100.7407
Скорегований R-квадрат	0.404807	Станд. відхил. залеж. змінної		1.506189
Станд. похибка регресії	1.162006	Інформ. критерій Акайке		3.165558
Сума квадратів похибок	141.7772	Критерій Шварца		3.240062
Логарифм правдоподібн.	-167.9401	F-статистика		37.38676
Статист. Дарбіна-Уотсона	1.996202	Ймовірність (F-стат.)		0.000000
Інвертовані AP корені	.55			
Інвертовані KC корені	-.06			

Вибрані статистичні характеристики адекватності цієї моделі:
 $R^2 = 0,416$; $J = CKП = 141,78$; $DW = 1,996$.

Характеристики якості однокрокового прогнозу:
 $SeCKП = 1,362$; $CAП = 1,016$; $CAПП = 1,005$; $U = 0,0067$.

Моделі AR(1) і ARKC(1,1) мають практично однакові характеристики адекватності та якості однокрокового прогнозу, а тому перевагу (при виборі з цих двох моделей) можна надати моделі AR(1), яка є простішою за структурою. Нижче подана порівняльна таблиця для всіх побудованих

моделей.

Врахування впливу на ІСЦ агрегату МЗ

Коефіцієнт кореляції між ІСЦ та агрегатом МЗ від'ємний: $-0,249$, тобто формальний взаємозв'язок між цими змінними незначний, але цікаво розглянути вплив МЗ на ІСЦ за допомогою моделі. Врахування регресора може покращити деякі характеристики моделі, а також врахувати наявний причинний зв'язок між вибраними змінними. Характеристики змішаної моделі: авторегресія $AR(1)$ плюс парна регресія подані в таблиці 2.12.

Таблиця 2.12. Результати оцінювання моделі ІСЦ: $AR(1)$ + регресор МЗ

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1996:02 2005:01				
Всього використано спостережень після корегування крайніх значень: 108				
Модель: $I=C(1)+C(2)*I(-1)+C(3)*M$				
	Оцінки коэф-в	Станд. похибка	t- статистика	Ймов.
C(1)	44.97869	6.920057	6.499757	0.0000
C(2)	0.554264	0.068261	8.119764	0.0000
C(3)	-2.82E-06	3.30E-06	-0.854463	0.3948
R-квадрат	0.419106	Середнє залеж. змінної		100.7407
Скорегований R- квадрат	0.408041	Станд. відхил. залеж. змінної		1.506189
Станд. похибка регресії	1.158844	Інформ. критерій Акайке		3.160108
Сума квадратів похибок	141.0066	Критерій Шварца		3.234612
Логарифм правдоподібн.	-167.6458	Стат. Дарбіна-Уотсона		1.919823

Спостерігається незначне покращення характеристик моделі та якості прогнозу порівняно з $AR(1)$, але коефіцієнт при МЗ дуже малий ($-2,82E-06$). Однак, формально він є значимим. Таким чином, на розглянутому часовому

інтервалі вплив агрегату МЗ на індекс споживчих цін незначний і ним можна знехтувати.

Можна припустити, що в обороті було недостатньо грошової маси для того, щоб її вплив на ІСЦ був істотним. З іншого боку, недостатній об'єм грошової маси в національній валюті компенсувався (і продовжує компенсуватись) “твердою” іноземною валютою, зокрема доларами США і, в деякій мірі, євро. Таким чином, слабка українська економіка позитивно впливає на курс долара США завдяки фактичному введенню його в частковий оборот на відносно великій території України. Встановити фактичне співвідношення між об'ємами національної та іноземної валют в обороті можливо, але для цього необхідно отримати додаткові статистичні дані і виконати спеціальне дослідження. Зокрема, необхідно мати дані щодо об'ємів реалізації торговельних операцій у валюті підприємствами всіх форм власності. Очевидно, що отримати такі дані для тіньового обороту непросто і це пов'язано з деяким ризиком.

Характеристики змішаної моделі: авторегресія $AR(1)$ плюс парна регресія подані в таблиці 2.13.

Метод оцінювання: метод найменших квадратів				
Скорегована часова вибірка даних: 1996:02 2005:01				
Всього використано спостережень після корегування крайніх значень: 108				
Модель: $I=C(1)+C(2)*I(-1)+C(3)*V$				
	Оцінки коєф-в	Станд. похибка	t- статистика	Ймов.

C(1)	45.53125	7.156626	6.362112	0.0000
C(2)	0.549611	0.070124	7.837741	0.0000
C(3)	-1.37E-05	1.64E-05	-0.833627	0.4064
R-квадрат	0.418913	Середнє залеж. змінної		100.7407
Скорегований R-квадрат	0.407844	Станд. відхил. залеж. змінної		1.506189
Станд. похибка регресії	1.159037	Інформ. критерій Акайке		3.160441
Сума квадратів похибок	141.0536	Критерій Шварца		3.234945
Логарифм правдоподібн.	-167.6638	Стат. Дарбіна-Уотсона		1.916289

О т р и м а н о

р і в н я н н я :

$$isc(k) = 45,53 + 0,55 \, isc(k-1) - (1,37E-05) \, vvp(k) + e(k)$$

з характеристиками: $R^2 = 0,419$; $J = СКП = 141,05$; $DW = 1,92$.

Х а р а к т е р и с т и к и я к о с т і о д н о к р о к о в о г о п р о г н о з у :
 $СеКП = 1,335$; $САП = 1,004$; $САПП = 0,993$; $U = 0,0066$.

Коефіцієнт при змінній *ВВП* невеликий і знаходиться на межі статистичної значимості, тобто вплив *ВВП* на *ІСЦ* незначний. Причиною такого незначного впливу може бути некоректний розподіл *ВВП*, який призводить до того, що більшість населення проживає на межі або навіть нижче межі бідності. Одночасно інша (менша) частина привласнює більшість благ і користується ними, але, очевидно, що це користування не приводить до позитивного впливу на *ІСЦ*. Тому і маємо, що *ВВП* зростає начебто значними темпами, а добробут населення сильно від нього відстає. Таким чином, можна зробити висновок, що на часовому періоді, який розглядається в даному прикладі, відбувався несправедливий розподіл суспільних благ, що проявилось, в даному випадку, в незначному впливі *ВВП* на споживчі ціни для більшості населення України. Побудована математична модель є формальним (об'єктивним) підтвердженням відомого факту нерівномірного розподілу матеріальних благ, вироблених в Україні.

В таблиці 2.14 зведені характеристики математичних моделей, побудованих для індексу споживчих цін, і характеристики однокрокових прогнозів, обчислених на основі цих моделей. Ця таблиця дає можливість оперативно порівняти результати моделювання та прогнозування, отримані за допомогою запропонованої методики, а також встановити можливість практичного використання отриманих результатів.

Таблиця 2.14. Результати моделювання і однокрокового прогнозування індексу оптових цін

Тип моделі	Характеристики моделі			Характеристики прогнозу			
	R^2	$\sum e^2(k)$	DW	Se_{KP}	$CAП$	$CAПП$	Коеф-т Тейла
AP(1)	0,415	141,99	1,931	1,360	1,020	1,008	0,0067
AP(3)	0,317	135,148	1,992	1,360	1,020	1,011	0,0068
AP(7)	0,346	127,244	1,811	1,360	1,012	1,002	0,0067
AP(12)	0,435	97,80	1,941	1,337	1,020	1,013	0,0066
APKC(1,1)	0,416	141,78	1,996	1,362	1,016	1,005	0,0067
AP(1)+M3	0,419	141,007	1,919	1,340	1,004	0,994	0,0066
AP(1)+ВВП	0,419	141,054	1,916	1,335	1,004	0,993	0,0066
Моделі для відхилень ІСЦ від середнього							
AP(1)	0,415	141,99	1,931	1,360	1,020	100,40	0,662
AP(13)	0,440	95,69	1,965	1,346	1,020	109,09	0,848

Результати моделювання, подані в таблиці 2.12, свідчать про те, що практично всі моделі, побудовані для ІСЦ (перші сім моделей), є придатними для прогнозування, оскільки коефіцієнт Тейла вимірюється тисячними долями. Характеристики однокрокового прогнозу для моделей різного порядку відрізняються несуттєво. Найкращі характеристики щодо прогнозування має модель AP(1)+ВВП. Для моделі AP(12) отримано найменше значення суми квадратів похибок моделі, але характеристики прогнозів, отриманих за цією моделлю, не кращі від інших.

Дві моделі, які побудовані для відхилень ІСЦ від середнього, мало

придатні для прогнозування. Про це свідчать високі значення коефіцієнта Тейла та середньої абсолютної похибки в процентах. Цей факт можна пояснити тим, що описати за допомогою АРКС процес, динаміка якого спостерігається у двох квадрантах, трудніше, ніж процес, який спостерігається в одному квадранті. Тому для прогнозування відхилень від середнього необхідно знайти іншу структуру математичної моделі, придатну для опису різно-знакових величин часового ряду.

Приклад 2.4. Прогнозування напрямку руху ціни біржового активу на основі статистичних даних.

Прогнозування руху цін активів на фінансових біржах знаходиться в центрі уваги багатьох дослідників, оскільки для прийняття коректних рішень щодо виконання торгових операцій необхідні високоякісні прогнози. Процеси ціноутворення на біржах мають, як правило, нестационарний стохастичний характер, що потребує застосування спеціальних процедур моделювання і прогнозування. Як правило, це процеси з детермінованими або стохастичними трендами; також можлива наявність гетероскедастичності. Прогнозування тренду та короткострокових коливань, які накладаються на нього, можна виконувати за допомогою моделей авторегресії з інтегрованим ковзним середнім (АРІКС), нейронних мереж, моделей байєсівської регресії, методу групового врахування аргументів (МГВА) та деяких інших типів моделей. Кожний метод має свої переваги та недоліки щодо практичного застосування. Одним із суттєвих недоліків більшості методів є наявність затримки в часі на 1-2 кроки оцінок прогнозу по відношенню до фактичного розвитку процесу ціноутворення. Використання нейромереж і МГВА супроводжуються значними обчислювальними витратами на навчання моделей; досить проблемною залишається „прозорість” моделей, отриманих за цими методами. Так, модель побудована за МГВА, буває складною щодо візуалізації отриманого результату і, відповідно, ускладнюється розуміння структури такої моделі. Відносно простими у використанні є моделі множинної регресії та АРІКС (авторегресія з інтегрованим ковзним середнім), для яких існує достатньо глибоко розроблена теорія для аналізу результатів моделювання і прогнозування. Крім того, для моделей АРІКС можна будувати функції прогнозування на довільну кількість кроків, що суттєво спрощує весь процес прогнозування та аналізу статистичних характеристик якості прогнозів.

Розглянемо можливість моделювання і прогнозування процесів ціноутворення за допомогою моделей множинної регресії з використанням індикаторів технічного аналізу. Моделі будуються на основі фактичних даних

щодо ціноутворення та індикаторів, які використовуються трейдерами при виконанні торговельних операцій. Для опису динаміки процесу вперше пропонується використати нелінійну модель логістичної регресії.

Постановка задачі. Необхідно побудувати адаптивні моделі для прогнозування напрямків руху нестационарного нормально розподіленого нестационарного процесу $\{y(k)\}$, $k=0,1,2, \dots$, $E[y(k)] \neq const$, $E\{y(k) - E[y(k)]\}^2 \neq const$, ціноутворення на міжнародному валютному ринку. При побудові моделей використати властивості індикаторів технічного аналізу, зокрема індикатора Pivot Point; встановити можливість застосування моделі логістичної регресії як інструменту прогнозування процесів з нелінійностями.

Індикатор технічного аналізу Pivot Point

Відомо, що досвідчені трейдери на біржах досить успішно використовують індикатори технічного аналізу, які формуються на основі даних щодо фактичного руху цін протягом визначеного (вибраного) проміжку часу. Одним із найпопулярніших індикаторів технічного аналізу, які використовуються як трейдерами, так і відомими технічними аналітиками фінансових структур, є індикатор *Pivot Point*. Вирази для розрахунку індикатора подано в таблиці 2.15.

Таблиця 2.15 Індикатор Pivot Points

№ п/п	Позначення	Вираз для обчислення
1	$R1$	$2 \text{ Pivot} - L$
2	$R2$	$\text{Pivot} + (H - L)$
3	$R3$	$H + 2 (\text{Pivot} - L)$
4	Pivot	$(H + L + C) / 3$
5	$S1$	$2 \text{ Pivot} - H$
6	$S2$	$\text{Pivot} - (H - L)$
7	$S3$	$L - 2 (H - \text{Pivot})$

В таблиці використано такі позначення: Н – найвища ціна минулого дня;

L – найнижча ціна минулого дня; C – ціна закриття минулого дня; Pivot – “точка” розвороту; S1 – перший рівень підтримки; S2 – другий рівень підтримки; S3 – третій рівень підтримки; R1 – перший рівень спротиву; R2 – другий рівень спротиву; R3 – третій рівень спротиву.

Принцип роботи індикатора такий: якщо ціна відкриття вища за Pivot і рух ціни починається донизу, то з точки зору технічного аналізу існує ймовірність того, що після досягнення значення Pivot ціна піде вгору. У випадку, коли цього не сталося і ціна продовжує прямувати донизу, з точки зору технічного аналізу існує ймовірність розвороту руху ціни або зупинки руху ціни донизу при підході до рівня S1. Така ж сама ситуація має місце для S2 та S3. Зазначимо, що до значення S3 ціна доходить дуже рідко. Але якщо таке трапляється і ціна опускається нижче рівня S3, то ймовірність майбутнього руху вниз є дуже незначною. Аналогічна ситуація має місце у випадку руху ціни угору (тільки замість S1, S2, S3 необхідно користуватись індикаторами R1, R2, R3).

Побудова регресійних моделей. У якості вхідних даних використано котирування щоденних мінімальних та максимальних цін відношення австралійського долара до канадського долара за період з 2007 по 2008 роки (249 значень) на міжнародному валютному ринку Форекс (FOREX). Структурна схема системи адаптивного формування оцінок прогнозів подана на рис. 2.4.

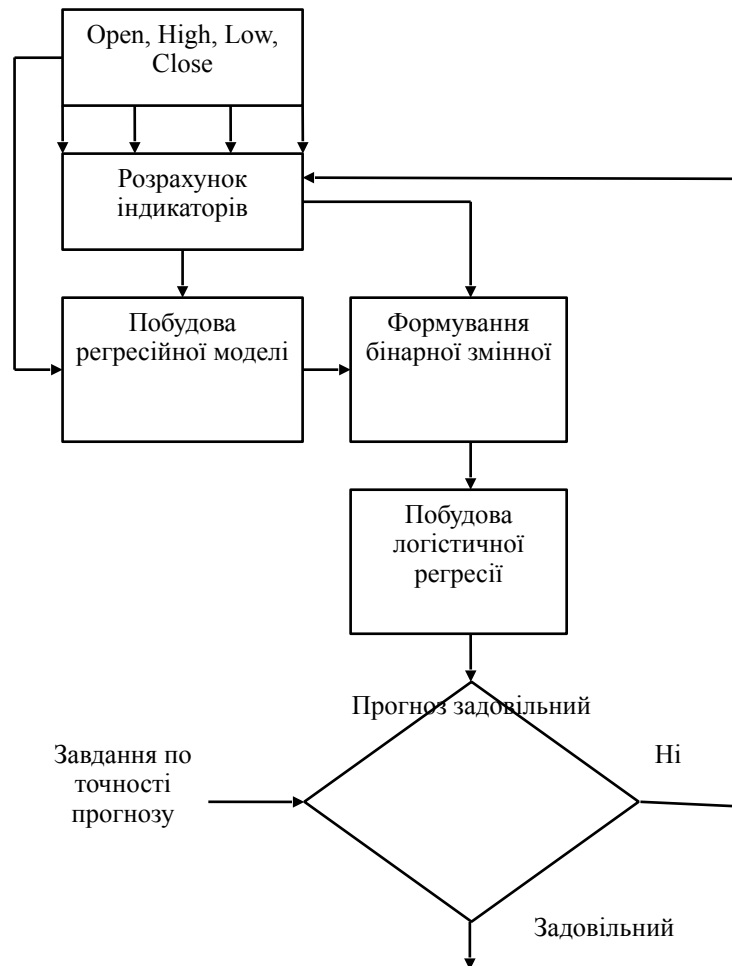


Рис. 2.4 Схема адаптивного прогнозування руху ціни

Вихідною величиною для адаптації моделі до даних є похибка прогнозу. Якщо вона перевищує заданий рівень, то виконується повторний вибір індикаторів (регресорів) для включення в праву частину моделі і процес прогнозування повторюється. Оцінка прогнозу (напрямку руху ціни) формується за допомогою моделі логістичної регресії, яка може описувати процеси з нелінійностями. Тобто вихід логістичної регресії вказує на напрямок руху процесу.

Спочатку побудуємо модель множинної регресії. Для прогнозування максимальної ціни на наступний день логічно вибрати у якості регресорів $R1$, $R2$, $R3$. Отримали таку модель:

$$y(k) = 0,0414 + 1,3466 \cdot R1(k) - 0,1195 \cdot R2(k) - 0,268 \cdot R3(k),$$

де k – дискретний час. В подану вище модель не включена авторегресійна

складова, оскільки це призводить до виродженості матриці вимірів. Побудована модель має такі основні статистичні характеристики якості:

$$R^2 = 0,9893; \quad SSR = 0,0095; \quad AIC = -7,306; \quad DW = 2,1317$$

де R^2 – коефіцієнт множинної детермінації; SSR – сума квадратів похибок моделі; AIC – інформаційний критерій Акайке; DW – статистика Дарбіна-Уотсона. Всі характеристики є задовільними. Побудована модель має такі статистичні характеристики якості однокрокового (історичного) прогнозу: $СКП = 0,0061$; $САП = 0,0044$; $САПП = 0,4955\%$; $U = 0,0034$, де $СКП$ – середньоквадратична похибка; $САП$ – середня абсолютна похибка; $САПП$ – середня абсолютна похибка в процентах; U – коефіцієнт Тейла.

Для прогнозування мінімальної ціни у якості регресорів візьмемо S1, S2,

S3. Модель має вигляд:

$$y(k) = -0,0458 + 1,9359 \cdot S1(k) - 0,8737 \cdot S2(k) + 0,0189 \cdot S3(k)$$

Побудована модель має такі основні статистичні характеристики адекватності: $R^2 = 0,9842$; $SSR = 0,0176$; $AIC = -6,6861$; $DW = 2,1242$.

Статистичні характеристики однокрокового прогнозу на історичному (навчальному) відрізку ряду статистичних даних: $СКП = 0.0084$; $САП = 0.0051$; $САПП = 0.5952\%$; $U = 0.0047$.

Побудовані моделі цілком мають прийнятні статистичні характеристики. Зазначимо, що кількість співпадань напрямків руху оцінок прогнозу для першої моделі дорівнює 171 (або 68,95%), а для другої – 183 (або 73,79%).

Побудова моделей логістичної регресії та дерева класифікацій

Оскільки в процесах формування цін біржових активів зустрічаються участки з нелінійністю довільного характеру, то для формального опису таких даних необхідно використовувати нелінійні моделі. Однією з досить простих моделей такого типу є логістична регресія, яка використана в даній роботі. Для прогнозування напрямку руху ціни побудуємо моделі логістичної регресії та дерева класифікації. Якщо в момент часу $t + 1$ ціна активу виявляється вищою ніж в момент часу t , то позначимо це зростання через „1”, а спадання відповідно через „0”. Ці значення використані як вхідні для моделі логістичної регресії та класифікаційного дерева. Такі ж позначення використаємо для

зростання і спадання відповідних вихідних значень індикатора *Pivot Point*, які позначимо $\hat{S}1$, $\hat{S}2$, $\hat{S}3$, $\hat{P}$, $\hat{R}1$, $\hat{R}2$, $\hat{R}3$ і подамо їх на вхід логістичної регресії і класифікаційного дерева. Для прогнозування напрямку руху максимальної ціни вибраної акції побудована модель має такий вигляд:

$$g_{\max 1}(x_1) = \frac{e^{x_1(k)}}{1 + e^{x_1(k)}}, \quad (2.140)$$

$$x_1(k) = -0,993 + 1,604 \cdot \hat{S}1(k) - 0,649 \cdot \hat{S}2(k) + 0,363 \cdot \hat{S}3(k) + 0,355 \cdot \hat{P}(k) -$$

$$- 0,298 \cdot \hat{R}1(k) - 0,217 \cdot \hat{R}2(k) - 0,278 \cdot \hat{R}3(k)$$

Порогове значення ймовірності вибираємо з умови мінімуму похибок 1-го та 2-го роду. У даному випадку за порогове вибрано значення 0,29; похибка першого роду складає 18, другого – 57, а кількість правильно спрогнозованих співпадань напрямку руху цінового процесу складає 173 (69,76%).

Для мінімальної ціни активу побудована така модель логістичної регресії:

$$g_{\min}(x_2) = \frac{e^{x_2(k)}}{1 + e^{x_2(k)}}, \quad (2.141)$$

$$x_2(k) = -0,139 + 1,21 \cdot \hat{S}1(k) - 0,979 \cdot \hat{S}2(k) - 0,472 \cdot \hat{S}3(k) - 0,22 \cdot \hat{P}(k) +$$

$$+ 0,423 \cdot \hat{R}1(k) + 0,577 \cdot \hat{R}2(k) - 0,01 \cdot \hat{R}3(k)$$

Для порогового значення ймовірності 0,42 похибка першого роду склала 21, другого – 63, а кількість коректно спрогнозованих співпадань напряму руху дорівнює 164 (66,13%).

Використовуючи для прогнозування дерево класифікацій (за алгоритмом CHAID), для процесу утворення максимальних цін і вибраного порогового значення 0,25 отримаємо похибку 1-го роду 27, 2-го роду – 50. Кількість співпадань (узгоджень) напрямків руху процесу склала 171 (68,95%). Для мінімальних значень отримані такі результати прогнозування: порогове значення 0,31, похибка 1-го роду 26, похибка 2-го роду – 62, кількість співпадань напрямків – 160 (64,66%).

Для покращення якості прогнозів у модель логістичної регресії та дерево класифікації введено значення прогнозів руху цін, отримані за допомогою

регресійних моделей. Для максимальної ціни модифікована модель логістичної регресії має вигляд:

$$g_{\max 2}(x_1) = \frac{e^{x_1(k)}}{1 + e^{x_1(k)}}, \quad (2.142)$$

$$x_1(k) = -1,267 + 1,071 \cdot \hat{S}1(k) - 0,384 \cdot \hat{S}2(k) + 0,142 \cdot \hat{S}3(k) + 0,206 \cdot \hat{P}(k) + \\ + 0,305 \cdot \hat{R}1(k) - 0,181 \cdot \hat{R}2(k) - 0,269 \cdot \hat{R}3(k) + 1,075 \cdot \hat{y}(k)$$

де $\hat{y}(k)$ – вихідна змінна регресійної моделі, що приймає значення 1 при прогнозі зростання ціни та 0 – при прогнозі спадання. При пороговому значенні ймовірності 0,46 похибка першого роду склала 28, другого – 36. Кількість співпадань напрямків руху ціни склала 184 (74,19%). При використанні дерева класифікації і пороговому значенні ймовірності 0,25 похибка першого роду склала 28, другого – 48; кількість співпадань напрямків руху ціни дорівнювала 172 (69,48%).

Для прогнозування мінімальної ціни побудована нова модель логістичної регресії у вигляді:

$$g_{\min 2}(x_2) = \frac{e^{x_2(k)}}{1 + e^{x_2(k)}}, \quad (2.143)$$

$$x_2(k) = -0,124 + 0,364 \cdot \hat{S}1(k) - 0,642 \cdot \hat{S}2(k) - 0,55 \cdot \hat{S}3(k) - 0,549 \cdot \hat{P}(k) - \\ - 0,282 \cdot \hat{R}1(k) + 0,599 \cdot \hat{R}2(k) - 0,451 \cdot \hat{R}3(k) + 2,974 \cdot \hat{y}(k)$$

При пороговому значенні ймовірності 0,47 похибка першого роду склала 46, другого – 17; кількість співпадань напрямків – 185 (74,7%). При використанні дерева класифікації і пороговому значенні ймовірності 0,47 похибка першого роду склала 50, другого – 13; кількість співпадань напрямків – 185 (74,6%). Результати прогнозів напрямку руху максимальної та мінімальної ціни представлено у таблицях 2.16 і 2.17, відповідно.

Таблиця 2.16 Результати прогнозування напрямку руху максимальної ціни

Тип моделі	Ймовірність співпадань напрямку прогнозу
Регресійна модель з індикаторами	68,95%
Логістична регресія з індикаторами	69,76%
Дерево класифікації з індикаторами	68,95%
Логістична регресія з індикаторами та прогнозом за регресійною моделлю	74,19%
Дерево класифікації з індикаторами та прогнозом за регресійною моделлю	69,48%

Таблиця 2.17 Результати прогнозування напрямку руху мінімальної ціни

Тип моделі	Ймовірність співпадань напрямку прогнозу
Регресійна модель з індикаторами	73,79%
Логістична регресія з індикаторами	66,13%
Дерево класифікації з індикаторами	64,66%
Логістична регресія з індикаторами та прогнозом за регресійною моделлю	74,7%
Дерево класифікації з індикаторами та прогнозом за регресійною моделлю	74,6%

Таким чином, в обох випадках найкращою виявилася модель логістичної регресії з використанням оцінок прогнозу за регресійною моделлю. Статистичні характеристики прогнозів свідчать про їх високу якість і можливість використання у практичних правилах торгівлі.

Також побудовано математичні моделі процесів формування максимальної та мінімальної цін австралійського долара по відношенню до канадського долара. Запропоновано метод побудови регресійних моделей з використанням схеми адаптивного прогнозування та індикаторів технічного аналізу. Для прогнозування напрямків руху ціни використано моделі, які прийнято відносити до інтелектуального аналізу даних – логістична регресія та

дерево класифікації. Найкращими моделями виявилися ті, що представляють собою симбіоз регресійної моделі та логістичної регресії. Зокрема, ймовірність коректного прогнозування руху максимальної ціни склала 74,19%, а мінімальної – 74,7%.

В подальших дослідженнях доцільно розглянути індикатори інших типів, а також застосувати удосконалені схеми адаптивного прогнозування з можливістю покращення якості попередньої обробки вхідних даних, що використовуються для побудови моделей.

Запитання і вправи до другого розділу

1. У чому полягає головна мета побудови моделі нелінійного або нестационарного процесу?
2. Які ключові відмінності між моделюванням лінійних та нелінійних процесів?
3. Чому у реальних економічних та фінансових даних частіше зустрічаються нестационарні процеси?
4. Що означає структурна та параметрична невизначеність у моделюванні?
5. Назвіть основні етапи побудови моделі за статистичними або експериментальними даними.
6. Які типові труднощі виникають при попередній оцінці структури моделі?
7. Чому важливо враховувати лагові змінні у побудові моделі?
8. Яким чином планування експерименту впливає на якість моделі?
9. У чому полягає різниця між детермінованими та стохастичними збуреннями?
10. Як враховуються збурення у моделі: адитивно чи мультиплікативно? Наведіть приклади.
11. Які існують підходи до моделювання нелінійних процесів?
12. Чим відрізняється структурний метод від функціонального методу побудови моделі?
13. У яких випадках застосовується функціональний підхід («чорна скринька»)?
14. Наведіть приклади економічних чи фінансових процесів, які краще моделювати структурним підходом.
15. Які методи оцінювання параметрів найчастіше використовуються для нелінійних моделей?

16. Чому метод найменших квадратів має обмеження у випадку нелінійних моделей?
17. Яку роль відіграють оптимізаційні алгоритми (генетичні, імунні) при оцінюванні параметрів?
18. Які вимоги висуваються до експериментальних даних, щоб модель була адекватною?
19. Як визначити, що побудована модель неадекватна реальному процесу?
20. Які є критерії якості моделі (статистичні показники)?
21. Чому іноді доводиться збирати додаткові експериментальні дані?
22. Як візуальний аналіз допомагає при виборі структури моделі?
23. У чому перевага адаптивних алгоритмів при побудові моделей?
24. Як комбінування кількох моделей може покращити результати прогнозування?
25. Чому більшість економічних процесів на практиці описуються саме нестационарними моделями?
26. Які особливості потрібно враховувати при моделюванні процесів з нелінійними трендами?
27. Як перевірити наявність сезонної компоненти у даних?
28. Які ризики виникають при некоректному виборі структури моделі?
29. У чому полягає ітераційний характер побудови адекватної моделі?
30. Як сучасні методи машинного навчання інтегруються з класичними методами моделювання часових рядів?

РОЗДІЛ 3. АНАЛІЗ СТАЦІОНАРНОСТІ ПРОЦЕСІВ АВТОРЕГРЕСІЇ З КОВЗНИМ СЕРЕДНІМ

3.1 Слабка і сильна стаціонарність

Процеси з одиничними коренями. Розглянемо докладніше властивості рівнянь, які ми вже застосовували вище для опису динаміки процесів різної природи, знаходили для них розв'язки та знаходили на їх основі функції прогнозування. Для аналізу стаціонарності застосуємо аналітичний підхід, який дає можливість узагальнити результати дослідження для вибраних типів рівнянь. Спочатку уточнимо, яке ж рівняння формально коректно можна називати авторегресією з ковзним середнім (АРКС)?

Рівняння виду

$$\begin{aligned} y(k) &= a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j \varepsilon(k-j) + \varepsilon(k) = \\ &= a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j) \end{aligned} \quad (3.1)$$
$$b_0 = 1, \quad \sum_{j=1}^q b_j = 1,$$

називають авторегресією з ковзним середнім, якщо корені його характеристичного рівняння лежать всередині одиничного кола на комплексній площині, а $\{\varepsilon(k)\}$ – випадковий процес [2, 51].

Якщо один або більше коренів характеристичного рівняння, записаного для (3.1), дорівнюють одиниці, то послідовність $\{y(k)\}$ називають *інтегрованою* або *процесом з одиничними коренями*, а рівняння (3.1) називають авторегресією з інтегрованим ковзним середнім (АРІКС). Процеси з одиничними коренями відносяться до класу *нестационарних процесів*. Оскільки такі процеси є досить характерними для виробничих технологій, економіки, фінансів, екології та іншим галузям, то їм буде приділена в подальшому значна увага. Нестационарні процеси такого типу особливо часто зустрічаються в перехідній економіці та відповідній фінансовій діяльності, для яких властива

висока нестаціонарна динаміка розвитку. Так, інфляція в Україні зросла за три роки (1991 – 1993 роки) від 15-20% до 10250% [10]. Таке катастрофічне зростання рівня інфляції супроводжувалось її повільним спадом до рівня 15-20% протягом наступних 8-10 років. Це приклад процесу з надзвичайно високою динамікою, а також позитивним і негативним трендом на різних етапах його розвитку.

Сильна і слабка стаціонарність

Розрізняють стаціонарність *слабку* і *сильну*. Слабку стаціонарність називають ще:

- стаціонарністю *по коваріації*;
- стаціонарністю *другого порядку*;
- стаціонарністю *в широкому сенсі*.

Формально стохастичний процес із скінченним середнім та дисперсією називають стаціонарним по коваріації, якщо для всіх k і $k-s$, ($k=0,1,2,\dots$; $s=0,1,2,\dots$) виконуються такі умови [23, 51]:

- *математичне сподівання* не залежить від часу:
-

$$E[y(k)] = E[y(k-s)] = \mu = \text{const} \quad (3.2)$$

- *дисперсія* залишається постійною для всього часового інтервалу, на якому досліджується процес:

$$E\{[y(k) - \mu]^2\} = E\{[y(k-s) - \mu]^2\} = \sigma_y^2 = \text{const} \quad (3.3)$$

$$\text{або } \text{var}[y(k)] = \text{var}[y(k-s)] = \sigma_y^2 = \text{const} ,$$

- *автоковаріація* залишається незмінною в часі для всього часового інтервалу:

$$\begin{aligned} E\{[y(k) - \mu][y(k-s) - \mu]\} = \\ = E\{[y(k-j) - \mu][y(k-j-s) - \mu]\} = \gamma(s) = \text{const} , \\ \text{або} \end{aligned} \quad (3.4)$$

Сильна стаціонарність не вимагає щоб математичне сподівання та дисперсія були постійними величинами на інтервалі дослідження.

Таким чином, слабка стаціонарність накладає більш строгі обмеження на поведінку випадкового процесу ніж сильна. В подальшому будемо розглядати ті процеси, які є стаціонарними по коваріації, тобто стаціонарні в слабкому смислі. Відповідно, процес будемо вважати *нестационарним*, якщо не виконується хоча б одна з умов (3.1.2)-(3.1.4). Виконання цих умов можна проаналізувати за допомогою відповідних статистичних критеріїв і даних стосовно розвитку досліджуваних процесів.

Зазначимо також, що основні статистичні характеристики процесу – математичне сподівання, дисперсію та коваріацію, можна обчислювати на основі конкретних значень часового ряду (*вибіркові характеристики*) або на основі математичної моделі, що описує досліджуваний процес (*теоретичні характеристики*).

При дослідженні стаціонарності процесів важливо знати необхідні та достатні умови їх стаціонарності. Це дає можливість визначити конкретні параметри та їх комбінації, які впливають на стаціонарність. Такі знання необхідні для дослідження характеристик нових моделей, визначення областей стійкості розв'язків і т. ін.

3.2 Необхідні та достатні умови стаціонарності процесу AP(1)

Розглянемо умови стаціонарності для стохастичного процесу AP(1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad (3.5)$$

де $\{\varepsilon(k)\}$ – послідовність білого шуму з нульовим середнім; $E[\varepsilon(k)] = 0$, та скінченною постійною дисперсією $E[\varepsilon^2(k)] = \sigma_\varepsilon^2$; $y(0) = y_0$ – детермінована початкова умова.

Повний розв'язок цього рівняння має вигляд [2, 51]:

$$y(k) = a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_0 + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) \quad (3.6)$$

Запишемо математичне сподівання для цього розв'язку:

$$E[y(k)] = a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_0, \quad (3.7)$$

а також математичне сподівання для (3.2.2), але при зміні моменту часу на s кроків вперед:

$$E[y(k+s)] = a_0 \sum_{i=0}^{k+s-1} a_1^i + a_1^{k+s} y_0. \quad (3.8)$$

Вирази (3.2.3) і (3.2.4) свідчать про те, що $E[y(k)] \neq E[y(k+s)]$, тобто математичне сподівання залежить від часу і послідовність $\{y(k)\}$ є нестационарною на відносно коротких проміжках часу при невеликих значеннях k . При $|a_1| < 1$ і $k \rightarrow \infty$ можна записати, що

$$\begin{aligned} \lim_{k \rightarrow \infty} [y(k)] &= \lim_{k \rightarrow \infty} \left[a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_0 + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) \right] = \\ &= \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i). \end{aligned} \quad (3.9)$$

Таким чином, безумовне математичне сподівання

$$\mu = E[y(k)] = \frac{a_0}{1-a_1}, \quad k \rightarrow \infty, \quad (3.10)$$

тобто при $k \rightarrow \infty$ процес буде стаціонарним: $E[y(k)] = E[y(k+s)] = \mu$.

Знайдемо вираз для теоретичної дисперсії при великих значеннях k за допомогою рівняння (3.11):

$$\begin{aligned}\text{var}[y(k)] &= E[(y(k) - \mu)^2] = E\left[\left(\sum_{i=0}^{\infty} a_1^i \varepsilon(k-i)\right)^2\right] = \\ &= E\left\{\left[\varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots\right] \times \left[\varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots\right]\right\} = \quad (3.11) \\ &= \sigma_{\varepsilon}^2 (1 + a_1^2 + a_1^4 + \dots) = \frac{\sigma_{\varepsilon}^2}{1 - a_1^2} = \text{const},\end{aligned}$$

оскільки $E[\varepsilon(k-i)\varepsilon(k-j)] = 0$ при $i \neq j$.

Запишемо вираз для автоковаріації:

$$\begin{aligned}\text{cov}[y(k) - \mu] &= E[(y(k) - \mu)(y(k-s) - \mu)] = \\ &= E\left\{\left[\varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots + a_1^s \varepsilon(k-s) + a_1^{s+1} \varepsilon(k-s-1) + \dots\right] \times \right. \\ &\quad \left. \times [\varepsilon(k-s) + a_1 \varepsilon(k-1-s) + a_1^2 \varepsilon(k-2-s) + \dots]\right\} = \quad (3.12) \\ &= \sigma_{\varepsilon}^2 a_1^s [1 + a_1^2 + a_1^4 + \dots] = \frac{\sigma_{\varepsilon}^2 a_1^s}{1 - a_1^2} = \text{const}\end{aligned}$$

для заданого значення параметра s .

Таким чином, стохастичний процес $AP(1)$ є стаціонарним при $|a_1| < 1$ і $k \rightarrow \infty$, тобто стаціонарність проявляється на довгих часових інтервалах.

Якщо врахувати однорідний розв'язок, то отримаємо такий вираз [2]:

$$y(k) = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) + A a_1^k \quad (3.13)$$

Тепер математичне сподівання для основної змінної в лівій частині моделі

$$E[y(k)] = \frac{a_0}{1 - a_1} + A \cdot E[a_1^k] \quad (3.14)$$

залежить від часу і послідовність $\{y(k)\}$ буде стаціонарною тільки в тому випадку, коли $Aa_1^k = 0$. Тобто для досягнення стаціонарності необхідно вибрати з початкових умов довільну константу $A = 0$. Величину Aa_1^k можна інтерпретувати як величину відхилення від довгострокової рівноваги.

Сформулюємо остаточно умови стаціонарності стохастичного процесу $AP(1)$:

а) модуль коефіцієнта $|a_1| < 1$ (корінь характеристичного рівняння);

б) однорідний розв'язок повинен дорівнювати нулю, тобто $A = 0$, або процес починається досить давно в минулому, що забезпечує збіжність $a_1^k \rightarrow 0$.

Умови стаціонарності процесу авторегресії довільного порядку

Отримані умови стаціонарності процесу $AP(1)$ можна досить легко розповсюдити на процес авторегресії довільного порядку $AP(p)$. Однорідний розв'язок для рівняння $AP(p)$ має вигляд:

$$y^h(k) = \sum_{i=1}^p A_i \alpha_i^k, \quad (3.15)$$

а якщо є кратні корені, то однорідний розв'язок приймає вигляд:

$$y^h(k) = \alpha^k \sum_{i=1}^m A_i k^{i-1} + \sum_{i=m+1}^p A_i \alpha_i^k, \quad (3.16)$$

де α – корінь характеристичного рівняння кратності m ; α_i , $i = m+1, \dots, p$ – різні дійсні корені. Очевидно, що середнє, дисперсія та коваріація, обчислені

для (3.2.7), будуть залежати від часу.

Таким чином, умовою стаціонарності процесу $AP(p)$ є рівність нулю його однорідного розв'язку.

3.3 Умови стаціонарності часткового розв'язку рівняння АРКС

Розглянемо задачу визначення умов стаціонарності на прикладі. На практиці досить часто використовують модель АРКС(2, 1):

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) + b_1 \varepsilon(k-1), \quad (3.17)$$

де $a_0 = 0$, тому що воно не впливає на стаціонарність. Оскільки для досягнення стаціонарності однорідний розв'язок повинен дорівнювати нулю, то знайдемо частковий розв'язок. Скористаємось пробною функцією для часткового розв'язку

$$y_{np}^p(k) = \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i), \quad (3.18)$$

і підставимо її в (3.19):

$$\begin{aligned} & \alpha_0 \varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots = \\ & = a_1 [\alpha_0 \varepsilon(k-1) + \alpha_1 \varepsilon(k-2) + \dots] + a_2 [\alpha_0 \varepsilon(k-2) + \alpha_1 \varepsilon(k-3) + \dots] + \\ & \quad + \varepsilon(k) + b_1 \varepsilon(k-1). \end{aligned} \quad (3.19)$$

Для того щоб отримати тотожність, прирівняємо коефіцієнти при однакових змінних:

$$\begin{aligned} \alpha_0 &= 1; \\ \alpha_1 &= a_1 \alpha_0 + b_1 \quad \text{або} \quad \alpha_1 = a_1 + b_1; \\ \alpha_n &= a_1 \alpha_{n-1} + a_2 \alpha_{n-2}. \end{aligned} \quad (3.20)$$

Характеристичне рівняння для $\alpha_i = a_1 \alpha_{i-1} + a_2 \alpha_{i-2}, \forall i \geq 2$ відповідає характеристичному рівнянню, записаному для однорідної частини (3.21):

$$\lambda^2 - a_1 \lambda - a_2 = 0. \quad (3.21)$$

Таким чином, якщо корені цього характеристичного рівняння знаходяться всередині одиничного кола на комплексній площині, то послідовність $\{\alpha_i\}$ – збіжний процес. Це і є умовою стаціонарності неоднорідного розв'язку рівняння АРКС(2, 1) [23].

Статистичні характеристики послідовності $\{y_{np}^p(k) = \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i)\}$

Для перевірки стаціонарності послідовності $\{y_{np}^p(k)\}$ знайдемо математичне сподівання, дисперсію та автоковаріацію для неоднорідного розв'язку, зумовленого наявністю випадкової величини $\varepsilon(k)$:

а) математичне сподівання:

$$E[y^p(k)] = E[\alpha_0 \varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots] = 0, \quad (3.22)$$

а також

$$E[y^p(k)] = E[y^p(k-s)] = 0, \quad \forall k, s. \quad (3.23)$$

б) дисперсія:

$$\begin{aligned} \text{var}[y^p(k)] &= E[\alpha_0 \varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots]^2 = \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha_i^2, \end{aligned} \quad (3.24)$$

тобто $\text{var}[y^p(k)] = \text{var}[y^p(k-s)], \quad \forall k, s$

в) автоковаріація:

$$s = 1:$$

$$\begin{aligned} \text{cov}[y^p(k), y^p(k-1)] &= E \{ [\varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots] \times \\ &\times [\varepsilon(k-1) + \alpha_1 \varepsilon(k-2) + \alpha_2 \varepsilon(k-3) + \dots] \} = \\ &= \sigma_\varepsilon^2 (\alpha_1 + \alpha_2 \alpha_1 + \alpha_3 \alpha_2 + \dots). \end{aligned} \quad (3.25)$$

$$s = 2:$$

$$\begin{aligned} \text{cov}[y^p(k), y^p(k-2)] &= E \{ [\varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots] \times \\ &\times [\varepsilon(k-2) + \alpha_1 \varepsilon(k-3) + \alpha_2 \varepsilon(k-4) + \dots] \} = \\ &= \sigma_\varepsilon^2 (\alpha_2 + \alpha_3 \alpha_1 + \alpha_4 \alpha_2 + \dots) \end{aligned}$$

і для довільного значення параметра s можна записати:

$$\text{cov}[y(k), y(k-s)] = \sigma_\varepsilon^2 (\alpha_s + \alpha_{s+1} \alpha_1 + \alpha_{s+2} \alpha_2 + \dots) = \text{const}, \quad (3.26)$$

тобто коваріація постійна і не залежить від часу k .

Таким чином, якщо корені характеристичного рівняння λ_i не лежать всередині одиничного кола на комплексній площині, то послідовність коефіцієнтів $\{\alpha_i\}$ не буде збіжною. Звідси випливає, що послідовність $\{y^p(k)\}$ також буде розбіжною.

Приклад 3.1. Визначимо умови стаціонарності для процесу АРКС(2,1):

$$y(k) = 1,6y(k-1) - 0,9y(k-2) + \varepsilon(k) + 0,5\varepsilon(k-1). \quad (3.27)$$

За методом невизначених коефіцієнтів знайдемо коефіцієнти

неоднорідного розв'язку, зумовленого наявністю $\varepsilon(k)$:

$$\begin{aligned}\alpha_0 &= 1; \\ \alpha_1 &= 1,6 + 0,5 = 2,1; \\ &\vdots\end{aligned}\tag{3.28}$$

$$\alpha_i = 1,6\alpha_{i-1} - 0,9\alpha_{i-2}, \quad \forall i \geq 2,$$

тобто (3.27), – це рівняння авторегресії другого порядку для коефіцієнтів неоднорідного розв'язку. Знайдемо розв'язок характеристичного рівняння для:

$$\begin{aligned}\lambda^2 - 1,6\lambda + 0,9 &= 0; \\ \lambda_{1,2} &= \frac{1}{2}(1,6 \pm \sqrt{2,56 - 3,6}) = \frac{1}{2}(1,6 \pm j1,02); \quad j = \sqrt{-1}.\end{aligned}\tag{3.29}$$

Розв'язок однорідного рівняння (3.3.7) в полярних координатах має вигляд:

$$\alpha^h(i) = \beta_1 r^i \cos(\theta i + \beta_2),\tag{3.30}$$

$$\text{де } r = \sqrt{-a_2} = \sqrt{0,9} = 0,949; \quad \cos(\theta) = \frac{a_1}{2(a_2)^{1/2}} = \frac{1,6}{2 \cdot 0,949} = 0,843, \text{ а звідси}$$

значення кута в радіанах: $\theta = 0,567$.

Тепер можна конкретизувати однорідний розв'язок таким чином:

$$\alpha^h(i) = \beta_1 (0,949)^i \cos(0,567i + \beta_2).\tag{3.31}$$

Константи β_1 і β_2 знайдемо за допомогою початкових умов:

$$\alpha_0 = 1, \alpha_1 = 2,1.$$

$$\begin{cases} 1 = \beta_1 \cos(\beta_2) \\ 2,1 = 0,949\beta_1 \cos(0,567 + \beta_2) \end{cases} \quad (3.32)$$

Звідси $\beta_1 = \frac{1}{\cos(\beta_2)}$; $\cos(\beta_2) - \frac{0,949}{2,1} \cos(0,567 + \beta_2) = 0$. Скориставшись тригонометричною тотожністю

$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta), \quad (3.33)$$

знайдемо, що $\beta_2 = -1,197$. Таким чином, розв'язок, із врахуванням початкових умов, приймає вигляд:

$$\alpha(i) = -\frac{1}{\cos(1,197)} \cdot 0,949^i \cos(0,567i - 1,197) \quad (3.34)$$

З (3.34) випливає, що загальна тенденція зміни коефіцієнтів α_i у часі буде в сторону зменшення їх значень. Деякі значення цих коефіцієнтів подано в таблиці 3.1.

Таблиця 3.1

i	0	1	2	3	4	5	3	7	8	9	10
α_i	1,0	2,1	2,43	2,043	1,03	-0,143	1,187	-1,783	-1,731	-1,22	-0,37

Якщо продовжити обчислення коефіцієнтів α_i , то побачимо, що цей процес збіжний – затухаючі гармонічні коливання. Звідси випливає, що процес $\{y(k)\}$ також повинен бути збіжним.

3.4 Узагальнення умов стаціонарності на АРКС(p, q) довільного

порядку

В цьому параграфі розглянемо задачу знаходження умов стаціонарності для процесів ковзного середнього та авторегресії довільного порядку.

Стаціонарність процесу $KC(\infty)$. Процес $KC(\infty)$ описується рівнянням

$$x(k) = \sum_{i=0}^{\infty} b_i \varepsilon(k-i), \quad (3.35)$$

де $\{\varepsilon(k)\}$ – випадковий процес білого шуму із скінченною постійною дисперсією σ_{ε}^2 (зазначимо, що ця умова забезпечується коректним застосуванням методу оцінювання параметрів моделі (3.36)). Необхідно встановити чи буде процес $x(k)$ стаціонарним за коваріацією. Запишемо вирази для математичного сподівання, дисперсії та коваріації процесу (3.36):

а) математичне сподівання:

$$\begin{aligned} \mu_x &= E[x(k)] = E[\varepsilon(k) + b_1 \varepsilon(k-1) + b_2 \varepsilon(k-2) + \dots] = 0; \\ \mu_x &= E[x(k)] = E[x(k-s)] = 0; \end{aligned} \quad (3.36)$$

б) дисперсія:

$$\begin{aligned} \text{var}[x(k)] &= E\{[\varepsilon(k) + b_1 \varepsilon(k-1) + b_2 \varepsilon(k-2) + \dots]^2\} = \\ &= E[\varepsilon(k)]^2 + b_1^2 E[\varepsilon(k-1)]^2 + b_2^2 E[\varepsilon(k-2)]^2 + \dots = \\ &= \sigma_{\varepsilon}^2 [1 + b_1^2 + b_2^2 + \dots]. \end{aligned} \quad (3.37)$$

Очевидно, що $\text{var}[x(k-s)] = \sigma_{\varepsilon}^2 [1 + b_1^2 + b_2^2 + \dots]$ і

$$\text{var}[x(k)] = \text{var}[x(k-s)] = \text{const} \quad \forall k, s. \quad (3.38)$$

Звідси випливає, що умовою стаціонарності процесу $\{x(k)\}$ є скінченність суми $\sum b_i^2$.

в) автоковаріація:

$$\begin{aligned} \text{cov}[x(k), x(k-s)] &= E \{ [\varepsilon(k) + b_1\varepsilon(k-1) + b_2\varepsilon(k-2) + \dots] \times \\ &\times [\varepsilon(k-s) + b_1\varepsilon(k-s-1) + b_2\varepsilon(k-s-2) + \dots] \} = \\ &= \sigma_\varepsilon^2 (b_s + b_1b_{s+1} + b_2b_{s+2} + \dots). \end{aligned} \quad (3.39)$$

Припустимо, що сума $(b_s + b_1b_{s+1} + b_2b_{s+2} + \dots)$ скінченна. В такому випадку коваріація залежить тільки від s і не залежить від k . Тобто, скінченність суми і буде однією з умов стаціонарності процесу $\{x(k)\}$.

Таким чином, необхідними і достатніми умовами стаціонарності процесу КС $^{(\infty)}$ є скінченність сум, що залежать від коефіцієнтів:

$$\sum b_i^2 \quad \text{і} \quad (b_s + b_1b_{s+1} + b_2b_{s+2} + \dots). \quad (3.40)$$

Оскільки остання сума повинна бути скінченою для всіх значень s , то умову скінченності суми $\sum b_i^2$ можна вважати надлишковою. Звідси можна зробити висновок, що будь-який процес ковзного середнього скінченного порядку буде стаціонарним.

Висновок: будь-який процес ковзного середнього, що має скінчений порядок, є стаціонарним.

Стаціонарність процесу AP(p)

Визначимо обмеження стаціонарності на коефіцієнти авторегресії

довільного порядку P . Частковий розв'язок рівняння

$$y(k) = a_0 + \sum_{i=0}^p a_i y(k-i) + \varepsilon(k) \quad (3.41)$$

має вигляд [23]:

$$y^p(k) = \frac{a_0}{1 - \sum_{i=1}^p a_i} + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i) \quad (3.42)$$

За аналогією з рівнянням другого порядку в даному випадку також можна знайти різницеве рівняння для обчислення коефіцієнтів α_i :

$$\alpha_i = a_1 \alpha_{i-1} + a_2 \alpha_{i-2} + \dots + a_p \alpha_{i-p}. \quad (3.43)$$

Якщо корені характеристичного рівняння, записаного для (3.4.7), знаходяться всередині одиничного кола на комплексній площині, то послідовність коефіцієнтів $\{\alpha_i\}$ буде збіжною, тобто сума $\sum \alpha_i^2$ буде скінченною. Звідси випливає, що для перевірки стаціонарності послідовності $\{y(k)\}$ можна скористатись рівнянням (3.4.3). Для цього знайдемо вирази для математичного сподівання, дисперсії та коваріації:

а) математичне сподівання:

$$E[y(k)] = E[y(k-s)] = \frac{a_0}{1 - \sum a_i} = const \quad (3.44)$$

Ця величина буде скінченною, якщо $1 - \sum a_i \neq 0$, тобто це умова того, що корені відповідного характеристичного рівняння знаходяться всередині одиничного кола.

б) Дисперсія:

$$\begin{aligned}
 \text{var}[y(k)] &= E \{ [\varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots]^2 \} = \\
 &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha_i^2 . \\
 \text{var}[y(k-s)] &= E \{ [\varepsilon(k-s) + \alpha_1 \varepsilon(k-s-1) + \alpha_2 \varepsilon(k-s-2) + \dots]^2 \} = \\
 &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha_i^2 .
 \end{aligned} \tag{3.45}$$

Таким чином, дисперсія буде скінченною, якщо сума $\sum \alpha_i^2$ є скінченною величиною.

в) Автоковаріація:

$$\begin{aligned}
 \gamma_s = \text{cov}[y(k), y(k-s)] &= E \{ [\varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots] \times \\
 &\times [\varepsilon(k-s) + \alpha_1 \varepsilon(k-s-1) + \alpha_2 \varepsilon(k-s-2) + \dots] \} = \\
 &= \sigma_\varepsilon^2 (\alpha_s + \alpha_1 \alpha_{s+1} + \alpha_2 \alpha_{s+2} + \dots) = \text{const}
 \end{aligned} \tag{3.46}$$

і не залежить від часу. Таким чином, процес $\text{AP}(p)$ буде стаціонарним, якщо відповідні суми коефіцієнтів будуть збіжними.

Аналіз процесу $\text{АРКС}(p, q)$ в цілому

Розглянемо тепер процес $\text{АРКС}(p, q)$ в цілому:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j), \quad b_0 = 0. \tag{3.47}$$

Запишемо неоднорідний розв'язок за допомогою лагового оператора:

$$y(k) = \frac{a_0}{1 - \sum_{i=1}^p a_i L^i} + \frac{\varepsilon(k)}{1 - \sum_{i=1}^p a_i L^i} + \frac{b_1 \varepsilon(k-1)}{1 - \sum_{i=1}^p a_i L^i} + \dots$$

Перший член у правій частині (3.4.9) є константою, а всі інші будуть стаціонарними, якщо корені поліномів $(1 - \sum a_i L^i)$ лежать всередині одиничного кола.

Таким чином, якщо послідовність $\{x(k)\}$, яка описується рівнянням (3.4.1), є стаціонарною, то і послідовність $\{y(k)\}$ буде стаціонарною за умови, що корені характеристичного рівняння для авторегресійної частини знаходяться всередині одиничного кола.

***Висновок:** якщо послідовність $\{x(k)\}$, яка описується рівнянням (3.4.1), стаціонарна, то і послідовність $\{y(k)\}$ буде стаціонарною за умови, що корені характеристичного рівняння для авторегресійної частини знаходяться всередині одиничного кола на комплексній площині.*

3.5 Теоретична автокореляційна функція

У попередньому параграфі ми встановили, що для оцінювання стаціонарності процесів необхідно обчислювати три основні статистичні характеристики: математичне сподівання, дисперсію і коваріацію. Ці статистичні параметри можна обчислювати безпосередньо за вибіркою, тобто знаходити *вибіркові характеристики*, або за допомогою математичної моделі – це *теоретичні характеристики*.

Розглянемо розрахунок теоретичних статистичних характеристик для моделі AP(1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad E[\varepsilon(k)] = 0,$$

$$E[\varepsilon(k)\varepsilon(k-j)] = \begin{cases} \sigma_\varepsilon^2, & j=0, \\ 0, & j \geq 1; \end{cases} \quad E[\varepsilon(k)\varepsilon(y-j)] = \begin{cases} \sigma_\varepsilon^2, & j=0, \\ 0, & j \geq 1. \end{cases} \quad (3.49)$$

Дисперсія: $\text{var}[y(k)] = \text{var}[a_1 y(k-1)] + \text{var}[\varepsilon(k)]$. Оскільки для

стаціонарного процесу, при $|a_1| < 1$, виконується умова, що $\text{var}[y(k)] = \text{var}[y(k-1)]$, то $\text{var}[y(k)](1 - a_1^2) = \text{var}[\varepsilon(k)]$. Звідси знайдемо вираз для дисперсії:

$$\sigma_y^2 = \text{var}[y(k)] = \frac{\sigma_\varepsilon^2}{1 - a_1^2} = \gamma_0, \quad |a_1| < 1 \quad (3.50)$$

Нагадаємо, що вибіркова дисперсія часового ряду $\{y(k)\}$ обчислюється за виразом:

$$\sigma_y^2 = \frac{1}{N-1} \sum_{k=1}^N [y(k) - \bar{y}]^2, \quad (3.51)$$

де $\bar{y}$ – вибіркове середнє значення ряду даних.

Коваріація: оскільки для стаціонарного процесу справедлива рівність: $\text{cov}[y(k), y(k-s)] = \text{cov}[y(k), y(k+s)]$, то зручніше користуватись останнім означенням. Запишемо рівняння AP(1) для моменту $(k+1)$:

$$\begin{aligned} y(k+1) &= a_1 y(k) + \varepsilon(k+1) = a_1 [a_1 y(k-1) + \varepsilon(k)] + \varepsilon(k+1), \\ \gamma(1) &= \text{cov}[y(k), y(k+1)] = \\ &= E \{ [a_1 y(k-1) + \varepsilon(k)] [a_1^2 y(k-1) + a_1 \varepsilon(k) + \varepsilon(k+1)] \} = \\ &= a_1^3 \text{var}[y(k-1)] + a_1 \text{var}[\varepsilon(k)] = a_1^3 \frac{\sigma_y^2}{1 - a_1^2} + a_1 \sigma_\varepsilon^2. \end{aligned} \quad (3.52)$$

Покладаючи $\sigma_y^2 \approx \sigma_\varepsilon^2$ (що не входить в протиріччя з означенням випадкових процесів згідно з теорією оцінювання), отримаємо:

$$\gamma(1) = \text{cov}[y(k), y(k+1)] = \frac{a_1^3 \sigma_\varepsilon^2 + a_1 \sigma_\varepsilon^2 - a_1^3 \sigma_\varepsilon^2}{1 - a_1^2} = \frac{a_1 \sigma_\varepsilon^2}{1 - a_1^2}. \quad (3.53)$$

Для довільного значення s тепер можна записати:

$$\gamma(s) = E\{[y(k) - \mu_y], [y(k-s) - \mu_y]\} = \frac{\sigma_\varepsilon^2 a_1^s}{1 - a_1^2}. \quad (3.54)$$

Автокореляційну функцію (АКФ) можна знайти шляхом нормування коваріаційної функції стосовно $\gamma(0) = \gamma_0$:

$$\rho(s) = \frac{\gamma(s)}{\gamma_0}; \quad \rho_0 = 1, \rho(1) = a_1, \rho(2) = a_1^2, \dots, \rho(s) = a_1^s. \quad (3.55)$$

Якщо $0 < a_1 < 1$, то АКФ експоненціально спадає, а при $-1 < a_1 < 0$ АКФ має коливальний спадаючий характер. Порівняння теоретичної та вибіркової АКФ також дає можливість визначити адекватність моделі процесу. Крім того, теоретична АКФ може бути використана для оцінювання стаціонарності процесів.

3.6 Обчислення кореляційних функцій за допомогою Юла-Уокера

АКФ процесу авторегресії другого порядку $AR(2)$

Систему рівнянь Юла-Уокера отримують шляхом множення обох частин рівняння, що описує динаміку процесу, на $y(k-s)$, $s = 0, 1, 2, \dots$. Помножимо ліву і праву частини рівняння

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k), \quad \{\varepsilon(k)\} \sim N(0, \sigma_\varepsilon^2), \quad (3.56)$$

на $y(k-s)$, $s=0,1,2,\dots$ і знайдемо математичне сподівання обох частин рівняння. Зазначимо, що перетин, a_0 , не впливає на кореляційну функцію, а тому його виключено з рівняння (3.3.1).

$$\begin{aligned} s=0: \quad E[y(k)y(k)] &= a_1 E[y(k-1)y(k)] + a_2 E[y(k-2)y(k)] + E[\varepsilon(k)y(k)], \\ s=1: \quad E[y(k)y(k-1)] &= a_1 E[y(k-1)y(k-1)] + a_2 E[y(k-2)y(k-1)] + \\ &\quad + E[\varepsilon(k)y(k-1)], \\ s=2: \quad E[y(k)y(k-2)] &= a_1 E[y(k-1)y(k-2)] + a_2 E[y(k-2)y(k-2)] + \\ &\quad + E[\varepsilon(k)y(k-2)], \end{aligned} \quad (3.57)$$

$$\exists s: \quad E[y(k)y(k-s)] = a_1 E[y(k-1)y(k-s)] + a_2 E[y(k-2)y(k-s)] + E[\varepsilon(k)y(k-s)].$$

$$E[y(k) y(k-s)] = E[y(k-s) y(k)] = \gamma_s, \quad (3.58)$$

$$\begin{cases} \gamma_0 = a_1\gamma_1 + a_2\gamma_2 + \sigma_\varepsilon^2 \\ \gamma_1 = a_1\gamma_0 + a_2\gamma_1 \\ \gamma_s = a_1\gamma_{s-1} + a_2\gamma_{s-2} \end{cases} \quad (3.59)$$

і поділимо два останніх рівняння на γ_0 (тобто виконаємо нормування):

$$\begin{cases} \rho_1 = a_1 \rho_0 + a_2 \rho_1, \\ \rho_s = a_1 \rho_{s-1} + a_2 \rho_{s-2}. \end{cases} \quad (3.60)$$

Оскільки $\rho_0 = 1$, то $\rho_1 = a_1 / (1 - a_2)$. Всі інші значення дискретної кореляційної функції можна знайти за допомогою рекурсивної формули:

$$\begin{aligned} \rho_s &= a_1 \rho_{s-1} + a_2 \rho_{s-2} \\ \text{або} \quad \rho(s) &= a_1 \rho(s-1) + a_2 \rho(s-2). \end{aligned} \quad (3.61)$$

Таким чином, отримаємо,

$$\rho(2) = \frac{a_1}{1 - a_2} + a_2, \quad \rho(3) = a_1 \left(\frac{a_1^2}{1 - a_2} + a_2 \right) + \frac{a_2 a_1}{1 - a_2} \quad (3.62)$$

АКФ процесу ковзного середнього 1-го порядку: КС(1)

Для процесу КС(1) рівняння Юла-Уокера мають вигляд:

$$\begin{aligned} \gamma_0 &= E[y(k) y(k)] = E \{ [\varepsilon(k) + b_1 \varepsilon(k-1)] [\varepsilon(k) + b_1 \varepsilon(k-1)] \} = \\ &= \sigma_\varepsilon^2 + b_1^2 \sigma_\varepsilon^2 = (1 + b_1^2) \sigma_\varepsilon^2, \\ \gamma_1 &= E[y(k) y(k-1)] = E \{ [\varepsilon(k) + b_1 \varepsilon(k-1)] [\varepsilon(k-1) + b_1 \varepsilon(k-2)] \} = \\ &= b_1 \sigma_\varepsilon^2, \\ \gamma_2 &= E[y(k) y(k-2)] = E \{ [\varepsilon(k) + b_1 \varepsilon(k-1)] [\varepsilon(k-2) + b_1 \varepsilon(k-3)] \} = 0 \\ \text{і} \quad \gamma_s &= 0. \end{aligned} \quad (3.63)$$

Поділимо γ_0, γ_1 на γ_0 і отримаємо:

$$\rho_0 = 1, \quad \rho_1 = \frac{b_1}{1+b_1^2}, \quad \rho_s = 0, \quad \forall s > 1, \quad (3.64)$$

тобто АКФ процесу КС(1) має тільки два значення, відмінних від нуля.

АКФ процесу АРКС(1,1)

Для рівняння

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) + b_1 \varepsilon(k-1) \quad (3.65)$$

необхідно знайти коваріаційну функцію γ_s і АКФ ρ_s . За означенням запишемо:

$$\begin{aligned} \gamma_s &= E \{ [y(k) - \mu_y] [y(k-s) - \mu_y] \}, \\ E[\varepsilon(k) y(k-s)] &= 0, \quad s \geq 1, \\ E[\varepsilon(k) y(k)] &= \sigma_\varepsilon^2, \\ E[\varepsilon(k-1) y(k)] &= E \{ \varepsilon(k-1) [a_1 y(k-1) + \varepsilon(k) + b_1 \varepsilon(k-1)] \} = \\ &= a_1 \sigma_\varepsilon^2 + b_1 \sigma_\varepsilon^2 = \sigma_\varepsilon^2 (a_1 + b_1). \end{aligned} \quad (3.66)$$

Рівняння Юла-Уокера у цьому випадку мають вигляд:

$$\begin{aligned} \gamma_0 &= E[y(k) y(k)] = a_1 E[y(k) y(k-1)] + \\ &+ E[y(k) \varepsilon(k)] + b_1 E[y(k) \varepsilon(k-1)] = \\ &= a_1 \gamma_1 + \sigma_\varepsilon^2 + b_1 (a_1 + b_1) \sigma_\varepsilon^2, \\ \gamma_1 &= E[y(k) y(k-1)] = a_1 E[y(k-1) y(k-1)] + \\ &+ E[y(k-1) \varepsilon(k)] + b_1 E[y(k-1) \varepsilon(k-1)] = \end{aligned}$$

В скороченій формі представлення отримані рівняння виглядають так:

$$\begin{aligned}
 \gamma_0 &= a_1 \gamma_1 + \sigma_\varepsilon^2 + b_1(a_1 + b_1) \sigma_\varepsilon^2, \\
 \gamma_1 &= a_1 \gamma_0 + b_1 \sigma_\varepsilon^2, \\
 \gamma_2 &= a_1 \gamma_1, \\
 \gamma_3 &= a_1 \gamma_2, \\
 &\vdots \\
 \gamma_s &= a_1 \gamma_{s-1}.
 \end{aligned} \tag{3.68}$$

Розв'язуючи цю систему рівнянь, отримаємо:

$$\gamma_0 = \frac{1 + b_1^2 + 2a_1b_1}{1 - a_1^2} \cdot \sigma_\varepsilon^2; \quad \gamma_1 = \frac{(1 + a_1b_1)(a_1 + b_1)}{1 + b_1^2 + 2a_1b_1}; \tag{3.69}$$

і коефіцієнти кореляції:

$$\rho_0 = 1, \quad \rho_1 = \frac{(1 + a_1b_1)(a_1 + b_1)}{1 + b_1^2 + 2a_1b_1}, \quad \rho_s = a_1 \rho_{s-1}, \quad \forall s > 1. \tag{3.70}$$

Оскільки $\rho_s = a_1 \rho_{s-1}$, то збіжність процесу залежить від значення a_1 .

Якщо $0 < a_1 < 1$, то збіжність гладка; а якщо $-1 < a_1 < 0$, то збіжність носить коливальний характер.

АКФ процесу АРКС(2,1)

Для рівняння АРКС(2,1):

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) + b_1 \varepsilon(k-1) \tag{3.71}$$

необхідно знайти коваріаційну функцію γ_s і АКФ ρ_s . За означенням,

$$\begin{aligned}
\gamma_s &= E \{ [y(k) - \mu_y] [y(k-s) - \mu_y] \}, \\
E[\varepsilon(k) y(k-s)] &= 0, \quad s \geq 1, \\
E[\varepsilon(k) y(k)] &= \sigma_\varepsilon^2,
\end{aligned} \tag{3.72}$$

$$\begin{aligned}
E[\varepsilon(k-1) y(k)] &= E \{ \varepsilon(k-1) [a_1 y(k-1) + \varepsilon(k) + b_1 \varepsilon(k-1)] \} = \\
&= a_1 \sigma_\varepsilon^2 + b_1 \sigma_\varepsilon^2 = \sigma_\varepsilon^2 (a_1 + b_1).
\end{aligned}$$

Рівняння Юла-Уокера для цього випадку мають вигляд:

$$\begin{aligned}
\gamma_0 &= E[y(k) y(k)] = a_1 E[y(k) y(k-1)] + a_2 E[y(k) y(k-2)] + \\
&\quad + E[y(k) \varepsilon(k)] + b_1 E[y(k) \varepsilon(k-1)] = \\
&= a_1 \gamma_1 + a_2 \gamma_2 + \sigma_\varepsilon^2 + b_1 (a_1 + b_1) \sigma_\varepsilon^2, \\
\gamma_1 &= E[y(k) y(k-1)] = a_1 E[y(k-1) y(k-1)] + a_2 E[y(k-1) y(k-2)] + \\
&\quad + E[y(k-1) \varepsilon(k)] + b_1 E[y(k-1) \varepsilon(k-1)] = \\
&= a_1 \gamma_0 + a_2 \gamma_1 + b_1 \sigma_\varepsilon^2, \\
\gamma_2 &= E[y(k) y(k-2)] = a_1 E[y(k-1) y(k-2)] + a_2 E[y(k-2) y(k-2)] + \\
&\quad + E[y(k-2) \varepsilon(k)] + b_1 E[y(k-2) \varepsilon(k-1)] = \\
&= a_1 \gamma_1 + a_2 \gamma_0, \\
\gamma_3 &= E[y(k) y(k-3)] = a_1 E[y(k-1) y(k-3)] + a_2 E[y(k-2) y(k-3)] + \\
&\quad + E[y(k-3) \varepsilon(k)] + b_1 E[y(k-3) \varepsilon(k-1)] = \\
&= a_1 \gamma_2 + a_2 \gamma_1.
\end{aligned} \tag{3.73}$$

□

$$\gamma_s = a_1 \gamma_{s-1} + a_2 \gamma_{s-2}.$$

Перейдемо до скороченої форми запису отриманого результату:

$$\begin{aligned}
 \gamma_0 &= a_1\gamma_1 + a_2\gamma_2 + \sigma_\varepsilon^2 [1 + b_1(a_1 + b_1)], \\
 \gamma_1 &= a_1\gamma_0 + a_2\gamma_1 + b_1 \sigma_\varepsilon^2, \\
 \gamma_2 &= a_1\gamma_1 + a_2\gamma_0, \\
 \gamma_3 &= a_1\gamma_2 + a_2\gamma_1, \\
 \gamma_4 &= a_1\gamma_3 + a_2\gamma_2, \\
 &\vdots \\
 \gamma_s &= a_1\gamma_{s-1} + a_2\gamma_{s-2}.
 \end{aligned} \tag{3.74}$$

Розв'язуючи цю систему рівнянь, знайдемо, що:

$$\gamma_0 = \frac{a_1(1+a_2)c_3 + c}{1 - [a_1 c_2(1+a_2) + a_2^2]}, \quad \gamma_1 = \frac{a_1\gamma_0 + c_1}{1 - a_2}, \tag{3.75}$$

де $c = \sigma_\varepsilon^2 [1 + b_1(a_1 + b_1)]$; $c_1 = b_1 \sigma_\varepsilon^2$; $c_2 = a_1 / (1 - a_2)$; $c_3 = c_1 / (1 - a_2)$, а

всі інші значення γ_s можна знайти за наведеною вище рекурсивною формулою.

У загальному випадку для процесу АРКС(p, q) коефіцієнти кореляції можна знайти за допомогою такої рекурсивної формули:

$$\rho_i = a_1\rho_{i-1} + a_2\rho_{i-2} + \dots + a_p\rho_{i-p}. \tag{3.76}$$

3.7 Часткова автокореляційна функція

Перед тим як розглядати часткову автокореляційну функцію визначимо деякі особливості звичайної АКФ. Так, АКФ процесу АР(1) обчислюється за виразами:

тобто значення $y(k)$ і $y(k-2)$ корельовані, хоча $y(k-2)$ в моделі відсутнє. З наведеного виразу також видно, що коефіцієнт кореляції ρ_2 між значеннями $y(k)$ і $y(k-2)$ дорівнює добутку коефіцієнтів кореляції між $y(k)$ і $y(k-1)$ та між $y(k-1)$ і $y(k-2)$:

$$\rho_2 = \rho_1 \cdot \rho_1 = \rho_1^2. \quad (3.78)$$

Очевидно, що такі “непрямі” кореляції наявні в АКФ будь-якого процесу авторегресії. В цілому це призводить до того, що АКФ не досить чітко відображає порядок авторегресії.

На відміну від АКФ, часткова АКФ (ЧАКФ) виключає вплив проміжних коефіцієнтів кореляції, тобто на значення коефіцієнта часткової кореляції ϕ_{ss} між $y(k)$ і $y(k-s)$ не впливають величини $y(k-1), \dots, y(k-s+1)$ (рис. 3.1).

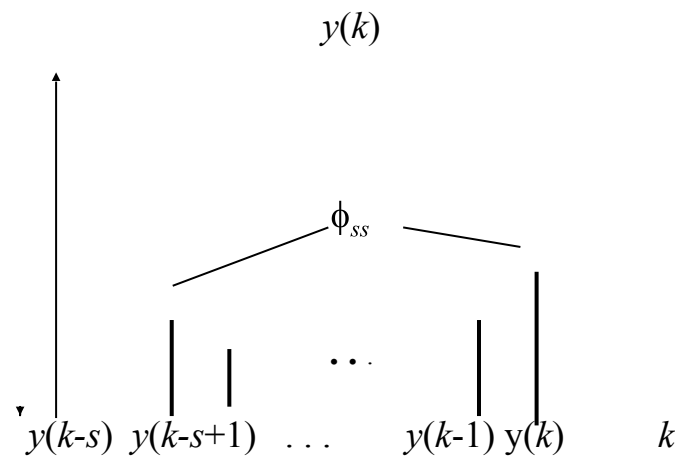


Рис.3.1. Коефіцієнт часткової кореляції між $y(k)$ і $y(k-s)$.

Для процесу $AR(1)$ коефіцієнт ЧАКФ між $y(k)$ і $y(k-2)$ дорівнює нулю за означенням. Звідси випливає, що в загальному випадку для процесу $AR(p)$ немає прямої кореляції між значеннями $y(k)$ і $y(k-s)$ при $s > p$:

$\phi_{ss} = 0, \forall s > p$. Ця властивість ЧАКФ є корисною при розв'язуванні задач ідентифікації моделей процесів $AP(p)$.

Однак, це не відноситься до процесів ковзного середнього. Так, для процесу $CC(1)$:

$$y(k) = \varepsilon(k) + b_1 \varepsilon(k-1) \quad (3.79)$$

можна записати, що $y(k) = (1 + b_1 L) \varepsilon(k)$ або, скориставшись властивістю оператора зсуву L , отримаємо:

$$\varepsilon(k) = \frac{y(k)}{1 + b_1 L} = y(k) - b_1 y(k-1) + b_1^2 y(k-2) - b_1^3 y(k-3) + \dots, \quad |b_1| < 1. \quad (3.80)$$

З цього виразу випливає, що значення $y(k)$ буде корельованим із всіма іншими, затриманими в часі значеннями, і ЧАКФ цього процесу не спадає різко до нуля.

Коефіцієнти ЧАКФ можна обчислити безпосередньо за допомогою значень часового ряду (вибіркова ЧАКФ). Процедура обчислень можна представити у вигляді двох кроків [2, 23]:

1. Сформулювати вибірку з відхилень від середнього:

$$\{y'(k)\} = \{y(k)\} - \mu_y. \quad (3.81)$$

2. Оцінити рівняння необхідного порядку

$$y'(k) = \phi_{11} y'(k-1) + e(k), \quad (3.82)$$

де $e(k)$ – похибка моделі, яка може і не бути білим шумом. Оскільки рівняння (3.7.1) не містить інших значень крім $y'(k)$ та $y'(k-1)$, то ϕ_{11} є коефіцієнтом часткової кореляції між цими значеннями.

За аналогією можна сформувати та оцінити рівняння другого порядку

$$y'(k) = \phi_{21}y'(k-1) + \phi_{22}y'(k-2) + e(k), \quad (3.83)$$

де ϕ_{22} – коефіцієнт часткової кореляції між $y'(k)$ і $y'(k-2)$, який не враховує впливу значення $y'(k-1)$.

На практиці ЧАКФ обчислюють за такими виразами [23]:

$$\begin{aligned} \phi_{11} = \rho_1, \quad \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}; \quad \phi_{ss} = \frac{\rho_s - \sum_{j=1}^{s-1} \phi_{s-1,j} \rho_{s-j}}{1 - \sum_{j=1}^{s-1} \phi_{s-1,j} \rho_j}, \quad (3.84) \\ \text{де } \phi_{sj} = \phi_{s-1,j} - \phi_{ss} \phi_{s-1,s-j}, \quad j = 1, 2, 3, \dots, s-1. \end{aligned}$$

Приклад обчислення ЧАКФ

Обчислити АКФ і ЧАКФ для процесу АРКС(1,1):

$$y(k) = -0,7y(k-1) + \varepsilon(k) - 0,7\varepsilon(k-1) \quad (3.85)$$

а) Коефіцієнти АКФ обчислимо за допомогою отриманих вище виразів:

$$\rho_0 = 1;$$

$$\rho_1 = \frac{(1 + a_1 b_1)(a_1 + b_1)}{1 + b_1^2 + 2a_1 b_1} = \frac{(1 + 0,49)(-0,7 - 0,7)}{1 + 0,49 + 2 \cdot 0,49} = -0,8445 \quad (3.86)$$

Тепер можна скористатись рекурсивною формулою $\rho_i = -0,7\rho_{i-1}$:

$$\rho_2 = 0,591; \quad \rho_3 = -0,414; \quad \rho_4 = 0,290; \quad \rho_5 = -0,203; \quad \rho_6 = 0,142;$$

$$\rho_7 = -0,010; \quad \rho_8 = 0,070; \quad \rho_9 = -0,049.$$

б) коефіцієнти ЧАКФ:

$$\phi_{11} = \rho_1 = -0,8445;$$

$$\phi_{22} = (\rho_2 - \rho_1^2) / (1 - \rho_1^2) = \frac{0,591 - (-0,8445)^2}{1 - (-0,8445)^2} = -0,425;$$

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = -1,204;$$

$$\phi_{33} = \frac{\rho_3 - \sum_{j=1}^2 \phi_{2,j} \rho_{3-j}}{1 - \sum_{j=1}^2 \phi_{2,j} \rho_j} = -0,262 \quad ; \quad \phi_{44} = \frac{\rho_4 - \sum_{j=1}^3 \phi_{3,j} \rho_{4-j}}{1 - \sum_{j=1}^3 \phi_{3,j} \rho_j} = -0,173 ,$$

де $\phi_{3j} = \phi_{2j} - \phi_{33}\phi_{2,2-j}$, а звідси $\phi_{3,1} = -1,315$; $\phi_{3,2} = -0,74$. Інші коефіцієнти мають такі значення:

$$\phi_{55} = -0,117; \quad \phi_{66} = -0,0681; \quad \phi_{77} = -0,056; \quad \phi_{88} = -0,039 .$$

В загальному випадку ЧАКФ стаціонарного процесу $APKC(p,q)$ повинна збігатись до нуля, починаючи з P -го значення, а швидкість збіжності залежить від значень коефіцієнтів такого полінома:

$$1 + b_1 L + b_2 L^2 + \dots + b_q L^q \quad (3.87)$$

Висновки щодо стаціонарних процесів

1. АКФ процесу $APKC(p,q)$ починає збігатись до нуля при значеннях зсуву $s = q$. Коефіцієнти АКФ задовольняють різницевому рівнянню:

$$\rho_i = a_1 \rho_{i-1} + a_2 \rho_{i-2} + \dots + a_p \rho_{i-p} \quad (3.88)$$

Оскільки корені характеристичного рівняння знаходяться всередині

одиничного кола на комплексній площині, то коефіцієнти АКФ будуть збігатись до нуля при значеннях часового зсуву $s \geq q$.

2. ЧАКФ процесу $APKC(p,q)$ починає збігатись до нуля при $s \geq p$.

Починаючи із значення $s = p$ коефіцієнти ЧАКФ за своїм характером відповідають коефіцієнтам АКФ.

Ці властивості кореляційних функцій дають можливість вибрати порядок моделі $APKC(p,q)$. Так, для деяких конкретних стаціонарних процесів АКФ і ЧАКФ мають такі значення:

- *білий шум*: всі $\rho_s = 0$, $\phi_{ss} = 0$;
- *процес $AP(1)$* при $-1 < a_1 < 0$: $\rho_s = a_1^s$ – експоненційно спадають; $\phi_{11} = \rho_1$; $\phi_{ss} = 0$, $s > 1$;
- *процес $AP(p)$* : коефіцієнти ρ_s збігаються до нуля, можливі коливання; значення ϕ_{ss} суттєво відрізняються від нуля при $s \leq p$ і $\phi_{ss} = 0$, $s > p$;
- *процес $KC(1)$* при $0 < b_1 < 1$: ρ_1 має від'ємне значення і $\rho_s = 0$, $s > 1$; $\phi_{ss} < 0$ і мають коливальну збіжність;
- *процес $KC(1)$* при $-1 < b_1 < 0$: ρ_1 має додатне значення і $\rho_s = 0$, $s > 1$; $\phi_{ss} > 0$ і мають коливальну збіжність;
- *процес $APKC(1,1)$* при $0 < a_1 < 1$: ρ_s мають гладку збіжність і $sign(\rho_1) = sign(a_1 + b_1)$; $\phi_{11} = \rho_1$, далі коливальна збіжність;
- *процес $APKC(1,1)$* при $-1 < a_1 < 0$: ρ_s мають коливальну збіжність і $sign(\rho_1) = sign(a_1 + b_1)$; $\phi_{11} = \rho_1$, далі гладка збіжність; $sign(\phi_{ss}) = sign(\phi_{11})$;
- *процес $APKC(p,q)$* : ρ_s мають гладку або коливальну збіжність, починаючи із значення $s > q$; ϕ_{ss} мають гладку або коливальну збіжність, починаючи з $s > p$.

Оцінювання статистичної значимості коефіцієнтів АКФ. Для того щоб

коректно визначити порядок моделі $APKC(p,q)$, необхідно користуватись відповідними статистиками. Для стаціонарного процесу значення r_s мають нормальний розподіл. Оскільки при оцінюванні порядку ми користуємось вибірковими статистичними характеристиками, то для встановлення статистичної значимості r_s (тобто, суттєвої відмінності від нуля значень АКФ) можна скористатись статистикою Бокса-Пірса або Люнга-Бокса.

Статистика Бокса-Пірса, або Q -статистика, розраховується за виразом [2, 23]:

$$Q = N \sum_{i=1}^s r_i^2 \quad (3.89)$$

Якщо досліджуваний процес $APKC$ стаціонарний, то статистика Q має розподіл χ^2 з s степенями свободи. Це досить наближена статистика, але її використовують для оцінювання стаціонарності *довгих вибірок* даних.

Для *коротких вибірок* запропонована статистика Люнга-Бокса:

$$Q_{LB} = N(N+2) \sum_{i=1}^s r_i^2 / (N-i) \quad (3.90)$$

яку використовують також при визначенні порядку процесу. Оскільки значення Q є пропорційним $\sum r_i^2$, то менші значення Q відповідають меншій ступені автокорельованості процесу, вибраного для дослідження. Очевидно, що існують інші можливості для аналізу стаціонарності процесів, представлених статистичними/експериментальними даними.

Запитання і вправи до третього розділу

1. Дайте означення процесу авторегресії з ковзним середнім? Поясніть можливу природу випадкового процесу $\{\varepsilon(k)\}$ у правій частині рівняння?

2. Які процеси називають інтегрованими? Наведіть приклади інтегрованих процесів.
3. Дайте означення слабкої стаціонарності? Чим відрізняється слабка стаціонарність від сильної?
4. В чому полягає різниця між теоретичними і вибірковими статистичними характеристиками? Як можна скористатись теоретичними статистичними характеристиками?
5. Поясніть різницю між кореляцією та коваріацією? Чи можна користуватись коваріацією з тією ж метою, що використовують кореляцію?
6. Обчисліть коваріаційну функцію $\text{cov}[x(k), x(k-s)]$ для процесу ковзного середнього:

$$x(k) = 0,15\gamma(k-1) + 0,25\gamma(k-2) + 0,35\gamma(k-3) + 0,25\gamma(k-4),$$
де $\gamma(k)$ – випадковий процес білого шуму:
$$E[\gamma(k)\gamma(l)] = \begin{cases} \sigma_\gamma^2, & k=l; \\ 0, & k \neq l. \end{cases}$$
7. Для авторегресійного процесу другого порядку

$$y(k) = a_0 + a_2 y(k-2) + \varepsilon(k)$$
 - (а) умовні математичні сподівання: $E_{k-2}[y(k)]; E_{k-1}[y(k)]; E_k[y(k+2)]$;
 - (б) коваріації: $\text{cov}[y(k), y(k-1)]; \text{cov}[y(k), y(k-2)]$;
 - (в) значення автокореляційної функції: $\rho(1), \rho(2), \rho(3)$.
 - (г) значення часткової автокореляційної функції ϕ_{11} і ϕ_{22} .
8. Побудуйте функцію прогнозування $\hat{y}(k+s) = E_k[y(k+s)]$ для процесу АРКС(1,2). Знайдіть аналітичний вираз для похибок оцінок прогнозів $f(s) = y(k+s) - \hat{y}(k+s)$ і теоретичну автокореляційну функцію для послідовності $\{f(s)\}$.
9. Поясніть умови стаціонарності процесу АР(1)? Наведіть приклад фінансового чи економічного процесу, для якого було б доцільно виконати відомі умови стаціонарності.
10. Яка на вашу думку загальна мета аналізу стаціонарності процесів? Якщо в результаті виконаного аналізу процес виявляється

нестационарним, то про це ще свідчить з точки зору побудови моделі процесу і оцінювання прогнозу для нього?

11. Для чого будують систему рівнянь Юла-Уокера?
12. Як можна інтерпретувати однорідний розв'язок, Aa_1^k , при аналізі стаціонарності процесів АРКС?
13. Яка умова стаціонарності процесу ковзного середнього довільного порядку?
14. Сформулюйте умови стаціонарності для процесу АРКС довільного порядку?
15. При якому значенні параметра зсуву s ЧАКФ процесу $АРКС(p,q)$ починає збігатись до нуля? Починаючи із якого значення коефіцієнти ЧАКФ за своїм характером відповідають коефіцієнтам АКФ?
16. Які статистики використовують для аналізу статистичної значущості АКФ?
17. Чи можна будувати функції прогнозування для нестационарних процесів?
18. Як можна аналізувати збіжність нестационарного процесу?
19. Чи може бути нестационарний процес нелінійним?

РОЗДІЛ 4. МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ ДЕТЕРМІНОВАНИХ І СТОХАСТИЧНИХ ТРЕНДІВ

4.1 Вступ до моделювання детермінованих та стохастичних трендів

Процеси, нестационарні стосовно дисперсії, описують рівняннями АРУГ (авторегресія з умовною гетероскедастичністю), АРУГ-М (АРУГ модифікована), УАРУГ (узагальнена АРУГ), стохастичними моделями волатильності та іншими модифікаціями моделей умовної дисперсії. Вони дають можливість описувати волатильність ряду, тобто ступінь його мінливості за допомогою рівняння динаміки дисперсії. А загальні тенденції руху того чи іншого процесу до зростання чи спадання характеризують його *трендом*. Так, зростання валового внутрішнього продукту (ВВП) розвинених країн на довгих часових інтервалах характеризується детермінованим позитивним (зростаючим) трендом, а ВВП країн з перехідною економікою досить часто характеризується негативним (спадаючим) трендом із значними коливаннями.

Можна сказати, що *детермінований тренд* описує довгострокові зміни

процесів, він вказує на загальний напрямок розвитку процесу. Він може мати досить високий порядок, наприклад, п'ятий. Якщо ж тренд відносно часто змінюється в часі в довільному напрямку, то його називають *стохастичним*. До таких процесів часто відносяться, наприклад, процеси формування цін на біржах, значні випадкові коливання курсів валют, зумовлені випадковими збуреннями, такими як фінансові кризи, локальні військові дії, енергетичні кризи і т. ін.

Для опису динаміки процесів з трендами можна застосовувати рівняння авторегресії з ковзним середнім. Це рівняння має таку узагальнену структуру:

$$y(k) = \text{дрейф} + \text{детермінований тренд} + \text{авторегресія} + \text{сезонна компонента} + \text{ковзне середнє}. \quad (4.1)$$

Подібна структура дає можливість описувати різні ефекти, які наявні в динамічних процесах, в тому числі і тренд. Розв'язки рівнянь АРКС містять складові, які описують тренди, детерміновані і нерегулярні компоненти процесу. Прикладами рівнянь, що використовуються для формального опису детермінованого тренду, можуть бути рівняння прямої, поліноми, експоненти та комбінації гармонічних складових:

$$\begin{aligned} y(k) &= a_0 + \sum_{i=1}^p a_i y(k-i) + c_1 t + \varepsilon(k), \\ y(k) &= a_0 + c_1 t + c_2 t^2 + \dots + c_m t^m + \varepsilon(k), \\ y(k) &= \exp(ck) + \varepsilon(k), \\ y(k) &= a_0 + \sum_{i=1}^p a_i \sin(\omega_0 i) + \sum_{i=1}^q b_i \cos(\omega_0 i) + \varepsilon(k), \end{aligned} \quad (4.2)$$

де P, q – порядки відповідних компонент моделі; очевидно, що подані вище приклади моделей можуть включати інші необхідні компоненти для опису складових досліджуваних процесів.

Вони описують відносно тривалі періоди стійкого спаду/зростання процесів або детерміновані коливання (останнє рівняння). Очевидно, що для

більш повного формального опису процесу подані вище рівняння необхідно розширювати відповідними складовими. Крім ВВП, тренд є характерним, наприклад, для таких процесів:

- зростання туристичної активності в деяких країнах;
- зростання добробуту в розвиненій економіці;
- неперервне зростання об'єму наукової інформації в глобальному масштабі;
- розвиток демографічних процесів (при цьому тренди можуть мати різний характер в локальному та глобальному масштабах);
- поведінка ринкових цін на енергоносії, зерно і т. ін.

Можна вказати на такі причини появи детермінованого тренду у фінансово-економічних процесах:

- вплив сталого розвитку науково-технічного прогресу на соціально-економічні процеси на довгих часових інтервалах (десятки і сотні років);
- поступальний розвиток технологій у промисловості та сільському господарстві;
- неперервне зменшення об'ємів наявних природних ресурсів – руди, нафти, газу;
- неперервне підвищення вимог до якості продукції;
- вплив кліматичних змін на результати ведення різних видів виробничої діяльності.

Поява *випадкового тренду* може бути зумовлена такими факторами:

- технологічні новинки, які швидко з'являються і створюють значний вплив на розвиток ФЕП;
- вплив особистості та окремих груп населення на протікання фінансово-економічних процесів в слабо розвинених країнах на відносно коротких часових інтервалах;
- нові інформаційні технології, які сприяють прискореному розвитку бізнесу та економіки;
- взаємодія соціально-політичних та фінансово-економічних процесів, що протікають в різних країнах світу;
- техногенні катастрофи, локальні та глобальні військові дії, природні аномалії.

Наприклад, 6-7 років засухи підряд може зменшити населення Землі в 3-4 рази. Чорнобильська катастрофа призвела до тяжких демографічних та соціально-економічних наслідків для України за відносно короткий період часу. Вона створила значний негативний вплив на розвиток соціально-економічних

процесів, здоров'я великих груп населення в Україні та Білорусі.

Випадкові тренди можна описати комбінаціями випадкових величин з параметрами, які відповідають параметрам фактичних випадкових процесів. Для формалізації опису випадкового тренду доцільно застосовувати імітаційне моделювання. В загальному випадку процес може містити як детерміновану, так і випадкову складові тренду. Розглянемо деякі можливості математичного опису випадкових трендів.

4.2 Модель випадкового кроку

Детермінований тренд називають також *глобальним трендом* процесу. Однак, на сьогодні існує тенденція до вироблення більш загального підходу до формального опису тренду, а саме, використання *локальних* моделей замість глобальних. При цьому тренд розглядають як стохастичну функцію часу. Одним із підходів до опису локального тренду є введення залежності коефіцієнтів моделі від часу:

$$y(k) = a(k) + d_1(k)k, \quad (4.3)$$

де $a(k)$ – локальна константа; $d_1(k)$ – коефіцієнт, який визначає локальний нахил тренду. Отримано результати моделювання, які показують, що такі функції є більш робастними ніж детерміновані функції поліноміального типу.

Альтернативним підходом опису локального тренду є використання рекурсивних рівнянь типу авторегресії, $y(k) = a_0 + y(k-1)$, або у варіанті:

$$y(k) = a(k) + y(k-1) + \varepsilon(k), \quad (4.4)$$

де $\varepsilon(k)$ – випадкова змінна, яку для спрощення аналізу можна прийняти послідовністю білого шуму з відомою (або оціненою) дисперсією. Це рівняння можна назвати рівнянням випадкового кроку із змінним в часі перетином $a(k)$. Нагадаємо, що випадковий процес типу білого шуму виникає у випадку

комбінованого впливу на основний досліджуваний процес множини випадкових збурень, що має місце, наприклад, у перехідній економіці.

Для повноти формального опису процесу модель (4.2.2) необхідно доповнити рівнянням, яке описує зміну в часі перетину, тобто $a(k)$. Наприклад, $a(k) = a(k-1) + v(k)$, де $v(k)$ – випадковий збурюючий процес. Зазначимо, що рівень (значення) $y(k)$ та швидкість зростання $a(k)$ можуть безпосередньо не спостерігатись (не вимірюватись). Наприклад, в моделі лінійного зростання спостереження часового ряду визначаються сумою $(y(k) + e(k))$, де $e(k)$ – похибка моделі, зумовлена рядом факторів, що виникають при її побудові. Зокрема, це збурення, похибки вимірів та обчислювальні похибки.

Модель випадкового кроку є однією з найпростіших, яка дає можливість описати випадковий тренд в деяких випадках. Вона має такий вигляд:

$$\begin{aligned} y(k) &= y(k-1) + \varepsilon(k), \\ \text{Або } \Delta y(k) &= \varepsilon(k), \end{aligned} \tag{4.5}$$

де $\varepsilon(k)$ – часто апроксимується білим шумом з нульовим середнім. Приростом значення основної змінної в даній моделі є випадкова величина. Знайдемо розв’язок цього рівняння ітераційним методом:

$$\begin{aligned} y(1) &= y_0 + \varepsilon(1), \\ y(2) &= y(1) + \varepsilon(2) = y_0 + \varepsilon(1) + \varepsilon(2), \\ &\vdots \\ y(k) &= y_0 + \sum_{i=1}^k \varepsilon(i). \end{aligned} \tag{4.6}$$

Розв’язок (4.6) називають *невласним (improper)*, оскільки він містить незважену суму випадкових величин. Розглянемо статистичні характеристики отриманого розв’язку.

Математичне сподівання: $E[y(k)] = y_0 = \text{const}$; умовне математичне сподівання:

$$y(k+1) = y(k) + \varepsilon(k+1),$$

$$E_k[y(k+1) | y(k), y(k-1), \dots, \varepsilon(k), \varepsilon(k-1), \dots] = E[y(k) + \varepsilon(k+1)] = y(k) \quad (4.7)$$

Функція прогнозування на довільну кількість кроків на основі розв'язку:

$$y(k) = y_0 + \sum_{i=1}^k \varepsilon(i),$$

$$y(k+s) = y_0 + \sum_{i=1}^{k+s} \varepsilon(i) = y_0 + \sum_{i=1}^k \varepsilon(i) + \sum_{i=k+1}^{k+s} \varepsilon(i) = y(k) + \sum_{i=k+1}^{k+s} \varepsilon(i), \quad (4.8)$$

$$E_k[y(k+s)] = y(k) + E_k \left[\sum_{i=k+1}^{k+s} \varepsilon(i) \right] = y(k).$$

Тобто величина прогнозу у даному випадку не залежить від кількості кроків. Знайдемо значення дисперсії для розв'язку $y(k)$. Для моменту k :

$$\begin{aligned} \text{var}[y(k)] &= E\{[y(k) - \mu_y][y(k) - \mu_y]\} = \\ &= E\left\{ \left[y(0) + \sum_{i=1}^k \varepsilon(i) - y(0) \right] \left[y(0) + \sum_{i=1}^k \varepsilon(i) - y(0) \right] \right\} = \quad (4.9) \\ &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k)][\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k)]\} = k\sigma_\varepsilon^2. \end{aligned}$$

Для моменту часу $(k-s)$:

$$\begin{aligned} \text{var}[y(k-s)] &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k-s)][\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k-s)]\} = \\ &= (k-s)\sigma_\varepsilon^2. \end{aligned} \quad (4.10)$$

Таким чином, $\text{var}[y(k)] \neq \text{const}$ і $\lim_{k \rightarrow \infty} \text{var}[y(k)] = \infty$.

Автокореляційна функція для цього процесу має вигляд:

$$\rho(s) = \frac{\gamma(s)}{\sqrt{\text{var}[y(k)] \text{var}[y(k-s)]}} = \frac{(k-s)\sigma^2}{\sqrt{k\sigma^2 (k-s)\sigma^2}} = \sqrt{\frac{k-s}{k}} \quad (4.11)$$

Тобто АКФ для процесу випадкового кроку є повільно спадаючою функцією.

Модель випадкового кроку плюс дрейф (зміщення або перетин). В цьому випадку до моделі випадкового кроку додається константа a_0 :

$$y(k) = a_0 + y(k-1) + \varepsilon(k) \quad (4.12)$$

За відомої початкової умови, $y(0) = y_0$, розв'язок рівняння (4.2.8) має вигляд:

$$y(k) = y_0 + a_0 k + \sum_{i=1}^k \varepsilon(k) \quad (4.13)$$

Таким чином, на $y(k)$ впливають дві нестационарні компоненти – лінійний детермінований тренд $a_0 k$ і стохастичний тренд $\sum \varepsilon(k)$. Математичне сподівання і умовне математичне сподівання мають вигляд:

$$E[y(k)] = y_0 + a_0 k, \quad E_k[y(k+s)] = y_0 + a_0(k+s) \quad (4.14)$$

Для того щоб отримати функцію прогнозування запишемо рівняння (4.2.9) для моменту $(k+s)$:

$$y(k+s) = y_0 + a_0(k+s) + \sum_{i=1}^{k+s} \varepsilon(k) = y(k) + a_0 s + \sum_{i=1}^s \varepsilon(k+i), \quad (4.15)$$

і умовне математичне сподівання

$$E_k[y(k+s)|k] = y(k) + a_0 s. \quad (4.16)$$

Таким чином, отримана функція прогнозування відрізняється від функції прогнозування для випадкового кроку (4.2.5) тим, що містить складову $a_0 s$.

Модель випадкового кроку з додатковою шумовою складовою. В цій моделі залежна змінна $y(k)$ визначається сумою стохастичного тренду та випадкової компоненти, тобто,

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= \mu(k-1) + \varepsilon(k), \end{aligned} \quad (4.17)$$

де $\{\eta(k)\}$, $\{\varepsilon(k)\}$ – незалежні процеси білого шуму з дисперсіями σ_η^2 та σ_ε^2 , відповідно; $E[\varepsilon(k)\eta(k-s)] = 0, \forall k, s$. Для початкової умови $\mu(0) = \mu_0$

розв'язок рівняння (4.2.12) має вигляд: $\mu(k) = \mu_0 + \sum_{i=1}^k \varepsilon(k)$, що представляє собою випадковий тренд для $y(k)$. Тепер повний розв'язок системи (4.2.11),

(4.2.12) приймає такий вигляд:
$$y(k) = \mu_0 + \sum_{i=1}^k \varepsilon(k) + \eta(k).$$

Використовуючи початкову умову $y(0) = y_0 = \mu_0 + \eta_0$, запишемо розв'язок у вигляді:

$$y(k) = y_0 - \eta_0 + \sum_{i=1}^k \varepsilon(k) + \eta(k) \quad (4.18)$$

Таким чином, безумовне середнє основного (досліджуваного) процесу $\{y(k)\}$ є константою $E[y(k)] = E[y(k+s)] = y_0 - \eta_0$. В кожний момент часу на цей процес впливає шумова компонента $\{\eta(k)\}$.

Дисперсія процесу $\{y(k)\}$ залежить від часу, оскільки $Var[y(k)] = k\sigma_\varepsilon^2 + \sigma_\eta^2$ і $Var[y(k-s)] = (k-s)\sigma_\varepsilon^2 + \sigma_\eta^2$. Так само як і у випадку інших моделей із стохастичним трендом $Var[y(k)] \rightarrow \infty$ при $k \rightarrow \infty$. Наявність шумової компоненти свідчить про те, що в даному випадку коефіцієнт кореляції між $y(k)$ та $y(k-s)$ є меншим ніж у випадку моделі чисто випадкового кроку, тобто автокореляційна функція буде спадати швидше ніж для моделі чисто випадкового кроку. Коваріація між $y(k)$ і $y(k-s)$ визначається таким чином:

$$\begin{aligned} cov[y(k)y(k-s)] &= E\{[y(k) - y_0 - \eta_0][y(k-s) - y_0 - \eta_0]\} = \\ &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k) + \eta(k)][\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k-s) + \eta(k-s)]\} = \\ &= (k-s)\sigma_\varepsilon^2, \end{aligned} \quad (4.19)$$

оскільки $\{\varepsilon(k)\}$ і $\{\eta(k)\}$ – незалежні послідовності. Таким чином, автокореляційну функцію $\rho(s)$ можна визначити за виразом:

$$\rho(s) = \frac{(k-s)\sigma_\varepsilon^2}{\sqrt{(k\sigma_\varepsilon^2 + \sigma_\eta^2)[(k-s)\sigma_\varepsilon^2 + \sigma_\eta^2]}}. \quad (4.20)$$

Із порівняння (4.2.14) з виразом (4.2.7) для коефіцієнтів автокореляції процесу чисто випадкового кроку, можна сказати, що коефіцієнти автокореляції для процесу випадкового кроку з додатковим шумом є завжди меншими при $\sigma_\eta^2 > 0$. Для того щоб записати функцію прогнозування для основного процесу

на s кроків, скористаємось рівнянням (4.2.13):

$$\begin{aligned} y(k+s) &= y_0 - \eta_0 + \sum_{i=1}^{k+s} \varepsilon(i) + \eta(k+s) = \\ &= y(k) - \eta(k) + \sum_{i=1}^s \varepsilon(k+i) + \eta(k+s) \end{aligned} \quad (4.21)$$

Умовне математичне сподівання для $y(k+s)$: $E_k[y(k+s)] = y(k) - \eta(k)$.

Таким чином, модель випадкового кроку з шумом містить тренд та нерегулярну компоненту, а прогноз складається з поточного значення $y(k)$, яке зменшується на випадкову величину $\eta(k)$. Постійною складовою $\{y(k)\}$ є стохастичний тренд у вигляді незваженої суми $\sum \varepsilon(k+i)$.

Випадковий крок + шум + дрейф

Модель випадкового кроку з шумом та дрейфом має вигляд:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= a_0 + \mu(k-1) + \varepsilon(k), \end{aligned} \quad (4.22)$$

де a_0 – константа (перетин); $\{\varepsilon(k)\}, \{\eta(k)\}$ – незалежні процеси білого шуму. В цьому випадку тренд $\mu(k)$ містить випадкову складову $\varepsilon(k)$ та детерміновану a_0 . Запишемо розв’язок для $\mu(k)$: $\mu(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i)$, що приводить до виразу для основної змінної $y(k)$ у вигляді:

$$y(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k) \quad (4.23)$$

Враховуючи початкову умову $y_0 = \mu_0 + \eta_0$, запишемо рівняння для $y(k)$:

$$y(k) = y_0 - \eta_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k) \quad (4.24)$$

Таким чином, основна змінна $y(k)$ представляє собою суму детермінованого і стохастичного трендів, а також чисто випадкової складової $\eta(k)$. Шумова складова не обов'язково повинна бути білим шумом. Рівняння (4.24) можна переписати у вигляді:

$$y(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + A(L) \eta(k) \quad (4.25)$$

де $A(L)$ – поліном стосовно оператора затримки (зсуву) L ; $A(L)\eta(k)$ – тут стаціонарний шумовий процес. Рівняння (4.25) називають *моделлю узагальненого тренду з нерегулярною складовою*.

Модель локального лінійного тренду. Модель локального лінійного тренду (ЛЛТ) поєднує в собі кілька процесів випадкового кроку з шумом:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= \mu(k-1) + \lambda(k) + \varepsilon(k), \\ \lambda(k) &= \lambda(k-1) + v(k), \end{aligned} \quad (4.26)$$

де $\{\eta(k)\}$, $\{\varepsilon(k)\}$, $\{v(k)\}$ – незалежні процеси білого шуму. В цьому випадку зміни тренду спричиняються процесом випадкового кроку та шумовою складовою, тобто, $\Delta\mu(k)$ складається з процесу випадкового кроку $\lambda(k)$ та білого шуму $\varepsilon(k)$. Можна показати, що розглянуті вище процеси (випадковий крок плюс шум та випадковий крок плюс дрейф і шум) представляють собою

окремі випадки лінійного локального тренду.

Знайдемо розв'язок для основної змінної $y(k)$. Спочатку запишемо його для $\lambda(k)$:

$$\lambda(k) = \lambda_0 + \sum_{i=1}^k v(i) \quad (4.27)$$

а також для $\mu(k)$:

$$\mu(k) = \mu(k-1) + \lambda_0 + \sum_{i=1}^k v(i) + \varepsilon(k) \quad (4.28)$$

або

$$\mu(k) = \mu_0 + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + (k-1)v(2) + (k-2)v(3) + \dots + v(k) \quad (4.29)$$

Оскільки $y_0 = \mu_0 + \eta_0$, то розв'язок для $y(k)$ має вигляд:

$$y(k) = y_0 + [\eta(k) - \eta_0] + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + (k-1)v(2) + \dots + v(k) \quad (4.30)$$

У цьому рівнянні можна спостерігати об'єднані властивості всіх інших моделей. Кожний елемент послідовності $\{y(k)\}$ містить детермінований тренд $(k(\lambda_0 + v(1)) + (k-1)v(2) + \dots + v(k))$, стохастичний тренд $(\sum \varepsilon(i))$ та нерегулярну компоненту $\eta(k)$. Звичайно, що в узагальненій формі моделі лінійного локального тренду нерегулярна компонента буде визначатися членом $A(L)\eta(k)$. Детермінований тренд залежить в цьому випадку від поточних та

минулих значень послідовності $\{v(k)\}$. Якщо в момент k сума $(\lambda_0 + v(1) + \dots + v(k))$ буде додатною, то коефіцієнт при k буде додатнім.

Очевидно, що в загальному випадку ця сума може бути додатною для деяких k , а для інших від'ємною, а тому тренд може мати відрізки з додатнім та від'ємним нахилом.

Функція прогнозування для ЛЛТ на s кроків:

$$y(k+s) = y_0 + [\eta(k+s) - \eta_0] + \sum_{i=1}^{k+s} \varepsilon(i) + (k+s)(\lambda_0 + v(1)) + (k+s-1)v(2) + (k+s-2)v(3) + \dots + v(k+s), \quad (4.31)$$

або

$$y(k+s) = y(k) + [\eta(k+s) - \eta(k)] + \sum_{i=1}^k \varepsilon(i) + s[\lambda_0 + v(1) + v(2) + \dots + v(k)] + \quad (4.32)$$

Умовне математичне сподівання:

$$E_k[y(k+s)] = [y(k) - \eta(k)] + s[\lambda_0 + v(1) + v(2) + \dots + v(k)]. \quad (4.33)$$

Нахил тренду визначається сумою $s[\lambda_0 + v(1) + v(2) + v(3) + \dots + v(k)]$.

4.3 Методи вилучення тренду

При моделюванні процесів з трендом, можна описати тренд за допомогою вибраної детермінованої чи випадкової функції або вилучити його зовсім з процесу і описувати математично тільки коливання, які накладаються на тренд. Для вилучення тренду застосовують два методи:

- обчислення різниць першого або вищого порядку (залежно від порядку тренду);
- описати тренд за допомогою вибраної функції з подальшим

використанням для опису стаціонарних коливань залишків від тренду.

Якщо процес має один одиничний корінь, то перші різниці будуть стаціонарними, а якщо більше, то для вилучення тренду необхідно обчислювати різниці вищих порядків.

Розглянемо спочатку процес випадкового кроку + шум. Як було показано вище, його розв'язок має вигляд:

$$y(k) = y(0) + a_0 k + \sum_{i=1}^k \varepsilon(i) \quad (4.34)$$

Перша різниця:

$$\begin{aligned} \Delta y(k) &= y(k) - y(k-1) = y_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) - \left(y_0 + a_0(k-1) + \sum_{i=1}^{k-1} \varepsilon(i) \right) = \\ &= a_0 + \varepsilon(k). \end{aligned} \quad (4.35)$$

Очевидно, що послідовність $\{\Delta y(k)\}$ буде стаціонарною, оскільки вона містить тільки константу і білий шум. Математичне сподівання для цього процесу:

$$\begin{aligned} E[\Delta y(k)] &= E[a_0 + \varepsilon(k)] = a_0, \\ \text{Var}[\Delta y(k)] &\equiv E[\Delta y(k) - a_0]^2 = E[\varepsilon(k)]^2 = \sigma_\varepsilon^2, \end{aligned} \quad (4.36)$$

і коваріація:

$$\begin{aligned} \text{Cov}[\Delta y(k), \Delta y(k-s)] &\equiv E[(\Delta y(k) - a_0)(\Delta y(k-s) - a_0)] = \\ &= E[\varepsilon(k) \varepsilon(k-s)] = 0. \end{aligned} \quad (4.37)$$

Оскільки середнє та дисперсія процесу – константи, а коваріація між $\Delta y(k)$ і $\Delta y(k-s)$ залежить тільки від s , то послідовність $\{\Delta y(k)\}$ є

стаціонарною.

Процес випадкового кроку + шум представляє собою досить цікавий випадок з точки зору простоти та характеристик цієї моделі. Перші різниці описуються рівнянням: $\Delta y(k) = \varepsilon(k) + \Delta \eta(k)$. Покажемо, що вони стаціонарні:

$$E[\Delta y(k)] = E[\varepsilon(k) + \Delta \eta(k)] = 0;$$

$$\begin{aligned} Var [\Delta y(k)] &\equiv E[\Delta y(k)]^2 = E\{[\varepsilon(k) + \Delta \eta(k)]^2\} = \\ &= E\{[\varepsilon(k)]^2 + 2\varepsilon(k)\Delta \eta(k) + [\Delta \eta(k)]^2\} = \\ &= \sigma_{\varepsilon}^2 + 2E[\varepsilon(k)\Delta \eta(k)] + E\{[\eta(k)]^2 - 2\eta(k)\eta(k-1) + [\eta(k-1)]^2\} = \\ &= \sigma_{\varepsilon}^2 + 2\sigma_{\eta}^2, \end{aligned} \quad (4.38)$$

$$\begin{aligned} Cov [\Delta y(k), \Delta y(k-1)] &= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)] [\varepsilon(k-1) + \eta(k-1) - \eta(k-2)]\} = \\ &= -\sigma_{\eta}^2, \end{aligned}$$

і коваріація:

$$\begin{aligned} Cov [\Delta y(k), \Delta y(k-s)] &= \\ &= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)] [\varepsilon(k-s) + \eta(k-s) - \eta(k-s-1)]\} = 0, \quad \forall s \geq 1. \end{aligned} \quad (4.39)$$

При $s = 1$, коефіцієнт кореляції між $\Delta y(k)$ і $\Delta y(k-1)$ дорівнює

$$\rho(1) = \frac{Cov [\Delta y(k), \Delta y(k-1)]}{Var [\Delta y(k)]} = -\frac{\sigma_{\eta}^2}{\sigma_{\varepsilon}^2 + 2\sigma_{\eta}^2}. \quad (4.40)$$

Можна показати, що $-0,5 \leq \rho(1) \leq 0$, а всі інші коефіцієнти кореляції є нульовими. Оскільки перші різниці $\Delta y(k)$ в даному випадку представляють собою процес КС(1), то процес випадкового кроку + шум можна формально визначити як АРІКС(0, 1, 1) (авторегресія з інтегрованим ковзним середнім). Зважаючи на те, що константа не впливає на кореляційну функцію, то додатково можна сказати, що розв'язок (4.2.17) також є процесом АРІКС(0, 1, 1).

А модель лінійного локального тренду є процесом АРІКС(0, 2, 2). Перша і друга різниці для цього процесу мають вигляд:

$$\begin{aligned}\Delta y(k) &= \Delta \mu(k) + \Delta \eta(k) = a(k) + \varepsilon(k) + \Delta \eta(k) \\ &\quad \text{i} \\ \Delta^2 y(k) &= \Delta a(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k) \\ &= \delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)\end{aligned}\quad (4.41)$$

Оскільки процес $a(k)$ нестационарний, то можна показати, що перші різниці процесу $y(k)$ нестационарні. Математичне сподівання, дисперсія і коваріація для других різниць $\Delta^2 y(k)$:

$$\begin{aligned}E[\Delta^2 y(k)] &= E[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)] = 0, \\ \text{Var}[\Delta^2 y(k)] &= E[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)]^2 = \\ &= E\{[\delta(k)]^2 + [\Delta \varepsilon(k)]^2 + [\Delta^2 \eta(k)]^2 + \\ &\quad + 2\delta(k)\Delta \varepsilon(k) + 2\delta(k)\Delta^2 \eta(k) + 2\Delta \varepsilon(k)\Delta^2 \eta(k)\} = \\ &= \sigma_\delta^2 + 2\sigma_\varepsilon^2 + 6\sigma_\eta^2, \\ \text{Cov}[\Delta^2 y(k), \Delta^2 y(k-1)] &= E\{[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)] \times \\ &\quad [\delta(k-1) + \Delta \varepsilon(k-1) + \Delta^2 \eta(k-1)]\} = \\ &= E\{[\delta(k) + \varepsilon(k) - \varepsilon(k-1) + \eta(k) - 2\eta(k-1) + \eta(k-2)] \times \\ &\quad [\delta(k-1) + \varepsilon(k-1) - \varepsilon(k-2) + \eta(k-1) \\ &\quad - 2\eta(k-2) + \eta(k-3)]\} = \\ &= -\sigma_\varepsilon^2 - 4\sigma_\eta^2, \\ \text{Cov}[\Delta^2 y(k), \Delta^2 y(k-2)] &= E\{[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)] \times \\ &\quad \times [\delta(k-2) + \Delta \varepsilon(k-2) + \Delta^2 \eta(k-2)]\} = \\ &= E\{[\delta(k) + \varepsilon(k) - \varepsilon(k-1) + \eta(k) - 2\eta(k-1) + \eta(k-2)] \times \\ &\quad \times [\delta(k-2) + \varepsilon(k-2) - \varepsilon(k-3) + \eta(k-2) \\ &\quad - 2\eta(k-3) + \eta(k-4)]\} =\end{aligned}\quad (4.42)$$

Всі інші члени будуть дорівнювати нулю. Таким чином, модель локального лінійного тренду відповідає процесу АРІКС(0, 2, 2). Можна також показати, що коефіцієнти кореляції знаходяться в межах: $-2/3 \leq \rho(1) \leq 0$, $0 \leq \rho(2) \leq 1/6$, а всі інші значення $\rho(s)$ нульові.

Загальний клас моделей АРІКС(p, d, q). Модель такого класу описується рівнянням АРІКС(p, d, q):

$$A(L)y(k) = B(L)\varepsilon(k), \quad (4.43)$$

де $A(L)$ і $B(L)$ – поліноми порядку p і q стосовно оператора L . Припустимо, що $A(L)$ має один одиничний корінь, а всі корені полінома $B(L)$ знаходяться поза межами одиничного кола (для зворотного характеристичного рівняння). Поліном $A(L)$ можна розкласти на складові $(1-L)A^*(L)$, де $A^*(L)$ – поліном порядку $p-1$. Оскільки $A(L)$ має тільки один одиничний корінь, то всі корені полінома $A^*(L)$ лежать за межами одиничного кола. Тепер рівняння (4.3.1) можна записати так:

$$(1-L)A^*(L)y(k) = B(L)\varepsilon(k) \quad (4.44)$$

Позначимо $y^*(k) = \Delta y(k)$ і перейдемо до запису:

$$A^*(L)y^*(k) = B(L)\varepsilon(k). \quad (4.454)$$

Послідовність $\{y^*(k)\}$, тобто перші різниці процесу, є стаціонарною оскільки всі корені полінома $A^*(L)$ знаходяться за межами одиничного кола.

Якщо $A(L)$ має два одиничних корені, то за аналогією можна показати, що другі різниці процесу $\{y(k)\}$ будуть стаціонарними. Таким чином, якщо процес має d одиничних коренів (тобто це процес $I(d)$), то його різниці порядку d будуть стаціонарними. Звідси випливає, що різниці порядку d процесу АРІКС(p, d, q) з d одиничними коренями будуть процесом АРКС(p, q).

Можливі проблеми з диференціюванням. Вище було показано, що шляхом диференціювання у дискретній формі модель нестаціонарного процесу можна перетворити в модель стаціонарного типу АРКС. Однак, це не означає, що всі нестаціонарні процеси можна трансформувати в стаціонарні типу АРКС шляхом диференціювання. Розглянемо, наприклад, модель процесу, який містить детермінований тренд та шумову складову:

$$y(k) = y(0) + \alpha_1 k + \varepsilon(k). \quad (4.46)$$

Перша різниця цього процесу $\Delta y(k)$ описується рівнянням:

$$\Delta y(k) = \alpha_1 + \varepsilon(k) - \varepsilon(k-1), \quad (4.47)$$

яке не є інвертованим процесом, оскільки воно не представляється авторегресією. Нагадаємо, що інвертованість стаціонарного процесу вимагає, щоб компонента ковзного середнього не мала одиничних коренів.

Коректним шляхом трансформування цієї моделі буде оцінювання спочатку такої регресії

$$y(k) = \alpha_0 + \alpha_1 k + \varepsilon(k). \quad (4.48)$$

Залишок цієї моделі $e(k) = y(k) - \hat{y}(k)$ буде стаціонарним. У більш загальному випадку ряд має поліноміальний тренд:

$$y(k) = \alpha_0 + \alpha_1 k + \alpha_2 k^2 + \alpha_3 k^3 + \dots + \alpha_n k^n + e(k). \quad (4.49)$$

де $\{e(k)\}$ – стаціонарний процес.

Порядок полінома можна визначити за допомогою стандартного t – тесту (визначення значимості коефіцієнтів моделі), F – тесту (визначення значимості групи коефіцієнтів), а також інформаційного критерію Акайке або Байєса-Шварца. Залишки моделі (4.3.3) будуть представляти собою стаціонарний процес, який описується моделлю АРКС.

Моделі, стаціонарні стосовно різниць та тренду. В и щ е м и розглянули два типи моделей нестационарних часових рядів: моделі із *стохастичними* та *детермінованими* трендами. У відповідності до термінології, що склалась в економетричній теорії, моделі (процеси), *стаціонарні стосовно різниць* (ССР), можна трансформувати в стаціонарні моделі шляхом диференціювання. Моделі (процеси), *стаціонарні стосовно тренду* (ССТ), можна перетворити в стаціонарні шляхом видалення детермінованого тренду.

При видаленні тренду може виникнути серйозна проблема, якщо застосувати некоректний метод. Вище було розглянуто приклад диференціювання рівняння: $y(k) = y_0 + \alpha_1 k + \varepsilon(k)$. Розглянемо, наприклад, такий процес, стаціонарний стосовно тренду:

$$A(L) y(k) = \alpha_0 + \alpha_1 k + e(k), \quad (4.50)$$

характеристичне рівняння якого $A(L)$ має корені, що лежать за межами одиничного кола (стаціонарний процес). Нехай залишки $e(k)$ мають форму процесу ковзного середнього: $e(k) = B(L)\varepsilon(k)$. Віднімання оцінки детермінованого часового тренду приводить до стаціонарної та інвертованої моделі АРКС. Однак, якщо скористатись позначеннями моделі (4.3.2), тобто,

$$A(L) y^*(k) = \alpha_1 + (1 - L) B(L) \varepsilon(k), \quad (4.51)$$

то бачимо, що перехід до перших різниць процесу СВТ призвело до появи

неінвертованого процесу з одиничним коренем в компоненті ковзного середнього цієї моделі: $(1-L)\varepsilon(k)$. Очевидно, що така ж проблема виникає при застосуванні моделі з поліноміальним трендом.

Таким чином, буде некоректно віднімати детермінований часовий тренд від процесу, стаціонарного стосовно різниць. Якщо при використанні моделі випадкового кроку + дрейф віднімати член $y_0 + a_0 k$ від кожного спостереження, то це не приведе до стаціонарності, оскільки стохастичний тренд не видаляється.

В загальному випадку введення компоненти детермінованого тренду в регресію у випадку, коли його немає, призводить до помилки вибору структури моделі, якщо процес дійсно має одиничний корінь. Може здатись, що детермінований тренд можна оцінити, використовуючи таку регресію. Однак, всі коефіцієнти моделі будуть в даному випадку лише статистичними артефактами в присутності нестационарних похибок моделі.

Приклад моделі з трендом: курс обміну валют (йена/долар США). За період з 1971 по 1991 рік курс йени зріс більше ніж на 60%. Згідно з деякими існуючими версіями моделей цей процес повинен мати стохастичну складову тренду. Цікаво розглянути наслідки вилучення тренду з даного ряду даних. Регресійна модель курсу обміну, побудована на основі квартальних даних, має вигляд:

$$\begin{array}{ll} y(k) = 0.8479 - 0.0064 \cdot k + e(k) & (4.52) \\ (91) & (-14.16) \end{array}$$

В дужках подано значення t -статистики, які свідчать про те, що коефіцієнти мають високу ступінь значимості. Залишки моделі $e(k)$ вже не містять тренд, оскільки він описаний членом $0,0064 \cdot k$. Однак, значення автокореляційної функції залишків не спадають до нуля навіть при зсуві $s = 16$ (тобто через 16 кварталів). Це свідчить про те, що вилучення детермінованого тренду не приводить до стаціонарності ряду залишків.

4.4 Стохастичні тренди та декомпозиція ряду

З наведеного вище прикладу випливає, що багато економічних часових рядів мають стохастичний тренд та нерегулярну (коливальну) компоненту. Важливою економетричною задачею є декомпозиція або розділення ряду на окремі компоненти. Це дає можливість визначити наявність тимчасових та постійних компонент ряду. Так, сучасна теорія споживання, згідно з якою індивідуальний дохід має постійні та перехідні компоненти, підкреслює необхідність такої декомпозиції.

Декомпозиція виконується прямолінійно у випадку, коли тренд детермінований. Так, лінійний тренд означає, що на кожному часовому інтервалі відбувається фіксований приріст/спад процесу. Вираховуючи детерміновану складову з процесу, отримуємо зміни, зумовлені нерегулярними компонентами.

Задача декомпозиції є значно складнішою у випадку, коли тренд стохастичний. Наприклад, припустимо, що необхідно визначити поточну фазу бізнесового циклу. Якщо тренд ВВП є стохастичним, то як можна визначити – буде значення ВВП над чи під трендом? У випадках, коли послідовність значень можна розкласти на постійні та стаціонарні компоненти, то задача може бути розв’язана. Для того щоб краще зрозуміти природу стохастичного тренду, необхідно зазначити, що на відміну від детермінованого, стохастичний тренд може зростати на фіксованому часовому інтервалі *в середньому*. Наприклад, розглянемо модель випадкового кроку + шум:

$$y(k) = y(k-1) + a_0 + \varepsilon(k). \quad (4.53)$$

Оскільки $E[\varepsilon(k)] = 0$, то *середня* зміна значення $y(k)$ визначається константою a_0 . Звичайно, що в будь-який момент часу k істинна зміна буде відрізнятись від константи a_0 на випадкову величину $\varepsilon(k)$. Однак, кожна послідовна зміна в $\{y(k)\}$ приводить до корегування її рівня, незалежно від того, чим ця зміна зумовлена – детермінованою чи випадковою складовою. Як було показано вище, модель випадкового кроку + дрейф не має нерегулярної компоненти, тобто, це модель чистого тренду. Цей факт виявився дуже корисним при виконанні аналізу часових рядів.

В роботі [51] показано, як розкласти процес АРІКС($p,1,q$) на суму випадкового кроку + дрейф і стаціонарну компоненту (тобто, на загальний тренд і нерегулярну складову). Розглянемо, як приклад, процес АРІКС(0,1,2):

$$y(k) = y(k-1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2). \quad (4.54)$$

Якщо $\beta_1 = \beta_2 = 0$, то це рівняння представляє собою модель випадкового кроку + дрейф. Введення в модель двох членів ковзного середнього додає до послідовності $\{y(k)\}$ нерегулярну компоненту. Першим кроком методики Беверіджа-Нельсона є формування функції прогнозування. Для спрощення записів введемо позначення: $e(k) = \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2)$, що приводить до рівняння: $y(k) = y(k-1) + a_0 + e(k)$. При відомій початковій умові y_0 розв'язок відносно $y(k)$ має вигляд:

$$y(k) = a_0 k + y_0 + \sum_{i=1}^k e(i). \quad (4.55)$$

Для моменту $k+s$ отримаємо:

$$y(k+s) = a_0(k+s) + y_0 + \sum_{i=1}^{k+s} e(i). \quad (4.55)$$

Підставимо (4.55) в (4.56) для того щоб виключити y_0 :

$$y(k+s) = a_0 s + y(k) + \sum_{i=1}^s e(k+i). \quad (4.56)$$

Тепер повернемося до аналізу процесу $\varepsilon(k)$:

$$\sum_{i=1}^s e(k+i) = \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i), \quad (4.57)$$

$$y(k+s) = a_0 s + y(k) + \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i).$$

Запишемо функції прогнозування для різних значень s . Оскільки $E_k[\varepsilon(k+i)] = 0, \forall i > 0$, то отримаємо:

$$\begin{aligned} E_k[y(k+1)] &= a_0 + y(k) + \beta_1 \varepsilon(k) + \beta_2 \varepsilon(k-1), \\ E_k[y(k+2)] &= 2a_0 + y(k) + (\beta_1 + \beta_2) \varepsilon(k) + \beta_2 \varepsilon(k-1), \\ &\quad \boxtimes \end{aligned} \quad (4.58)$$

$$E_k[y(k+s)] = s a_0 + y(k) + (\beta_1 + \beta_2) \varepsilon(k) + \beta_2 \varepsilon(k-1).$$

Тобто всі прогнози при $s > 1$ обчислюються за виразом: $s a_0 + y(k) + (\beta_1 + \beta_2) \varepsilon(k) + \beta_2 \varepsilon(k-1)$. Таким чином, функція прогнозування збігається до лінійної функції від горизонту прогнозування s . Нахил цієї функції визначається коефіцієнтом a_0 , а рівень виразом: $y(k) + (\beta_1 + \beta_2) \varepsilon(k) + \beta_2 \varepsilon(k-1)$. Цей стохастичний рівень можна вважати трендом в момент k , який вже позначався через $\mu(k)$. Прогнозоване значення $E_k[y(k+s)]$ визначається цим трендом та детермінованою складовою $a_0 s$. Отриманий результат дає можливість зробити такі висновки:

1. Вказаний тренд визначається як умовне математичне сподівання граничного значення функції прогнозування. В нестрогому сенсі він є “довгостроковим” прогнозом. Значення цього прогнозу буде відрізнятися в кожний момент часу при надходженні нових значень послідовності. В

будь-який момент k нерегулярна компонента ряду представляє собою різницю між $y(k)$ і трендом $\mu(k)$. Таким чином, використовуючи (4.59), знайдемо, що нерегулярна (коливальна) компонента ряду визначається за виразом:

$$\begin{aligned} y(k) - E_k[y(k+s)] + sa_0 &= -(\beta_1 + \beta_2)\varepsilon(k) - \beta_2\varepsilon(k-1), \\ \text{або} \\ y(k) - \mu(k) &= -(\beta_1 + \beta_2)\varepsilon(k) - \beta_2\varepsilon(k-1). \end{aligned} \quad (4.59)$$

Звідси видно, що в будь-який момент часу тренд та нерегулярні компоненти є максимально корельованими.

2. За визначенням, $\varepsilon(k)$, що має дисперсію σ_ε^2 , є інноваційним процесом для $y(k)$. Оскільки зміна тренду при зміні $\varepsilon(k)$ визначається сумою $1 + \beta_1 + \beta_2$, то дисперсія інновацій в тренді може перевищити дисперсію самого $y(k)$. Якщо $(1 + \beta_1 + \beta_2)^2 > 1$, то тренд буде в більшій степені мінливим ніж $y(k)$, оскільки від'ємна кореляція між трендом і нерегулярними компонентами створює ефект згладжування для послідовності $\{y(k)\}$.

3. Тренд представляє собою процес випадкового кроку + дрейф. Позначимо тренд в момент k через $\mu(k)$ і отримаємо: $\mu(k) = y(k) + (\beta_1 + \beta_2)\varepsilon(k) + \beta_2\varepsilon(k-1)$. Таким чином, приріст тренду

$$\begin{aligned} \Delta\mu(k) &= \Delta y(k) + (\beta_1 + \beta_2)\Delta\varepsilon(k) + \beta_2\Delta\varepsilon(k-1) = \\ &= [y(k) - y(k-1)] + (\beta_1 + \beta_2)\varepsilon(k) - \beta_1\varepsilon(k-1) - \beta_2\varepsilon(k-2). \end{aligned} \quad (4.60)$$

Оскільки $y(k) - y(k-1) = a_0 + \varepsilon(k) + \beta_1\varepsilon(k-1) + \beta_2\varepsilon(k-2)$, то

$$\Delta\mu(k) = a_0 + (1 + \beta_1 + \beta_2)\varepsilon(k).$$

Таким чином, $\mu(k) = \mu(k-1) + a_0 + (1 + \beta_1 + \beta_2)\varepsilon(k)$; тобто, тренд в момент

k складається з дрейфу a_0 та інновації у вигляді білого шуму $(1 + \beta_1 + \beta_2)\varepsilon(k)$.

В роботі [37] показано, як виділити тренд та нерегулярну компоненти з даних. Модель процесу оцінюється за відомою методикою Бокса-Дженкінса. Після диференціювання даних належним чином ідентифікована та оцінена модель АРКС містить оцінки коефіцієнтів a_0, β_1 і β_2 . Потім обчислюють значення $\varepsilon(k)$ і $\varepsilon(k-1)$ на основі моделі АРКС як похибки однокрокових прогнозів $y(k)$ і $y(k-1)$, відповідно. Підставляючи оцінки $\hat{\beta}_1, \hat{\beta}_2, \hat{\varepsilon}(k)$ і $\hat{\varepsilon}(k-1)$ в рівняння (4.60), отримаємо значення нерегулярної компоненти. Повторюючи розглянуту обчислювальну процедуру для кожного моменту часу k , отримаємо всю нерегулярну послідовність. З рівняння (4.4.8) видно, що ця нерегулярна компонента є різницею між $y(k)$ і трендом, тобто тепер можна легко отримати складову тренду.

4.5 Вилучення стохастичного тренду із узагальненої моделі АРІКС($p,1,q$)

В цьому розділі розглянемо можливості виділення стохастичного тренду з класу процесів, яким присвячено дану главу. Для прикладу розглянемо наступну модель АРІКС:

$$y(k) = a_0 + y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-2) + \beta_2 \varepsilon(k-2) + \dots \quad (4.61)$$

Введемо позначення

$$e(k) = \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2) + \beta_3 \varepsilon(k-3) + \dots \quad (4.62)$$

і, по аналогії з розглянутими вище прикладами, запишемо розв'язок рівняння (4.63) у вигляді:

$$y(k+s) = y(k) + a_0 s + \sum_{i=1}^s e(k+i) \quad (4.63)$$

У відповідності з рівнянням (4.62) останню складову в рівнянні (4.63) представимо у вигляді:

$$\sum_{i=1}^s e(k+i) = \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i) + \beta_3 \sum_{i=1}^s \varepsilon(k-3+i) + \dots \quad (4.64)$$

Використовуючи можливість розкладання сум, наявних в (4.64), можна записати функцію прогнозу на s кроків:

$$E_k[y(k+s)] = y(k) + a_0 s + \left(\sum_{i=1}^s \beta_i \right) \varepsilon(k) + \left(\sum_{i=2}^{s+1} \beta_i \right) \varepsilon(k-1) + \left(\sum_{i=3}^{s+2} \beta_i \right) \varepsilon(k-2) + \dots \quad (4.65)$$

Для того щоб визначити структуру стохастичного тренду покладемо $s \rightarrow \infty$:

$$\begin{aligned} \lim_{s \rightarrow \infty} E_k[y(k+s)] &= y(k) + a_0 s + \left(\sum_{i=1}^{\infty} \beta_i \right) \varepsilon(k) + \left(\sum_{i=2}^{\infty} \beta_i \right) \varepsilon(k-1) + \\ &+ \left(\sum_{i=3}^{\infty} \beta_i \right) \varepsilon(k-2) + \dots \end{aligned} \quad (4.66)$$

З іншого боку, прогноз на s кроків можна представити у вигляді тотожності:

$$y(k+s) = \Delta y(k+s) + \Delta y(k+s-1) + \Delta y(k+s-2) + \dots + \Delta y(k+1) + y(k) \quad (4.67)$$

$$\begin{aligned} \text{Наприклад, } y(k+2) &= \Delta y(k+2) + \Delta y(k+1) + y(k) = \\ &= y(k+2) - y(k+1) + y(k+1) - y(k) + y(k) = y(k+2). \end{aligned}$$

Тепер прогноз на основі умовного математичного сподівання

$$\begin{aligned}
\lim_{s \rightarrow \infty} E_k[y(k+s)] &= \lim_{s \rightarrow \infty} E_k\{[y(k+s) - y(k+s-1)] + [y(k+s-1) - y(k+s-2)] + \dots \\
&\quad + [y(k+2) - y(k+1)] + [y(k+1) - y(k)] + y(k)\} = \\
&= \lim_{s \rightarrow \infty} E_k[\Delta y(k+s) + \Delta y(k+s-1) + \dots + \Delta y(k+2) + \Delta y(k+1)] + y(k)
\end{aligned} \tag{4.68}$$

Таким чином, для того щоб побудувати прогноз за виразом (4.68), необхідно знайти значення $y(k+1), \dots, y(k+s)$. Останній вираз і (4.63) дають можливість знайти нерегулярну компоненту (тобто, стохастичний тренд) як

$$\begin{aligned}
Irt &= y(k) - \left[\lim_{s \rightarrow \infty} [E_k(y(k+s))] + a_0 s \right] = \\
&= - \lim_{s \rightarrow \infty} E_k[\Delta y(k+s) + \Delta y(k+s-1) + \dots + \Delta y(k+2) + \Delta y(k+1)] - s a_0
\end{aligned} \tag{4.69}$$

На основі наведених викладок можна сформулювати таку послідовність виділення стохастичного тренду.

Крок 1. Використовуючи ряд вимірів $\{y(k)\}$, обчислити послідовність перших різниць $\{\Delta y(k)\}$ і побудувати модель АРКС(p, q) для даного процесу.

Крок 2. Для кожного моменту часу $k = 1, 2, \dots, N$ обчислити за допомогою отриманої на попередньому кроці моделі АРКС(p, q) прогнози $E_k[\Delta y(k+s)]$, тобто, для $\exists k$ і s .

Використовуючи спрогнозовані значення, для кожного k обчислити суму:

$$E_k[\Delta y(k+s) + \Delta y(k+s-1) + \dots + \Delta y(k+1)] + y(k) \tag{4.70}$$

Наприклад, при $s = 10$ для $k = 1$ необхідно обчислити суму:

$$\zeta_1 = E_1[\Delta y(11) + \Delta y(10) + \dots + \Delta y(2)] + y(1) \tag{4.71}$$

Також, у відповідності до (4.68), можна записати:

$$y(1) + E_1 [\Delta y(11) + \Delta y(10) + \dots + \Delta y(2)] = E_1 [y(11)]. \quad (4.72)$$

Звідси можна записати, що стохастична частина тренду буде представлена в момент $k = 1$ наступним виразом:

$$E_1 [y(11)] - a_0 s, \quad (4.73)$$

де $a_0 s$ – детермінована частина тренду. Аналогічно, можна записати суму для моменту $k = 2$:

$$\zeta_2 = E_2 [\Delta y(12) + \Delta y(11) + \dots + \Delta y(3)] + y(2). \quad (4.74)$$

Для ряду, що складається з N спостережень, компонента випадкового тренду в момент часу $k = N$ визначається сумою:

$$\zeta_N = E_N [\Delta y(N+10) + \Delta y(N+9) + \dots + \Delta y(N+1)] + y(N). \quad (4.75)$$

В результаті отримаємо послідовність значень сум:

$$\{\zeta(k)\} = [\zeta_1, \zeta_2, \dots, \zeta_N]. \quad (4.76)$$

Крок 3. Формування нерегулярної компоненти ряду:

$$Irt(k) = -E_k [\Delta y(k+10) + \Delta y(k+9) + \dots + \Delta y(k+1)]. \quad (4.77)$$

Для більшості рядів значення s можна брати невеликим, наприклад, $s = 2$ для процесу, що описується рівнянням:

$$y(k) = y(k-1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2), \quad (4.78)$$

оскільки всі прогнози для $s > 2$ будуть рівними нулю. Якщо модель АРКС, що побудована на першому кроці, містить коефіцієнти авторегресії, які повільно спадають, то s необхідно брати відносно великим, таким щоб $E_k[y(k+s)] \rightarrow a_0$.

Приклад 4.2. Розглянемо модель курсу обміну валют

$$\Delta y(k) = -0,0113 + \varepsilon(k) + 0,2973 \varepsilon(k-1). \quad (4.79)$$

Прогноз на один крок: $E_k[\Delta y(k+1)] = -0,0113 + 0,2973 \varepsilon(k)$. Таким чином, для кожного значення k суму, що визначається виразом

$$E_k[\Delta y(k+100) + \Delta y(k+99) + \dots + \Delta y(k+1)], \quad (4.80)$$

можна замінити на $-100 \cdot 0,0113 + 0,2973 \varepsilon(k)$ при $s = 100$. Так, для першого використаного спостереження стохастичну частину тренду можна представити у вигляді: $y(1) + 0,2973 \varepsilon(1)$. Повторення цього виразу в кожний момент часу дає можливість визначити детерміновану і випадкову компоненти тренду.

4.6 Розкладання тренду для моделі випадкового кроку шум+дрейф

Як було показано вище, модель випадкового кроку + шум + дрейф має вигляд:

де $\eta(k)$, $\varepsilon(k)$ – некорельовані процеси білого шуму. Запишемо розв’язок для моментів k і $(k + s)$:

$$\begin{aligned} y(k) &= \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k) \\ y(k + s) &= \mu_0 + a_0(k + s) + \sum_{i=1}^{k+s} \varepsilon(i) + \eta(k + s) \end{aligned} \quad (4.82)$$

З рівняння (4.82) знайдемо:

$$\mu_0 = y(k) - a_0 k - \sum_{i=1}^k \varepsilon(i) - \eta(k) \quad (4.83)$$

і підставимо цей вираз в (4.84):

$$y(k + s) = y(k) + a_0 s + \sum_{i=1}^s \varepsilon(k + i) + \eta(k + s) - \eta(k) \quad (4.84)$$

Запишемо рівняння (4.85) для моменту $y(k + 1)$:

$$y(k + 1) = \mu_0 + a_0(k + 1) + \sum_{i=1}^{k+1} \varepsilon(i) + \eta(k + 1) \quad (4.85)$$

і знайдемо першу різницю:

$$\begin{aligned} \Delta y(k + 1) &= y(k + 1) - y(k) = \\ &= \mu_0 + a_0 k + a_0 + \sum_{i=1}^k \varepsilon(i) + \varepsilon(k + 1) + \eta(k + 1) - \end{aligned} \quad (4.86)$$

Знайдемо математичне сподівання, дисперсію і коваріацію для $\Delta y(k)$:

$$\begin{aligned}
 E[\Delta y(k)] &= E[a_0 + \varepsilon(k) + \eta(k) - \eta(k-1)] = a_0. \\
 \text{var}[\Delta y(k)] &= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)][\varepsilon(k) + \eta(k) - \eta(k-1)]\} = \\
 &= \sigma_\varepsilon^2 + 2\sigma_\eta^2 = \text{const.} \\
 \text{cov}[\Delta y(k), \Delta y(k-1)] &= \\
 &= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)][\varepsilon(k-1) + \eta(k-1) - \eta(k-2)]\} = \\
 &= -\sigma_\eta^2. \\
 \text{cov}[\Delta y(k), \Delta y(k+s)] &= 0, \quad k \geq 2.
 \end{aligned} \tag{4.87}$$

Знайдені характеристики дозволяють представити послідовність $\{\Delta y(k)\}$ через компоненти білого шуму як

$$\Delta y(k) = a_0 + e(k) + \beta_1 e(k-1), \tag{4.88}$$

де $e(k)$ – послідовність, яка включає в себе обидві компоненти: $\{\varepsilon(k)\}$ і $\{\eta(k)\}$, які необхідно розділити.

Якщо скористатись рівнянням (4.87) для описання вихідного процесу, то у відповідності до (4.88) тільки $\text{cov}[\Delta y(k), \Delta y(k-1)] \neq 0$. Таким чином, основні статистичні характеристики процесу (4.89) мають вигляд:

$$\begin{aligned}
 \text{Var}[\Delta y(k)] &= \text{Var}[e(k) + \beta_1 e(k-1)] = \\
 &= E\{[e^2(k) + 2\beta_1 e(k)e(k-1) + \beta_1^2 e^2(k-1)]\} = \\
 &= (1 + \beta_1^2) \text{Var}[e(k)] = (1 + \beta_1^2) \sigma_\varepsilon^2. \\
 \text{Cov}[\Delta y(k), \Delta y(k-1)] &= E\{[e(k) + \beta_1 e(k-1)][e(k-1) + \beta_1 e(k-2)]\} =
 \end{aligned} \tag{4.89}$$

Тепер можна прирівняти значення статистичних характеристик процесу, отримані раніше:

$$\sigma_{\varepsilon}^2 + 2\sigma_{\eta}^2 = (1 + \beta_1^2)\sigma_e^2 \quad \text{і} \quad -\sigma_{\eta}^2 = \beta_1\sigma_e^2. \quad (4.90)$$

Звідси випливає, що

$$\begin{aligned} \sigma_{\eta}^2 &= -\beta_1 \sigma_e^2, \\ \sigma_{\varepsilon}^2 &= (1 + \beta_1^2)\sigma_e^2 - 2\sigma_{\eta}^2 = (1 + \beta_1^2)\sigma_e^2 + 2\beta_1\sigma_e^2 = \\ &= \sigma_e^2 (1 + \beta_1)^2. \end{aligned} \quad (4.91)$$

Тепер вилучимо самі послідовності $\{\varepsilon(k)\}$ і $\{\eta(k)\}$. З рівняння (4.91) запишемо прогноз на один крок:

$$E_k[y(k+1)] = y(k) + a_0 - \eta(k) \quad (4.92)$$

і для (4.93):

$$E_k[\Delta y(k+1)] = a_0 + \beta_1 e(k). \quad (4.93)$$

Умовний прогноз для $y(k+1) = y(k) + a_0 + e(k+1) + \beta_1 e(k)$ має вигляд:

$$E_k[y(k+1)] = y(k) + a_0 + \beta_1 e(k). \quad (4.94)$$

Оскільки функції прогнозування (4.92) і (4.93) повинні бути еквівалентними, то $y(k) + a_0 - \eta(k) = y(k) + a_0 + \beta_1 e(k)$

З рівняння для перших різниць $\Delta y(k) = a_0 + \varepsilon(k) + \Delta \eta(k)$ знайдемо, що

$$\varepsilon(k) = \Delta y(k) - a_0 + \beta_1 \Delta \eta(k), \quad (4.96)$$

тобто, послідовність $\{\varepsilon(k)\}$ можна оцінити за допомогою відомих значень $\Delta y(k)$, а також оцінок a_0 і $\Delta \eta(k)$. Таким чином, за допомогою виразів (4.6.21) і (4.6.22) можна розкласти ряд, у якому кореляція між трендом і нерегулярною компонентою дорівнює нулю.

Поданий вище приклад є ілюстрацією наступного важливого моменту. Для того щоб розкласти ряд на компоненти *випадковий крок + дрейф* та *стаціонарну нерегулярну компоненту*, необхідно визначити коефіцієнт кореляції між інноваціями тренду і нерегулярною компонентою. Загалом, було розглянуто два підходи до розкладання процесу АРІКС(0,1,1). Для того щоб застосувати метод Беверіджа-Нельсона, інновації повинні бути повністю корельованими, тобто необхідно виконати умову:

$$E[\varepsilon(k)\eta(k)]/\sigma_\varepsilon\sigma_\eta = 1, \quad (4.97)$$

а другий підхід накладає обмеження: $E[\varepsilon(k)\eta(k)] = 0$. В дійсності, коефіцієнт кореляції може приймати значення в інтервалі $[-1, +1]$. Без виконання додаткової умови стосовно кореляції між інноваціями, тренд і стаціонарну компоненту ідентифікувати неможливо в тому сенсі, що невістачає одного рівняння. Цей результат приводить до необхідності застосування складніших моделей, оскільки завжди необхідно розділяти складові ряду, які протікають одночасно. Проблема є досить складною, тому що економічна теорія не завжди забезпечує існування взаємозв'язків між двома розглянутими інноваційними процесами. Однак, необхідно зазначити, що без наявності апріорної інформації щодо взаємозв'язку між інноваціями тренду та стаціонарними компонентами, *розмежування ряду на випадковий крок + дрейф і стаціонарну компоненту є не єдиним*.

Наприклад, який результат буде отримано у випадку, коли послідовності

$\varepsilon(k)$ і $\eta(k)$ не корельовані, і ми будемо некоректно застосовувати методику Беверіджа-Нельсона з метою отримання тимчасових та постійних складових? Очевидно, що ретроспективні прогнози (всередині вибірки) є інваріантними до форми розкладання ряду. Рівняння (4.6.8) має форму представлення АРІКС(0,1,1), яку необхідно ідентифікувати на першому кроці методики Беверіджа-Нельсона. При цьому немає можливості визначити коректність припущення щодо повністю корельованих інновацій. Це питання не має нічого спільного з коректною формою моделі АРІКС; задача тут полягає в тому, щоб знайти метод розділення інновацій тренду і нерегулярної компоненти.

Що можна визначити в результаті некоректного розділення дисперсій? Застосовуючи декомпозицію Беверіджа-Нельсона до моделі АРІКС(0,1,1), $y(k) = y(k-1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k-1)$, можна визначити, що нерегулярна компонента буде дорівнювати добутку $-\beta_1$ на всю інноваційну послідовність для $y(k)$, тобто на $(\varepsilon(k) + \eta(k))$. Якщо інновації є, фактично, некорельованими, то істинним значенням нерегулярної компоненти в момент k буде $\eta(k)$. Ділема полягає в цьому випадку у тому, що немає можливості ідентифікувати “істинну” модель на основі вибірових даних.

Таким чином, в загальному випадку декомпозицію ряду можна виконувати при двох різних припущеннях щодо корельованості інновацій. Розглянемо можливості декомпозиції логарифму валового внутрішнього продукту (ВВП). Використовуючи методику Беверіджа-Нельсона, знайдемо модель АРІКС(1,1,0) такого вигляду [51]:

$$\Delta y(k) = 0,0043 + 0,395 \Delta y(k-1) + \varepsilon(k), \quad \sigma_\varepsilon = 0,0108. \quad (4.98)$$

Припускаючи, що інновації тренду та нерегулярної компоненти є некорельованими, знайдемо рівняння:

$$y(k) = \mu(k) + A(L) \eta(k),$$

$$\text{де} \quad \Delta \mu(k) = 0,0076 + \varepsilon(k), \quad \sigma_\varepsilon = 0,0061, \quad (4.99)$$

$$A(L) \eta(k) = (1 - 1,487 L + 0,569 L^2) \eta(k), \quad \sigma_\eta = 0,0079.$$

Короткострокові прогнози, отримані за обома моделями є досить близькими. Стандартна похибка однокрокового прогнозу, отриманого за другою моделлю, є дещо меншою ніж за першою, тобто, $(\sigma_\varepsilon^2 + \sigma_\eta^2)^{1/2} \approx 0,0098$ (що є меншим 0,0108). Однак, властивості довгострокових прогнозів є досить різними. Наприклад, записуючи рівняння для $\Delta y(k)$ у вигляді передаточної функції

$$\Delta y(k) = \frac{0,0043 + \varepsilon(k)}{1 - 0,395L}, \quad (4.100)$$

можна знайти імпульсну характеристику цього процесу, сума коефіцієнтів якої складає 1,59 (всі коефіцієнти додатні). Таким чином, зростання інновації на одиницю приведе до зростання логарифму ВВП в 1,59 рази. Однак, друга модель поводить себе інакше – вона має характеристики моделі АРІКС(0,1,2), для якої сума коефіцієнтів імпульсної характеристики складає тільки 0,63. Всі коефіцієнти, починаючи з лагу 4, є від’ємними. Таким чином, зростання інновації на одиницю спричинить набагато більший вплив на $y(k)$ на короткостроковому часовому інтервалі ніж на довгостроковому.

4.7 Стохастичні тренди та декомпозиція ряду

З наведеного вище прикладу випливає, що багато економічних часових рядів мають стохастичний тренд та нерегулярну (коливальну) компоненту. Важливою економетричною задачею є декомпозиція або розділення ряду на окремі компоненти. Це дає можливість визначити наявність тимчасових та постійних компонент ряду. Так, сучасна теорія споживання, згідно з якою індивідуальний дохід має постійні та перехідні компоненти, підкреслює необхідність такої декомпозиції.

Декомпозиція виконується прямолінійно у випадку, коли тренд детермінований. Так, лінійний тренд означає, що на кожному часовому інтервалі відбувається фіксований приріст/спад процесу. Вираховуючи детерміновану складову з процесу, отримаємо зміни, зумовлені нерегулярними компонентами.

Задача декомпозиції є значно складнішою у випадку, коли тренд стохастичний. Наприклад, припустимо, що необхідно визначити поточну фазу бізнесового циклу. Якщо тренд ВВП є стохастичним, то як можна визначити – буде значення ВВП над чи під трендом? У випадках, коли послідовність значень можна розкласти на постійні та стаціонарні компоненти, то задача може бути розв’язана. Для того щоб краще зрозуміти природу стохастичного тренду, необхідно зазначити, що на відміну від детермінованого, стохастичний тренд зростає на фіксованому часовому інтервалі *в середньому*. Наприклад, розглянемо модель випадкового кроку + шум:

$$y(k) = y(k-1) + a_0 + \varepsilon(k). \quad (4.101)$$

Оскільки $E[\varepsilon(k)] = 0$, то *середня* зміна значення $y(k)$ визначається константою a_0 . Звичайно, що в будь-який момент часу k істинна зміна буде відрізнятися від константи a_0 на випадкову величину $\varepsilon(k)$. Однак, кожна послідовна зміна в $\{y(k)\}$ приводить до корегування її рівня, незалежно від того, чим ця зміна зумовлена – детермінованою чи випадковою складовою. Як було показано вище, модель випадкового кроку + дрейф не має нерегулярної компоненти, тобто, це модель чистого тренду. Цей факт виявився дуже корисним при виконанні аналізу часових рядів.

В роботі [23] показано, як розкласти процес АРІКС($p,1,q$) на суму випадкового кроку + дрейф і стаціонарну компоненту (тобто, на загальний тренд і нерегулярну складову). Розглянемо, як приклад, процес АРІКС(0,1,2):

$$y(k) = y(k-1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2). \quad (4.102)$$

Якщо $\beta_1 = \beta_2 = 0$, то це рівняння представляє собою модель випадкового кроку + дрейф. Введення в модель двох членів ковзного середнього додає до послідовності $\{y(k)\}$ нерегулярну компоненту. Першим кроком методики Беверіджа-Нельсона є формування функції прогнозування. Для спрощення записів введемо позначення: $e(k) = \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2)$, що приводить

до рівняння: $y(k) = y(k-1) + a_0 + e(k)$. За відомої початкової умови y_0 розв'язок відносно $y(k)$ має вигляд:

$$y(k) = a_0 k + y_0 + \sum_{i=1}^k e(k) \quad (4.103)$$

Для моменту $k+s$ отримаємо:

$$y(k+s) = a_0(k+s) + y_0 + \sum_{i=1}^{k+s} e(i) \quad (4.104)$$

Підставимо (4.103) в (4.104) для того щоб виключити y_0 :

$$y(k+s) = a_0 s + y(k) + \sum_{i=1}^s e(k+i) \quad (4.105)$$

Тепер повернемося до аналізу процесу $\varepsilon(k)$:

$$\begin{aligned} \sum_{i=1}^s e(k+i) &= \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i) \\ y(k+s) &= a_0 s + y(k) + \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i) \end{aligned} \quad (4.106)$$

Запишемо функції прогнозування для різних значень s . Оскільки $E_k[\varepsilon(k+i)] = 0, \forall i > 0$, то отримаємо:

$$E_k[y(k+1)] = a_0 + y(k) + \beta_1 \varepsilon(k) + \beta_2 \varepsilon(k-1),$$

Тобто, всі прогнози при $s > 1$ обчислюються за виразом: $sa_0 + y(k) + (\beta_1 + \beta_2)\varepsilon(k) + \beta_2\varepsilon(k-1)$. Таким чином, функція прогнозування збігається до лінійної функції від горизонту прогнозування s . Нахил цієї функції визначається коефіцієнтом a_0 , а рівень виразом: $y(k) + (\beta_1 + \beta_2)\varepsilon(k) + \beta_2\varepsilon(k-1)$. Цей стохастичний рівень можна вважати трендом в момент k , який вже позначався через $\mu(k)$. Прогнозоване значення $E_k[y(k+s)]$ визначається цим трендом та детермінованою складовою a_0s . Отриманий результат дає можливість зробити наступні висновки:

1. Вказаний тренд визначається як умовне математичне сподівання граничного значення функції прогнозування. В нестрогому сенсі він є “довгостроковим” прогнозом. Значення цього прогнозу буде відрізнятися в кожний момент часу при надходженні нових значень послідовності. В будь-який момент k нерегулярна компонента ряду представляє собою різницю між $y(k)$ і трендом $\mu(k)$. Таким чином, використовуючи (4.7.7), знайдемо, що нерегулярна (коливальна) компонента ряду визначається за виразом:

$$\begin{aligned} y(k) - E_k[y(k+s)] + sa_0 &= -(\beta_1 + \beta_2)\varepsilon(k) - \beta_2\varepsilon(k-1) \\ \text{або} & \\ y(k) - \mu(k) &= -(\beta_1 + \beta_2)\varepsilon(k) - \beta_2\varepsilon(k-1). \end{aligned} \tag{4.108}$$

Звідси видно, що в будь-який момент часу тренд та нерегулярні компоненти є максимально корельованими.

2. За означенням, $\varepsilon(k)$, що має дисперсію σ_ε^2 , є інноваційним процесом для $y(k)$. Оскільки зміна тренду при зміні $\varepsilon(k)$ визначається сумою $1 + \beta_1 + \beta_2$, то дисперсія інновацій в тренді може перевищити дисперсію самого $y(k)$. Якщо

$(1 + \beta_1 + \beta_2)^2 > 1$, то тренд буде в більшій степені мінливим ніж $y(k)$, оскільки від’ємна кореляція між трендом і нерегулярними компонентами створює ефект згладжування для послідовності $\{y(k)\}$.

3. Тренд представляє собою процес випадкового кроку + дрейф. Позначимо тренд в момент k через $\mu(k)$ і отримаємо: $\mu(k) = y(k) + (\beta_1 + \beta_2)\varepsilon(k) + \beta_2\varepsilon(k-1)$. Таким чином, приріст тренду дорівнює величині:

$$\begin{aligned}\Delta\mu(k) &= \Delta y(k) + (\beta_1 + \beta_2)\Delta\varepsilon(k) + \beta_2\Delta\varepsilon(k-1) = \\ &= [y(k) - y(k-1)] + (\beta_1 + \beta_2)\varepsilon(k) - \beta_1\varepsilon(k-1) - \beta_2\varepsilon(k-2).\end{aligned}$$

$$\text{Оскільки } y(k) - y(k-1) = a_0 + \varepsilon(k) + \beta_1\varepsilon(k-1) + \beta_2\varepsilon(k-2), \quad (4.109)$$

то

$$\Delta\mu(k) = a_0 + (1 + \beta_1 + \beta_2)\varepsilon(k).$$

Таким чином, $\mu(k) = \mu(k-1) + a_0 + (1 + \beta_1 + \beta_2)\varepsilon(k)$; тобто, тренд в момент k складається з дрейфу a_0 та інновації у вигляді білого шуму $(1 + \beta_1 + \beta_2)\varepsilon(k)$.

В роботі [23] показано, як виділити тренд та нерегулярну компоненти з даних. Модель процесу оцінюється за методикою Бокса-Дженкінса. Після диференціювання даних належним чином ідентифікована та оцінена модель АРКС містить оцінки коефіцієнтів a_0, β_1 і β_2 . Потім обчислюють значення $\varepsilon(k)$ і $\varepsilon(k-1)$ на основі моделі АРКС як похибки однокрокових прогнозів $y(k)$ і $y(k-1)$, відповідно. Підставляючи оцінки $\hat{\beta}_1, \hat{\beta}_2, \hat{\varepsilon}(k)$ і $\hat{\varepsilon}(k-1)$ в рівняння (4.7.8), отримаємо значення нерегулярної компоненти. Повторюючи розглянуту обчислювальну процедуру для кожного моменту часу k , отримаємо всю нерегулярну послідовність. З рівняння (4.7.8) видно, що ця нерегулярна компонента є різницею між $y(k)$ і трендом, тобто тепер можна легко отримати складову тренду.

4.8 Виділення стохастичного тренду із узагальненої моделі АРІКС($p,1,q$)

В цьому розділі розглянемо можливості виділення стохастичного тренду з класу процесів, яким присвячено дану главу. Для прикладу розглянемо наступну модель АРІКС:

$$y(k) = a_0 + y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-2) + \beta_2 \varepsilon(k-2) + \dots \quad (4.110)$$

Введемо позначення

$$e(k) = \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2) + \beta_3 \varepsilon(k-3) + \dots \quad (4.111)$$

і, по аналогії з розглянутими вище прикладами, запишемо розв'язок рівняння (4.8.1) у вигляді:

$$y(k+s) = y(k) + a_0 s + \sum_{i=1}^s e(k+i) \quad (4.112)$$

У відповідності з рівнянням (4.112) останню складову в рівнянні (4.113) представимо у вигляді:

$$\sum_{i=1}^s e(k+i) = \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i) + \beta_3 \sum_{i=1}^s \varepsilon(k-3+i) + \dots \quad (4.113)$$

Використовуючи можливість розкладання сум, наявних в (4.8.4), можна записати функцію прогнозу на s кроків:

$$E_k[y(k+s)] = y(k) + a_0 s + \left(\sum_{i=1}^s \beta_i \right) \varepsilon(k) + \left(\sum_{i=2}^{s+1} \beta_i \right) \varepsilon(k-1) + \left(\sum_{i=3}^{s+2} \beta_i \right) \varepsilon(k-2) + \dots \quad (4.114)$$

Для того щоб визначити структуру стохастичного тренду покладемо $s \rightarrow \infty$:

$$\begin{aligned} \lim_{s \rightarrow \infty} E_k[y(k+s)] = & y(k) + a_0 s + \left(\sum_{i=1}^{\infty} \beta_i \right) \varepsilon(k) + \left(\sum_{i=2}^{\infty} \beta_i \right) \varepsilon(k-1) + \\ & + \left(\sum_{i=3}^{\infty} \beta_i \right) \varepsilon(k-2) + \dots \end{aligned} \quad (4.115)$$

З іншого боку, прогноз на s кроків можна представити у вигляді тотожності:

$$\begin{aligned} y(k+s) = & \Delta y(k+s) + \Delta y(k+s-1) + \Delta y(k+s-2) + \dots + \Delta y(k+1) + y(k) \\ & \cdot \\ \text{Наприклад, } y(k+2) = & \Delta y(k+2) + \Delta y(k+1) + y(k) = \\ = & y(k+2) - y(k+1) + y(k+1) - y(k) + y(k) = y(k+2). \end{aligned} \quad (4.116)$$

Тепер прогноз на основі умовного математичного сподівання

$$\begin{aligned} \lim_{s \rightarrow \infty} E_k[y(k+s)] = & \lim_{s \rightarrow \infty} E_k \{ [y(k+s) - y(k+s-1)] + [y(k+s-1) - y(k+s-2)] + \dots \\ & + [y(k+2) - y(k+1)] + [y(k+1) - y(k)] + y(k) \} = \\ = & \lim_{s \rightarrow \infty} E_k [\Delta y(k+s) + \Delta y(k+s-1) + \dots + \Delta y(k+2) + \Delta y(k+1)] + y(k) \end{aligned} \quad (4.117)$$

Таким чином, для того щоб побудувати прогноз за виразом (4.8.8), необхідно знайти значення $y(k+1), \dots, y(k+s)$. Останній вираз і (4.8.3) дають можливість знайти нерегулярну компоненту (тобто, стохастичний тренд) у вигляді:

На основі поданих вище викладок можна сформулювати таку методику виділення стохастичного тренду.

Крок 1. Використовуючи ряд вимірів $\{y(k)\}$, обчислити послідовність перших різниць $\{\Delta y(k)\}$ і побудувати модель АРКС(p, q) для даного процесу.

Крок 2. Для кожного моменту часу $k = 1, 2, \dots, N$ обчислити за допомогою отриманої на попередньому кроці моделі АРКС(p, q) прогнози $E_k[\Delta y(k + s)]$, тобто для $\exists k$ і s .

Використовуючи спрогнозовані значення, для кожного k обчислити суму:

$$E_k[\Delta y(k + s) + \Delta y(k + s - 1) + \dots + \Delta y(k + 1)] + y(k). \quad (4.119)$$

Наприклад, при $s = 10$ для $k = 1$ необхідно обчислити суму:

$$\zeta_1 = E_1[\Delta y(11) + \Delta y(10) + \dots + \Delta y(2)] + y(1). \quad (4.120)$$

Також, у відповідності до (4.8.8), можна записати:

$$y(1) + E_1[\Delta y(11) + \Delta y(10) + \dots + \Delta y(2)] = E_1[y(11)]. \quad (4.121)$$

Звідси можна записати, що стохастична частина тренду буде представлена в момент $k = 1$ таким виразом:

$$E_1[y(11)] - a_0 s \quad (4.122)$$

де $a_0 s$ – детермінована частина тренду. Аналогічно, можна записати суму для моменту $k = 2$:

$$\zeta_2 = E_2 [\Delta y(12) + \Delta y(11) + \dots + \Delta y(3)] + y(2). \quad (4.123)$$

Для ряду, що складається з N спостережень, компонента випадкового тренду в момент часу $k = N$ визначається сумою:

$$\zeta_N = E_N [\Delta y(N + 10) + \Delta y(N + 9) + \dots + \Delta y(N + 1)] + y(N). \quad (4.124)$$

В результаті отримаємо послідовність значень сум:

$$\{\zeta(k)\} = [\zeta_1, \zeta_2, \dots, \zeta_N]. \quad (4.125)$$

Крок 3. Формування нерегулярної компоненти ряду:

$$Irt(k) = -E_k [\Delta y(k + 10) + \Delta y(k + 9) + \dots + \Delta y(k + 1)]. \quad (4.126)$$

Для більшості рядів значення s можна брати невеликим, наприклад, $s = 2$ для процесу, що описується рівнянням:

$$y(k) = y(k - 1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k - 1) + \beta_2 \varepsilon(k - 2), \quad (4.127)$$

оскільки всі прогнози для $s > 2$ будуть рівними нулю. Якщо модель АРКС, що побудована на першому кроці, містить коефіцієнти авторегресії, які повільно спадають, то s необхідно брати відносно великим, таким щоб $E_k[y(k + s)] \rightarrow a_0$.

Приклад 4.1. Розглянемо модель курсу обміну валют

$$\Delta y(k) = -0,0113 + \varepsilon(k) + 0,2973 \varepsilon(k - 1). \quad (4.128)$$

Прогноз на один крок: $E_k [\Delta y(k+1)] = -0,0113 + 0,2973 \varepsilon(k)$. Таким чином, для кожного значення k суму, що визначається виразом

$$E_k [\Delta y(k+100) + \Delta y(k+99) + \dots + \Delta y(k+1)], \quad (4.129)$$

можна замінити на $-100 \cdot 0,0113 + 0,2973 \varepsilon(k)$ при $s = 100$. Так, для першого використаного спостереження стохастичну частину тренду можна представити у вигляді: $y(1) + 0,2973 \varepsilon(1)$. Повторення цього виразу в кожний момент часу дає можливість визначити детерміновану і випадкову компоненти тренду.

4.9 Розкладання тренду для моделі випадкового кроку+шум+дрейф

Як було показано вище, модель випадкового кроку + шум + дрейф має вигляд:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= a_0 + \mu(k-1) + \varepsilon(k), \\ E[\eta(k)\varepsilon(k)] &= 0, \end{aligned} \quad (4.130)$$

де $\eta(k)$, $\varepsilon(k)$ – некорельовані процеси білого шуму. Запишемо розв’язок для моментів часу k і $(k+s)$:

$$\begin{aligned} y(k) &= \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k), \\ y(k+s) &= \mu_0 + a_0(k+s) + \sum_{i=1}^{k+s} \varepsilon(i) + \eta(k+s). \end{aligned} \quad (4.131)$$

З рівняння (4.131) знайдемо:

$$\mu_0 = y(k) - a_0 k - \sum_{i=1}^k \varepsilon(i) - \eta(k) \quad (4.132)$$

і підставимо цей вираз в (4.133):

$$y(k+s) = y(k) + a_0 s + \sum_{i=1}^s \varepsilon(k+i) + \eta(k+s) - \eta(k) \quad (4.133)$$

Запишемо рівняння (4.134) для моменту $y(k+1)$:

$$y(k+1) = \mu_0 + a_0(k+1) + \sum_{i=1}^{k+1} \varepsilon(i) + \eta(k+1) \quad (4.134)$$

і знайдемо першу різницю:

$$\begin{aligned} \Delta y(k+1) &= y(k+1) - y(k) = \\ &= \mu_0 + a_0 k + a_0 + \sum_{i=1}^k \varepsilon(i) + \varepsilon(k+1) + \eta(k+1) - \\ &\quad - \mu_0 - a_0 k - \sum_{i=1}^k \varepsilon(i) - \eta(k) = \\ &= a_0 + \varepsilon(k+1) + \eta(k+1) - \eta(k). \end{aligned} \quad (4.135)$$

Знайдемо математичне сподівання, дисперсію і коваріацію для $\Delta y(k)$:

$$\begin{aligned} E[\Delta y(k)] &= E[a_0 + \varepsilon(k) + \eta(k) - \eta(k-1)] = a_0. \\ \text{var}[\Delta y(k)] &= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)][\varepsilon(k) + \eta(k) - \eta(k-1)]\} = \end{aligned}$$

Знайдені характеристики дають можливість представити послідовність $\{\Delta y(k)\}$ через компоненти білого шуму таким чином:

$$\Delta y(k) = a_0 + e(k) + \beta_1 e(k-1), \quad (4.137)$$

де $e(k)$ – послідовність, яка включає в себе обидві компоненти: $\{\varepsilon(k)\}$ і $\{\eta(k)\}$, які необхідно розділити.

Якщо скористатись рівнянням (4.136) для описання вихідного процесу, то у відповідності до (4.137) тільки $\text{cov}[\Delta y(k), \Delta y(k-1)] \neq 0$. Таким чином, основні статистичні характеристики процесу (4.138) мають вигляд:

$$\begin{aligned} \text{Var} [\Delta y(k)] &= \text{Var} [e(k) + \beta_1 e(k-1)] = \\ &= E \{ [e^2(k) + 2\beta_1 e(k)e(k-1) + \beta_1^2 e^2(k-1)] \} = \\ &= (1 + \beta_1^2) \text{Var} [e(k)] = (1 + \beta_1^2) \sigma_e^2. \end{aligned} \quad (4.138)$$

$$\begin{aligned} \text{Cov} [\Delta y(k), \Delta y(k-1)] &= E \{ [e(k) + \beta_1 e(k-1)] [e(k-1) + \beta_1 e(k-2)] \} = \\ &= E [\beta_1 e(k-1) e(k-1)] = \beta_1 \sigma_e^2. \end{aligned}$$

Тепер можна прирівняти значення статистичних характеристик процесу, отримані раніше:

$$\sigma_\varepsilon^2 + 2\sigma_\eta^2 = (1 + \beta_1^2) \sigma_e^2 \quad ; \quad -\sigma_\eta^2 = \beta_1 \sigma_e^2. \quad (4.139)$$

Звідси випливає, що

$$\begin{aligned} \sigma_\eta^2 &= -\beta_1 \sigma_e^2, \\ \sigma_\varepsilon^2 &= (1 + \beta_1^2) \sigma_e^2 - 2\sigma_\eta^2 = (1 + \beta_1^2) \sigma_e^2 + 2\beta_1 \sigma_e^2 = \end{aligned} \quad (4.140)$$

Тепер вилучимо самі послідовності $\{\varepsilon(k)\}$ і $\{\eta(k)\}$. З рівняння (4.140) запишемо прогноз на один крок:

$$E_k[y(k+1)] = y(k) + a_0 - \eta(k) \quad (4.141)$$

і для (4.142):

$$E_k[\Delta y(k+1)] = a_0 + \beta_1 e(k). \quad (4.142)$$

Умовний прогноз для $y(k+1) = y(k) + a_0 + e(k+1) + \beta_1 e(k)$ має вигляд:

$$E_k[y(k+1)] = y(k) + a_0 + \beta_1 e(k). \quad (4.143)$$

Оскільки функції прогнозування (4.9.18) і (4.9.20) повинні бути еквівалентними, то $y(k) + a_0 - \eta(k) = y(k) + a_0 + \beta_1 e(k)$,

$$\text{або} \quad \eta(k) = -\beta_1 e(k). \quad (4.144)$$

З рівняння для перших різниць $\Delta y(k) = a_0 + \varepsilon(k) + \Delta \eta(k)$ знайдемо, що

$$\varepsilon(k) = \Delta y(k) - a_0 + \beta_1 \Delta \eta(k), \quad (4.145)$$

тобто послідовність $\{\varepsilon(k)\}$ можна оцінити за допомогою відомих значень $\Delta y(k)$, а також оцінок a_0 і $\Delta \eta(k)$. Таким чином, за допомогою виразів (4.9.21) і (4.9.22) можна розкласти ряд, у якому кореляція між трендом і нерегулярною компонентою дорівнює нулю.

Поданий вище приклад є ілюстрацією наступного важливого моменту.

Для того щоб розкласти ряд на компоненти *випадковий крок + дрейф* та *стаціонарну нерегулярну компоненту*, необхідно визначити коефіцієнт кореляції між інноваціями тренду і нерегулярною компонентою. Загалом, було розглянуто два підходи до розкладання процесу АРІКС(0,1,1). Для того щоб застосувати методику Беверіджа-Нельсона, інновації повинні бути повністю корельованими, тобто необхідно виконати умову:

$$E[\varepsilon(k)\eta(k)]/\sigma_{\varepsilon}\sigma_{\eta}=1, \quad (4.146)$$

а другий підхід накладає обмеження: $E[\varepsilon(k)\eta(k)]=0$. В дійсності, коефіцієнт кореляції може приймати значення в інтервалі $[-1,+1]$. Без виконання додаткової умови стосовно кореляції між інноваціями, тренд і стаціонарну компоненту ідентифікувати неможливо в тому сенсі, що невиває одного рівняння. Цей результат приводить до необхідності застосування складніших моделей, оскільки завжди необхідно розділяти складові ряду, які протікають одночасно. Проблема є досить складною, тому що економічна теорія не завжди забезпечує існування взаємозв'язків між двома розглянутими інноваційними процесами. Однак, необхідно зазначити, що без наявності апріорної інформації щодо взаємозв'язку між інноваціями тренду та стаціонарними компонентами, *розмежування ряду на випадковий крок + дрейф і стаціонарну компоненту є не єдиним*.

Наприклад, який результат буде отримано у випадку, коли послідовності $\varepsilon(k)$ і $\eta(k)$ не корельовані, і ми будемо некоректно застосовувати методику Беверіджа-Нельсона з метою отримання тимчасових та постійних складових? Очевидно, що ретроспективні прогнози (всередині вибірки) є інваріантними до форми розкладання ряду. Рівняння (4.9.8) має форму представлення АРІКС(0,1,1), яку необхідно ідентифікувати на першому кроці методики Беверіджа-Нельсона. При цьому немає можливості визначити коректність припущення щодо повністю корельованих інновацій. Це питання не має нічого спільного з коректною формою моделі АРІКС; задача тут полягає в тому, щоб знайти метод розділення інновацій тренду і нерегулярної компоненти.

Що можна визначити в результаті некоректного розділення дисперсій? Застосовуючи декомпозицію Беверіджа-Нельсона до моделі АРІКС(0,1,1),

$y(k) = y(k-1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k-1)$, можна визначити, що нерегулярна компонента буде дорівнювати добутку $-\beta_1$ на всю інноваційну послідовність для $y(k)$, тобто, на $(\varepsilon(k) + \eta(k))$. Якщо інновації є, фактично, некорельованими, то істинним значенням нерегулярної компоненти в момент k буде $\eta(k)$. Ділема полягає в даному випадку в тому, що немає можливості ідентифікувати “істинну” модель на основі вибіркових даних.

Таким чином, в загальному випадку декомпозицію ряду можна виконувати при двох різних припущеннях щодо корельованості інновацій. Розглянемо можливості декомпозиції логарифму валового внутрішнього продукту (ВВП). Використовуючи методику Беверіджа-Нельсона, знайдемо модель АРІКС(1,1,0) наступного вигляду:

$$\Delta y(k) = 0,0043 + 0,395 \Delta y(k-1) + \varepsilon(k), \quad \sigma_\varepsilon = 0,0108. \quad (4.147)$$

Припускаючи, що інновації тренду та нерегулярної компоненти є некорельованими, знайдемо рівняння:

$$y(k) = \mu(k) + A(L) \eta(k),$$

де $\Delta \mu(k) = 0,0076 + \varepsilon(k), \quad \sigma_\varepsilon = 0,0061,$ (4.148)

$$A(L) \eta(k) = (1 - 1,487L + 0,569L^2) \eta(k), \quad \sigma_\eta = 0,0079.$$

Короткострокові прогнози, отримані за обома моделями є досить близькими. Стандартна похибка однокрокового прогнозу, отриманого за другою моделлю, є дещо меншою ніж за першою, тобто, $(\sigma_\varepsilon^2 + \sigma_\eta^2)^{1/2} \approx 0,0098$ (що є меншим 0,0108). Однак, властивості довгострокових прогнозів є досить різними. Наприклад, записуючи рівняння для $\Delta y(k)$ у вигляді передаточної функції

$$\Delta y(k) = \frac{0,0043 + \varepsilon(k)}{1 - 0,395L}, \quad (4.149)$$

можна знайти імпульсну характеристику цього процесу, сума коефіцієнтів якої складає 1,59 (всі коефіцієнти додатні). Таким чином, зростання інновації на одиницю приведе до зростання логарифму ВВП в 1,59 рази. Однак, друга модель поводить себе інакше – вона має характеристики моделі АРІКС(0,1,2), для якої сума коефіцієнтів імпульсної характеристики складає тільки 0,63. Всі коефіцієнти, починаючи з лагу 4, є від’ємними. Таким чином, зростання інновації на одиницю спричинить набагато більший вплив на $y(k)$ на короткостроковому часовому інтервалі ніж на довгостроковому.

Запитання і вправи до четвертого розділу

1. Поясніть, якими функціями можна описати детермінований тренд? Які функції, на вашу думку, необхідно застосовувати в першу чергу?
2. Запишіть математичну модель, яка відповідає такій узагальненій структурі процесу:

$$y(k) = \text{дрейф} + \text{детермінований тренд} + \text{авторегресія} + \\ + \text{сезонна компонента} + \text{ковзне середнє}.$$

3. Детермінований тренд вказує на довгострокові зміни процесу, а на які зміни вказує стохастичний тренд? Чому його називають стохастичним?
4. Яким чином можна встановити характер тренду – детермінований чи стохастичний?
5. Який тренд можна описати за допомогою моделі випадкового кроку з константою: $y(k) = a_0 + y(k-1) + \varepsilon(k)$?
6. Які можливі причини виникнення детермінованого та випадкового трендів? Наведіть приклади з економіки перехідного періоду.
7. Який розв’язок різницевого рівняння називають невластним? Від чого залежить напрям розвитку такого процесу? У якому випадку такий розв’язок може бути корисним (наведіть приклад процесу)?
8. Модель випадкового кроку з додатковою випадковою (шумовою) складовою має вигляд:

$$y(k) = \mu(k) + \eta(k), \\ \mu(k) = \mu(k-1) + \varepsilon(k),$$

де $\{\eta(k)\}, \{\varepsilon(k)\}$ – незалежні процеси білого шуму з дисперсіями σ_{η}^2 та σ_{ε}^2 , відповідно; $E[\varepsilon(k)\eta(k-s)] = 0, \forall k, s$. Запишіть прогноз змінної $y(k)$ на п'ять кроків.

9. Якщо модель лінійного локального тренду

$$y(k) = \mu(k) + \eta(k),$$

$$\mu(k) = \mu(k-1) + \lambda(k) + \varepsilon(k),$$

$$\lambda(k) = \lambda(k-1) + v(k),$$

де $\{\eta(k)\}, \{\varepsilon(k)\}, \{v(k)\}$ – незалежні процеси білого шуму, застосувати до формалізованого опису фактичного (реального) випадкового процесу, то яким чином можна визначити параметри випадкових послідовностей $\{\eta(k)\}, \{\varepsilon(k)\}, \{v(k)\}$?

10. Дайте інтерпретацію графіку процесу, який моделюється за допомогою моделі лінійного локального тренду, за умови, що процеси $\{\eta(k)\}, \{\varepsilon(k)\}, \{v(k)\}$ мають нормальний розподіл з нульовим середнім та одиничною дисперсією.

11. Чи є різниця між вилученням тренду із процесу за допомогою різниць і за допомогою детермінованої моделі? Якщо є, то в чому вона полягає? Поясніть результати на прикладі.

12. Дайте пояснення скороченню: АРІКС(p, d, q)? Яким чином можна знайти оцінку параметра d ?

13. Запропонуйте послідовність побудови моделі процесу із стохастичним трендом.

14. Яка мета вилучення тренду із процесу?

РОЗДІЛ 5. АНАЛІЗ КОІНТЕГРОВАНИХ ПРОЦЕСІВ

5.1. Задача моделювання нестационарних коінтегрованих процесів

В багатьох випадках побудови математичних моделей на основі часових рядів ми зустрічаємось з нестационарністю у вигляді часового тренду, яка

одночасно наявна у кількох часових рядах, для яких будується модель. Загалом відомо, що тренд вказує на відносно довгострокові зміни процесу.

При цьому під трендом, як і раніше, будемо розуміти поточне середнє значення процесу, яке може бути отримане, наприклад, за допомогою процедури фільтрації, описане за допомогою полінома або іншої детермінованої функції. На короткому часовому інтервалі його можна оцінити за допомогою формули для обчислення поточного середнього, а в іншому випадку – за допомогою поліноміальних або інших детермінованих функцій. Стохастичний тренд можна описати за допомогою лінійної комбінації випадкових процесів.

В загальному випадку задача коректного математичного опису тренду може бути не дуже простою. Якщо для часового ряду даних з трендом побудувати модель авторегресії з ковзним середнім, то вона може виявитись коректною з точки зору параметрів, що використовуються для визначення її адекватності (коефіцієнт детермінації наближається до 1, статистика Дарбіна-Уотсона близька до 2, коефіцієнти (параметри) моделі статистично значимі у відповідності із значеннями t – статистики і т. ін.), але така модель не відтворює істинні причинні зв'язки між змінними, а тільки формально констатує наявність спільного тренду.

Від нестационарного ряду можна перейти до стаціонарного шляхом знаходження перших різниць або різниць більш високого порядку. Але при переході до часового ряду у вигляді різниць в ньому залишається лише та інформація, що відповідає короткостроковим змінам (коливанням) процесу. Тобто *втрачається інформація щодо довгострокових змін процесу*, яка міститься в тих складових змінної, що втрачаються при переході до різниць. Задача вилучення тренду за допомогою різниць чи детермінованої функції вирішується в кожному випадку аналізу даних окремо, оскільки наперед невідомо, який саме тренд міститься у процесі – детермінований чи випадковий (стохастичний). Остаточний вибір методу побудови моделі відбувається на етапі аналізу моделі для залишків, які ми отримуємо після формального опису (або вилучення) тренду.

Тому в загальному випадку існує проблема коректного моделювання процесів (часових рядів), що містять позитивні чи негативні тренди (залежно від напрямку його руху – вгору або вниз), а також інші складові, які приводять до відхилення від стаціонарності.

Одним із підходів до коректного формалізованого математичного опису

нестационарних процесів з трендом є так званий *коінтеграційний підхід* (до аналізу даних), який дає можливість побудувати модель для кількох нестационарних процесів (як мінімум двох), що розглядаються в межах однієї постановки задачі. *Якщо ряди нестационарні, але коінтегровані, то їх лінійна комбінація є стаціонарним процесом.* Зокрема, цей підхід дає можливість додатково будувати моделі корегування похибок (МКП), що містять механізм корегування похибок (математичного опису) стосовно довгострокових ефектів, які містяться у процесах, що моделюються.

Концепція коінтегрованості змінних передбачає існування довгострокового зв'язку між значеннями відповідних змінних. Тобто припускається існування спільної врівноваженої (довгострокової) траєкторії руху (поведінки) цих змінних, від якої вони можуть відхилятися на коротких проміжках часу, але економічні механізми в цілому діють таким чином, що рівновага повторно відтворюється і зберігається на довгих часових інтервалах шляхом корегування відповідних відхилень від врівноваженого стану.

Якщо процеси, що розглядаються спільно (в одній моделі), коінтегровані, то можна побудувати відповідну модель корегування похибки, що має такі характеристики:

- вона одночасно *відображає короткострокові та довгострокові аспекти* динаміки досліджуваних показників;
- забезпечує *побудову коректної регресії*;
- *не потребує попереднього розподілу змінних* на ендогенні та екзогенні;
- *відповідає основним припущенням економетрики.*

Попередня обробка даних при побудові економетричних моделей включає, як правило, чисельне диференціювання вихідних часових рядів за допомогою різниць з метою видалення тренду. Проте задача ускладнюється у випадку аналізу багатовимірних часових рядів, що містять нестационарні змінні. У цьому випадку для опису нестационарних змінних будують лінійні комбінації декількох нестационарних змінних, які називають коінтегрованими.

Наявність коінтеграційних співвідношень є характерною для багатьох процесів у економіці, особливо для процесів економіки перехідного періоду (ЕПП). Відомо, що більшість процесів ЕПП відрізняються високою динамікою, нестационарністю, наявністю інтегруючих компонент та нелінійності.

Для ілюстрації концепції коінтеграційного підходу розглянемо рівняння, що описує *попит на грошову масу*:

де $m(k)$ – довгостроковий попит на грошову масу; $p(k)$ – рівень споживчих цін; $y(k)$ – реальний прибуток; $r(k)$ – процентна ставка центрального банку; $e(k)$ – збурення у вигляді (теоретично) стаціонарного випадкового процесу; β_i – коефіцієнти рівняння.

Якщо припустити, що попит на грошову масу відповідає пропозиції на неї, то можна одержати експериментальні дані, які характеризують пропозицію грошової маси, рівень цін, реальний прибуток (для його вимірювання можна скористатися реальним внутрішнім валовим продуктом (ВВП)), а також відповідними значеннями процентних ставок центрального банку.

З логічних міркувань очевидно, що значення параметрів мають бути такими: $\beta_1 = 0$, $\beta_2 > 0$, $\beta_3 < 0$, що повинно бути підтверджено експериментальними даними. Також очевидно, що послідовність $\{e(k)\}$ повинна включати ті складові, що не ввійшли у пояснюючі змінні.

Для того щоб застосовувана теорія мала сенс, необхідно щоб будь-яке відхилення попиту на грошову масу мало короткочасний характер. Основна вимога до послідовності $\{e(k)\}$ полягає у тому, що вона повинна бути стаціонарною, тому що у випадку наявності стохастичного тренду похибки моделі будуть накопичуватися і надалі неможливо буде усувати відхилення від ринкової рівноваги.

Проблема, з якою доводиться зіштовхуватися у даному випадку, полягає у тому, що реальний ВВП, пропозиція грошової маси, рівень цін і процентна ставка можуть бути нестаціонарними інтегрованими процесами типу $I(1)$ (лінійний тренд). А це означає, що значення кожної із перерахованих змінних може істотно коливатися і не повертатися до очікуваного довгострокового (асимптотичного) рівня. Теорія коінтеграційного аналізу ґрунтується на припущенні, що існує така комбінація нестаціонарних змінних, яка є стаціонарною. Запишемо рівняння (5.1) у такій формі:

$$e(k) = m(k) - \beta_0 - \beta_1 p(k) - \beta_2 y(k) - \beta_3 r(k) \quad (5.2)$$

Оскільки послідовність $\{e(k)\}$ повинна бути стаціонарною, то лінійна комбінація у правій частині (5.2) також повинна бути стаціонарною. Цей приклад ілюструє основну ідею коінтеграційного методу: *рівновага процесів, що включають нестаціонарні змінні, потребує існування стаціонарної комбінації цих змінних*. Наведений приклад із попитом на грошову масу є тільки однією ілюстрацією стаціонарної комбінації нестаціонарних змінних.

Формально *коінтеграційний аналіз* починається із розгляду множини економічних змінних, що знаходяться у певній довгостроковій рівновазі:

$$\beta_1 x_1(k) + \beta_2 x_2(k) + \dots + \beta_n x_n(k) = 0 \quad (5.3)$$

Якщо позначити вектори параметрів і вимірів через β і $\mathbf{x}$, відповідно, то рівняння (5.1.3) можна представити у векторному вигляді:

$$\mathbf{e}(k) = \beta \mathbf{x}(k), \quad \beta = [\beta_1, \beta_2, \dots, \beta_n]^T, \quad \mathbf{x} = [x_1(k), x_2(k), \dots, x_n(k)]^T \quad (5.4)$$

Введемо таке *означення коінтеграції*:

Компоненти вектора $\mathbf{x}(k)$ є *коінтегрованими* порядку d, b (що позначається так: $\mathbf{x}(k) \sim CI(d, b)$), якщо

1. Всі компоненти вектора $\mathbf{x}(k)$ є інтегрованими процесами порядку d .
2. Існує такий вектор параметрів $\beta = [\beta_1, \beta_2, \dots, \beta_n]^T$, що лінійна комбінація $\beta_1 x_1(k) + \beta_2 x_2(k) + \dots + \beta_n x_n(k)$ є інтегрованою послідовністю порядку $(d - b)$, де $b > 0$. Тобто b показує наскільки зменшується ступінь інтегрованості результуючого процесу після об'єднання змінних в одній моделі.

Вектор параметрів β називають *коінтеграційним вектором*. Якщо в рівнянні (5.1.1) рівень цін, реальний прибуток і процентна ставка є

інтегрованими процесами порядку $I(1)$, а лінійна комбінація змінних

$$e(k) = m(k) - \beta_0 - \beta_1 p(k) - \beta_2 y(k) - \beta_3 r(k) \quad (5.5)$$

стаціонарна, то ці змінні є інтегрованими порядку $(1,1)$. Вектори вимірів і параметрів мають вигляд:

$$\mathbf{x}^T(k) = [m(k) \ 1 \ p(k) \ y(k) \ r(k)], \quad \beta^T = [1 - \beta_0 - \beta_1 - \beta_2 - \beta_3] \quad (5.6)$$

Відхилення від довгострокової ринкової рівноваги визначається випадковою складовою $e(k)$. Оскільки $\{e(k)\}$ – стаціонарна послідовність, то відхилення від рівноваги носять короткочасний характер.

Чотири важливих зауваження щодо означення коінтеграції

1. Коінтеграція відноситься до утворення *лінійної комбінації нестаціонарних змінних* (процесів). Теоретично допускається також існування довгострокової нелінійної залежності між елементами множини інтегрованих змінних. Проте, на сьогодні теорія коінтеграційного аналізу не дає можливості просто встановлювати нелінійні зв'язки [23]. Слід також зазначити, що коінтеграційний вектор не єдиний. Якщо $(\beta_1, \beta_2, \beta_3, \dots, \beta_n)$ – коінтеграційний вектор, то для будь-якого ненульового значення λ вектор $(\lambda\beta_1, \lambda\beta_2, \lambda\beta_3, \dots, \lambda\beta_n)$ також є коінтеграційним вектором. Взагалі *коефіцієнт при одній із змінних вибирають рівним одиниці* з метою нормування коінтеграційного вектора. Наприклад, для нормування вектора стосовно змінної x_1 обирають значення $\lambda = 1/\beta_1$.

2. *Всі змінні повинні бути інтегрованими одного порядку.* Проте, це не означає, що всі однаково інтегровані змінні є коінтегрованими. Зазвичай множина змінних, що мають порядок інтегрування $I(d)$, не є коінтегрованою. Відсутність коінтеграції означає, що відсутня довгострокова рівновага між вибраними змінними, тобто характер їхньої зміни може істотно різнитися у часі. *Якщо змінні інтегровані, але мають різний порядок інтегрування, то вони не*

можуть бути коінтегрованими. Припустимо змінна $x_1(k)$ має порядок інтегрованості $I(d_1)$, а змінна $x_2(k)$ має порядок інтегрованості $I(d_2)$, де $d_2 > d_1$. Можна показати, що будь-яка лінійна комбінація змінних x_1, x_2 буде мати порядок інтегрованості $I(d_2)$.

Необхідно зазначити, що *термін «рівновага» тут не зовсім вдалий*, оскільки економісти-теоретики й економісти-математики вкладають у нього дещо різний зміст. Так, економісти-теоретики зазвичай розуміють під рівновагою рівність між необхідними і реальними результатами транзакцій, попитом і пропозицією і т. ін.

Економетристи використовують цей термін для відображення взаємовідносин між нестационарними змінними. Коінтеграція не потребує щоб довгостроковий зв'язок (рівновага) підтримувався за рахунок ринкових впливів або за рахунок деяких визначених правил поведіння індивідуумів. У більшості випадків *під рівновагою припускають існування причинних, поведінкових або просто співвідношень безпосереднього взаємозв'язку між змінними, що мають тренди однакового характеру.*

3. Якщо вектор $\mathbf{x}(k)$ складається з n компонент, то може існувати $n-1$ лінійно незалежних коінтеграційних векторів. Тобто якщо вектор $\mathbf{x}(k)$ складається з двох компонент, то існує тільки один коінтеграційний вектор. Число коінтеграційних векторів називають *рангом коінтеграції* вектора $\mathbf{x}(k)$. Так, нехай в поданому вище прикладі обсяг випуску грошової маси скорочується при збільшенні обсягу внутрішнього валового продукту (ВВП) і збільшується при зменшенні виробництва ВВП. Це правило можна формально записати так:

$$\begin{aligned} m(k) &= \gamma_0 - \gamma_1[y(k) + p(k)] + e_1(k) = \\ &= \gamma_0 - \gamma_1 y(k) - \gamma_1 p(k) + e_1(k) \end{aligned} \quad (5.7)$$

де $\{e_1(k)\}$ – стаціонарний процес залишків моделі, що описує випуск в оборот грошової маси.

Для моделі потреби у грошовій масі (5.1.1) існує два коінтеграційних вектори для таких змінних як пропозиція грошової маси, рівень цін, реальний прибуток і процентна ставка. Їх можна представити у вигляді коінтеграційної матриці, що складається з двох векторів:

$$B = \begin{bmatrix} 1 & -\beta_0 & -\beta_1 & -\beta_2 & -\beta_3 \\ 1 & -\gamma_0 & \gamma_1 & \gamma_1 & 0 \end{bmatrix} \quad (5.8)$$

Обидві лінійні комбінації, що задаються добутком $Bx(k)$, є стаціонарними. Таким чином, ранг коінтеграції для $x(k)$ дорівнює двом. У випадку наявності множини коінтеграційних векторів може виявитися, що неможливо ідентифікувати скорочену форму взаємозв'язків.

4. У переважній більшості випадків дослідження з коінтеграції фокусуються на багатовимірних процесах, кожен із яких має один одиничний корінь. Причина полягає у тому, що традиційно регресійний аналіз часових рядів застосовується до рядів із нульовим порядком інтегрованості, тобто $I(0)$, а також тому, що існує обмежена кількість економічних змінних, порядок інтегрованості яких перевищує одиницю. Хоча для економіки перехідного періоду досить характерним є інший порядок для більшості фактичних процесів [2].

У випадку, коли не виникає невизначеності, багато авторів використовують термін «коінтеграція» у випадку, коли змінні мають порядок коінтегрованості $CI(1,1)$. Очевидно, що існує також ряд інших можливостей такого типу. Наприклад, множини змінних перемінного рівня інтегрованості $I(2)$ можуть бути коінтегрованими порядку $CI(2,1)$ і при цьому може існувати лінійна комбінація, що задовольняє рівню інтегрованості $I(1)$.

5.2 Моделі коінтеграції

5.2.1 Коінтеграція і тренди

В процесі аналізу коінтегрованих змінних було встановлено, що вони мають однакові стохастичні тренди. Це полегшує розуміння сутності коінтегрованості. З метою більш докладного розгляду цього явища повернемося

до випадку, коли вектор $\mathbf{x}(k)$ містить тільки дві змінних: $\mathbf{x}(k) = [y(k) \ z(k)]^T$. Якщо не розглядати циклічні і сезонні ефекти, то кожен із змінних можна розкласти на випадковий крок і нерегулярну складову (не обов'язково білий шум), тобто у випадку двох змінних маємо:

$$\begin{aligned} y(k) &= \mu_y(k) + \varepsilon_y(k), \\ z(k) &= \mu_z(k) + \varepsilon_z(k), \end{aligned} \tag{5.9}$$

де $\mu_i(k)$ – процес випадкового кроку, що представляє собою перемінний тренд у момент k ; $\varepsilon_i(k)$ – нерегулярна стаціонарна компонента змінної.

Якщо послідовності $\{y(k)\}$, $\{z(k)\}$ є коінтегрованими порядку (1,1), то повинні існувати такі ненульові значення коефіцієнтів β_1, β_2 , для яких лінійна комбінація $\beta_1 y(k) + \beta_2 z(k)$ буде стаціонарною, тобто:

$$\begin{aligned} \beta_1 y(k) + \beta_2 z(k) &= \beta_1 [\mu_y(k) + \varepsilon_y(k)] + \beta_2 [\mu_z(k) + \varepsilon_z(k)] = \\ &= [\beta_1 \mu_y(k) + \beta_2 \mu_z(k)] + [\beta_1 \varepsilon_y(k) + \beta_2 \varepsilon_z(k)]. \end{aligned} \tag{5.10}$$

Для того щоб комбінація $\beta_1 y(k) + \beta_2 z(k)$ була стаціонарною, складова $[\beta_1 \mu_y(k) + \beta_2 \mu_z(k)]$ повинна дорівнювати нулю. Якщо у рівнянні (5.10) залишається один із двох трендів, то лінійна комбінація $\beta_1 y(k) + \beta_2 z(k)$ також буде мати тренд. Оскільки друга складова у правій частині останнього рівняння є стаціонарною, то необхідною і достатньою умовою коінтегрованості $CI(1,1)$ послідовностей $\{y(k)\}$, $\{z(k)\}$ є

$$\beta_1 \mu_y(k) + \beta_2 \mu_z(k) = 0 \tag{5.11}$$

Очевидно, що значення змінних $\mu_x(k), \mu_y(k)$ будуть неперервно змінюватися в часі. Оскільки передбачається, що обидва коефіцієнти β_1, β_2 повинні відрізнятися від нуля, то (5.12) буде виконуватися тільки у тому випадку, якщо виконується таке співвідношення:

$$\mu_y(k) = -\mu_z(k) \frac{\beta_2}{\beta_1}. \quad (5.12)$$

Єдиною можливістю забезпечення існування цієї рівності при ненульових значеннях коефіцієнтів β_1, β_2 є рівність стохастичних трендів. Таким чином, якщо два стохастичних процеси $\{y(k)\}, \{z(k)\}$, інтегровані порядку $I(1)$, то вони повинні мати однакові стохастичні тренди для того, щоб бути коінтегрованими порядку $(1,1)$.

У поданому вище прикладі послідовності $\{y(k)\}, \{z(k)\}$ задовольняли рівнянням:

$$y(k) = \mu(k) + \varepsilon_y(k); \quad z(k) = \mu(k) + \varepsilon_z(k), \quad (5.13)$$

де $\mu(k) = \mu(k-1) + \varepsilon(k)$; $\varepsilon_y(k)$, $\varepsilon_z(k)$, $\varepsilon(k)$ – незалежно розподілені послідовності білого шуму.

За означенням, $\mu(k)$ представляє собою процес випадкового кроку, що є стохастичним трендом для послідовностей $\{y(k)\}, \{z(k)\}$. При цьому $\mu_0 = 0$, а послідовності $\{\varepsilon_y(k)\}, \{\varepsilon_z(k)\}, \{\varepsilon(k)\}$ (по двадцять відліків кожна) є псевдовипадковими числами. Віднімання $z(k)$ від $y(k)$ приводить до стаціонарної послідовності вигляду:

Використовуючи термінологію коінтеграційного аналізу, можна сказати, що добуток вектора $\mathbf{x}(k) = [y(k) \ z(k)]^T$ на коінтеграційний вектор $\beta = [1 \ -1]$ дозволяє отримати стаціонарну послідовність $\varepsilon(k) = \varepsilon_y(k) - \varepsilon_z(k)$. Параметри коінтеграційного вектора повинні бути такими, щоб тренд був вільним від лінійної комбінації змінних. Будь-яка інша лінійна комбінація цих двох змінних буде містити тренд, тому коінтеграційний вектор повинен бути єдиним. Наприклад, комбінація $\beta_3 y(k) + \beta_4 z(k)$ не буде стаціонарною, якщо не буде виконуватися рівність $\beta_3 / \beta_4 = \beta_1 / \beta_2$.

Коінтеграційним вектором для вектора змінних $\mathbf{x}(k) = [y(k) \ z(k) \ w(k)]^T$ буде вектор $\mathbf{c} = [1 \ 1 \ -1]^T$. В результаті одержимо стаціонарну лінійну комбінацію $y(k) + z(k) - w(k)$, для якої можна записати такий результат:

$$\begin{aligned} y(k) + z(k) - w(k) &= [\mu_y(k) + \varepsilon_{y(k)}] + [\mu_z(k) + \varepsilon_z(k)] - [\mu_w(k) + \varepsilon_w(k)] = \\ &= \varepsilon_y(k) + \varepsilon_z(k) - \varepsilon_w(k). \end{aligned} \quad (5.15)$$

Цей приклад ілюструє загальну ідею коінтеграції, яка полягає в тому, що коінтеграція виникає у тому випадку, коли *тренд однієї змінної може бути представлений у вигляді лінійної комбінації трендів інших змінних*. В таких випадках завжди можна знайти такий вектор β , що лінійна комбінація $\beta_1 y(k) + \beta_2 z(k) + \beta_3 w(k)$ не містить тренд. Цей результат легко узагальнюється на випадок n змінних. Розглянемо векторну модель

$$\mathbf{x}(k) = \mu(k) + \varepsilon(k) \quad (5.16)$$

де $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \dots \ x_n(k)]^T$ – n -вимірний вектор основних змінних;

$\mu(k) = [\mu_1(k) \ \mu_2(k) \ \dots \ \mu_n(k)]^T$ – n -вимірний вектор стохастичних трендів;
 $\varepsilon(k)$ – вектор випадкових компонент такої ж розмірності. Якщо необхідно виразити тренд однієї із змінних у вигляді лінійної комбінації інших трендів, то повинен існувати такий вектор β , що

$$\beta_1 \mu_1(k) + \beta_2 \mu_2(k) + \dots + \beta_n \mu_n(k) = 0. \quad (5.17)$$

Якщо помножити (5.17) на вектор β , то отримаємо

$$\beta^T \mathbf{x}(k) = \beta^T \mu(k) + \beta^T \varepsilon(k) \quad (5.18)$$

Зважаючи на те, що $\beta^T \mu(k) = 0$, то $\beta^T \mathbf{x}(k) = \beta^T \varepsilon(k)$. Отже, лінійна комбінація $\beta^T \mathbf{x}(k)$ є стаціонарною. Цей результат легко узагальнюється на випадок, коли існує множина лінійних взаємозв'язків між трендами, тобто

$$\mathbf{B} \mu(k) = 0, \quad (5.19)$$

де $\mathbf{B}[r \times n]$ – матриця, що складається з елементів β_{ij} . Наприклад, якщо для n змінних існує два незалежних коінтеграційних вектори, то матриця $\mathbf{B}$ має вигляд:

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2n} \end{bmatrix} \quad (5.20)$$

Ще одну стаціонарну лінійну комбінацію для вектора $\mathbf{x}(k)$ можна отримати шляхом віднімання із першого рядка елементи другого рядка

помножені на β_{1i} / β_{2i} . Проте, у цій комбінації буде тільки $n-1$ ненульових коефіцієнтів у векторі $\mathbf{x}(k)$. У загальному випадку, якщо існує r коінтеграційних векторів для n змінних, то буде існувати коінтеграційний вектор для кожної підмножини з $(n-r)$ змінних.

5.1.2. Коінтеграція і корегування похибок

Принциповою відмінною властивістю коінтегрованих змінних є те, що на траєкторію їхньої зміни у часі впливають будь-які відхилення від довгострокової рівноваги. Тобто, якщо необхідно повернути систему до стану довгострокової рівноваги, то керуючі зміни, щонайменше, повинні реагувати на величину порушення цієї рівноваги. Наприклад теорія, що описує структуру процесу формування процентних ставок, припускає існування довгострокової рівноваги між короткочасними і довгостроковими значеннями ставок. Якщо різниця між значеннями короткострокових і довгострокових процентних ставок є істотною порівняно із довгостроковим співвідношенням, то короткочасна процентна ставка зрештою повинна піднятися щодо довгострокової ставки. Очевидно, що ця різниця може бути перекрита (а) шляхом збільшення короткочасної ставки та/або зниження довгострокової ставки; (б) шляхом збільшення довгострокової ставки при одночасному і більш істотному збільшенні короткочасної ставки; (в) шляхом зниження довгострокової ставки, але з одночасним незначним зниженням короткочасної ставки. При цьому без повного опису динаміки моделі неможливо визначити яка з можливостей є більш прийнятною. Очевидно те, що на короткочасну динаміку повинно впливати відхилення від довгострокового взаємозв'язку.

Модель динаміки, що логічно випливає з наведених міркувань, називають *моделлю корегування похибок* (відхилення від рівноваги). У моделі корегування похибок (МКП) на короткочасну динаміку змінних у системі впливають відхилення від рівноваги. Якщо припустити, що обидва типи процентних ставок є процесами порядку $I(1)$, то проста МКП, що придатна для опису структури процесу формування процентних ставок, має вигляд:

$$\Delta r_S(k) = \alpha_S [r_L(k-1) - \beta r_S(k-1)] + \varepsilon_S(k), \quad \alpha_S > 0;$$

де $r_L(k)$, $r_S(k)$ – довгострокова і короткострокова процентна ставка, відповідно; збурення у вигляді процесів білого шуму, що можуть бути корельовані між собою; $\alpha_S, \alpha_L, \beta$ – додатні параметри.

Відповідно до опису, *короткочасна і довгострокова процентні ставки змінюються у відповідності із впливом стохастичного збурення* (поданого змінними $\varepsilon_S(k)$, $\varepsilon_L(k)$), а також відповідно до впливу відхилення від довгострокової рівноваги у попередньому періоді. Якщо за інших рівних умов це відхилення є додатним (тобто, $r_L(k-1) - \beta r_S(k-1) > 0$), то короткочасна процентна ставка збільшиться, а довгострокова зменшиться. Довгострокова рівновага досягається при $r_L(k) = \beta r_S(k)$.

Подана вище модель ілюструє взаємозв'язок між МКП і коінтегрованими змінними. За припущенням, різниця $\Delta r_S(k)$ є стаціонарною, що забезпечує порядок інтегрованості для лівої частини рівняння (5.2.7). Очевидно, що ліва частина (5.2.8) також повинна мати ступінь інтегрованості. Якщо послідовність $\varepsilon_S(k)$ стаціонарна, то звідси випливає, що лінійна комбінація $r_L(k-1) - \beta r_S(k-1)$ також повинна бути стаціонарною. Таким чином, розглянуті процентні ставки повинні бути коінтегрованими з вектором коінтеграції, рівним $[1 \ -\beta]$. Аналогічні міркування справедливі стосовно рівняння (5.2.8). Важливим моментом тут є те, що представлення у вигляді МКП потребує щоб дві розглянуті змінні були коінтегрованими порядку $CI(1,1)$.

Отриманий результат принципово не зміниться, якщо сформулювати більш загальну модель шляхом введення затриманих у часі значень змінних в обидва рівняння:

$$\Delta r_L(k) = a_{20} - \alpha_L[r_L(k-1) - \beta r_S(k-1)] + \sum a_{21}^i \Delta r_S(k-1) + \sum a_{22}^i \Delta r_L(k-1) + \varepsilon_L(k) \quad (5.22)$$

де $\varepsilon_S(k)$, $\varepsilon_L(k)$ і всі члени, які включають $\Delta r_S(k-1)$ і $\Delta r_L(k-1)$, є стаціонарними. Таким чином, лінійна комбінація процентних ставок $[r_L(k-1) - \beta r_S(k-1)]$ також повинна бути стаціонарною.

З докладного розгляду рівнянь (5.21) і (5.22) випливає, що вони дуже схожі на векторні авторегресійні моделі (ВАР). Ця модель на основі двох змінних із корекцією похибок представляє собою двовимірну векторну авторегресійну модель для перших різниць, розширену членами корекції похибок $\alpha_S[r_L(k-1) - \beta r_S(k-1)]$ і $-\alpha_L[r_L(k-1) - \beta r_S(k-1)]$.

Зазначимо, що α_L і α_S інтерпретують як параметри *швидкості налагодження*. Чим більше значення має параметр α_S , тим більшою буде реакція змінної $r_S(k)$ на відхилення від довгострокової рівноваги у попередньому періоді. З іншого боку, дуже маленькі значення α_S припускають, що короткочасна процентна ставка нечутлива до порушення рівноваги у попередньому періоді.

Для того щоб на послідовність $\{\Delta r_S(k)\}$ не впливала послідовність значень довгострокової процентної ставки, всі коефіцієнти α_S , a_{12}^i повинні бути рівними нулю. Таким чином, для того щоб між коінтегрованими змінними була відсутня каузальність (причинний зв'язок) Грейнджера, необхідно виконати ще одну умову – коефіцієнт швидкості налагодження повинен бути рівним нулю. Очевидно, що принаймні один із коефіцієнтів швидкості налагодження повинен бути ненульовим. Якщо обидва коефіцієнти α_L, α_S рівні нулю, то співвідношення довгострокової рівноваги не буде існувати і модель не буде коінтегрованою або ж моделлю корекції похибки.

Отриманий результат можна узагальнити на модель, що включає n змінних. Формально вектор $\mathbf{x}(k) = [x_1(k) \ x_2(k) \ \dots \ x_n(k)]^T$ має представлення у вигляді МКП, якщо його можна описати у такий спосіб:

$$\Delta \mathbf{x}(k) = \mathbf{f}_0 + \mathbf{F} \mathbf{x}(k-1) + \mathbf{F}_1 \Delta \mathbf{x}(k-1) + \mathbf{F}_2 \Delta \mathbf{x}(k-2) + \dots + \mathbf{F}_p \Delta \mathbf{x}(k-p) + \varepsilon(k) \quad (5.23)$$

де $\mathbf{f}_0$ – вектор постійних складових (зміщень) розмірності $[n \times 1]$;

$\mathbf{F}$ – матриця елементів f_{jk} таких, що один або більше елементів $f_{jk} \neq 0$;

$\mathbf{F}_i$ – матриці коефіцієнтів $f_{jk}(i)$ вимірності $[n \times n]$;

$\varepsilon(k)$ – вектор збурень вимірності $[n \times 1]$.

Зазначимо, що збурення для різних змінних можуть бути корельовані між собою, тобто $\text{cov}[\varepsilon_i(k)\varepsilon_j(k)] \neq 0$.

Нехай усі компоненти (змінні) вектора $\mathbf{x}(k)$ мають порядок інтегрованості $I(1)$. Якщо для цих змінних існує представлення у вигляді МКП типу (5.23), то обов'язково буде мати місце стаціонарна лінійна комбінація змінних із порядком інтегрованості $I(0)$. Розв'язок рівняння (5.23) відносно $\mathbf{F}\mathbf{x}(k-1)$ має вигляд:

$$\mathbf{F}\mathbf{x}(k-1) = \Delta\mathbf{x}(k) - \mathbf{f}_0 - \sum \mathbf{F}_i \Delta\mathbf{x}(k-i) - \varepsilon(k). \quad (5.24)$$

Оскільки кожна складова у правій частині є стаціонарною, то добуток $\mathbf{F}\mathbf{x}(k-1)$ також повинний бути стаціонарним. Зважаючи на те, що матриця $\mathbf{F}$ є постійною, то кожний рядок цієї матриці є вектором коінтеграції для вектора $\mathbf{x}(k)$. Наприклад, множення першого рядка на вектор $\mathbf{x}(k)$ має вигляд:

$$f_{11}x_1(k-1) + f_{12}x_2(k-1) + \dots + f_{1n}x_n(k-1). \quad (5.25)$$

Оскільки кожний процес $x_i(k-1)$ є інтегрованим порядку $I(1)$, то вектор коефіцієнтів $[f_{11} \ f_{12} \ \dots \ f_{1n}]^T$ повинен бути вектором коінтеграції для вектора $\mathbf{x}(k)$.

Ключовим моментом рівняння (5.23) є наявність у ньому матриці $\mathbf{F}$. При

цьому необхідно зазначити таке:

- Якщо всі елементи матриці $\mathbf{F}$ рівні нулю, то рівняння (5.23) є звичайним векторним авторегресійним рівнянням щодо перших різниць. У цьому випадку представлення для корекції похибки відсутнє, тому що $\Delta \mathbf{x}(k)$ не реагує на відхилення від довгострокової рівноваги у попередньому періоді.
- Якщо один або більше елементів матриці $\mathbf{F}$ відрізняється від нуля, то $\Delta \mathbf{x}(k)$ реагує на відхилення від довгострокової рівноваги у попередньому періоді. Отже, оцінювати $\mathbf{x}(k)$ як векторне авторегресійне рівняння щодо перших різниць буде некоректно, якщо вектор $\mathbf{x}(k)$ має представлення у вигляді МКП. Виключення із розгляду виразу $\mathbf{F}\mathbf{x}(k-1)$ призводить до похибки формального опису процесу, якщо $\mathbf{x}(k)$ представляється у вигляді МКП типу (5.23).

Для того щоб дослідити взаємозв'язок між коінтеграцією і корекцією похибок, необхідно розглянути властивості простої векторної авторегресійної моделі типу:

$$\begin{aligned}x(k) &= a_{11}x(k-1) + a_{22}y(k-1) + \varepsilon_x(k), \\y(k) &= a_{21}x(k-1) + a_{22}y(k-1) + \varepsilon_y(k),\end{aligned}\tag{5.26}$$

де $\varepsilon_x(k), \varepsilon_y(k)$ – збурення у вигляді білого шуму, що можуть бути корельовані між собою; постійні зміщення для спрощення запису не розглядаються. Використовуючи оператори затримки, останні рівняння можна записати у вигляді:

$$\begin{aligned}(1 - a_{11}L)x(k) - a_{12}Ly(k) &= \varepsilon_x(k), \\-a_{21}Lx(k) + (1 - a_{22}L)y(k) &= \varepsilon_y(k).\end{aligned}\tag{5.27}$$

Наступним кроком є знаходження розв'язку цих рівнянь стосовно

$x(k)$, $y(k)$. Перепишемо останню систему у матричному вигляді:

$$\begin{bmatrix} (1-a_{11}L) & -a_{12}L \\ -a_{21}L & (1-a_{22})L \end{bmatrix} \begin{bmatrix} x(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} \varepsilon_x(k) \\ \varepsilon_y(k) \end{bmatrix}. \quad (5.28)$$

Використовуючи правило Крамера для знаходження оберненої матриці, запишемо рішення для $x(k)$, $y(k)$:

$$\begin{aligned} x(k) &= \frac{(1-a_{22})\varepsilon_x(k) + a_{12}L\varepsilon_y(k)}{(1-a_{11}L)(1-a_{22}L) - a_{12}a_{21}L^2}, \\ y(k) &= \frac{a_{21}L\varepsilon_x(k) + (1-a_{11}L)\varepsilon_y(k)}{(1-a_{11}L)(1-a_{22}L) - a_{12}a_{21}L^2}. \end{aligned} \quad (5.29)$$

Таким чином, система рівнянь (5.24), (5.25) перетворена в два різницевих рівняння другого порядку, що мають однакові зворотні характеристичні рівняння виду:

$$(1-a_{11}L)(1-a_{22}L) - a_{12}a_{21}L^2 = 0. \quad (5.30)$$

Розв'язуючи це рівняння стосовно L , одержимо два корені, які можна перетворити у корені звичайного характеристичного рівняння таким чином:

$\lambda = 1/L$. Тобто «пряме» характеристичне рівняння має вигляд:

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0. \quad (5.31)$$

Оскільки обидва рівняння (5.29) і (5.30) мають одне і те ж характеристичне рівняння, то рівняння (5.31) визначає часові характеристики для обох змінних.

Щодо цих характеристик можна зробити такі зауваження:

- Якщо обидва корені характеристичного рівняння (λ_1, λ_2) лежать всередині одиничного кола, то розв'язки рівнянь (5.2.14), (5.2.15) відносно $x(k), y(k)$ будуть стійкими. Для достатньо великих значень k або при таких початкових умовах, що забезпечують нульове однорідне рішення, зазначена умова усталеності забезпечує стаціонарність змінних. Ці змінні не можуть бути коінтегрованими порядку $CI(1,1)$, оскільки вони є стаціонарними.
- Якщо один із коренів лежить за межами одиничного кола, то розв'язок має розбіжний характер. Жодна із різниць цих змінних не буде стаціонарною і вони не можуть бути $CI(1,1)$. Також, якщо обидва характеристичних корені є одиничними, другі різниці кожної змінної будуть стаціонарними. Оскільки обидва процеси є процесами типу $I(2)$, ці змінні не можуть мати порядок коінтегрованості $CI(1,1)$.
- Як випливає з (5.2.12), (5.2.13), при $a_{12} = a_{21} = 0$ рішення буде тривіальним. Для того щоб $\{x(k)\}, \{y(k)\}$ були процесами з одиничним коренем, необхідно, щоб $a_{11} = a_{22} = 1$. При цьому $\lambda_1 = \lambda_2 = 1$ і обидві змінні змінюються у часі, не досягаючи рівноважного взаємозв'язку на довгих часових інтервалах; таким чином, вони не можуть бути коінтегрованими.
- Для того щоб процеси $\{x(k)\}, \{y(k)\}$ були коінтегрованими порядку $CI(1,1)$, необхідно, щоб один із характеристичних коренів був одиничним, а інший за модулем меншим одиниці. В такому випадку обидві змінні будуть мати однакові стохастичні тренди, а перша різниця кожної змінної буде стаціонарною. Наприклад, якщо $\lambda_1 = 1$, то (5.2.14) прийме вигляд:

$$x(k) = \frac{(1 - a_{22}L)}{(1 - L)(1 - \lambda_2 L)} \varepsilon_x(k) + \frac{a_{12}L}{(1 - L)(1 - \lambda_2 L)} \varepsilon_y(k)$$

або

(5.32)

$$(1 - L)y(k) = \Delta y(k) = \frac{1 - a_{22}L}{1 - \lambda_2 L} \varepsilon_x(k) + \frac{a_{12}L}{1 - \lambda_2 L} \varepsilon_y(k),$$

який є стаціонарним при $|\lambda_2| < 1$.

Таким чином, для того щоб забезпечити умову коінтегрованості змінних, наприклад $CI(1,1)$, необхідно покласти один із характеристичних коренів рівним нулю, а інший – меншим одиниці за модулем. Для того щоб більший із коренів дорівнював одиниці, необхідно виконати умову (з рівняння (5.2.16)):

$$0,5(a_{11} + a_{22}) + 0,5 \sqrt{(a_{11}^2 + a_{22}^2) - 2a_{11}a_{22} + 4a_{12}a_{21}} = 1. \quad (5.33)$$

Після спрощення це співвідношення приймає вигляд:

$$a_{11} = [(1 - a_{22}) - a_{12}a_{21}] / (1 - a_{22}). \quad (5.34)$$

Тепер розглянемо другий характеристичний корінь. Оскільки a_{12} і/або a_{21} повинні відрізнятися від нуля, якщо змінні коінтегровані, то з умови $|\lambda_2| < 1$ випливає, що

$$a_{22} > -1, \text{ і } a_{12}a_{21} + (a_{22})^2 < 1. \quad (5.35)$$

Умови (5.2.17)-(5.2.18) представляють собою обмеження, які необхідно накласти на коефіцієнти рівнянь (5.2.12)-(5.2.13) для досягнення коінтегрованості розглянутих змінних порядку $CI(1,1)$. Для того щоб визначити вплив цих обмежень на характер розв'язку рівнянь, перепишемо (5.2.12) – (5.2.13) у вигляді:

$$\begin{bmatrix} \Delta x(k) \\ \Delta y(k) \end{bmatrix} = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ y(k-1) \end{bmatrix} + \begin{bmatrix} \varepsilon_x(k) \\ \varepsilon_y(k) \end{bmatrix}. \quad (5.36)$$

З (5.2.17) випливає, що $a_{11} - 1 = -a_{12}a_{21} / (1 - a_{22})$. Тому після деяких

перетворень представимо (5.2.20) у такій формі:

$$\begin{aligned}\Delta x(k) &= -[a_{12}a_{21}/(1-a_{22})]x(k-1) + a_{12}y(k-1) + \varepsilon_x(k), \\ \Delta y(k) &= a_{21}x(k-1) - (1-a_{22})y(k-1) + \varepsilon_y(k).\end{aligned}\tag{5.37}$$

Рівняння (5.2.21), (5.2.22) представляють собою модель корекції похибки. Якщо обидва коефіцієнти a_{12} і a_{21} відрізняються від нуля, вектор коінтегрування можна нормувати відносно будь-якої змінної. Нормуючи відносно $x(k)$, одержуємо:

$$\begin{aligned}\Delta x(k) &= \alpha_x[x(k-1) - \beta y(k-1)] + \varepsilon_x(k), \\ \Delta y(k) &= \alpha_y[x(k-1) - \beta y(k-1)] + \varepsilon_y(k),\end{aligned}\tag{5.38}$$

де $\alpha_x = -a_{12}a_{21}/(1-a_{22})$; $\beta = (1-a_{22})/a_{21}$; $\alpha_y = a_{21}$.

Очевидно, що $x(k)$ і $y(k)$ будуть змінюватися у відповідь на відхилення від довгострокової рівноваги у попередньому періоді: $x(k-1) - \beta y(k-1)$. Якщо $x(k-1) = \beta y(k-1)$, то $x(k)$ і $y(k)$ будуть змінюватися тільки під впливом розмірів $\varepsilon_x(k)$, $\varepsilon_y(k)$. Якщо ж $\alpha_x < 0$ і $\alpha_y > 0$, то $x(k)$ буде зменшуватися, а $y(k)$ буде збільшуватися у відповідь на позитивне відхилення від стану довгострокової рівноваги.

Можна легко переконатися, що умови (5.2.18), (5.2.19) гарантують ненульове значення коефіцієнта β ($\beta \neq 0$) і принаймні один із параметрів швидкості налагоджування (тобто, α_x, α_y) не буде дорівнювати нулю. Крім того, форма останньої моделі така ж, як і моделі (5.2.7), (5.2.8), розглянутої вище.

Хоча обидва коефіцієнти a_{12}, a_{21} не можуть дорівнювати нулю, цікавий

окремий випадок має місце, якщо дорівнює нулю один із цих коефіцієнтів. Наприклад, якщо покласти $a_{12} = 0$, то коефіцієнт швидкості налагодження $\alpha_x = 0$. У цьому випадку $x(k)$ змінюється тільки під впливом $\varepsilon_x(k)$, тому що $\Delta x(k) = \varepsilon_x(k)$. Послідовність $\{y(k)\}$ виконує усі функції корегування щоб усунути можливе відхилення від довгострокової рівноваги.

Таким чином, по відношенню до розглянутої моделі можна зробити такі висновки:

1. *Обмеження, які необхідні для того щоб змінні були коінтегрованими порядку $CI(1,1)$, гарантують існування моделі корекції похибок.* В розглянутому прикладі обидві послідовності $\{x(k)\}, \{y(k)\}$ є процесами з одиничним коренем, однак лінійна комбінація $x(k) - \beta y(k)$ є стаціонарною, а нормований вектор коінтеграції має вигляд: $[1, -(1 - a_{22})/a_{21}]$. Змінні моделі мають представлення у вигляді корекції похибки із коефіцієнтами швидкості налагодження $\alpha_x = -a_{12}a_{21}/(1 - a_{22})$ і $\alpha_y = a_{21}$. Встановлено також, що модель корекції похибки для змінних типу $I(1)$ обов'язково припускає існування коінтеграції. Цей факт підтверджує *теорему представлення Грейнджера*, яка стверджує, що для будь-якої множини змінних типу $I(1)$ представлення у вигляді корекції похибок і коінтеграції є еквівалентними.
2. *Коінтеграція потребує накладання обмежень на коефіцієнти векторної авторегресійної моделі.* Наприклад, нехай $\mathbf{z}(k) = [x(k) \ y(k)]^T$ і $\varepsilon(k) = [\varepsilon_x(k) \ \varepsilon_y(k)]^T$. Тепер систему (5.2.20) можна представити у вигляді:

$$\Delta \mathbf{z}(k) = \mathbf{F} \mathbf{z}(k-1) + \varepsilon(k). \quad (5.39)$$

Очевидно, що оцінити векторну авторегресійну модель (VAR) для

коінтегрованих змінних, використовуючи тільки перші різниці, неможливо. Якщо у рівнянні (5.2.23) виключити з розгляду $\mathbf{F}z(k-1)$, то в такий спосіб із моделі буде виключена частина, відповідальна за корекцію похибки. Необхідно також зазначити, що рядки матриці $\mathbf{F}$ не будуть лінійно незалежними, якщо змінні моделі коінтегровані. Якщо помножити кожний елемент першого рядка на $-(1-a_{22})/a_{12}$, одержимо другий рядок. Таким чином, детермінант $\mathbf{F}$ буде дорівнювати нулю, а змінні $x(k)$, $y(k)$ мають представлення у вигляді корекції похибки у вигляді (5.2.21), (5.2.22).

Розглянутий приклад ілюструє можливість використання рангу матриці $\mathbf{F}$ для встановлення факту коінтегрованості послідовностей $x(k)$, $y(k)$. Порівняємо детермінант $\mathbf{F}$ із характеристичним рівнянням (5.2.16). Якщо найбільший характеристичний корінь дорівнює одиниці ($\lambda_1 = 1$), то детермінант $\mathbf{F}$ дорівнює нулю і ранг $\mathbf{F}$ дорівнює одиниці. Для того щоб матриця $\mathbf{F}$ мала нульовий ранг, необхідно покласти $a_{11} = 1$, $a_{22} = 1$ і $a_{12} = a_{21} = 0$. Векторна авторегресійна модель, представлена рівняннями (5.2.12), (5.2.13), зводиться у цьому випадку до простого вигляду: $\Delta x(k) = \varepsilon_x(k)$, $\Delta y(k) = \varepsilon_y(k)$. В даному випадку послідовності $\{x(k)\}$, $\{y(k)\}$ є послідовностями з одиничним коренем без коінтеграційного вектора. Нарешті, якщо матриця $\mathbf{F}$ має повний ранг, то жодний із характеристичних коренів не може дорівнювати одиниці і послідовності $\{x(k)\}$, $\{y(k)\}$ спільно стаціонарні.

а. В загальному випадку обидві змінні в коінтегрованій системі будуть реагувати на відхилення від довгострокової рівноваги. Проте можливо, що один із параметрів (але не обидва) швидкості налагодження буде дорівнювати нулю. У цьому випадку відповідна змінна не реагує на відхилення від довгострокової рівноваги, а інша змінна здійснює усе необхідне налагодження. Отже, в коінтегрованій системі необхідно по-новому інтерпретувати каузальність Грейнджера [23, 28]. В коінтегрованій системі послідовність $\{y(k)\}$ не буде впливати за Грейнджером на послідовність $\{x(k)\}$, якщо затримані значення $\Delta y(k-i)$

не входять у рівняння для $\Delta x(k)$ і якщо $x(k)$ не реагує на відхилення від довгострокової рівноваги. Наприклад, у коінтегрованій системі (5.2.9), (5.2.10) послідовність $\{r_L(k)\}$ не впливає, за Грейнджером, на послідовність $\{r_S(k)\}$, якщо усі коефіцієнти $a_{12}(i) = 0$ і $\alpha_S = 0$.

3. Методика побудови моделі коінтегрованих процесів

Побудова моделі коінтегрованих процесів складається з кількох етапів, першим з яких є тестування на стаціонарність. Якщо ряди нестационарні, то необхідно визначити порядок їх інтегрованості. Однаковий порядок інтегрованості дозволяє перейти до перевірки рядів на коінтегрованість. Якщо ряди виявляються коінтегрованими, то можна будувати модель корекції похибки.

Етап 1: перевірка ряду на стаціонарність

При перевірці на стаціонарність спочатку необхідно візуально дослідити часовий ряд. Нагадаємо, що слабка стаціонарність (яка частіше всього використовується на практиці) означає, що середнє значення, дисперсія та коваріація часового ряду даних не змінюються в часі. Досить часто вже попереднє візуальне дослідження дозволяє визначити наявність лінійного чи нелінійного тренду. Стаціонарний ряд має нульовий порядок інтегрованості, що формально записується так: $\{y(k)\} \sim I(0)$.

Порядком інтегрованості є число, яке показує скільки разів необхідно застосувати до часового ряду оператор перших різниць, щоб отримати стаціонарний ряд.

За означенням, часовий ряд має *одиничний корінь* або порядок інтеграції 1, тобто $\{y(k)\} \sim I(1)$, якщо його перші різниці $\Delta y(k) = y(k) - y(k-1)$ утворюють стаціонарний ряд $\{\Delta y(k)\} \sim I(0)$.

Часовий ряд має *два одиничних корені* або порядок інтеграції 2, якщо для досягнення стаціонарності необхідно обчислити його другі різниці:

$$\begin{aligned} \Delta^2 y(k) &= \Delta y(k) - \Delta y(k-1) = y(k) - y(k-1) - [y(k-1) - y(k-2)] = \\ &= y(k) - 2y(k-1) + y(k-2), \end{aligned} \quad (5.40)$$

де $\{\Delta^2 y(k)\} \sim I(0)$. В загальному випадку часовий ряд може мати довільний порядок інтегрованості $\{y(k)\} \sim I(\text{int})$, хоча для усталеної економіки інтегрованість, як правило, не перевищує значення 2. Для встановлення факту існування нестационарності (існування одиничних коренів) запропоновано ряд тестів.

Перевірка факту наявності нестационарності (тест Дікі-Фуллера)

Після виконання візуального контролю необхідно застосувати формальні тести на стаціонарність, які дають можливість переконатись в її існуванні. Розглянемо порядок застосування тесту Дікі-Фуллера.

За допомогою цього критерію визначають яку величину має коефіцієнт a_1 в рівнянні

$$y(k) = a_1 y(k-1) + \varepsilon(k), \quad (5.41)$$

тобто, $a = 1$ чи $a < 1$. Якщо $a = 1$, то дані містять одиничний корінь і ступінь інтегрованості дорівнює буде $I(1)$. Якщо ж $0 < a_1 < 1$, то ряд стаціонарний, тобто має степінь інтегрованості $I(0)$. Для фінансово-економічних процесів значення $a_1 > 1$ не є характерним, тому що такі значення означають наявність в процесах різко зростаючих (спадаючих) ефектів. Виникнення таких процесів є малоімовірним, оскільки фінансово-економічне середовище є достатньо інерційним і не дозволяє змінним приймати нескінченно великі значення за короткі проміжки часу.

Нагадаємо, що застосування МНК до розв'язування задачі оцінювання коефіцієнтів моделі часового ряду передбачає скінченність дисперсії залишків (похибок) $e(k)$ моделі. Наявність нестационарності призводить до порушення цього припущення. Наприклад, розглянемо таке рівняння:

$$\begin{aligned} y(k) &= y(k-1) + e(k) = [y(k-2) - e(k-1)] + e(k) = \dots \\ &= y(0) + e(k) + e(k-1) + e(k-2) + \dots + e(1). \end{aligned} \quad (5.42)$$

Оскільки залишки $e(k)$ незалежні і мають постійну дисперсію, то дисперсія $y(k)$ зростає до нескінченності при $k \rightarrow \infty$. В такому випадку для формального опису динаміки ряду можна скористатись таким рівнянням:

$$\begin{aligned} y(k) - y(k-1) &= a_1 y(k-1) - y(k-1) + e(k), \\ &\text{або} \\ \Delta y(k) &= b y(k-1) + e(k), \end{aligned} \quad (5.43)$$

де $b = a_1 - 1$.

Якщо $b = 0$, то ряд містить одиничний корінь і має степінь інтегрованості $I(1)$, а ряд $\{\Delta y(k)\}$ може бути вже стаціонарним. Якщо ж $b < 0$, то $a < 1$ і стаціонарним буде сам ряд $\{y(k)\}$.

В рівнянні $y(k) = a_1 y(k-1) + \varepsilon(k)$ відсутнє середнє значення (перетин) і опис тренду. Якщо включити середнє, то воно приймає вигляд:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) \quad (5.44)$$

або

$$\Delta y(k) = a_0 + a_1 y(k-1) - y(k-1) + \varepsilon(k) = a_0 + b y(k-1) + \varepsilon(k). \quad (5.45)$$

Із врахуванням тренду останнє рівняння приймає вигляд:

$$y(k) = a_0 + a_1 k + a_2 y(k-1) + \varepsilon(k), \quad (5.46)$$

де k – дискретний час. Це рівняння можна записати для першої різниці

$$y(k) - y(k-1) = a_0 + a_1 k + a_1 y(k-1) - y(k-1) + \varepsilon(k) \quad (5.47)$$

або

$$\Delta y(k-1) = a_0 + a_1 k + b y(k-1) + \varepsilon(k). \quad (5.48)$$

Для такої моделі було б некоректно застосовувати t – статистику з метою визначення значимості коефіцієнта b , оскільки застосування регресії для оцінювання цього коефіцієнта передбачає, що $b < 0$ ($a_1 < 1$). Тобто при $b \approx 0$ великий процент оцінок за t – статистикою не буде прийматися як значущий, а нульова гіпотеза щодо існування одиничного кореня буде часто відкидатись.

Крім того, одиничні корені *робастні* (зберігаються і можуть бути виявлені) при різних ступенях гетероскедастичності, але можуть виникати проблеми з автокореляцією залишків моделі. В умовах наявності автокореляції залишків задача тестування на стаціонарність розв'язується за допомогою розширеного тесту Дікі-Фуллера. При використанні цього методу значення залежної змінної вводяться в рівняння регресії з великими значеннями лагу, достатніми для того щоб уникнути автокореляції залишків. Це рівняння може мати такий вигляд:

$$\Delta y(k) = a_0 + b y(k-1) + c_1 \Delta y(k-1) + c_2 \Delta y(k-2) + \dots + c_n \Delta y(k-n) + \varepsilon(k) \quad (5.49)$$

Форма критерію значущості залежить від виду моделі, що тестується, тобто, чи введено в модель середнє значення і член, який описує тренд.

Нульова гіпотеза без середнього

При тестуванні рівняння

$$\Delta y(k) = b y(k-1) + c_1 \Delta y(k-1) + c_2 \Delta y(k-2) + \dots + c_n \Delta y(k-n) + \varepsilon(k), \quad (5.50)$$

тобто середнє відсутнє, гіпотеза записується так:

$H_0: b = 0$ – ряд нестационарний;

$H_1: b < 0$ – ряд стаціонарний.

Нульова гіпотеза відкидається, якщо статистика b/SE_b має від'ємне значення, яке менше за критичне значення, взяте з таблиці Дікі-Фуллера. Критичні значення для рівнів значимості $\alpha = 1\%$ і $\alpha = 5\%$ дорівнюють $-2,58$ і $-1,95$, відповідно [23, 28]. Якщо нульова гіпотеза приймається, то ряд $\{y(k)\}$ – це випадкове блукання без перетину (константи в рівнянні).

В більш загальному вигляді цього критерію враховується розмір вибірки N , що досягається шляхом обчислення модифікованого критичного значення по формулі:

$$\tau_{\infty} + \frac{\tau_1}{N} + \frac{\tau_2}{N^2}, \quad (5.51)$$

де $\tau_{\infty} = -2,57$ ($\alpha = 1\%$) або $\tau_{\infty} = -1,94$ ($\alpha = 5\%$);

$\tau_1 = -1,96$ ($\alpha = 1\%$) або $\tau_1 = -0,398$ ($\alpha = 5\%$);

$\tau_2 = -10,04$ ($\alpha = 1\%$) або $\tau_2 = 0$ ($\alpha = 5\%$);

(значення τ табульовані Маккінномом в 1991 [28]).

Нульова гіпотеза з середнім значенням

Перевірка рівняння $\Delta y(k) = a_0 + b y(k-1) + e(k)$ із врахуванням можливої автокореляції залишків (як це було показано вище) ґрунтується на використанні того ж статистичного критерію, що і для рівняння без середнього, і тієї ж формули критичних значень, але при таких значеннях τ :

$\tau_{\infty} = -3,43$ ($\alpha = 1\%$) або $\tau_{\infty} = -2,86$ ($\alpha = 5\%$);

$\tau_1 = -6,00$ ($\alpha = 1\%$) або $\tau_1 = -2,74$ ($\alpha = 5\%$); (5.52)

$\tau_2 = -29,25$ ($\alpha = 1\%$) або $\tau_2 = -8,36$ ($\alpha = 5\%$).

Нульова гіпотеза за наявності середнього та тренду

В даному випадку застосовується така ж процедура, що і вище, але при таких значеннях τ :

$$\begin{aligned}
\tau_{\infty} &= -3,96 \ (\alpha = 1\%) \quad \text{або} \quad \tau_{\infty} = -3,41 \ (\alpha = 5\%); \\
\tau_1 &= -8,35 \ (\alpha = 1\%) \quad \text{або} \quad \tau_1 = -4,04 \ (\alpha = 5\%); \\
\tau_2 &= -47,44 \ (\alpha = 1\%) \quad \text{або} \quad \tau_2 = -17,83 \ (\alpha = 5\%).
\end{aligned}
\tag{5.53}$$

Розширений тест Дікі-Фуллера

Для того щоб скористатись розширеним тестом ДФ, необхідно побудувати наступне рівняння регресії:

$$\Delta y(k) = a_0 + a_1 k + b y(k-1) + \sum_{i=1}^p c_i \Delta y(k-i) + \varepsilon(k)
\tag{5.54}$$

де a_0, a_1, b, c_i – невідомі коефіцієнти регресії. Якщо всі коефіцієнти $c_i = 0, i = 1, 2, \dots, p$, то рівнянням (5.3.1) можна скористатись для застосування тесту ДФ, інакше необхідно використати розширений тест ДФ. На практиці рекомендують застосовувати тест РДФ з числом затриманих у часі значень P меншим 10% числа спостережень, тобто $p < 0,1N$, де N – довжина (потужність) часового ряду. При використанні тесту ДФ і РДФ важливо правильно задати структуру моделі, зокрема, необхідно визначити чи потрібно включати члени a_0 і $a_1 k$.

В роботі [51] пропонується просте евристичне правило: якщо візуально з графіка **не** можна зробити висновок про наявність тренду, то в модель (5.3.1) необхідно включати тільки константу (перетин) a_0 , навіть якщо значення коливаються навколо нуля. Якщо візуальний аналіз ряду свідчить про наявність тренду, то в модель (5.3.1) необхідно ввести a_0 і $a_1 k$.

За допомогою базової моделі (5.3.1) тестуються такі гіпотези:

$$\begin{aligned}
H_0 : b = 0, \text{ або часовий ряд нестационарний: } \{y(k)\} \sim I(\text{int}), \text{ int} > 0; \\
H_1 : b < 0, \text{ або часовий ряд стационарний: } \{y(k)\} \sim I(0), \text{ int} = 0.
\end{aligned}
\tag{5.55}$$

Нульова гіпотеза відкидається, якщо отримана оцінка коефіцієнта $\hat{b} < 0$ та обчислена τ – статистика Маккіннона (для тестування на наявність одиничного кореня) за абсолютною величиною більша за величину критичного значення цієї статистики при вибраному рівні значимості α . Формально це можна записати так:

$$|\tau| = \left| \frac{\hat{b}}{SE_{\hat{b}}} \right| \geq |\tau_{крит}| \quad (5.56)$$

на рівні значимості α ; $SE_{\hat{b}}$ – стандартна похибка оцінки $\hat{b}$.

Визначення порядку інтегрованості нестационарного ряду. Якщо на першому етапі встановлено, що ряд нестационарний, то необхідно визначити порядок його інтегрованості. Базовим рівнянням для тесту РДФ в даному випадку є наступне:

$$\Delta^2 y(k-1) = a_0 + b \Delta y(k-1) + \sum_{i=1}^p c_i \Delta^2 y(k-i) + \varepsilon(k) \quad (5.57)$$

Це рівняння використовується для перевірки на стаціонарність перших різниць $\Delta y(k)$ і, як правило, не включає тренду. Ймовірність наявності тренду в других різницях дуже низька. За допомогою моделі (5.3.2) тестуються наступні гіпотези:

- $H_0 : b = 0$, або часовий ряд перших різниць нестационарний: $\{\Delta y(k)\} \sim I(int)$, $int > 0$;
- $H_1 : b < 0$, або часовий ряд перших різниць стаціонарний: $\{\Delta y(k)\} \sim I(0)$, тобто ряд $\{y(k)\}$ має порядок інтегрованості 1.

Нульова гіпотеза щодо нестационарності відкидається, якщо отримана оцінка $\hat{b} < 0$ та обчислена τ – статистика Маккіннона більша за абсолютну величину критичного значення цієї статистики на вибраному рівні значимості. Якщо ж нульова гіпотеза не відкидається, то необхідно продовжити тестування

далі для наступних різниць.

При коінтеграційному моделюванні необхідно мати щонайменше два часових ряди, наприклад, $\{y(k)\}$ і $\{x(k)\}$. Тому розглянуту вище процедуру тестування необхідно повторити для ряду $\{x(k)\}$.

Якщо виявиться, що обидва ряди нестационарні та мають однаковий порядок інтегрованості, то може існувати така лінійна комбінація цих змінних, що є стаціонарним рядом:

$$e(k) = y(k) - \beta x(k), \quad (5.58)$$

тобто $\{e(k)\} \sim I(0)$. Це свідчить про те, що змінні $\{y(k)\}$ і $\{x(k)\}$ є коінтегрованими, а вектор параметрів $\theta = [1 \ -\beta]^T$ називають вектором коінтеграції (його суть розглянута в попередньому параграфі).

Етап 2: перевірка ряду на коінтегрованість

Довгостроковий зв'язок між змінними можна виразити формально через коінтегрованість. Якщо цей довгостроковий зв'язок існує, то можна побудувати відповідну модель коінтеграції змінних, що отримала назву моделі корегування похибки. Вона описує короткострокові та довгострокові аспекти взаємодії змінних. На другому етапі виконується перевірка часових рядів на коінтегрованість. Це можна зробити за допомогою кількох тестів, наприклад, тесту Інгла-Грейнджера у випадку двох змінних, але найбільш загальним на сьогоднішній день є тест Йохансена. Всі тести можна застосовувати у випадку, коли всі нестационарні ряди, для яких будується одна модель, мають однаковий порядок інтегрованості.

Дві змінні $x(k)$ та $y(k)$ вважають коінтегрованими, якщо вони мають однаковий порядок інтегрованості, а їх лінійна комбінація

$$\begin{aligned} [x(k) - \alpha - \beta y(k)] &\sim I(0), \\ [y(k) - \alpha' - \beta' x(k)] &\sim I(0), \end{aligned} \quad (5.59)$$

є стаціонарним рядом.

Ці комбінації відображають існування довгострокового рівноважного

зв'язку між змінними $x(k)$ і $y(k)$.

Покажемо як користуватись тестом Інгла-Грейнджера для перевірки змінних $x(k)$ і $y(k)$ на коінтегрованість, який застосовується до залишків рівнянь (5.3.3) та (5.3.4):

$$\begin{aligned} e_y(k) &= y(k) - \alpha - \beta x(k), \\ e_x(k) &= x(k) - \alpha' - \beta' x(k). \end{aligned} \quad (5.60)$$

Нехай обидва ряди $x(k)$ і $y(k)$ мають порядок інтегрованості 1. Базовим для виконання тесту Інгла-Грейнджера є регресійне рівняння виду:

$$\Delta e(k) = b e(k-1) + \sum_{i=1}^p c_i \Delta e(k-i) + \varepsilon(k) \quad (5.61)$$

де $\Delta e(k) = e(k) - e(k-1)$, $\varepsilon(k)$ – випадковий член, природа якого приблизно така ж, як і в звичайній регресії; $b, c_i, i=1, \dots, p$ – коефіцієнти регресії, які необхідно оцінити на основі часового ряду $\{e(k)\}$, який формується із залишків, що визначаються за допомогою (5.3.5) або (5.3.6).

За допомогою моделі (5.3.7) тестуються такі гіпотези:

$$\begin{aligned} - H_0 : \hat{b} &= 0, \text{ або часові ряди } x(k) \text{ і } y(k) \text{ не коінтегровані; при} \\ &\text{цьому } e(k) \sim I(1); \\ - H_1 : \hat{b} &< 0, \text{ або часові ряди } x(k) \text{ і } y(k) \text{ коінтегровані; при} \\ &\text{цьому } e(k) \sim I(0). \end{aligned} \quad (5.62)$$

Нульова гіпотеза H_0 відкидається, якщо оцінка $\hat{b} < 0$, а обчислена τ – статистика Маккіннона більша по модулю за абсолютну величину критичного значення цієї статистики на вибраному рівні значимості α . Якщо нуль-гіпотеза не приймається, то часові ряди коінтегровані, тобто $e(k) \sim I(0)$.

Етап 3: побудова моделі корегування похибки

У випадку коінтегрованості змінних $x(k)$ і $y(k)$, для них може бути побудована модель корегування похибки, яка поєднує динаміку змінних на коротких проміжках часу з довгостроковим врівноваженим зв'язком і має такий вигляд:

$$\begin{aligned}\Delta x(k) &= a_{10} + \sum_{i=1}^p b_{1i} \Delta x(k-i) + \sum_{i=1}^p c_{1i} \Delta y(k-i) + \lambda_1 e(k-1) + \varepsilon_1(k) \\ \Delta y(k) &= a_{20} + \sum_{i=1}^p b_{2i} \Delta y(k-i) + \sum_{i=1}^p c_{2i} \Delta x(k-i) + \lambda_2 e_x(k-1) + \varepsilon_2(k)\end{aligned}, \quad (5.63)$$

Коефіцієнти λ_1, λ_2 в наведених рівняннях називають швидкістю пристосування (корегування). Вони показують на скільки відсотків відхилення від рівноваги корегується в поточний момент часу (миттєво) і, відповідно, решта похибок $(1 + \lambda_1) \times 100\%$ та $(1 + \lambda_2) \times 100\%$ (при $\lambda_1 < 0, \lambda_2 < 0$) корегуються в інших періодах. Необхідними умовами забезпечення динамічної стійкості моделей (5.3.8) та (5.3.9) є такі: $-1 \leq \lambda_1 \leq 0, -1 \leq \lambda_2 \leq 0$.

Модель (5.3.8), (5.3.9) завжди коректна завдяки тому, що попередні етапи її побудови забезпечують виконання припущень щодо стаціонарності її змінних, тобто

$$\begin{aligned}\Delta x(k) \sim I(0), \quad \Delta y(k) \sim I(0), \quad e_x(k-1) \sim I(0), \quad e_y(k-1) \sim I(0), \\ \varepsilon_1(k) \sim I(0), \quad \varepsilon_2(k) \sim I(0) \text{ за припущенням.}\end{aligned} \quad (5.64)$$

Етап 4: приклад застосування моделі корекції похибки

Розглянемо процес стягування податку на додану вартість (ПДВ), який є одним із основних податків української фіскальної системи і складає приблизно 30% надходжень до зведеного державного бюджету України [24]. Законом “Про податок на додану вартість” в редакції від 3 квітня 1997 року визначено ставку ПДВ 20% [10]. В окремих випадках суб’єкти господарської діяльності можуть звільнятися від сплати ПДВ; крім того, існують деякі товари, що не обкладаються цим податком. Експортерам товарів та послуг ПДВ повертається,

а при імпорті товари оподатковуються за звичайною ставкою. Податок сплачується щоквартально та/або щомісячно в залежності від обороту.

Реальні надходження ПДВ свідчать про їх різке скорочення на протязі 1993-1994 років, що пояснюється значним спадом виробництва (до 35-40% у порівнянні з 1990 роком). Після різкого спаду ситуація дещо стабілізувалася і доходи від ПДВ набули передбачуваного характеру. В процесі надходжень від ПДВ наявний сезонний ефект, тобто, протягом останнього місяця року спостерігається значне збільшення надходжень, а в перші місяці року надходження, як правило, менші ніж в середньому за рік. Звідси витікає, в моделі необхідно врахувати сезонний ефект. Крім того, при моделюванні економічних процесів необхідно визначити чи є процес стаціонарним або коінтегрованим.

Розширений тест Дікі-Фуллера показав, що фактичні надходження ПДВ є стаціонарними на рівні значимості $\alpha = 5\%$. Цей результат дещо суперечливий, оскільки більшість макроекономічних показників мають порядок інтегрованості $I(1)$, а пояснюється він наявністю чіткого сезонного ефекту. Тому при побудові моделі необхідно відфільтрувати сезонну компоненту.

В результаті застосування РДФ до даних, звільнених від сезонної компоненти, встановлено, що ряд надходжень ПДВ нестационарний, а аналіз різниць показав, що порядок інтегрованості $I(1)$. Такий само порядок інтегрованості має ряд значень ВВП. В загальному випадку не можна застосовувати звичайні регресійні методи до нестационарних рядів, але якщо змінні коінтегровані, то результат буде коректним.

Тестування за методом Йохансена показало, що фактичні надходження ПДВ та ВВП коінтегровані. Це означає, що можливі тимчасові відхилення від певного тренду, але на довгих часових інтервалах вказані змінні мають сталий зв'язок. Таким чином, для спільного математичного описання процесів надходження ПДВ та виробництва ВВП можна застосувати модель корегування похибок, яка враховує короткострокові ефекти та довгострокові зв'язки.

Запитання і вправи до п'ятого розділу

1. Які процеси називають інтегрованими? Що означає порядок інтегрованості? Як формально позначають інтегрованість та коінтегрованість?
2. Наведіть приклад коінтегрованих процесів в макроекономіці? Поясніть суть взаємодії цих процесів між собою?

3. Наявність коінтегрованості макроекономічних процесів – це позитивний макроекономічний ефект чи негативний?
4. Чи є відмінності в термінах: „інтегровані процеси”, „процеси з трендом”, нестационарні процеси” та „процеси з одиничними коренями”? Якщо є, то поясніть, які конкретно?
5. Якими методами можна вилучити тренд з процесу? Чи є ці методи повністю рівнозначними за остаточним результатом? Чи потрібно взагалі вилучати тренд з наявних даних і для чого?
6. Знайдіть аналітично другі різниці для такої моделі:

$$y(k) = a_0 + a_1 y(k-1) + b_1 k + b_2 k^2$$
; дайте інтерпретацію отриманому результату. В чому полягає принципова відмінність отриманого результату (після обчислення других різниць) від початкової моделі?
7. Що означає коінтегрованість процесів? Яким умовам повинні задовольняти коінтегровані процеси? Наведіть приклад макроекономічних процесів (або процесів іншої природи), які можуть бути коінтегрованими?
8. Якщо процеси мають однаковий порядок інтегрованості, чи обов’язково вони будуть коінтегрованими?
9. Поясніть суть тестування на інтегрованість за Дікі-Фулерам? В чому полягає відмінність розширеного тесту Дікі-Фулера від звичайного?
10. Яку статистику використовують при тестуванні на інтегрованість? Чи можна використати для цієї мети t – статистику Стюдента?
11. Поясніть послідовність процедури побудови моделі коінтегрованих процесів? Яку роль відіграє у цій процедурі модель корегування похибок? Конкретизуйте, про які похибки йде мова у даному випадку?
12. Яка мета побудови моделі коінтегрованих процесів, які складові включає ця модель? Поясніть на прикладі.
13. За якої умови вектор параметрів математичної моделі $\beta = [\beta_1, \beta_2, \dots, \beta_n]^T$ можна вважати коінтеграційним вектором? Як визначають ранг коінтеграції вектора змінних $\mathbf{x}(k)$?
14. Сформулюйте задачу імітаційного моделювання коінтегрованих процесів за допомогою лінійної комбінації випадкових процесів? Поясніть існування можливості побудови такої моделі?

15. У чому полягає мета побудови моделі корегування похибок? Яка її принципова відмінність від звичайної регресії?
16. Поясніть на прикладі можливість використання рангу матриці F для встановлення факту коінтегрованості інтегрованих послідовностей $x(k)$, $y(k)$.
17. Побудуйте модель коінтеграції для валового внутрішнього продукту України та іншого вибраного вами макроекономічного процесу (скористайтесь щомісячними даними за період 01.1995 – 12.2005, або за інший часовий інтервал). Поясніть важливість факту наявності коінтегрованості між макроекономічними процесами перехідного періоду? Яку задачу можна розв'язати завдяки наявності ефекту коінтегрованості?
18. Яку можливість надає факт наявності коінтегрованості вибраних для аналізу процесів з точки зору їх менеджменту (управління)?
19. Як трактувати коінтегрованість – це позитивне явище у макроекономіці чи негативне?
20. Чи можливе настання ефекту коінтеграції в інших галузях крім макроекономіки? Які статистичні дані необхідні для аналізу наявності ефекту коінтеграції?
21. Яка мета побудови моделі стохастичного тренду?
22. Чи можна будувати для одного досліджуваного процесу моделі детермінованого і стохастичного трендів? Якою може бути мета такого дослідження?

РОЗДІЛ 6. АНАЛІЗ ГЕТЕРОСКЕДАСТИЧНИХ ПРОЦЕСІВ

6.1 Загальні відомості про гетероскедастичні процеси

Гетероскедастичними називають процеси із змінною в часі дисперсією, а *гомоскедастичними* – процеси із сталою дисперсією на відрізку часу, що розглядається при моделюванні та прогнозуванні. Гетероскедастичні процеси відносяться до широкого класу нестационарних процесів, який включає в себе такі процеси:

- процеси із детермінованими і стохастичними трендами;
- процеси із змінною в часі дисперсією;
- процеси, які характеризуються змінним в часі математичним сподіванням та змінною дисперсією одночасно;
- процеси із змінною коваріацією.

Оскільки аналіз гетероскедастичних процесів (ГСП) спрямовується на побудову моделей умовної (змінної у часі) дисперсії, то такі моделі характеризуються наявністю нелінійності (щонайменше) квадратичного типу. Таким чином, гетероскедастичні процеси одночасно є нестационарними і нелінійними стосовно змінних. При виконанні досліджень стосовно аналізу ГСП потрібно коректно робити постановку задачі, яка може бути досить складною для виконання.

Дисперсію та стандартне відхилення часто використовують як *міру ризику* при дослідженні фінансово-економічних процесів, а тому цій проблемі приділяється значна увага в спеціальній літературі. При дослідженні фінансових процесів дисперсію та стандартне відхилення використовують як *міру волатильності* (мінливості) процесу. Волатильність – один із ключових параметрів, які використовують для формування множини правил торгівлі на біржах. Для прогнозування волатильності на основі статистичних даних будують відповідні математичні моделі, структура і характеристики яких будуть розглянуті нижче.

В технічних системах (процесах) дисперсія – *міра розсіювання вимірів*; вона може характеризувати:

- поточний стан механізмів і машин, технологічних процесів; ця характеристика активно використовується в діагностичних системах;
- інтенсивність випадкових збурень, що впливають на технічні

системи;

- інтенсивність похибок (шумів) вимірів і т. ін.;
- ступінь розсіювання значень вимірів у відповідному просторі.

Тому дисперсія – це надзвичайно важливий статистичний параметр з точки зору дослідження поточного та прогнозування майбутнього стану досліджуваної системи.

У діагностичних системах різного призначення дисперсію також використовують як один із параметрів, на основі значень яких приймають діагностичні рішення.

В соціальних дослідженнях дисперсію і стандартне відхилення використовують *для порівняння між собою характеристик різних соціальних груп* або різних множин індивідуумів, що відносяться до однієї й тієї ж соціальної групи.

Оскільки дисперсія може змінюватись у часі, то важливо мати таку математичну модель, яка дозволить коректно описувати динаміку дисперсії та прогнозувати її значення на один або більше кроків наперед. Це дасть можливість покращувати якість рішень по відношенню до управління процесами із змінною в часі дисперсією. Наприклад, коректно формулювати правила прийняття рішень щодо купівлі/продажу акцій на біржі, коректно оцінювати міру кредитного та інших ризиків в банківській діяльності. В технічних системах дисперсія та стандартне відхилення – це міри відхилення розмірів від заданих, ступінь зношеності механізму, рівень шуму і т. ін. Так, дисперсію і стандартне відхилення використовують в системах аналізу якості продукції при побудові контрольних карт, при аналізі точності розмірів і т. ін.

Формально процес вважається гетероскедастичним, якщо

$$Var[\varepsilon(k)] = \sigma_{\varepsilon}^2 \neq const. \quad (6.1)$$

Припущення щодо гомоскедастичності означає, що варіація кожної випадкової величини $\varepsilon(k)$ навколо її математичного сподівання залишається сталою величиною незалежно від значень факторів. Тобто, не є функцією часу.

Гетероскедастичність означає, що дисперсія процесу зменшується чи збільшується в часі, або є більш складною функцією часу. Тобто вона може змінюватися за досить складним законом, який і потрібно знайти при створенні моделі процесу. Іноді використовують припущення, що гетероскедастичність має таку відносно просту форму:

$$\sigma_{\varepsilon(k)}^2 = c^2 x^2(k), \quad (6.2)$$

де c – константа, яку необхідно оцінити на основі статистичних (експериментальних) даних та вибраного методу оцінювання параметрів; $x(k)$ – випадкова змінна, яка визначає дисперсію (впливає на дисперсію) процесу.

Однак, у випадку коли послідовність – це послідовність різних за значеннями чисел, то умовна дисперсія при відомих значеннях визначається так:

$$Var[y(k+1) | x(k)] = x^2(k) \sigma^2. \quad (6.3)$$

Таким чином, умовна дисперсія величини залежить від конкретної реалізації змінної. Оскільки можна виміряти в момент, то дисперсію величини можна визначити як умовну при відомому значенні. Якщо величина $x^2(k)$ велика (мала), то дисперсія також буде великою (малою). З цього випливає, що за наявності додатної кореляції між значеннями послідовності (тобто є тенденція, що за великим значенням знову іде велике значення), то умовна дисперсія послідовності також матиме тенденцію до послідовної додатної кореляції.

Недоліком такої моделі є те, що потрібно дійсно знайти таку змінну, яка суттєво впливає на дисперсію процесу і в основному визначає її, що не завжди можливо з практичної точки зору (такої змінної може просто не існувати). Крім

того, первинний (вихідний) часовий ряд необхідно обробляти таким чином, щоб результуючий ряд (після обробки) мав постійну дисперсію. В поданому прикладі (6.1.2) покладається, що послідовність має постійну дисперсію. Якщо ця умова не виконується, то необхідно зробити додаткову обробку даних.

Замість того щоб шукати одну змінну, яка визначає дисперсію процесу $\{y(k)\}$, і робити додаткові перетворення даних, можна одночасно моделювати (формально описувати) середнє значення та дисперсію ряду. Перш ніж переходити до розгляду цієї моделі, зазначимо, що використання умовного прогнозування має значні переваги перед використанням безумовного.

Прогнозування на основі умовного та безумовного середнього.

Розглянемо оцінювання моделі стаціонарного процесу авторегресії із ковзним середнім (АРКС, в даному випадку це просто АР(1)):

$$\begin{aligned} |a_1| < 1, \\ E[\varepsilon(k)] = 0, \quad E[\varepsilon^2(k)] = \sigma_\varepsilon^2, \quad E[\varepsilon(k)y(k-l)] = 0, \quad l > 0, \\ E[\varepsilon(k)y(k)] \approx \sigma_\varepsilon^2. \end{aligned} \quad (6.4)$$

Необхідно оцінити одно кроковий прогноз величини $y(k+1)$, тобто знайти оцінку $\hat{y}(k+1) = E_k[y(k+1)]$. Умовним математичним сподіванням для процесу АР(1) є

$$E_k[y(k+1)] = a_0 + a_1 y(k). \quad (6.5)$$

Якщо використати це *умовне середнє* для прогнозування значення, то дисперсія похибки прогнозу визначиться так:

$$E_k \{ [y(k+1) - a_0 - a_1 y(k)]^2 \} = E_k [\varepsilon^2(k+1)] = \sigma_\varepsilon^2. \quad (6.6)$$

Знайдемо *безумовне середнє*, тобто:

і при $E[y(k)] = E[y(k-1)]$ маємо, що $E[y(k)] = \frac{a_0}{1-a_1}$. Воно відображає середнє значення для послідовності на нескінченному часовому інтервалі, яке дорівнює в даному випадку $a_0/(1-a_1)$, що випливає з розв'язку рівняння АР(1), що має вигляд:

$$y(k) = a_0 \sum_{i=0}^{\infty} a_1^i + y_0 a_1^k + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) \quad (6.8)$$

а при $k \rightarrow \infty$ маємо:

$$y(k) = a_0 \sum_{i=0}^{\infty} a_1^i + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i), \quad |a_1| < 1 \quad (6.9)$$

Таким чином, дисперсія похибки безумовного прогнозу визначається за виразом:

$$\begin{aligned} E\{[y(k+1) - a_0/(1-a_1)]^2\} &= \\ &= E\{[\varepsilon(k+1) + a_1\varepsilon(k) + a_1^2\varepsilon(k-1) + a_1^3\varepsilon(k-2) + \dots]^2\} = \\ &= \sigma^2/(1-a_1^2), \quad |a_1| < 1. \end{aligned} \quad (6.10)$$

Послідовність в правій частині останнього виразу з'явилась у зв'язку із використанням розв'язку рівняння першого порядку для знаходження безумовного середнього. Оскільки то *безумовний прогноз має більшу дисперсію ніж умовний*. В зв'язку з цим, а також завдяки тому факту, що умовний прогноз ґрунтується на відомих поточних та минулих значеннях ряду, *віддають перевагу умовному прогнозуванню*.

Якщо дисперсія послідовності змінюється в часі, то тенденцію зміни цього параметра можна описати за допомогою моделі АРКС. Наприклад, позначимо через оцінки залишків (похибок) моделі першого порядку . Для

такого випадку умовна дисперсія основної змінної визначається так:

$$\text{Var} [y(k+1) | y(k)] = E_k \{ [y(k+1) - a_0 - a_1 y(k)]^2 \} = E_k [\hat{\varepsilon}^2(k+1)]. \quad (6.11)$$

6.2 Методи оцінювання умовної дисперсії

Для формального опису часового ряду можна скористатись такою моделлю:

$$y(k) = f(y(k-1), \dots, y(k-p)) + \sigma(y(k-1), \dots, y(k-p)) \varepsilon(k), \quad (6.12)$$

де $f(\cdot)$ – авторегресійна функція; $\sigma(\cdot)$ – функція умовного стандартного відхилення; $\{\varepsilon(k)\} \sim N(0, 1)$ – тобто випадковий процес з нульовим середнім та одиничною дисперсією, який не залежить від $\{y(k-1), \dots, y(k-p)\}$. Такі параметри розподілу забезпечуються належним нормуванням вимірів, наприклад, так:

$$y_{\text{норм}}(k) = \frac{y(k) - \bar{y}}{\sigma} \quad (6.13)$$

Модель (6.2.1) дає можливість враховувати різні значення лагів основної змінної. Очевидно, що функції $f(\cdot)$ і $\sigma(\cdot)$ можуть залежати від різної кількості лагів. Припустимо, що функція $f(\cdot)$ оцінена, тобто відома оцінка $\hat{f}(\cdot)$ і відома різниця:

$$r(k) = y(k) - \hat{f}(y(k-1), \dots, y(k-p)). \quad (6.14)$$

Введемо позначення

$$f(k) = f(v(k-1), \dots, v(k-p)), \quad \hat{f}(k) = \hat{f}(v(k-1), \dots, v(k-p)). \quad (6.15)$$

і піднесемо до квадрату ліву і праву частини рівняння (6.2.2):

$$r^2(k) = (\hat{f}(k) - f(k))^2 + \sigma^2(y(k-1), \dots, y(k-p))\varepsilon^2(k) - 2(\hat{f}(k) - f(k))\sigma(y(k-1), \dots, y(k-p))\varepsilon(k). \quad (6.16)$$

Середнє значення останнього члена в (6.2.3) наближено дорівнює нулю за постановкою задачі. Локальне середнє цього добутку, як правило, має такий же порядок як і перший член цієї рівності.

Припустимо, що залишок $\hat{f}(k) - f(k) = O_p(b_N)$ для деякого значення $b_N \rightarrow 0$. Тоді похибки оцінювання функції $f(\cdot)$ впливають на оцінки дисперсії $\sigma^2(\cdot)$ пропорційно $O_p(b_N^2)$. Очевидно, що у багатьох випадках цією величиною можна знехтувати. Таким чином, значення дисперсії $\sigma^2(\cdot)$ можна наближено оцінити навіть не знаючи функції f [37, 49]. Оцінку умовної дисперсії, знайдену таким методом, називають прогнозованою оцінкою (oracle estimator).

Для того щоб максимально уникнути згаданого вище евристичного аргументу, розглянемо випадок, коли $p = 1$. У цьому випадку

$$b_N = h^2 + 1/(Nh)^{1/2} \quad (6.17)$$

для локальної лінійної моделі з шириною смуги h [37]. Для досить великого діапазону значень h величина $b_N^2 = o(N^{-2/5})$, або навіть $b_N^2 = o(N^{-1/2})$. Однак функція $\sigma^2(\cdot)$ може бути оцінена з точністю $b_N^2 = O(N^{-2/5})$ або $b_N^2 = O(N^{-1/2})$, якщо $\sigma^2(\cdot)$ – стала величина. Таким чином, похибкою $O(b_N^2)$ можна знехтувати при оцінюванні $\sigma^2(\cdot)$.

Наведений вище евристичний аргумент свідчить також про те, що вибір

параметра згладжування h для регресії f не критичний з точки зору оцінювання $\sigma^2(\cdot)$, оскільки для широкого діапазону значень похибкою оцінювання $\hat{f}$ можна знехтувати.

Таким чином, згідно (6.2.3) можна записати таку наближену рівність:

$$r^2(k) \approx \sigma^2(y(k-1), \dots, y(k-p)) \varepsilon^2(k). \quad (6.18)$$

З цього рівняння можна отримати умовне математичне сподівання у вигляді:

$$E_{k-1}[r^2(k) | y(k-1), \dots, y(k-p)] \approx \sigma^2(y(k-1), \dots, y(k-p)). \quad (6.19)$$

Таким чином, величину $\sigma^2(\cdot)$ на можна розглядати як непараметричну регресію $r^2(k)$ на $\{y(k-1), \dots, y(k-p)\}$. Далі для обчислення оцінки $\sigma^2(\cdot)$ можна застосувати локальний МНК [37].

Метод псевдоправдоподібності для оцінювання умовної дисперсії

Розглядаючи випадковий процес $\{\varepsilon(k)\} \sim N(0, 1)$ та ігноруючи похибки оцінювання функції $\hat{f}$, можна записати функцію умовної щільності для $r(k)$ за умови існування вимірів $\{y(k-1), \dots, y(k-p)\}$ таким чином:

$$\frac{1}{\sigma(y(k-1), \dots, y(k-p)) \sqrt{2\pi}} \exp \left[-\frac{r^2(k)}{2\sigma^2(y(k-1), \dots, y(k-p))} \right]. \quad (6.20)$$

Після логарифмування і додавання правдоподібності отримаємо таку функцію псевдоправдоподібності:

$$\sum_{k=p+1}^N \left[-\log \sigma^2(y(k-1), \dots, y(k-p)) - \frac{r^2(k)}{\sigma^2(y(k-1), \dots, y(k-p))} \right]. \quad (6.21)$$

Константа $\sqrt{2\pi}$ тут випущена з розгляду, а логарифм правдоподібності помножено на 2. Це не впливає на оцінку $\sigma(\cdot)$, яка далі обчислюється за відомим алгоритмом методу максимальної правдоподібності за умови явно заданої структури моделі для $\sigma^2(\cdot)$.

Локальна процедура найменших квадратів і оцінювання локальної функції псевдоправдоподібності – це два поширені методи оцінювання умовної дисперсії. Їх можна застосовувати для непараметричного оцінювання $\sigma^2(\cdot)$ з використанням локального або глобального моделювання умовної дисперсії $\sigma^2(\cdot)$.

Одновимірний випадок оцінювання умовної дисперсії

Для ілюстрації розглянутих вище пропозицій стосовно оцінювання умовної дисперсії розглянемо одновимірну модель.

Нехай $\{(X(k), Y(k)): k = 1, \dots, n\}$ – строго стаціонарний процес, що відповідає такій моделі:

$$y(k) = f(x(k)) + \sigma(x(k)) \varepsilon(k), \quad (6.22)$$

При $p = 1$ авторегресія

$$y(k) = f(y(k-1), \dots, y(k-p)) + \sigma(y(k-1), \dots, y(k-p)) \varepsilon(k) \quad (6.23)$$

відповідає рівнянню $y(k) = x(k-1)$, $n = N-1$. Нехай $\hat{f}$ локальна лінійна оцінка регресійної моделі, отримана на основі даних: $\{y(k), x(k)): k = 1, \dots, n\}$ з використанням ядра K і смуги пропускання h_2 . Застосовуючи непараметричний регресійний підхід, можна записати:

$$E_{k-1}[r^2(k) | y(k-1), \dots, y(k-p)] \approx \sigma^2(y(k-1), \dots, y(k-p)) \quad (6.24)$$

з використанням лінійної моделі на основі ядра W і смуги пропускання h_1 , отримаємо оцінку $\hat{\sigma}^2(x)$ для кожного значення x , тобто значення $\hat{\sigma}^2(x)$ обчислюється за умови локальної лінійної підгонки непараметричної моделі до даних: $\{(Y(k), r^2(k)) : k = 1, \dots, n\}$.

Приклад 6.1. У цьому прикладі розглядається процес формування доходності $R(k)$ для вибраних біржових акцій. Динаміку цього процесу можна описати, наприклад, моделлю Васічека [37]:

$$dR(t) = (\alpha_0 + \alpha_1 R(t))dt + \sigma dw(t); \quad (6.25)$$

або моделлю:

$$dR(t) = (\alpha_0 + \alpha_1 R(t))dt + \sigma R^{1/2}(t) dw(t); \quad (6.26)$$

або моделлю:

$$dR(t) = (\alpha_0 + \alpha_1 R(t))dt + \sigma R^\gamma(t) dw(t), \quad (6.27)$$

де $w(k)$ – процес стандартного одновимірного броунівського руху. Броунівським рухом називають гаусівський процес з нульовим середнім, нульовою початковою умовою і коваріаційною функцією

$$E [w(t) w(\tau)] = \min(t, \tau). \quad (6.28)$$

Ці однорідні стосовно часу моделі фактично являють собою окремі випадки такої непараметричної моделі:

$$dR(t) = \mu(R(t))dt + \sigma(R(t)) dw(t). \quad (6.29)$$

Функцію $\mu(\cdot)$ називають моментальним доходом; а $\sigma(\cdot)$ – це функція волатильності.

Нехай $\{X(k), k = 1, \dots, n + 1\}$ – квартальні спостереження доходності цінного паперу у відповідні дискретні моменти часу $k = 1, \dots, n + 1$, тобто $X(k) = R(k)$. Для щотижневих даних і при використанні за одиницю часу один

рік можна записати поточний час таким чином: $t_k = t_0 + \frac{k}{52}$. Введемо такі позначення:

$$y(k) = x(k + 1) - x(k), \quad z(k) = w(t_{k+1}) - w(t_k), \quad \Delta_k = t_{k+1} - t_k. \quad (6.30)$$

Оскільки значення броунівського руху незалежні, то незалежні прирости цього процесу мають такий розподіл: $\{z(k)\} \sim N(0, \Delta(k))$. Дискретну версію рівняння (6.2.7) можна представити таким чином:

$$\begin{aligned} y(k) &\approx \mu(x(k)) \Delta(k) + \sigma(x(k)) z(k) = \\ &= \mu(x(k)) \Delta(k) + \sqrt{\Delta(k)} \varepsilon(k), \quad k = 1, \dots, n, \end{aligned} \quad (6.31)$$

де $\{\varepsilon(k)\}$ – послідовність незалежно розподілених випадкових величин з розподілом: $\{\varepsilon(k)\} \sim N(0, 1)$. Похибка дискретизації має другий порядок у випадку, коли період дискретизації вимірів короткий. Навіть при використанні щомісячних даних похибка дискретизації залишається незначною. Застосування імітаційного моделювання (генеруються значення $\varepsilon(k)$ і $\Delta(k)$) з використанням фактичних даних $(y(k), x(k))$ до рівняння (6.2.8) дає можливість знайти оцінки величин $\hat{\sigma}(\cdot)$ і $\hat{\sigma}^2(\cdot)$ та довірчі інтервали для них.

Встановлено, що для цього конкретного прикладу кореляція між $\log(x)$ і

$\log[\hat{\sigma}(x)]$ становить 0,89. На основі цього можна записати таку модель:

$$\sigma(x) = \alpha x^{\beta} \quad (6.32)$$

Підставляючи цю параметричну форму у рівняння (6.2.5), отримаємо функцію псевдоправдоподібності у вигляді:

$$\mathbb{J}(\alpha, \beta) = \sum_{k=1}^n [-\log(\alpha^2 x^{2\beta}(k) - r^2(k)/(\alpha^2 x^{2\beta}(k)))] \quad (6.33)$$

де $r(k) = y(k) - \hat{u}(x(k)) \Delta(k)$. Для кожного заданого значення параметра β максимум правдоподібності досягається за виконання умови [26]:

$$\hat{\alpha}^2(\beta) = \frac{1}{n} \sum_{k=1}^n \frac{r^2(k)}{x^{2\beta}(k)} \quad (6.34)$$

Таким чином, необхідно максимізувати скалярну функцію $\mathbb{J}(\hat{\alpha}(\beta), \beta)$. В результаті максимізації цієї функції знайдені такі оцінки параметрів: $\hat{\alpha} = 0,051$ і $\hat{\beta} = 0,87$. В цілому, отримано два варіанти розв'язку задачі: непараметричний і параметричний.

6.3 Прогнозні моделі динаміки умовної дисперсії

Умовна дисперсія досліджуваного процесу $E_k[\varepsilon^2(k+1)] = \sigma_{\varepsilon}^2$ може бути постійною величиною і може динамічно змінюватись у часі. У зв'язку з цим виникає задача побудови моделі динаміки дисперсії для її формального опису, прогнозування і подальшого використання оцінок прогнозів для розв'язування згаданих вище практичних задач.

Якщо всі коефіцієнти дорівнюють нулю, то оцінка дисперсії буде просто константою. Інакше, умовна дисперсія для описується рівнянням (6.3.1). Це

рівняння можна використати для прогнозування умовної дисперсії на один крок таким чином:

$$E_k[\hat{\varepsilon}^2(k+1)] = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + \alpha_2 \hat{\varepsilon}^2(k-2) + \dots + \alpha_q \hat{\varepsilon}^2(k-q) \quad (6.35)$$

Оскільки (6.3.1) може бути побудоване за умови, що $Var[y(k)] \neq const$, то (6.3.1) називають *авторегресійним умовно гетероскедастичним* (АРУГ) рівнянням. Залишки (збурення), які використовуються у рівнянні (6.3.1), можуть бути отримані на основі рівнянь регресії, авторегресії або авторегресії із ковзним середнім низького порядку. Очевидно, що для АРУГ моделі можна знайти багато потенційних можливостей практичного і теоретичного застосування, наприклад, для прогнозування волатильності.

Окрім рівняння типу (6.3.1) можна вибрати і складніші форми описання поведінки дисперсії. Наприклад, майже ніколи наперед невідомо як впливає збурення на процес – адитивно чи мультиплікативно.

$$\begin{aligned} E[\varepsilon(k)] &= E[v(k)(\alpha_0 + \alpha_1 \varepsilon^2(k-1))^{1/2}] = \\ &= E[v(k)] E[\alpha_0 + \alpha_1 \varepsilon^2(k-1)]^{1/2} = 0. \end{aligned} \quad (6.36)$$

При $E[v(k)v(k-1)] = 0$ маємо, що

Аналогічно можна легко показати, що умовне середнє процесу дорівнює нулю. Якщо $\varepsilon(k)$ – незалежні величини і $\varepsilon(k-1)$ то умовне середнє для $\varepsilon(k)$ можна знайти як

$$E_{k-1}[\varepsilon(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = E_{k-1}[v(k)] E_{k-1}[\alpha_0 + \alpha_1 \varepsilon^2(k-1)]^{1/2} = 0. \quad (6.37)$$

Очевидно, що процес, який визначається виразом (6.3.4), впливає на умовну дисперсію. Оскільки дисперсія $\varepsilon(k)$ то умовна дисперсія для $\varepsilon(k)$ визначається так:

$$E_{k-1}[\varepsilon^2(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = \alpha_0 + \alpha_1 \varepsilon^2(k-1). \quad (6.38)$$

Звідси видно, що умовна дисперсія змінної залежить від значення. Якщо остання змінна приймає велике значення, то і умовна дисперсія в момент також буде мати велике значення. Рівняння (6.3.6) формально є авторегресією першого порядку, але відрізняється від звичайної авторегресії тим, що коефіцієнти α_0, α_1 повинні мати певні обмеження. Для того щоб забезпечити додатність дисперсії, обидва коефіцієнти повинні бути додатними. Крім того, для забезпечення стійкості (збіжності) процесу авторегресії необхідно щоб виконувалась така умова:

Рівняння (6.3.3) – (6.3.6) ілюструють загальні властивості будь-якого процесу АРУГ. Структура АРУГ моделі така, що умовне та безумовне середнє значення похибок цієї моделі дорівнює нулю (фактично, воно має дорівнювати нулю для адекватної моделі). Окрім того, послідовність є послідовно некорельованою, оскільки для Головним моментом в даному випадку є те, що елементи послідовності зв'язані між собою через їх другий центральний момент, тобто умовну дисперсію. Якщо значення елемента суттєво відрізняється від нуля, що приводить до відносно великого значення, то дисперсія змінної має тенденцію до зростання. Таким чином, умовна гетероскедастичність процесу свідчить про те, що послідовність також відноситься до процесу типу АРУГ.

Яким же чином структура похибки впливає на послідовність. Суттєвим моментом є те, що структура похибки процесу АРУГ і автокореляційні параметри послідовності взаємодіють між собою. Про це свідчать численні обчислювальні експерименти. Цьому явищу можна дати пояснення на інтуїтивному рівні. Будь-яке різке зростання модуля змінної викликає збільшення дисперсії послідовності; при цьому чим більше значення має коефіцієнт тим більшим буде вплив на дисперсію. Більше того, чим більшим буде авторегресійний параметр тим більшим буде вплив правої частини на I чим більшою буде тенденція до зміни середнього значення послідовності тим більшою буде її дисперсія.

Умовні значення середнього та дисперсії для послідовності можна записати так:

$$\begin{aligned} \text{Var} [y(k) | y(k-1), y(k-2), \dots] &= E_{k-1} [y(k) - a_0 - a_1 y(k-1)]^2 = \\ &= E_{k-1} [\varepsilon^2(k)] = \alpha_0 + \alpha_1 \varepsilon^2(k-1). \end{aligned} \quad (6.39)$$

Оскільки $\varepsilon(k)$ не можуть приймати від'ємних значень, то мінімальним значенням умовної дисперсії буде α_0 . Будь-яке ненульове значення змінної позитивно зв'язане з умовною дисперсією змінної через коефіцієнт Безумовні значення середнього та дисперсії для $y(k)$ можна отримати за допомогою розв'язку різницевого рівняння, яке описує процес та математичного сподівання.

Оскільки $E[\varepsilon(k)] = 0, \forall k$, то безумовне математичне сподівання для останнього рівняння матиме вигляд:

$$Var[y(k)] = \sum_{i=0}^{\infty} a_1^{2i} Var[\varepsilon(k-i)]. \quad (6.40)$$

Якщо врахувати те, що безумовна дисперсія для $\varepsilon(k)$ є величиною постійною, тобто $Var[\varepsilon(k)] = Var[\varepsilon(k-1)] = Var[\varepsilon(k-2)] = \dots = \alpha_0 / (1 - \alpha_1)$, то

$$Var[y(k)] = \frac{\alpha_0}{(1 - \alpha_1)} \frac{1}{(1 - a_1^2)}. \quad (6.41)$$

Модель гетероскедастичного процесу (6.3.4) може бути розширена до довільного порядку і записана у вигляді:

$$\varepsilon(k) = v(k) \left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) \right)^{1/2}. \quad (6.42)$$

В цьому рівнянні на змінну $\varepsilon(k)$ впливають всі значення від $\varepsilon(k-q)$ до $\varepsilon(k)$ а тому умовну дисперсію можна розглядати як процес авторегресії порядку q .

6.4 Узагальнена авторегресійна умовно гетероскедастична модель (УАРУГ)

Подальшим розширенням моделі АРУГ є опис умовної дисперсії як процесу АРКС. Умовна дисперсія $h(k)$ визначається за виразом:

Оскільки процес визначено як білий шум, який є некорельованим із значеннями, то умовне і безумовне середнє для дорівнюють нулю.

Умовна дисперсія змінної визначається як Вибіркову умовну дисперсію $h_e(k)$ можна наближено визначити за виразом:

$$h_e(k) = \frac{1}{k-1} \sum_{i=1}^k [y(i) - \bar{y}_e(k)]^2, \quad k = 2, 3, \dots, N, \quad (6.44)$$

де $\bar{y}_e(k)$ – вибіркове умовне середнє.

Узагальнена модель АРУГ, яку називають УАРУГ(p, q), складається з двох компонент – авторегресії та ковзного середнього стосовно дисперсії гетероскедастичного процесу. Процес першого порядку отримуємо при, його можна формально визначити як процес УАРУГ(0,1). Якщо всі коефіцієнти то модель УАРУГ(p, q) еквівалентна моделі АРУГ(q). Для того щоб забезпечити скінченність умовної дисперсії, корені характеристичного рівняння, записаного для (6.4.1), повинні знаходитися всередині кола одиничного радіусу на комплексній площині.

Основним характерним моментом УАРУГ моделі є те, що збурення, яке діє на процес є процесом АРКС. Тому можна очікувати, що залишки (похибки) моделі АРКС (попередньо побудованої моделі процесу) будуть відповідати за своїми характеристиками гетероскедастичному процесу.

Це твердження можна пояснити таким чином. Нехай — процес АРКС. Якщо модель АРКС адекватна процесу, то автокореляційна функція (АКФ) і часткова автокореляційна функція (ЧАКФ) залишків повинні вказувати на те, що це процес білого шуму. З іншого боку, АКФ *квадратів* залишків можна використати для попереднього (наближеного) визначення порядку процесу УАРУГ.

Останнє рівняння формально схоже на процес АРКС(p, q) стосовно послідовності. Якщо умовна гетероскедастичність в процесі наявна, то корелограма повинна вказувати на цей факт.

Крок 1:

Побудувати просту модель низького порядку АРКС (або АР) для основної досліджуваної послідовності і обчислити додатковий ряд із квадратів похибок побудованої моделі. Обчислити вибірккову (на основі значень елементів ряду)

дисперсію залишків

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{k=1}^N \hat{\varepsilon}^2(k), \quad (6.45)$$

де кількість елементів ряду залишків.

Крок 2:

Обчислити вибірку автокореляційну функцію квадратів залишків за виразом:

$$(6.46)$$

Крок 3:

Для вибірок досить великої довжини стандартне відхилення для змінної приблизно прирівнюють значенню $N^{1/2}$ [28, 51]. Значення змінної, які суттєво відрізняються від нуля, вказують на наявність процесу УАРУГ. Для того щоб визначити, наскільки є відмінними від нуля значення вибіркової АКФ (тобто чи вони значимі в статистичному сенсі), можна застосувати статистику Льюнга-Бокса. Якщо значення некорельовані між собою, то статистика має в асимптотиці розподіл з ступенями свободи. За нульову гіпотезу можна прийняти те, що значення некорельовані між собою. Відмова від цієї нульової гіпотези еквівалентна тому, що процес АРУГ чи УАРУГ в конкретному випадку наявний.

Більш формально перевірку на наявність гетероскедастичності можна здійснити за допомогою такої двокрокової процедури:

Крок 1:

За допомогою МНК побудувати модель $AP(p)$ для основної змінної:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \dots + a_n y(k-p) + \varepsilon(k). \quad (6.47)$$

При цьому вибір кращої моделі відбувається за відомою методикою

Бокса-Дженкінса. Частіше всього на цьому кроці будують модель першого або другого порядку.

Крок 2:

Побудувати додатковий ряд із квадратів похибок моделі $\hat{\varepsilon}^2(k) = e^2(k)$, де $e(k) = y(k) - \hat{y}(k)$. Оцінка $\hat{y}(k)$ визначається за побудованою моделлю.

Якщо ефект гетероскедастичності відсутній, то оцінки коефіцієнтів повинні бути нульовими або близькими до нуля. При цьому коефіцієнт детермінації буде мати, як правило, досить низьке значення. Якщо вибрати за нуль-гіпотезу, що процес АРУГ відсутній, то при довжині ряду статистика повинна мати розподіл χ^2 . Якщо значення NR^2 досить високе, то відмова від нуль-гіпотези, що всі коефіцієнти дорівнюють нулю, еквівалентна відмові від нуль-гіпотези, що процес АРУГ (УАРУГ) відсутній. З іншого боку, якщо має відносно невелике значення, то можна зробити висновок про відсутність ефекту АРУГ.

Недоліки моделей АРУГ та УАРУГ, пов'язані з їх застосуванням

Частіше всього моделі такого типу застосовують до аналізу та прогнозування фінансових процесів, зокрема процесів формування цін на біржах. Розглянемо деякі недоліки застосування цих моделей саме по відношенню до таких процесів та задач, які при цьому розв'язуються.

1. З емпіричних досліджень відомо, що *дохід від акцій знаходиться у від'ємній кореляції із змінами волатильності* (ступеня мінливості ряду). Це явище отримало назву “леверидж” або “*леверидж-ефект*” (leverage-effect). Зниження ринкової вартості акціонерного капіталу збільшує відношення *взятого кредиту до власного капіталу* і, таким чином, підвищує ризик інвестицій в ту чи іншу компанію. Формально це проявляється у збільшенні волатильності процесу. В результаті майбутні значення волатильності від'ємно корелюють з поточним доходом (віддачею) від акцій.

Це явище можна описати таким чином: від'ємні інновації (або нев'язки), тобто $\varepsilon(k) = y_{\text{факт.}}(k) - y_{\text{очікуване}}(k) < 0$, що трактують як “невтішні новини” – це ситуація, в якій фактичний доход є нижчим очікуваного; це призводить до зростання волатильності. Навпаки, додатні інновації, $\varepsilon(k) > 0$, тобто “хороші новини” – це ситуація, коли фактичний доход є вищим за очікуваний, ведуть до зменшення волатильності.

В даному випадку мова йде про від'ємну кореляцію між $\varepsilon(k)$ і $h(k+1)$ і ця кореляція не враховується в моделях АРУГ і УАРУГ. Дійсно, $h(k)$ – це функція власних затриманих значень та значень $\varepsilon^2(k)$, а тому вона не залежить від зміни знака $\varepsilon(k)$, а тільки від її абсолютних значень. В результаті від'ємність або додатність похибок моделі не впливають на умовну дисперсію. Якщо послідовність $\{\varepsilon(k)\}$ має симетричний розподіл, то майбутнє значення умовної дисперсії не корелюється з поточною похибкою прогнозування.

2. По відношенню до узагальнених умовно гетероскедастичних процесів не узгоджено визначення стаціонарності, тобто сильно стаціонарний процес УАРУГ не завжди буде слабо стаціонарним. Таким чином, виникає проблема визначення стаціонарності таких процесів. Означення слабкої стаціонарності узгоджується з процесами АР та АРКС, але не узгоджується з псевдолінійними моделями АРУГ та УАРУГ. Так, для лінійних процесів формально чітко визначено, що слабка стаціонарність означає наступне: $E[y(k)] = \text{const}$, $E[y^2(k)] = \text{const}$ та $E[y(k)y(k-s)] = \text{const}$.

3. Обмеження області допустимих значень параметрів α і β додатніми значеннями створюють додаткові труднощі з оцінюванням моделей АРУГ.

6.5 Деякі інші моделі гетероскедастичних процесів

Вказаних недоліків не має експоненціальна модель УАРУГ (ЕУАРУГ). В цій моделі логарифм умовної дисперсії визначається за допомогою функції нормованих похибок $g(\cdot)$:

$$\log[h(k)] = c_0 + \sum_{i=1}^{\infty} c_i g[y(k-i)], \quad (6.48)$$

$$g(y) = \alpha y(k) + \beta [|y(k)| - E|y(k)|],$$

де $E[g(y)] = 0$, α, β – параметри моделі; $y(k)$ – основна модельована величина.

Оскільки $E[g(y)] = 0$, то $\log[h(k)]$ – це процес лінійної фільтрації, а рівняння (6.5.1) – ковзне середнє нескінченного порядку. На практиці для умовної дисперсії застосовують представлення типу АРКС(p, q):

$$\log[h(k)] = c_0 + \frac{1 + \alpha(L)}{1 - \beta(L)} g[y(k)] \quad (6.49)$$

де поліноми $1 + \alpha(L)$ і $1 - \beta(L)$ не повинні мати спільних коренів. Ніякі обмеження на значення параметрів, зв'язані з вимогою невід'ємності умовної дисперсії, в рівняннях (6.5.1) – (6.5.3) не накладаються. Обидві складові функції $g(y)$, яка визначається за виразом (6.5.2), мають нульове середнє. Якщо процес $\{y(k)\}$ має симетричний розподіл, то вони ортогональні.

В інтервалі $0 < y(k) < +\infty$ функція $g(y)$ лінійна по $y(k)$ з коефіцієнтом нахилу $\alpha + \beta$, а в інтервалі $-\infty < y(k) < 0$ вона лінійна по $y(k)$ з коефіцієнтом нахилу $\alpha - \beta$. Таким чином, умовна дисперсія реагує асиметрично на несподівані падіння і підйоми ринку.

Член $\alpha y(k)$ реагує на зміни знака $\varepsilon(k)$. Нехай $\alpha < 0, \beta = 0$. В такому випадку інновація умовної дисперсії додатня (від'ємна), якщо похибка прогнозу доходності від'ємна (додатна).

Член $\beta[|y(k)| - E|y(k)|]$ представляє в моделі “ефект абсолютного значення” в дусі моделей АРУГ. Нехай $\alpha = 0, \beta > 0$. Тепер інновація логарифму умовної дисперсії функції $g(y)$ додатна (від'ємна), якщо абсолютне значення $y(k)$ є більшим (меншим) очікуваного.

Якщо розподіл процесу $\{y(k)\}$ симетричний, то $E[y(k) \cdot |y(k)|] = 0$ і

$$\text{Cov}[\log(h(k)), y(k - i)] = c_k \alpha \quad (6.50)$$

Ефект левериджу має місце при $\alpha < 0$.

Ще одна форма моделі ЕУАРУГ має вигляд:

$$\log[h(k)] = \alpha_0 + \sum_{i=1}^p \alpha_i \frac{|\varepsilon(k-i)|}{\sqrt{h(k-i)}} + \sum_{i=1}^p \beta_i \frac{\varepsilon(k-i)}{\sqrt{h(k-i)}} + \sum_{i=1}^q \gamma_i \log[h(k-i)] + v(k) \quad (6.51)$$

В цій моделі є “стандартна” частина, яка враховує модуль інновації $\varepsilon(k)$, і частина, яка враховує знак цієї інновації. Значення $\varepsilon(k)$ нормовані стосовно волатильності, що дає можливість зменшити великі імпульсні значення цієї величини у випадку їх наявності. Значення умовної дисперсії логарифмовані з метою згладжування. Практичне використання моделі [6.5.4] свідчить про можливість досягти високої якості прогнозів умовної дисперсії за її допомогою [2]. Тому вона часто використовується у системах підтримки прийняття рішень, наприклад, в діагностичних системах.

Модель стохастичної волатильності

В моделі такого типу мінлива структура дисперсії описується стохастичним, наприклад, нормально розподіленим процесом. Структура моделі стохастичної волатильності може бути представлена таким чином:

$$\begin{aligned} y(k) &= a_0 + a_1 y(k-1) + \varepsilon(k), \\ \varepsilon(k) &= \sqrt{h(k)} v(k), \quad \{v(k)\} \sim N(0,1), \\ \ln[h(k+1)] &= \alpha_0 + \alpha_1 \ln[h(k)] + \eta(k), \\ \{\eta(k)\} &\sim N(0, \sigma_\eta^2), \end{aligned} \quad (6.52)$$

де $y(k)$ – основна змінна; $\varepsilon(k)$, $\eta(k)$ – незалежні стохастичні процеси білого шуму; $h(k)$ – умовна дисперсія; $a_0, a_1, \alpha_0, \alpha_1$ – коефіцієнти моделі. Процес авторегресії першого порядку (6.5.5) з інновацією $\varepsilon(k)$ пояснює можливе існування авторегресії в процесі $\{y(k)\}$. Для оцінювання таких моделей застосовують метод максимальної правдоподібності, метод моментів та Байєсів підхід в поєднанні з процедурами Монте Карло.

6.6 Врахування великих імпульсних значень в моделях ГСП

Моделі АРУГ дають можливість врахувати ефекти, наявні в часових рядах, які не можуть бути описані математично за допомогою інших підходів до моделювання. Так, узагальнена модель АРУГ використовується для опису даних, що мають надлишковий куртозис. Відомо, модель УАРУГ можна використати також для опису похибок моделювання, що мають t -розподіл (модель УАРУГ- t). Однак, існують ще й інші проблеми з досягненням високого ступеня адекватності моделей такого типу, наприклад, у тих випадках, коли ряд містить адитивні складові, які суттєво перевищують поточне середнє значення (викиди або імпульси), а також ближні (сусідні) значення. Така проблема існує при побудові моделей для цін акцій на біржах, при моделюванні доходів від цінних паперів, а також при моделюванні екологічних і деяких технологічних процесів. Тому в цьому параграфі пропонується підхід до побудови моделей УАРУГ, які враховують вказаний ефект, що дозволяє суттєво підвищити адекватність моделі.

Розглянемо часовий ряд $\{y(k)\}$, $k = 1, 2, \dots, N$, який може бути описаний з деякою ступінню адекватності моделлю авторегресії з ковзним середнім АРКС (p, q) , тобто

$$y(k) = \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j) \quad (6.53)$$

де $\varepsilon(k)$ – випадкова послідовність типу білого шуму з нульовим середнім та дисперсією σ_ε^2 . В моделі (6.6.1) немає константи a_0 , оскільки припускається, що вона будується на основі вимірів $y(k) - \mu_y$ (тобто відхилень від середнього), де μ_y – середнє значення ряду. Припустимо також, що рівняння (6.6.1) має збіжний розв’язок, тобто корені його характеристичного рівняння знаходяться в колі одиничного радіуса. Рівняння (6.6.1) можна записати також із застосуванням лінійного оператора зсуву L так:

де L – оператор затримки в часі, $Ly(k) = y(k-1)$.

Нехай модель, що враховує значні адитивні імпульсні складові часового ряду, має вигляд:

$$z(k) = y(k) + \beta u(k, l) \quad (6.55)$$

де $u(k, l) = \begin{cases} 1, & k = l, \\ 0, & k \neq l. \end{cases}$ Тобто фактичні спостереження описуються рівнянням (6.6.2), а не рівнянням (6.6.1). Розглянемо приклад, коли часовий ряд містить одну адитивну складову (значний імпульс), що суттєво перевищує поточне середнє значення.

При оцінюванні моделі АРКС(p, q) на основі такого ряду залишки моделі можна описати таким чином:

$$\varepsilon(k) = \frac{A(L)}{B(L)} y(k) = D(L) y(k) \quad (6.56)$$

де $D(L) = A(L) / B(L) = (1 - d_1 L - d_2 L^2 - \dots)$. Значення порядку авторегресії p та ковзного середнього q можна визначити в процесі побудови моделі (6.6.1). Якщо ряд містить значний імпульс, то рівняння (6.6.3) приймає вигляд:

$$\hat{\varepsilon}(k) = \varepsilon(k) + \beta D(L) u(k, l) \quad (6.57)$$

яке можна розглядати як модель регресії для $\hat{\varepsilon}(k)$:

$$\hat{\varepsilon}(k) = \beta \omega(k) + \varepsilon(k) \quad (6.58)$$

$$\text{де } \omega(k) = \begin{cases} 0, & k < l, \\ 1, & k = l, \text{ і } \omega(k+j) = -D(j), \quad k > l, \quad j = 1, 2, \dots \end{cases}$$

Тепер вплив суттєвого імпульсного значення на модель в момент $k = l$ можна представити у вигляді:

$$\hat{\beta}(l) = \frac{\sum_{k=l}^N \varepsilon(k) \omega(k)}{\sum_{k=l}^N \omega^2(k)} \quad (6.59)$$

Функцію $\hat{\beta}(l)$ можна представити у вигляді, зручному для тестування її значимості. Для цього необхідно визначити оцінку дисперсії залишків, отриманих при побудові моделі (6.6.1). При цьому бажано, щоб оцінка $\hat{\beta}(l)$ не була суттєво зміщеною внаслідок наявності імпульсів у вихідному (первісному) ряді значень. Оцінка дисперсії залишків може бути отримана різними способами. Один з них полягає в тому, що дисперсія обчислюється для залишків моделі, з яких вилучається спостереження (імпульс) при $k = l$. Якщо позначити цю дисперсію через $\hat{\sigma}_e^2$, то можна записати статистичний параметр у вигляді:

$$\hat{t} = \frac{\hat{\beta}(l) / \hat{\sigma}_e}{\left[\sum_{k=l}^N \omega^2(k) \right]^{1/2}} \quad (6.60)$$

Вплив імпульсного значення ряду вважається значним, якщо $\hat{t} > c$, де c – деяке порогове значення. Оскільки в асимптотиці $\hat{t}$ має нормальний розподіл, то порогове значення c можна визначити з таблиць для t – розподілу. На основі емпіричних досліджень пропонується використовувати для цього величину

$c = 4$. Якщо $\hat{\tau} > c$, можна скоригувати спостереження $z(k)$ в момент $k = l$ і, таким чином, перейти до ряду $y(k)$ (без імпульсного значення), використовуючи величину $\hat{\beta}$, отриману за допомогою (6.6.6), тобто: $y(k) = z(k) - \hat{\beta}u(k, l)$.

Розглянутий підхід до врахування імпульсних значень часового ряду даних поширюється на випадок наявності множини імпульсів. Тобто, процедура повторюється для кожного імпульсу до тих пір, поки $\hat{\tau}$ – статистика не стане незначимою. На останньому кроці параметри моделі оцінюють заново для всіх спостережень.

Розглянемо застосування запропонованої вище методики до побудови моделей гетероскедастичних процесів. Нехай $p(k)$ – ціна акцій на біржі або індекс біржових цін. Сформуємо ряд з логарифмів перших різниць як $d(k) = \log[p(k)] - \log[p(k-1)]$ і запишемо можливу структуру узагальненої авторегресійної умовно гетероскедастичної моделі у вигляді:

$$d(k) = \eta(k) h^{1/2}(k), \quad (6.61)$$

$$h(k) = \alpha_0 + \alpha_1 d^2(k-1) + \gamma_1 h(k-1),$$

де $h^{1/2}(k)$ – стандартне відхилення для умовної дисперсії, обчисленої на основі ряду даних. Якщо зробити припущення, що $\eta(k)$ – некорельована нормально розподілена послідовність, то рівняння (6.6.8), (6.6.9) представляють собою модель УАРУГ. Але якщо $\eta(k)$ має умовний t – розподіл, то модель (6.6.8), (6.6.9) називають УАРУГ – t . Для покращення опису ряду з імпульсними значеннями скористаємось стандартною моделлю УАРУГ, але будемо коригувати її із врахуванням наявності імпульсних значень в ряді $\eta(k)$.

Рівняння (6.6.9) описує умовну мінливість величин $d(k)$, а тому ним можна скористатись для прогнозування мінливості цін, доходів і т. п.

Представимо модель УАРУГ у вигляді:

$$d^2(k) = \alpha_0 + (\alpha_1 + \gamma_1)d^2(k-1) + x(k) - \gamma_1x(k-1) \quad (6.62)$$

де $x(k) \equiv d^2(k) - h(k)$. Рівняння (6.6.10) – це модель типу АРКС(1,1) для $d^2(k)$. Зазначимо, що ряд $x(k)$ можна розглядати як гетерогенний ряд залишків. Тепер алгоритм побудови моделі із врахування наявності імпульсних значень можна представити у вигляді таких кроків:

Крок 1. Обчислити оцінки параметрів рівнянь (6.6.8), (6.6.9) методом максимальної правдоподібності. В результаті отримаємо оцінки $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\gamma}_1$, а також оцінки рядів $\hat{h}(k), \hat{\eta}(k)$. Це дозволить знайти ряд $\hat{x}(k) = d^2(k) - \hat{h}(k)$, який використовується на наступному кроці при визначенні наявності імпульсних значень.

Крок 2. З рівняння (6.6.3) випливає, що поліном $\hat{D}(L) = (1 - (\hat{\alpha}_1 + \hat{\gamma}_1)L) / (1 - \hat{\gamma}_1L)$. Тепер для $\forall k = l$ необхідно оцінити рівняння регресії (6.6.10), і, використовуючи оцінки залишків $\hat{z}(k)$, обчислити $\hat{\beta}_z(l)$. Незважаючи на те, що ряд $\hat{x}(k)$ гетерогенний, необхідно обчислити безумовну дисперсію для $\hat{x}(k)$ без врахування імпульсних значень ряду. Тобто дисперсія похибки моделі обчислюється при нульовому значенні спостереження при $k = l$. Щоб врахувати зміну в часі дисперсії величини $\hat{x}(k)$, її необхідно обчислювати для вибраного рухомого вікна, а не для вибірки в цілому. Статистики $\hat{\tau}$, що визначаються виразом (6.6.7), обчислюються для $\hat{x}(k)$ і порівнюються із значенням $c = 4$.

Крок 3. Спостереження $\hat{x}(k)$ при $k = l$ та найбільшому значенні статистики $\hat{\tau}$, яке повинне перевищувати $c = 4$, замінюється на $\hat{x}^*(k)$ (так як це

показано в рівнянні (6.6.2)) з використанням вагового коефіцієнта $\beta(l)$ (вираз (6.6.6)).

Крок 4. За допомогою ряду $\hat{x}^*(k)$ та $\hat{h}(k)$ побудувати ряд $d^{*2}(k)$, тобто обчислити його за виразом: $d^{*2}(k) = \hat{x}^*(k) + \hat{h}(k)$ для моменту часу $k = l$. Скоригувати ряд даних стосовно імпульсних значень таким чином:

$$d^*(k) = \begin{cases} d(k), & k \neq l, \\ \text{sign}[d(k)] \cdot [d^{*2}(k)]^{1/2}, & k = l. \end{cases} \quad (6.63)$$

Наведений вираз показує, що $\text{sign}[d^2(k)] = \text{sign}[d(k)]$ при $k = l$. З цього випливає, що для членів ряду, які коригуються у відповідності до наявності імпульсного значення, знак не змінюється.

Крок 5. Після обчислення значень ряду $d^*(k)$ перейти на крок 1 і повторювати всю наведену процедуру до тих пір, поки виконується нерівність: $\hat{\tau} > 4$, тобто, до тих пір, поки не будуть оброблені всі члени ряду, що мають суттєві імпульсні значення. Після обробки даних отримуємо остаточні значення оцінок $\hat{\alpha}_0^*$, $\hat{\alpha}_1^*$, $\hat{\gamma}_1^*$ і ряди даних $d^*(k)$, $\hat{h}^*(k)$. За допомогою цих оцінок параметрів та рядів обчислюється прогноз на один крок, тобто $\hat{h}^*(N+1)$, за допомогою рівняння (6.6.9).

Якщо $d(k)$ в рівнянні (6.6.8) представляє собою корельований процес, то перед використанням запропонованого п'ятикрокового алгоритму необхідно відфільтрувати детерміновану динаміку процесу за допомогою моделі АР або АРКС. В результаті отримаємо залишки моделі $\varepsilon(k)$ так само, як це було раніше при побудові моделі АРУГ. Після цього отримані залишки подаються на вхід розглянутого алгоритму. Однак, необхідно пам'ятати, що коефіцієнт корекції β в рівнянні (6.6.2) необхідно застосовувати до початкових (вихідних) даних $d(k)$, а не до залишків $\varepsilon(k)$, оскільки структура моделі АРКС може бути іншою після застосування корекції внаслідок наявності суттєвих імпульсних значень.

Розглянемо побудову моделі УАРУГ для ряду $d(k)$ та моделі УАРУГ для скоригованого відносно імпульсних значень ряду $d^*(k)$. Для прикладу візьмемо дані, які характеризують щотижневий дохід біржі сільгосппродуктів за 1996-1999 роки. Процедура обробки статистичних даних включає оцінювання параметрів математичних моделей на основі щотижневих чотирирічних даних, які використовуються для обчислення оцінки однокрокового прогнозу для дисперсії $h(k)$. Потім з вибірки даних видаляється перше значення (за перший тиждень всього періоду) і одне значення додається в кінці вибірки (наступний тиждень); знову оцінюються параметри моделей і обчислюється однокроковий прогноз. Таким чином, запропонований метод корегування впливу суттєвих імпульсних значень застосовується послідовно до всіх значень вибірки.

На основі цього часового ряду оцінюються моделі $AR(p)$, потім УАРУГ(1,1), а також модель для доходу, що має структуру:

$$d(k) = \mu + a_1 d(k-1) + a_2 d(k-2) + \varepsilon(k), \quad (6.64)$$

$$\varepsilon(k) = \eta(k) h^{1/2}(k), \quad \eta(k) \sim N(0,1),$$

$$h(k) = \alpha_0 + \alpha_1 \varepsilon^2(k-1) + \gamma_1 h(k-1).$$

Результати оцінювання наведено в таблиці 6.1. Параметри моделей оцінені за методом максимальної правдоподібності. Діагностичні статистичні параметри Q_1, Q_2 , які наведені в двох останніх стовпчиках таблиці, свідчать про те, що моделі мають достатню ступінь адекватності. Параметри $Q_1(10)$ і $Q_2(10)$ – це статистики Лjung-Бокса, які обчислені для визначення ступеня кореляції залишків моделі. Критичне значення для Q – статистики визначається за допомогою таблиць для розподілу χ^2 . На наступному кроці до оцінених моделей застосовується метод корекції моделі із врахуванням наявності суттєвих імпульсних значень ряду.

Таблиця 6.1. Результати оцінювання моделей $AR(p)$ та УАРУГ(1,1).

Вибірка даних	Параметри моделі					
	$\hat{\mu}$	$\hat{a}_1$	$\hat{a}_2$	α_0	α_1	γ_1
1996-1999	0,0034 (1,97)	0,0017 (2,05)	0,179 (2,73)	0,0002 (2,18)	0,274 (4,61)	0,468 (4,97)
1997-2000	0,0015 (2,36)	0,0011 (1,99)	0,196 (2,57)	0,0003 (3,19)	0,385 (3,27)	0,652 (4,09)

В таблиці 6.2. подані результати оцінювання тих же моделей, але вже із врахуванням корекції внаслідок наявності суттєвих імпульсних значень, тобто модель побудована для $d^*(k)$.

Таблиця 6.2. Результати оцінювання моделей AP(p) та АРУГ(1,1) після корекції.

Вибірка даних	Параметри моделі					
	$\hat{\mu}$	$\hat{a}_1$	$\hat{a}_2$	α_0	α_1	γ_1
1996-1999	0,0043 (2,45)	0,0021 (2,39)	0,353 (3,19)	0,0002 (2,18)	0,118 (4,02)	0,694 (5,37)
1997-2000	0,0012 (2,44)	0,0009 (1,78)	0,105 (2,69)	0,0002 (3,19)	0,157 (3,06)	0,774 (4,51)

Порівняно з таблицею 6.1. оцінки параметрів α_1 і γ_1 змінились суттєво. Наприклад, для періоду даних 1996 – 1999 значення α_1 зменшилось від 0,274 до 0,118, а значення γ_1 збільшилось від 0,468 до 0,694. Однак сума коефіцієнтів $\alpha_1 + \gamma_1$ суттєво не змінилась, що свідчить про те, що загалом динаміка зміни доходів є робастною по відношенню до імпульсних значень. На основі отриманих даних моделювання можна зробити висновок, що застосування корекції даних приводить до більш стійких оцінок параметрів для різних часових вибірок.

6.7 Приклади побудови математичних моделей гетероскедастичних процесів

Приклад 6.2. Модель інфляції.

Моделі типу АРУГ та УАРУГ зручні тим, що дозволяють оцінити (спрогнозувати) дисперсію процесу в будь-який конкретний момент часу на заданому часовому інтервалі. Далі оцінкою дисперсії можна скористатись для обчислення міри ризику чи міри розсіювання цін, інфляції і т. ін.

Очевидно, що як для державних, так і для приватних підприємств завжди корисно знати рівень інфляції як в поточний момент часу, так і в майбутньому, оскільки рівень інфляції враховується в багатьох розрахунках, наприклад, при визначенні величини заробітної плати.

Нехай $E_k [\pi(k+1)]$ – умовне математичне сподівання очікуваної інфляції в момент $k + 1$, $\sigma^2_{\pi(k)}$ – умовна дисперсія інфляції. Як правило, міру ризику оцінюють за допомогою умовної дисперсії.

Відомі поквартальні дані щодо інфляції в Англії за період з другого кварталу 1958 р. по другий квартал 1977р. (всього 80 значень). Введемо наступні позначення:

$P(k)$ – логарифм індексу споживчих цін;

$W(k)$ – логарифм індексу номінальної (середньої) заробітної плати.

Рівень інфляції: $\pi(k) = p(k) - p(k-1)$.

Реальний прибуток: $r(k) = w(k) - p(k)$.

Перший етап побудови моделі

За часовим рядом отримали таку модель:

$$\pi(k) = 0,0257 + 0,394 \pi(k-1) + 0,408 \pi(k-4) - 0,404 \pi(k-5) + 0,0559 r(k-1) + \varepsilon_1(k), \quad (6.65)$$

де $h(k) = 0,000089$ – дисперсія процесу $\{\varepsilon_1(k)\}$; член $0,0559r(k-1)$ збільшує інфляцію в поточний момент часу; введення в модель значень $\pi(k-4)$, $\pi(k-5)$ зумовлено сезонними ефектами.

Тестування моделі АРУГ(4) дало величину $NR^2 = 15,2$. Критичне значення величини χ^2 (при $t = 4$) дорівнює 13,28. Тому в даному випадку можна зробити висновок про наявність процесу АРУГ:

$$\varepsilon_1^2(k) = \alpha_0 + \alpha_1 [0,4 \varepsilon_1^2(k-1) + 0,3 \varepsilon_1^2(k-2) + 0,2 \varepsilon_1^2(k-3) + 0,1 \varepsilon_1^2(k-4)]. \quad (6.66)$$

Два параметри α_0 і α_1 були визначені з умови невід’ємності та

стаціонарності, тобто: $\alpha_0 > 0$, $0 < \alpha_1 < 1$.

Необхідно відзначити, що коефіцієнти 0,4; 0,3; 0,2; 0,1 вибрані довільно за припущенням, що найбільше значення має перший коефіцієнт (максимальна кореляція між значеннями $\varepsilon^2(k)$ та $\varepsilon^2(k-1)$).

Другий етап побудови моделі

Таким чином, процедура оцінювання коефіцієнтів має двоступінчатий характер (у відповідності з кроком 5 методики, наведеної вище):

1 крок:

- спочатку оцінюються коефіцієнти рівняння (6.7.1) і залишки;
- на основі залишків оцінюються коефіцієнти α_0 і α_1 рівняння (6.7.2);

2 крок:

- уточнення коефіцієнтів моделі (6.7.1) після знаходження коефіцієнтів моделі (6.7.2).

Остаточно отримана модель:

$$\begin{aligned}\pi(k) &= 0,0328 + 0,162 \pi(k-1) + 0,264 \pi(k-4) - 0,325 \pi(k-5) + \\ &\quad + 0,0707 r(k-1) + \varepsilon_2(k), \\ h(k) &= \varepsilon_2^2(k) + 1,4 E -5 + 0,955 [0,4\varepsilon_2^2(k-1) + 0,3\varepsilon_2^2(k-2) + \\ &\quad + 0,2\varepsilon_2^2(k-3) + 0,1\varepsilon_2^2(k-4)];\end{aligned}\tag{6.67}$$

$h(k)$ – передбачена на один крок дисперсія похибки моделі;

$\alpha_1 = 0,955$ – свідчить про те, що дисперсія похибок значна;

$$\sigma_{\varepsilon^2} = \text{const}; E [\varepsilon^2(k)] = 0.$$

Спільним моментом побудови подібних моделей є використання лінійно спадаючих коефіцієнтів рівняння умовної дисперсії з метою врахування довготривалої пам'яті, яка, як правило, притаманна емпіричним даним. Якщо структура цього рівняння оцінюється одним з відомих методів, то це часто призводить до порушення обмеження невід'ємності.

Приклад 6.3. Моделювання утворення цін на продукцію фірми.

Вхідні та вихідні змінні виробничої фірми схематично зображені на рис. 6.1. Вхідними керуючими впливами є очікувана ціна на продукцію і витрати на виробництво, а вихідними – об'єм виробництва та прибуток фірми. Побудуємо математичну модель прибутку фірми із врахуванням дисперсії ціни на її продукцію.

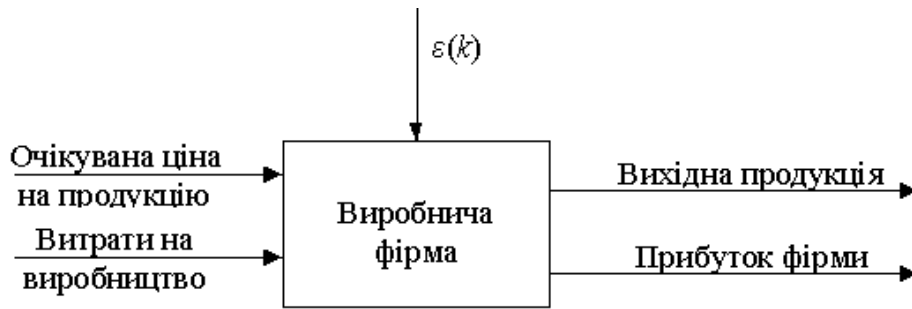


Рис. 6.1. Вхідні та вихідні змінні виробничої фірми

Нехай $q(k)$ – прибуток фірми в момент часу k ;

$p^e(k)$ – очікувана ціна продукції фірми в момент k , отримана на основі інформації в момент часу $(k-1)$ (тобто, $p^e(k) = E_{k-1}[p(k)]$);

$h(k)$ – очікувана дисперсія ціни на продукцію фірми в момент k , на основі інформації в момент часу $(k-1)$;

$pp(k-1)$ – затрати на виробництво одиниці продукції в момент $(k-1)$;

$p(k-1)$ – об'єм виробництва продукції в момент часу $(k-1)$;

$\varepsilon_1(k)$ – випадкові збурення доходу фірми.

Для формального опису прибутку вибрана така структура моделі:

$$q(k) = a_0 + a_1 p^e(k) + a_2 h(k) + a_2 pp(k-1) + a_4 p(k-1) + a_5 q(k-4) + \varepsilon_1(k). \quad (6.68)$$

Для обчислення $p^e(k)$ і $h(k)$ будуються окремі моделі на основі квартальних даних.

Величина $pp(k-1)$ враховує витрати на виробництво одиниці продукції за попередній квартал (вона зменшує прибуток). Величина $p(k-1)$ сприяє зростанню прибутку. Значення $q(k-4)$ береться з запізнюванням 4, щоб врахувати той же квартал попереднього року.

Що цікавого в цій моделі – вплив умовної дисперсії ціни продукції на прибуток фірми. Змінна $p^e(k)$ оцінюється на основі ціни в попередньому кварталі. Якщо динаміка ціни не дуже висока, то це робиться просто.

Якщо ціна змінюється досить швидко, то виробник, який уникає ризику, намагається знизити об'єм виробництва. Для прогнозу рівня ціни на продукцію використовується модель:

$$(1 - \beta_1 L - \beta_2 L^2 - \beta_3 L^3 - \beta_4 L^4) p(k) = \beta_0 + \varepsilon_2(k). \quad (6.69)$$

Оцінена модель має вигляд:

$$(1 - 0,511L - 0,129 L^2 - 0,130 L^3 - 0,138 L^4) p(k) = 1,632 + \varepsilon_2(k). \quad (6.70)$$

Тепер підставимо $p^e(k) = p(k)$ в (7.7.5).

$$h(k) = 1,353 + 0,162 \varepsilon_2^2(k-1) + 0,591 h(k-1). \quad (6.71)$$

Значення, обчислені в (6.7.6) і (6.7.7) підставляються в (6.7.4) для оцінки рівня цін на продукцію. Коефіцієнти в (6.7.6) і (6.7.7) відносяться до прибутку птахофабрики з вирощування бройлерів.

Для отримання $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_5$ необхідно мати 5 рядів змінних:

$$q(k) = 2,767p^e(k) - 0,521h(k) + 4,325pp(k-1) + 1,887p(k-1) + 0,603q(k-4) + \varepsilon_1(k). \quad (6.72)$$

Другий підхід до прогнозу ціни базується на відомому виразі для ковзного середнього

$$p^e(k) = \alpha p(k-1) + (1-\alpha) p(k-1), \quad (6.73)$$

$$p^e(k) = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i p(k-1-i).$$

По аналогії з цим рівнянням можемо записати рівняння для дисперсії похибок прогнозованої ціни:

$$h(k) = \beta \sum_{i=0}^{\infty} (1-\beta)^i [p(k-1-i) - p^e(k-1-i)]^2, \quad 0 < \beta < 1. \quad (6.74)$$

Очевидно, що другий підхід являється більш формальним (не пов'язаним з вхідними даними).

Приклад 6.4. Модель прибутку від продажу майна або цінних паперів (модель АРУГ-М). Даний підхід до моделювання може бути використаний у випадках, коли необхідно прогнозувати доход на ринку майна або цінних

паперів.

При виконанні торговельних операцій з нерухомістю та цінними паперами необхідно оцінювати можливу премію за ризик, а також міру ризику за затримку продажу майна з метою збільшення премії. Плату за ризик будемо визначати як зростаючу функцію умовної дисперсії прибутку від продажу майна.

Для описання приросту прибутку від затримки продажу ризикованого майна або цінних паперів скористаємось рівнянням:

$$y(k) = \mu(k) + \varepsilon(k) \quad (6.75)$$

де $y(k)$ – приріст прибутку за рахунок затримки продажу ризикованого майна при умові, що ціна на нього буде зростати; $\mu(k)$ – величина прибутку, яка необхідна для того щоб не схильний до ризику агент міг затримати продаж майна на деякий, відносно строк; $\varepsilon(k)$ – випадкове коливання прибутку на даному інтервалі часу.

Очікуваний прибуток визначається як умовне математичне сподівання

$$E_{k-1}[y(k)] = \mu(k) \quad (6.76)$$

Припустимо, що плата за ризик є зростаючою функцією умовної дисперсії випадкового процесу $\varepsilon(k)$. Тобто, чим більшою є умовна дисперсія процесу $\varepsilon(k)$, тим вищою повинна бути компенсація за затримку продажу. Це можна виразити за допомогою рівняння:

$$\mu(k) = \beta + \gamma h(k), \quad \gamma > 0 \quad (6.77)$$

де $h(k)$ – умовна дисперсія процесу $\varepsilon(k)$; $\beta = const$. Для описання умовної дисперсії виберемо модель АРУГ:

$$h(k) = \alpha_0 + \sum_{i=1}^n \alpha_i \varepsilon^2(k-i) \quad (6.78)$$

Рівняння (6.7.8) – (6.7.10) утворюють модель АРУГ-М. З цієї моделі випливає, що дисперсія процесу $y(k)$ залежить від умовної дисперсії $h(k)$. При $\alpha_1 = \alpha_2 = \dots = \alpha_q = 0$ модель АРУГ-М буде описувати постійну додаткову плату за ризик.

Прикладом оцінювання моделі (6.7.8) є

$$\begin{aligned} y(k) &= 0,185 + \varepsilon_1(k), \\ \mu(k) &= -0,0368 + 0,745h(k) + \varepsilon_2(k), \\ h(k) &= 0,0031 + 1,73[0,4\varepsilon_1^2(k-1) + 0,3\varepsilon_1^2(k-2) + 0,2\varepsilon_1^2(k-3) + 0,1\varepsilon_1^2(k-4)] \end{aligned} \quad (6.79)$$

де коефіцієнт 0,185 – характеризує міру прибутку в процентах за квартал (при використанні квартальних даних). Значення коефіцієнта $\alpha_1 = 1,73$ свідчить про те, що безумовна дисперсія буде досить великою. З першого і другого рівняння випливає, що плата за ризик буде більшою, якщо майно або цінні папери знаходяться під впливом меншого ризику (в другому рівнянні, яке безпосередньо враховує умовну дисперсію, константа менша ніж у першому).

Приклад 6.5. Дисперсія як міра валютного ризику банку. Оцінюванню валютного ризику приділяють в банківській діяльності значну увагу. Він представляє собою *можливість фінансових втрат внаслідок негативних змін (падіння) курсів іноземних валют* або внаслідок невідповідності встановлених банком курсів ринковій кон'юнктурі (наприклад, масовий продаж валюти). Для визначення можливих фінансових втрат можна застосувати методику їх розрахунку, яка отримала назву VaR (Value-at-Risk).

Значення VaR – це величина капіталу, яка з визначеною ймовірністю покриє можливі збитки протягом деякого визначеного періоду часу.

Застосування VaR дає можливість стверджувати: “Ми впевнені на $X\%$,

що наші втрати не будуть перевищувати значення Y на протязі наступних N днів”. Тобто, розраховане значення VaR можна інтерпретувати як таке, що вказує на можливі втрати із заданою ймовірністю, а для компенсації втрат необхідно мати відповідний резервний капітал. Для пояснення VaR можна скористатись визначенням *квантиля*.

Квантиль розподілу визначається парою величин: це *деяке вибране значення випадкової змінної x та відповідна йому ймовірність c* . Для квантиля x значення c визначається так:

$$c = F(x) = \int_{-\infty}^x f(u) du \quad (6.80)$$

Тобто, ймовірність того, що $x_i < x$ дорівнює c : $p(x_i < x) = c$. Ймовірність того, що $x_i > x$ визначається як $p(x_i > x) = 1 - c$. Квантиль позначають як $Q(X, c)$. Очевидно, що **медіана** представляє собою 50%-й квантиль.

З точки зору теорії статистики VaR можна інтерпретувати як точку зрізу на розподілі втрат (прибутків) таку, що втрати не будуть мати місце з ймовірністю, скажемо, $p = 95\%$. Якщо $f(u)$ – розподіл втрат (прибутків), то визначення VaR зв’язано з виразом:

$$F(x) = \int_{-\infty}^x f(u) du = 1 - p \quad (6.81)$$

де P – ймовірність того, що втрат не буде (ймовірність, зв’язана з правим хвостом розподілу). Тепер VaR можна визначити як відхилення між очікуваним значенням та квантилем:

$$VaR(c) = E[X] - Q(X, c) \quad (6.82)$$

де c – ймовірність втрат, зв'язана з лівим хвостом розподілу; $Q(x, c)$ – відповідний квантиль розподілу.

Методика VaR для оцінювання валютного ризику

Дана методика використовується для оцінки валютних ризиків, тобто, ризиків, пов'язаних з використанням іноземних валют. Для оцінки валютних позицій використовують курс НБУ.

Вихідні (початкові) дані, необхідні для розрахунку VaR:

- значення курсів НБ по валютах за останні 60 робочих днів (за 60 днів можна зібрати необхідну за об'ємом статистику відносно курсів валют); наприклад, можна розглянути три найбільш вживаних валюти: USD, EUR і RUR;
- значення відкритих валютних позицій для вибраних валют, тобто, щоденні об'єми операцій для кожної валюти.

Розглянемо виконання обчислювальних операцій методики у вигляді послідовності декількох кроків.

Крок 1. Першим кроком реалізації методики є збір статистичних даних щодо курсів та відкритих валютних позицій.

Крок 2. На другому кроці розраховуються темпи зміни курсів валют за виразом:

$$x^i(k) = \ln \left(\frac{Курс^i(k)}{Курс^i(k-1)} \right), \quad k = 1, 2, 3, \dots, N, \quad i = 1, 2, \dots, n, \quad (6.83)$$

де $Курс^i(k)$ – значення курсу i -ї валюти в момент k ; N – загальне число значень статистичних даних; n – число валют. З обчислених значень динаміки курсів формують матрицю вимірів $\mathbf{X}$, вимірністю $[N \times n]$.

Крок 3. Розрахунок коваріаційної та кореляційної матриць для $\mathbf{X}$. Елементи кореляційної матриці можна розрахувати в такій послідовності:

- Нормування елементів матриці $\mathbf{X}$:

$$x_{ij}^n = \frac{x_{ij} - \mu_i}{\sqrt{\text{var}[x_i]}} = \frac{x_{ij} - \mu_i}{\sigma_{x_i}}, \quad i = 1, \dots, n; \quad j = 1, \dots, N, \quad (6.84)$$

де μ_i – середнє значення динаміки зміни i -ї валюти, тобто, i -го стовпчика матриці $\mathbf{X}$; $\text{var}[x_i]$ – дисперсія динаміки курсу i -ї валюти; σ_{x_i} – відповідне стандартне відхилення. В результаті такого нормування отримаємо матрицю нормованих значень $\mathbf{X}_H$ динаміки курсів валют. Діагональні елементи коваріаційної матриці $\mathbf{C}$ представляють собою дисперсії $\sigma_i, i=1, \dots, n$ відповідних динамік курсів валют.

– Обчислення кореляційної матриці $\mathbf{R}[n \times n]$:

$$\mathbf{R} = \mathbf{X}_H^T \mathbf{X}_H. \quad (6.85)$$

Крок 4. Прогнозування волатильності. Деякі методики обчислення прогнозу пропонують наступний вираз: $\sigma_i^{\text{прогн.}} = a_i \sigma_i$, де $a_i = F(\mathbf{R}, \sigma_i)$. Так, пропонуються, наприклад, такі значення: $a = 0,5 \div 3,0$. При цьому чим ближче значення a до 1, тим нижчою є волатильність валюти.

Коректним підходом до прогнозування волатильності є застосування рівнянь гетероскедастичних процесів, які розглянуті вище. Кращі результати, як правило, можна отримати за допомогою рівнянь УАРУГ та ЕУАРУГ.

Крок 5. Розрахунок величини VaR. Значення VaR визначають за виразом:

$$\text{VaR}_i = \alpha \sigma_i^{\text{прогн.}} V_i \sqrt{N}, \quad i=1, \dots, n, \quad (6.86)$$

де α – квантиль для довірчого інтервалу (для інтервалу 95% він складає 1,65, а для 99% дорівнює 2,33); $\sigma_i^{\text{прогн.}}$ – значення прогнозу волатильності, отримане на попередньому кроці; V_i – об'єм операцій для i -ї валютної позиції на визначений момент часу, який визначають за виразом:

Обчислені значення VaR_i утворюють вектор

$$VaR = [VaR_1 \ VaR_2 \ ... \ VaR_n]^T, \quad (6.88)$$

який використовують для обчислення загального значення $VaR_{заг.}$:

$$VaR_{заг.} = \sqrt{VaR^T \cdot \mathbf{R} \cdot VaR}. \quad (6.89)$$

Значення $\sigma_i^{прогн.}$ можуть корегуватись у бік зростання в залежності від того, яка є додаткова інформація щодо можливих змін курсів.

Виконані розрахунки показують, що можливі втрати (наступного дня) по валютних позиціях не перевищують значення VaR_i . Існують модифіковані варіанти розглянутої методики, які ґрунтуються на інших розподілах випадкових величин.

Приклад 6.6. Нелінійне оцінювання моделі УАРУГ (GARCH) за допомогою розширеного фільтра Калмана. Пропонується трансформування моделі УАРУГ у простір станів і формулюється алгоритм оптимального оцінювання станів нелінійного фінансового процесу у формі фільтра Калмана.

Розглянемо дискретну модель нелінійного нестационарного процесу у просторі станів:

$$\mathbf{x}(k) = \mathbf{F}(\mathbf{x}(k-1), \mathbf{u}(k-1), \theta) + \sigma(\mathbf{u}(k-1), \theta) \mathbf{w}(k), \quad (6.90)$$

$$\mathbf{z}(k) = \mathbf{H}(\mathbf{x}(k), \mathbf{u}(k), \theta) + \mathbf{v}(k),$$

де $\mathbf{x}(k) \in \mathbb{R}^n$ – вектор стану досліджуваної системи (процесу); $\mathbf{u}(k) \in \mathbb{R}^r$ – вектор вхідних керуючих впливів; $\mathbf{z}(k) \in \mathbb{R}^m$ – вектор зашумлених спостережень; $\mathbf{w}(k) \in \mathbb{R}^n$ – вектор зовнішніх випадкових впливів (збурень) на

досліджувану систему; $\mathbf{v}(k) \in \mathbb{R}^m$ – похибки (шум) вимірів; – вектор параметрів моделі, який у загальному випадку невідомий; $\mathbf{F}(\cdot) \in \mathbb{R}^n$, $\sigma(\cdot) \in \mathbb{R}^{n \times n}$, $\mathbf{H}(\cdot) \in \mathbb{R}^m$ – нелінійні функції, які характеризують динаміку, волатильність і зв'язок вектора стану системи з вектором вимірів.

Характеристики випадкових процесів у моделі (6.7.11) – (6.7.12) такі:

- $\mathbf{w}(k)$ – вектор білих гаусівських шумів з нульовим середнім, $E[\mathbf{w}(k)] = 0$, і коваріаційною матрицею $E[\mathbf{w}(k) \cdot \mathbf{w}^T(k)] = \mathbf{Q} \delta(k - j)$; $\delta(k - j)$ – це одинична матриця при $k = j$ і нульова матриця при $k \neq j$;
- $\mathbf{v}(k)$ – вектор білих гаусівських шумів з нульовим середнім, $E[\mathbf{v}(k)] = 0$, і коваріаційною матрицею $E[\mathbf{v}(k) \cdot \mathbf{v}^T(k)] = \mathbf{R} \delta(k - j)$;
- процеси $\mathbf{w}(k)$ і $\mathbf{v}(k)$ – некорельовані між собою.

Модель УАРУГ(1, 1) має такий вигляд:

$$x(k) = \sigma(k)\varepsilon(k), \quad (6.91)$$

$$h(k) = \alpha_0 + \alpha_1 x^2(k-1) + \beta_1 h(k-1),$$

де $x(k)$ – доходність фінансового активу, а $\sigma(k) > 0$ – її волатильність; $\{\varepsilon(k)\}$ – гаусівська послідовність незалежних однаково розподілених величин; $E[\varepsilon(k)] = 0$, $E[\varepsilon(k)\varepsilon(j)] = 0$, $k \neq j$; параметри $\alpha_0 > 0$, $\alpha_1 \geq 0$ і $\beta_1 \geq 0$; значення $x(0)$ не залежить від послідовності $\varepsilon(k)$, $k > 0$; $h(k) = \sigma^2(k)$.

Представлення (6.7.13) і (6.7.14) у просторі станів таке:

$$\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} = \begin{bmatrix} f_1(x_1(k-1), x_2(k-1), \theta) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \varepsilon(k), \quad (6.92)$$

$$z(k) = x(k) = x_2(k) \sqrt{x_1(k)},$$

де $\theta = [\alpha_0, \alpha_1, \beta_1]$; $f_1(x_1(k), x_2(k), \theta) = \alpha_0 + \alpha_1 x_2^2(k) x_1(k) + \beta_1 x_1(k)$;

$$x_1(k) = \sigma^2(k) \text{ і } x_2(k) = x(k) / \sigma(k) = \frac{\sigma(k) \varepsilon(k)}{\sigma(k)} = \varepsilon(k).$$

Зазначимо, що представлення (6.7.15), (6.7.16) містить нелінійність у рівнянні спостережень (6.7.6).

Розширений фільтр Калмана (РФК) призначений для використання узагальненої моделі досліджуваних стохастичних процесів і ґрунтується на лінеаризації нелінійної моделі у просторі станів.

Функції $\mathbf{F}(\cdot)$ і $\mathbf{H}(\cdot)$ моделі у просторі станів (6.7.11) і (6.7.12) необхідні для обчислення оцінок прогнозів вектора стану і представлення зв'язку вектора вимірів із вектором стану. Оцінювання однокрокового прогнозу на основі моделі виконується таким чином:

$$\hat{\mathbf{x}}(k | k-1) = \mathbf{F}(\hat{\mathbf{x}}(k-1 | k-1), \mathbf{u}(k-1), k, \theta). \quad (6.93)$$

Для того щоб обчислити коваріаційну матрицю похибок оцінок прогнозів вектора стану необхідно знайти спочатку матрицю динаміки $\mathbf{A}$, елементами якої будуть частинні похідні від функції $\mathbf{F}$ (Якобіан). Вона обчислюється у кожному циклі реалізації рівнянь фільтра. Тобто матриця $\mathbf{A}$ – це лінеаризована версія нелінійної функції $\mathbf{F}$ у околі значення поточної оцінки.

Коваріаційна матриця похибок оцінок вектора стану обчислюється за виразом:

$$\mathbf{P}(k | k-1) = \mathbf{A}(k-1) \mathbf{P}(k-1 | k-1) \mathbf{A}^T(k-1) + \mathbf{Q}. \quad (6.94)$$

Після обчислення поточної оцінки стану необхідно врахувати останній вимір і обчислити інновацію:

$$\mathbf{v}(k) = \mathbf{z}(k) - \mathbf{H}(\hat{\mathbf{x}}(k | k-1), \mathbf{u}(k), \theta), \quad (6.95)$$

і оновлена коваріаційна матриця похибок оцінок вектора стану:

$$\mathbf{S}(k | k - 1) = \mathbf{C}(k) \mathbf{P}(k | k - 1) \mathbf{C}^T(k - 1) + \mathbf{R}, \quad (6.96)$$

де $\mathbf{C}$ – лінеаризована версія функції $\mathbf{H}$ у околі поточної оцінки.

Далі обчислюється оптимальний коефіцієнт фільтра за виразом:

$$\mathbf{K}(k | k - 1) = \mathbf{P}(k | k - 1) \mathbf{C}^T(k) \mathbf{S}^{-1}(k | k - 1), \quad (6.97)$$

поточна оцінка вектора стану

$$\hat{\mathbf{x}}(k | k) = \hat{\mathbf{x}}(k | k - 1) + \mathbf{K}(k) v(k), \quad (6.98)$$

і апостеріорна коваріаційна матриця похибок оцінок вектора стану:

$$\mathbf{P}(k | k) = (\mathbf{I} - \mathbf{K}(k) \mathbf{C}(k)) \mathbf{P}(k | k - 1). \quad (6.99)$$

Перехідна матриця станів і матриця спостережень (лінеаризовані версії функцій $\mathbf{F}$ і $\mathbf{H}$ визначаються за виразами:

$$\mathbf{A}(k - 1) = \left. \frac{\partial \mathbf{F}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}(k - 1, k - 1), u(k - 1)}; \quad \mathbf{C}(k) = \left. \frac{\partial \mathbf{H}}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}(k | k - 1)}. \quad (6.100)$$

Нелінійне оцінювання параметрів

Для оцінювання невідомого вектора параметрів θ , застосуємо метод максимальної правдоподібності (ММП). За умови наявності вимірів виходу системи $\mathbf{Y} = [\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(k), \dots, \mathbf{y}(N)]$ функцію правдоподібності можна записати як функцію спільної щільності ймовірностей:

$$L(\theta, \mathbf{Y}_N) = p(\mathbf{Y}_N | \theta), \quad (6.101)$$

або у еквівалентній формі:

$$L(\theta, \mathbf{Y}_N) = p(\mathbf{y}(0) | \theta) \prod_{k=1}^N p(\mathbf{y}(k) | \mathbf{Y}_{k-1}, \theta) \quad (6.102)$$

Оскільки опис динаміки стохастичної системи (6.7.11), (6.7.12) пов'язаний із гаусівськими процесами білого шуму, то за припущення деяких умов регулярності протікання цих процесів спільну функцію щільності розподілу ймовірностей $p(\mathbf{y}(k) | \mathbf{Y}_{k-1}, \theta)$ можна подати таким чином:

$$L(\theta, \mathbf{Y}_N) = \frac{p(\mathbf{y}(0) | \theta)}{(2\pi)^{m/2}} \prod_{k=1}^N \frac{g(k)}{(\det(\mathbf{S}(k | k-1)))^{1/2}}, \quad (6.103)$$

де $g(k) = \exp\{-\frac{1}{2} \mathbf{v}^T(k) \mathbf{S}^{-1}(k | k-1) \mathbf{v}(k)\}$; $\mathbf{v}(k)$ – процес інновацій, визначений рівнянням (6.7.19); $\hat{\mathbf{y}}(k | k-1) = E[\mathbf{y}(k) | \mathbf{Y}_{k-1}, \theta]$ – умовне середнє процесу $\mathbf{y}(k)$ за умови, що наявні дані $\mathbf{y}(0), \mathbf{y}(1), \dots, \mathbf{y}(k-1)$ і θ ; $\mathbf{S}(k | k-1)$ – коваріаційна матриця похибок оцінок вектора стану, визначена рівнянням (6.7.20).

Використовуючи функцію правдоподібності

$$l(\theta) = -\ln[L(\theta, \mathbf{Y}_k | \mathbf{y}(0))], \quad (6.104)$$

оцінку максимальної правдоподібності $\hat{\theta}$ знайдемо шляхом розв'язку такої задачі нелінійної оптимізації:

$$\hat{\theta} = \arg \min_{\theta} (l(\theta)) \quad (6.105)$$

Зазначимо, що значення інновацій $\mathbf{v}(k)$ і коваріаційної матриці $\mathbf{S}(k | k - 1)$ визначаються алгоритмом фільтра Калмана на кожному циклі його реалізації. Таким чином, якість оцінок вектора параметрів $\hat{\theta}$ залежить від коректності функціонування розширеного оптимального фільтра. Імітаційне моделювання розглянутого алгоритму фільтрації та оцінювання параметрів моделі УАРУГ свідчить про можливість практичного застосування такої процедури моделювання нестационарних гетероскедастичних процесів у фінансах.

У одному з прикладів імітаційного моделювання використано такі початкові дані для моделі УАРУГ(1, 1): $\theta = [\alpha_0, \alpha_1, \beta_1] = [1,35; 0,37; 0,64]$; коваріація випадкових збурень для стану у алгоритмі оптимального фільтра $Q=0,15$; довжина згенерованої вибірки даних склала $N = 1500$. Оцінки параметрів, обчислені за розглянутою вище процедурою, мають значення, подані у табл. 6.3.

Таблиця 6.3 Оцінки параметрів моделі і їх середньоквадратичні похибки (СКП)

Параметр	α_0	α_1	β_1
Середнє	1,367	0,359	0,655
СКП	0,0078	0,00095	0,00087

Середньоквадратичні похибки оцінок станів, усереднені за всією вибіркою склали 2,785 і 2,185, відповідно. Середня абсолютна похибка у процентах (САПП) оцінок однокрокових прогнозів умовної дисперсії на навчальній вибірці склала 8,73%. Моделювання цього ж процесу у системі Eviews дало можливість отримати САПП = 7,34%, тобто тут отримано дещо кращий результат за рахунок додаткового попереднього згладжування даних.

Запитання і вправи до шостого розділу

1. Поясніть різницю між гетероскедастичними і гомоскедастичними процесами? Наведіть приклад гетероскедастичного процесу з макроекономіки України.
2. В чому полягає різниця між умовними та безумовними статистичними характеристиками випадкового процесу? Порівняйте вирази для розрахунку безумовної та умовної дисперсій; в чому

принципова різниця між отриманими в результаті розрахунку значеннями?

3. Чому для прогнозування значень випадкових значень змінної та її дисперсії використовують умовне, а не безумовне математичне сподівання?

4. Для чого необхідно прогнозувати значення дисперсії? Наведіть приклади з різних галузей діяльності людини?

5. Що таке *волатильність* (*volatility*) процесу? Як можна її оцінити? Наведіть приклад використання волатильності.

6. Запишіть функцію прогнозування для дисперсії на довільне число кроків за допомогою моделі: $x(k) = \alpha_0 + \alpha_1 x(k-1) + \alpha_2 x(k-2) + v(k)$, де $x(k) = \varepsilon^2(k)$; $\{v(k)\}$ – процес білого шуму з нульовим середнім.

7. Поясніть різницю між моделлю АРУГ та узагальненою авторегресійною моделлю гетероскедастичного процесу:

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i) + v(k)$$

8. Які існують можливості для знаходження (обчислення) оцінок умовної дисперсії $h(k)$?

9. До якого класу відносяться моделі АРУГ і УАРУГ з точки зору їх лінійності? Якими методами можна оцінювати їхні параметри?

10. Сформулюйте своїми словами методику побудови моделі УАРУГ. Поясніть, яка мета побудови на першому кроці авторегресійної моделі процесу низького порядку? Яку роль відіграє при побудові моделі УАРУГ автокореляційна функція процесу $\{\varepsilon^2(k)\}$?

11. Поясніть суть статистики Льюнга-Бокса? Який розподіл мають значення цієї статистики і чому?

12. Які недоліки мають моделі АРУГ та УАРУГ, пов'язані з їх застосуванням?

13. В чому полягає суть леверидж-ефекту? Як можна його врахувати в моделі умовної дисперсії?

14. Які переваги та недоліки має експоненціальна модель УАРУГ?

15. Запишіть стохастичну модель волатильності, поясніть її структуру? В чому полягають її переваги перед іншими моделями гетероскедастичних процесів?
16. В чому полягає суть методу врахування великих імпульсних значень при побудові математичної моделі гетероскедастичного процесу?
17. Наведіть приклади процесів з екології, фінансів та інших галузей, в яких можуть зустрічатись великі імпульсні значення?
18. Якщо будувати модель процесу без врахування наявності великих імпульсних значень, до чого це призведе?
19. Яким чином можна врахувати вплив динаміки дисперсії ринкової ціни на прибуток виробничої фірми? Запишіть відповідну математичну модель.
20. Як можна формально (математично) змодельовати плату за ризик при виконанні ринкових фінансових операцій, інвестуванні і т. ін.?
21. Поясніть суть методики VaR, яка застосовується для оцінювання валютного або іншого фінансового ризику? Який недолік має стандартизована методика VaR?
22. Поясніть суть методики оцінювання ризику CVaR? У чому полягає її перевага перед VaR?
23. В чому полягає суть байєсівського методу оцінювання параметрів стохастичної моделі волатильності, поданого в розділі 6.6? Сформулюйте коротко процедуру програмної реалізації цього методу.
24. Яким чином оцінюють апріорні та апостеріорні розподіли оцінок параметрів регресійної моделі?
25. Чи можна застосувати для оцінювання параметрів моделей гетероскедастичних процесів нелінійний МНК і метод максимальної правдоподібності (ММП)? Які обмеження має нелінійний МНК?
26. Поясніть суть методу Монте-Карло для марковських ланцюгів для оцінювання параметрів моделей? До оцінювання параметрів яких моделей його можна застосовувати?
27. Який метод оцінювання параметрів є незалежним від структури моделі і може бути використаний для оцінювання моделей гетероскедастичних процесів?

РОЗДІЛ 7. АНАЛІЗ НЕСТАЦІОНАРНИХ ТА НЕЛІНІЙНИХ ПРОЦЕСІВ ЗА МЕТОДОМ ПОДІБНИХ ТРАЄКТОРІЙ

7.1 Принцип і алгоритм ітераційного прогнозування

Розглянемо метод непараметричного прогнозування для одновимірної часової послідовності з незначним рівнем шуму, який не потребує побудови моделі в явному вигляді. Цей метод придатний для прогнозування лінійних і нелінійних, а також нестационарних процесів. На практиці часто вважається, що послідовність генерується нелінійною динамічною системою, яка задана такими рівняннями:

$$\begin{aligned} z_{t+1} &= h(z_t, u_t), \\ y_t &= g(z_t), \end{aligned} \tag{7.1}$$

де $h(z_t, u_t)$ та $g(z_t)$ – нелінійні незалежні від часу функції, які мають обмежені похідні; $u_t \in R^{d_s}$ – деяка функція змінних t ; $y_t \in R^1$ і $z_t \in R^{d_s}$. Вважається, що $h(z_t, u_t)$, $g(z_t)$, u_t та d_s – невідомі випадкові процеси, а прогноз має бути побудований виключно на згенерованій системою часовій послідовності $[y_1, y_2, \dots, y_n]$.

Локальне моделювання. Локальні моделі використовуються для розв’язання задачі прогнозування шляхом пошуку частин часових послідовностей, які є досить подібними до тієї частини послідовності, яка знаходиться безпосередньо перед прогнозною точкою. За прогноз обирається оцінка середнього значення приросту, що має місце безпосередньо за схожими частинами послідовностей.

Попередні дослідження показали, що прогнози, отримані за допомогою методів, основаних на локальних моделях, – кращі або схожі до результатів, отриманих за іншими аналітично обґрунтованими методами. Перед використанням локальних моделей необхідно прийняти декілька рішень. Наприклад, математично визначити поняття „локальний”, як побудувати вхідний вектор, який тип локальної моделі використати, як обчислювати

точність моделі та отриманих результатів прогнозування, а також розв'язати деякі інші проблеми.

В кожному окремому випадку є різні можливості та пропозиції щодо того як прийняти те чи інше рішення. Однак загального, всебічно обгрунтованого методу, який поєднував би успішні результати різних досліджень, поки що нема. При використанні локальних моделей також досить часто випускається з розгляду проблема визначення співвідношення між точністю моделі та вартістю обчислень. Нижче будуть розглянуті деякі ідеї щодо використання локальних моделей у прогнозуванні нелінійних та нестационарних процесів.

Обмеження, пов'язані з обчисленнями. Фактично всі нелінійні моделі (локальні та глобальні) мають параметри, які потрібно визначити в процесі побудови моделі чи під час її використання. При аналізі функціонування систем типу чорний ящик може бути така ситуація, коли аналітику про процес невідомо нічого, що могло б допомогти вибрати (оцінити) ці параметри. В ідеалі, параметри вибираються з умови максимізації деякої міри точності моделі. Хоча процес обчислення сьогодні не вимагає значних витрат та доступний як ніколи, глобальна максимізація за всіма параметрами моделі майже у всіх випадках залишається майже недосяжним рішенням внаслідок ряду очевидних причин. Природна ж альтернатива – нелінійна оптимізація – в загальному випадку недосяжна через наявність значного числа локальних мінімумів.

В деяких випадках можуть існувати ефективні методи знаходження допустимих значень частини цих параметрів. Також існуючі теоретичні підходи до аналізу складних процесів дозволяють знайти прийнятні оцінки або, значно частіше, обмеження на значення параметрів. На жаль ці пропозиції в більшості випадків не можна застосувати до всіх параметрів моделі.

Досить часто виникають ситуації, коли точність моделі дуже чутлива до одних параметрів і менш чутлива до значень інших параметрів. Досвідчений аналітик може вибрати прийнятні значення цих параметрів. Однак, процес вибору значень параметрів є суб'єктивним і не допускає узагальнення. Отже точність (адекватність) моделі може сильно залежати від вміння та досвідченості аналітика.

Часто виконується груба оптимізація шляхом порівняння та вибору кращої моделі для декількох варіантів значень її параметрів. Використання такої оптимізації обмежується алгоритмічними труднощами та обчислювальними ресурсами. В результаті, як правило, при цьому страждає точність моделі.

Реконструкція станів системи. Стан багатьох динамічних систем можна достатньо точно реконструювати за допомогою скінченої вибірки (вікна) часової послідовності:

$$x_t = [y_t, y_{t-T_s}, \dots, y_{t-(m-1)T_s}], \quad (7.2)$$

де m – розмір вікна; T_s – період дискретизації вимірів. На сьогодні цей підхід узагальнено на широкий клас лінійних і нелінійних систем.

Вибрані часові вікна даних широко використовуються як вхідні вектори для лінійних і нелінійних динамічних систем. Теорема Тейкенса забезпечує ґрунтовний теоретичний базис такому підходу. Ця теорема припускає, що часовий ряд нескінчений та не містить шуму. Однак на практиці подібні умови ніколи не виконуються. Тому вибір параметрів та періоду дискретизації T_s є критичним по відношенню до точності реконструкції станів системи (ступені відповідності одновимірної моделі векторній).

Для вирішення цієї проблеми запропоновано методи визначення величин m і T_s для скінченої часової послідовності з шумовою складовою та без неї. Загальна мета цих методів – знайти такі значення вказаних параметрів, що мінімізують розмірність часових проміжків m без погіршення точності реконструкції моделі. Хоча оптимізоване реконструювання є досить ефективним з обчислювальної точки зору, максимізація точності є необов'язковою.

Теорема Тейкенса застосовується тільки до нескінченно довгих часових послідовностей, але очевидно, що при достатньо великих значеннях m точність реконструкції є нечутливою до значення m при скінчених вибірках даних без шуму (очевидно, що це ідеалізація постановки задачі). Найважливішим правилом, яким треба користуватись при виборі співвідношення між розмірністю часових проміжків m та періодом дискретизації T_s , є довжина вікна:

$$w = m \times T_s. \quad (7.3)$$

Отже широкий діапазон значень m та T_s , при достатньо великому w та не дуже малих значеннях m , як правило, дозволяє забезпечити відносно точне реконструювання станів системи. Якщо період дискретизації T_s в окремих випадках немає можливості варіювати, то проблеми з вибором належного значення m частіше всього відсутні. Одним із важливих питань при застосуванні підходу на основі локальних моделей є, також, вибір належної метрики для пошуку необхідних сегментів даних.

Вибір метрики. Локальні моделі, що розглядаються в даному розділі, будуються на основі використання тільки найближчих точок на k найближчих сусідніх сегментах траєкторії розвитку досліджуваного процесу.

Вибір належної міри подібності, іншими словами метрики, є важливим рішенням, якому, зазвичай, не завжди приділяється належна увага. Широко розповсюджена метрика – квадрат евклідової відстані між вхідним вектором x_t та цільовим вектором q :

$$d_E(q_t, x_t) = (q_t - x_t)^T (q_t - x_t). \quad (7.4)$$

В даному випадку маємо один параметр – m , розмірність. Вибір метрики більш загального вигляду з більшим числом параметрів може неоднозначно вплинути на точність моделі. Наприклад, зважена евклідова відстань

$$d_\Lambda(q_t, x_t) = (q_t - x_t)^T \Lambda (q_t - x_t). \quad (7.5)$$

може використовуватись як метрика, якщо $\Lambda \in R^{m \times n}$ – додатно напіввизначена матриця. Однак, короткі часові послідовності часто бувають занадто малими для визначення оцінок m^2 параметрів. З іншого боку, оптимізація такої кількості параметрів для великих вибірок може бути не вигідною з обчислювальної точки зору.

Іноді для побудови метрики пропонується використовувати експоненційно зважену діагональну матрицю:

де $0 < \lambda < 1$. Така метрика є інтуїтивно привабливою, тому що ближчим до прогнозованої точки компонентам вектора присвоєна експоненціально більша вага. Це особливо доцільно для хаотичних систем, де з часом сусідні стани експоненційно розбігаються. Оскільки стандартна евклідова метрика, d_E , є підмножиною d_λ , то оптимальний вибір λ не зменшить точність моделі, а навпаки збільшить її. Однак, необхідно конкретизувати, яка оптимізація ставиться за мету в кожному конкретному випадку аналізу даних.

Метрика d_λ визначається параметрами λ та m . При досить великих m та досить малих λ вплив найбільш віддалених компонентів x_i буде незначним. Таким чином, якщо модель будується при досить великому значенні m , на точність моделі впливає лише значення λ .

Оскільки точність моделі залежить від λ , зростають обчислювальні витрати при знаходженні найкращого значення λ . Це і є критичним моментом такого підходу. Більшість нелінійних моделей, такі, наприклад, як нейромережі, нелінійна оптимізація з адаптивними параметрами в загальному випадку потребують значних об'ємів обчислень для знаходження глобального екстремуму за структурними параметрами (наприклад, число нейронів в схованому шарі). На відміну від таких моделей, запропоновану нижче модель можна побудувати та досить ефективно реалізувати глобальну оптимізацію за її параметрами, що дає можливість спростити процедуру побудови та використання моделі.

Вибір параметрів часових проміжків (довжина вікна, період дискретизації). Можна вибрати прийнятне значення розміру вікна при візуальному дослідженні часової послідовності. За відсутності шуму, типова одновимірна часова послідовність, згенерована динамічною системою, що описана рівняннями (7.1.1), буде містити періодичні коливання приблизно однакового періоду. Розмір вікна w треба вибрати таким чином, щоб воно охоплювало декілька коливань. Більше за розмірами вікно потребує більших обчислень при незначному підвищенні точності. При малій довжині часової послідовності велике вікно значно скорочує кількість точок, які можуть бути використані для побудови моделі:

$$n_c = n - m T_s. \quad (7.7)$$

І навпаки, занадто мале вікно зменшить точність моделі. Нехай визначено розмір вікна w . Треба вибрати значення періоду дискретизації T_s . Єдиним виправданням вибору (в тих випадках, коли це можливо) найменшого значення T_s є точність d_λ . Наприклад, якщо часова послідовність була отримана дискретизацією неперервного у часі сигналу з частотою F , то зважена квадратична похибка апроксимації ISE_λ :

$$ISE_\lambda(q_t, x_t) = \int_{tF - (m-1)T_s F}^{tF} \lambda^{v-tF} (q_v - x_v)^2 dv, \quad (7.8)$$

$$d_\lambda(q_t, x_t) = \sum_{i=0}^{m-1} \lambda^i (q_{t-iT_s} - x_{t-iT_s})^2 \approx \frac{1}{F} ISE_\lambda(q_t, x_t).$$

Для фіксованого значення w , вибір T_s впливає на точність ISE_λ . Малі значення збільшать точність ISE_λ , однак точність моделі не обов'язково збільшиться. З іншого боку, зменшення T_s означає збільшення m , а отже зростання об'єму обчислень. Беручи до уваги це протиріччя, період дискретизації T_s треба обирати якомога меншим з обмеженнями щодо

обчислювальних можливостей та колінеарності вимірів.

Ітераційна версія прямого прогнозування. Розглянемо побудову ітераційної версії алгоритму так званого прямого прогнозування. Нехай існує послідовність $[y_1, y_2, \dots, y_n]$ і необхідно зробити прогноз на P кроків вперед. Прямий метод передбачає побудову моделі для прогнозування безпосередньо величини y_{n+p} на основі вхідного вектора x_n .

Ітераційне прогнозування – це метод за яким будується модель для прогнозування P разів на один крок. На основі цієї моделі оцінюється значення y_{n+1} за значеннями x_n , яке потім використовується для оцінювання першої компоненти вектора x_{n+1} :

$$\hat{x}_{n+1} = [\hat{y}_{n+1}, y_{n+1-\tau}, \dots, y_{n+1-(m-1)T_s}]. \quad (7.9)$$

Цей процес повторюється P разів і в результаті отримують прогнозоване значення $\hat{y}_{n+1}$.

Продовжуються природні дискусії на тему який з методів кращий – прямий чи ітеративний. На інтуїтивному рівні зрозуміло, що ітеративний метод при прогнозуванні поведінки більшості нелінійних систем на коротких часових послідовностях дає кращі результати. Також зрозуміло, що для нелінійних систем функція, яка прогнозує на P кроків вперед (пряме прогнозування), буде значно складнішою від функції, яка прогнозує на один крок.

Ітеративне прогнозування має недолік, пов'язаний з накопиченням похибки у вхідному векторі. Тому, починаючи з деякої ітерації, вхідний вектор не буде відповідати станам системи, що моделюється. В такому випадку найкращим прогнозом за критерієм мінімуму суми квадратів буде просте середнє значення сигналу, $\bar{y}$:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (7.10)$$

Однак, незважаючи на погіршення точності вхідного вектора, ітераційний метод дає такий само прогностичний результат з точки зору дисперсії похибки прогнозу. В результаті дисперсії похибок прогнозів на багато кроків вперед і на кілька кроків вперед будуть однаковими. Однак, при ітераційному прогнозуванні середньоквадратична похибка може бути значно більшою ніж при використанні методу прямого прогнозування, чи використанні середнього значення $\bar{Y}$.

Для глобальних моделей пряме прогнозування є більш витратним з обчислювальної точки зору, тому що в даному випадку для отримання прогнозу на $1, 2, \dots, P$ кроків необхідно побудувати P моделей або P функцій типу:

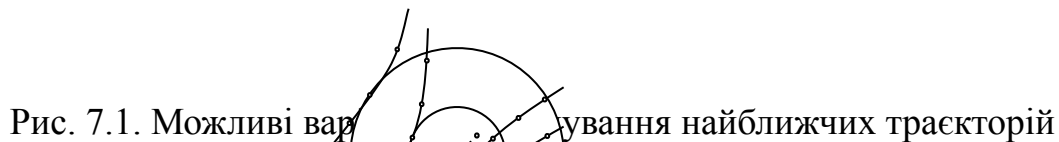
$$\begin{aligned}\hat{y}_{n+1} &= f_1(x_n), \\ \hat{y}_{n+2} &= f_2(x_n), \\ &\dots \\ \hat{y}_{n+p} &= f_p(x_n).\end{aligned}\tag{7.11}$$

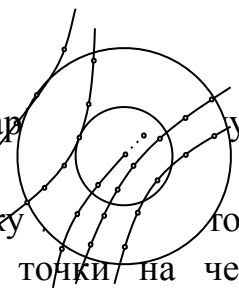
З іншого боку, ітераційний метод прогнозування потребує побудови лише однієї моделі. Для прогнозування за допомогою локальних моделей можна використовувати ті ж самі k найближчих траєкторій як базис для усіх P кроків прогнозування. Для локальних моделей найбільші обчислювальні витрати вимагає операція знаходження найближчої траєкторії, тому прямий підхід є більш прийнятним з обчислювальної точки зору. Це зумовлено тим, що ітераційний метод потребує знаходження найближчої траєкторії для кожного з P вхідних векторів.

Незважаючи на більші обчислювальні витрати, ітераційний метод також часто використовується завдяки високій точності прогнозування на коротких часових послідовностях. Однак його потрібно використовувати з обережністю при довгостроковому прогнозуванні, тому що можна отримати прогноз, гірший за середнє значення $\bar{Y}$. Необхідно зазначити також, що задачі довгострокового прогнозування частіше всього розв'язуються з великим наближенням, що є

характерним практично для всіх методів прогнозування. Тобто в таких випадках задача прогнозування фактично трансформується, в деякій мірі, в задачу передбачення.

Алгоритм пошуку найближчої траєкторії. Однією з проблем прогнозування часової послідовності є те, що існує можливість збільшення кількості вимірів за рахунок зменшення періоду дискретизації чи за допомогою інтерполяції; тобто можна отримати більше даних, не додаючи при цьому нової інформації. Це створює проблему, оскільки у дискретизованого з надто великою частотою сигналу усі найближчі траєкторії будуть знаходитись у часовій послідовності поруч одна до одної. Крім того, при високій частоті дискретизації можуть виникнути проблеми з мультиколінеарністю, яка є наближеною до попередньої проблеми.



Наприклад, на рисунку  точок найближчого сегменту траєкторії розташовані ближче ніж точки на четвертому з найближчих сегментів траєкторії. Ця проблема вирішується пошуком найближчих сегментів траєкторії, замість пошуку найближчих точок траєкторії. До речі, можна модифікувати існуючі алгоритми пошуку найближчих точок траєкторії для знаходження найближчих сегментів. Перевага такого підходу полягає у тому, що відносно алгоритму пошуку найближчих точок вже зроблено багато досліджень.

Опишемо алгоритм пошуку найближчих траєкторій за наявності алгоритму пошуку найближчих точок. Нехай знайдено точку x_i , яка знаходиться ближче за k -у найближчу точку, знайдену алгоритмом раніше. Пошук найближчої траєкторії розглянемо у вигляді таких кроків:

– Підрахувати відстані до попередніх точок послідовно від точки x_i (x_{i-1} , x_{i-2} , ...) з метою пошуку найближчого локального мінімуму. Повторити цю процедуру для точок, що йдуть після x_i (x_{i+1} , x_{i+2} , ...). Локальний мінімум позначимо за $x_{\min}$. Це буде найближча точка в даному сегменті траєкторії.

– Виключити інші точки цього сегменту з подальшого розгляду. Для цього

будемо підраховувати відстані до попередніх точок послідовно від точки x_i ($x_{i-1}, x_{i-2}, \dots$) до тих пір, поки не досягнемо локального максимуму або відстань не перевищить відстань до найближчої k -ї точки, знайденої раніше. Позначимо цю точку через $x_{\max}$.

- Виключити з подальшого розгляду точки, що знаходяться між $x_{\min}$ та $x_{\max}$.
- Повторити попередній крок для точок, що йдуть після $x_{\min}$.
- Замінити k -у найближчу точку точкою $x_{\min}$ і продовжити пошук найближчих точок.

Алгоритми пошуку найближчих точок. Для подолання обчислювальних проблем при прямому підході було запропоновано багато алгоритмів пошуку найближчої точки. Майже всі вони поділяються на дві групи: (1) алгоритми розбивки по осях та (2) алгоритми, що ґрунтуються на нерівності трикутника. Перші розбивають m -вимірний простір на гіперкуби і визначають нижню границю відстані між вибраною точкою та всіма точками кожного з гіперкубів. Якщо знайдена нижня границя є більшою за відстань до знайденої на попередньому кроці k -ї найближчої точки, то до розгляду не приймається увесь гіперкуб і не витрачається час на підрахунок відстані до кожної точки.

В алгоритмах, що ґрунтуються на нерівності трикутника, для знаходження нижньої границі відстані до сусідніх точок використовують нерівність трикутника. Якщо нижня границя є більшою за відстань до знайденої на попередньому кроці k -ї найближчої точки, то така точка виключається з розгляду без подальших витрат на підрахунок відстані до цієї точки. Також можна виключати з розгляду підмножини точок за допомогою визначення нижньої границі відстані до оточуючої гіперсфери.

В просторах з невеликим числом вимірів обидва типи алгоритмів є набагато ефективнішими за прямий підхід. В цьому випадку час пошуку більшості з цих алгоритмів дорівнює $O(\log n_c)$, час та об'єм пам'яті для обробки – $O(n_c \log n_c)$, де n_c – число точок в множині даних, що розглядаються.

В багатовимірному просторі, скажімо при $m > 15$, обидва алгоритми стають неефективними та значно погіршується продуктивність пошуку. На практиці ці алгоритми можуть вимагати значно більшого об'єму обчислень через значне число кроків, зумовлене необхідністю пошуку та обчислення нижніх границь.

У випадку, коли всі точки вибірки належать підпростору R^m з малою розмірністю, алгоритм, що ґрунтується на нерівності трикутника, має переваги над алгоритмами розбивки по осях. Його час пошуку наближається до випадку вибірок з невеликою розмірністю. Якщо розмірність підпростору залишається сталою, то час пошуку збільшується лінійно, пропорційно m . Це особливо помітно у випадку, коли відліки даних беруть з часової послідовності, згенерованої нелінійною динамічною системою, тому що точки часто належать множині з невеликою розмірністю. Ця особливість алгоритму, що ґрунтується на нерівності трикутника, не висвітлюється в літературі, де розглядаються алгоритми, та в літературі з прогнозування часових послідовностей.

Алгоритм прогнозування на основі пошуку найближчої сусідньої точки.

Так само як і інші алгоритми, що ґрунтуються на нерівності трикутника, наведений нижче алгоритм пошуку найближчого сусіда (АПНС) потребує вибору метрики для побудови дерева. Прийmemo за метрику корінь квадратний метрики, представленої виразом (7.1.2). Квадратний корінь необхідний для того, щоб можна було скористатись нерівністю трикутника:

$$d(q, x_i) = d_{\Delta}(q, x_i)^{1/2}. \quad (7.12)$$

АПНС складається з двох етапів: етапу ініціалізації та етапу пошуку. На етапі ініціалізації рекурсивно будується дерево пошуку. Кожна підмножина точок послідовно розбивається на c підмножин до моменту, коли розмір підмножини не перевищує n_T .

Для розбиття на підмножини можна використати будь-який алгоритм кластеризації. В АПНС використовується наступний алгоритм: нехай необхідно розбити множину з n_p точок на c відокремлених підмножин. Тепер можна поступити таким чином:

- за x_0 вибираємо довільну точку;
- знаходимо точку множини, що знаходиться найдалше від x_0 , ця точка є центром першого кластера;
- знаходимо точку множини, що знаходиться найдалше від найближчого з центрів, знайдених до цього; ця точка є центром наступного кластера;
- повторюємо попередній крок до знаходження c центрів кластерів;
- об'єднуємо точки в кластери за найближчим центром; підраховуємо центр мас для кожного кластера точок;
- знаходимо точки $M_1, M_2, \dots, M_c$, що є найближчими до центрів мас кожного з кластерів;
- співставляємо кожну точку з найближчим центром M_i ; позначаємо множину точок, співставлених з M_i , через S_i ;
- для кожної множини S_i підраховуємо радіус як відстань від M_i до найдалшої точки множини S_i :

$$r_i = \max d(M_i, x_i). \quad (7.13)$$

- для кожної множини S_i , яка складається менше ніж з n_T точок, підраховуємо відстань $d(M_i, x_i)$ від центральної точки M_i до кожної з точок множини S_i .

Незважаючи на велику кількість кроків алгоритму кластеризації, об'єм обчислень визначається через $O(n_p)$.

Під час фази пошуку використовується алгоритм пошуку границі на гілці першого рівня. Спочатку алгоритм підраховує відстань від точки, що розглядається, до кожної з центральних точок $d(M_i, q)$ на верхньому рівні дерева пошуку. Потім шукають наступний вузол, який є представником меншої підмножини точок за умови зростання $d(M_i, q)$. Цей процес продовжується до

тих пір, поки алгоритм не знайде кінцевий вузол дерева.

Є кілька правил, які дають можливість уникнути пошуку вузла після підрахунку k відстаней; ці правила такі:

П1: Жодна з точок множини S_i не може бути k -ю найближчою точкою до точки q , якщо

$$d(x_k, q) < d(M_i, q) - r_i, \quad (7.14)$$

де x_k – k -а найближча точка, знайдена на попередньому кроці.

П2: Жодний з вузлів останнього рівня не може бути найближчою точкою, якщо $d(x_k, q) < |d(M_i, q) - d(M_i, x_j)|$.

Також запропоновано кілька додаткових правил, що дозволяють вивести з розгляду більше точок, ніж запропоновані правила. Однак ці правила погано спрацьовують у багатовимірних просторах при значних розмірностях.

Найближча точка в часовій послідовності. Якщо розглядаються точки, отримані з часової послідовності, то час пошуку за алгоритмом АПНС можна зменшити. Наприклад, якщо для вхідного вектора x_t знайдено k перших найближчих точок $[x_1, x_2, \dots, x_k]$, то послідовність точок $[x_{1+1}, x_{2+1}, \dots, x_{k+1}]$ зазвичай буде дуже близькою до x_{t+1} . Якщо підраховувати відстань до цих точок заздалегідь (перед пошуком в дереві), то найменша границя, що використовується в правилах відсіву точок 1 та 2, буде значно вужчою. При цьому більше точок не будуть прийматись до уваги, і таким чином зменшиться об'єм обчислень.

Інтерполяція. Як зазначено вище, алгоритм пошуку найближчої траєкторії знаходить найближчу точку серед точок k найближчих сегментів траєкторії. Якщо часова послідовність представляє собою досить грубо дискретизований, неперервний у часі сигнал, що є виходом динамічної системи, то знайти найближчу точку найближчої траєкторії можна лише з малою точністю (наближено). У випадку, коли послідовність отримують дискретизацією з більшою частотою, або ж, що еквівалентно, часова послідовність є виходом дискретної у часі динамічної системи, то найближча точка траєкторії буде знайдена з більшою точністю і прогнозування буде

виконано точніше. При занадто малій частоті дискретизації сигналу алгоритмом пошуку найближчих траєкторій можуть бути знайдені невірні траєкторії. Проілюструємо сказане малюнком:

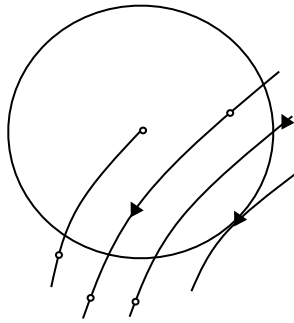


Рис. 7.2. Можливості знаходження невірних траєкторій

На практиці можна значно підвищити частоту дискретизації з досить великою точністю, якщо вихідна послідовність квантована занадто грубо. Однак, пропорційно збільшенню кількості точок у часовій послідовності підвищується і об'єм обчислень при реалізації алгоритму пошуку найближчої траєкторії. Правило, яким необхідно користуватись в цьому питанні, таке: виконати інтерполяцію таким чином, щоб частота дискретизації перевищувала частоту Найквіста (теорема Котельникова-Шеннона) в 5-10 разів.

Апроксимація локальними функціями. Для виконання прогнозу після знаходження k найближчих сегментів траєкторії необхідно побудувати локальну модель. Це стандартна проблема апроксимації за винятком одного важливого зауваження: вимірність вхідного вектора (m) є порівняною з довжиною послідовності (k), що використовується для побудови моделі. Більшість методів нелінійного моделювання не можуть бути застосовані за таких умов. Вони мають забагато вільних параметрів і потребують більше k точок для побудови моделі. Також не можна використовувати методи, що потребують об'ємні обчислення, тому що кожний раз необхідно будувати нову модель.

Локальні лінійні моделі. Локальні лінійні моделі стали популярним вибором завдяки тому, що їх можна швидко побудувати використовуючи короткі вибірки. Деякі автори запропонували перед побудовою лінійної моделі проектувати вхідні вектори на підпростір. Частіше за все для знаходження l -вимірного підпростору, в якому вхідні вектори мають найбільшу дисперсію, будується лінійна модель та генерується прогноз.

Недоліком локальних моделей є проблема їх застосування у випадку розривності. Цей недолік згадується в результатах досліджень найчастіше. Невелика зміна вхідного вектора може змінити k -у найближчу траєкторію, що в результаті призводить до зовсім іншої моделі. Ця проблема може бути вирішена завдяки зменшенню впливу найближчих траєкторій на локальну модель. Наприклад, можна будувати лінійні моделі, використовуючи зважену лінійну регресію, де найдаліші траєкторії мають найменшу вагу.

Розроблені різноманітні вагові функції. Загальноприйнятим є те, що точність моделі не залежить від типу вибраної вагової функції. При цьому прийнятним вибором є квадратична функція:

$$w(i) = \left(1 - \frac{d_i^2}{d_k^2}\right)^2, \quad (7.15)$$

де d_i – відстань до i -ї найближчої точки. Однією з переваг біквадратичної вагової функції є те, що вагова функція для k -ї найближчої точки дорівнює нулю, що робить локальну модель неперервною. Така функція досить часто використовується на практиці.

Локально стійкі (сталі) моделі. Трохи рідше використовують локальні сталі моделі. Як і локально-лінійні моделі, їх можна зробити неперервними, використовуючи гладкі вагові функції. Завдяки тому, що сталі моделі мають тільки одну ступінь свободи, вони мають меншу варіацію, ніж лінійні моделі, але більше відхилення для тієї ж кількості точок, K . При використанні оптимального K жоден з методів не є більш точним порівняно з іншими. Лінійні моделі вимагають трохи більше обчислень, але в більшості випадків об'єм додаткових обчислень значно менший за траєкторії, вартості обчислень яких потребує пошуку K найближчих траєкторій.

Хоча локально сталі моделі достатньо точні при розв'язуванні задачі

інтерполяції, в загальному випадку вони дають поганий результат при екстраполяції. Враховуючи те, що локальні моделі будуються на зваженому середньому, вихід ніколи не вийде за межі точок, з яких вираховувалось середнє. В просторах з великою кількістю вимірностей це може бути суттєвим недоліком через те, що майже кожна точка знаходиться поза цими межами. Наприклад, на малюнку сірим колом показано прогнозоване значення, згенероване як середнє наступних за найближчими траєкторіями точок.

Одним із шляхів вирішення цієї проблеми є прогнозування першої різниці $y_{t+1} - y_t$ замість y_{t+1} . Кінцеве значення прогнозу знаходиться інтегруванням поточного значення.



На рис. 7.3 показано прогнозування першої різниці і подальше інтегрування (інтегруюче прогнозування) і усереднююче прогнозування середнього сегментів найближчих траєкторій (усереднююче прогнозування).

Для великих вибірок даних ці методи мають однакову продуктивність і точність. Однак, виходячи з емпіричних порівнянь, можна стверджувати, що інтегруюче прогнозування дає трохи точніший результат на коротких вибірках, а усереднююче прогнозування – на довгих. Мабуть це має місце тому, що усереднююче прогнозування є менш чутливим до накопичення похибок у вхідному векторі.

Кількість найближчих траєкторій. Багато дослідників показують, що точність моделі чутлива до кількості найближчих траєкторій K , що робить цей параметр кандидатом на глобальну оптимізацію. Це рішення є досить таки критичним, тому глобальну оптимізацію можна виконувати лише за декількома параметрами.

Оптимізація значення K є набагато ефективнішою за оптимізацію параметра метрики λ з двох причин. По-перше, пошук найближчої точки можна виконати для декількох значень параметра λ без побудови нових дерев пошуку алгоритмом АПНС. Кожне нове значення λ потребує перебудови дерева пошуку.

По-друге, в загальному випадку для часових послідовностей з малим

рівнем шуму оптимальне значення параметра k є малим. Тому для пошуку оптимального значення необхідно розглянути невелику кількість різних значень k . З іншого боку, для знаходження оптимального λ треба розглянути набагато більше значень, оскільки λ знаходиться у неперервному проміжку від 0 до 1.

7.2 Аналіз точності прогнозу

Запропоновано багато методів обчислення точності моделі. Для локальних моделей популярним є метод послідовного відкидання точок з обчисленням похибки прогнозу.

Цей метод починається з побудови моделі для прогнозування на 1 крок вперед, використовуючи всі, окрім однієї з K точок. Підраховується похибка прогнозування відкинутої точки і процес повторюється для всіх точок. Можна використати будь-який з існуючих способів вимірювання похибки (абсолютна похибка, квадратична похибка і коефіцієнт кореляції використовуються найчастіше).

Усереднена похибка є оцінкою очікуваної похибки. Наприклад, якщо вибрана квадратична похибка, то усереднена похибка визначається як:

$$\begin{aligned} MSE_1^{\Delta} &= \frac{1}{n_c} \sum_{i=1}^{n_c} (y_{i+1} - f_{i+1}^{-(i+1)}(x_i))^2 \approx E_t[(y_{t+1} - f_{t+1}^{-(t+1)}(x_t))^2] \\ &\approx E_t[(y_{t+1} - f_{t+1}(x_t))^2], \end{aligned} \quad (7.16)$$

де $f_{i+1}^{-(i+1)}(x_i)$ – модель, що побудована при відкинутих $(i+1)$ точках для прогнозування на один крок вперед.

Цей метод є особливо привабливим для локальних моделей, тому що за ним можна ефективно виконати обчислення без побудови n_c різних моделей. Навпаки, при пошуку k найближчих точок відкинуті точки можна просто не розглядати. Подальші переваги можуть бути отримані одноразовим виконанням пошуку найбільшого значення k . Пізніше це допоможе для найменших значень k будувати локальні моделі без виконання додаткових пошуків для пошуку оптимального значення k , що приводить до збільшення об'єму обчислень.

Можливість ефективного обчислення усередненої похибки є значною

перевагою локальних моделей. Багато стратегій глобального моделювання, такі як нейромережі, потребують великих об'ємів обчислень для побудови однієї моделі. Для цих стратегій побудова n_c різних моделей є, як правило, неприпустимою з обчислювальної точки зору.

Недоліком використання однокрокової усередненої похибки як способу вимірювання точності моделі є те, що вона не враховує чутливість моделі до похибок у вхідному векторі, що з'являються при ітераційному прогнозуванні. Тобто параметри, що мінімізують однокрокову усереднену похибку не співпадають з параметрами, що мінімізують p -крокову усереднену похибку.

Краще вибирати такий спосіб вимірювання похибки, який вказує на вартість отримання низькоякісного прогнозу у тому застосуванні, у якому буде використовуватись модель.

7.3 Моделювання ітераційного алгоритму АПНС

Методика моделювання ітераційного алгоритму. Задача моделювання алгоритму полягає у визначенні параметрів алгоритму прогнозування. Запропонований алгоритм для прогнозування потребує визначення наступних параметрів: λ , m , w , де

- w – розмір вікна, або вимірність простору в якому будується траєкторія руху системи; цей параметр визначає кількість значимих відліків при пошуку найближчої точки траєкторії руху системи і, разом з цим, вимірність міри, визначеної в цьому просторі;

- λ – коефіцієнт, який визначає ступінь важливості врахування кожного попереднього відліку при порівнянні точок;

- m – кількість найближчих точок, пошук яких виконується на кожному кроці і на основі яких виконується прогнозування.

Програмна реалізація алгоритму у вигляді пакету програм nlPredictor.exe дозволяє оцінити прогноз як за алгоритмом АПНС, так і за його модифікацією, що при пошуку найближчої точки виконує відносне порівняння. Це означає побудову локальної моделі з попереднім видаленням тренду. Тренд в цьому випадку може мати довільний характер. Модифікація алгоритму також є одним з параметрів запропонованого алгоритму прогнозування, вибір якого – частина процесу моделювання.

При моделюванні алгоритму використовувалась оптимізаційна процедура, реалізована в пакеті програм nlPredictor.exe. Проводилась глобальна оптимізація параметрів моделі, при цьому вказувалось межі зміни параметрів, точність (крок, з яким змінювалось значення параметру), модифікація алгоритму, яка використовується при прогнозуванні, а також частина вибірки, яка використовувалась для оптимізації параметрів.

Отже, при моделюванні алгоритму використовувалась наступна послідовність кроків:

- 1) вибір вихідної (початкової) послідовності;
- 2) вибір модифікації методу прогнозування;
- 3) визначення частини послідовності для оптимізації параметрів;
- 4) глобальна оптимізація параметрів моделі;
- 5) виконання імітаційного прогнозування і визначення точності прогнозування на основі ряду статистичних показників.

Результати моделювання ітераційного алгоритму

1. Авіаперевезення. Розглянемо ряд, який містить помісячну інформацію щодо об'єму міжнародних пасажирських авіаперевезень перевезень (рис. 7.4).

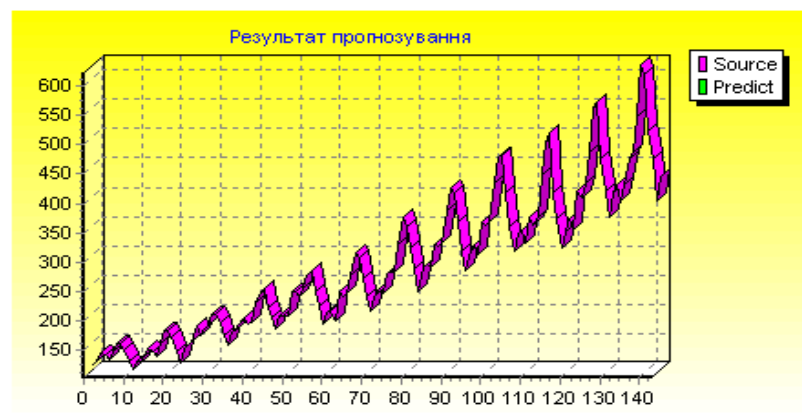


Рис.7.4. Статистичні дані щодо міжнародних авіаперевезень

Ряд має явно виражений сезонний характер, містить тренд. Виходячи з характеру ряду, для прогнозування була вибрана друга модифікація методу (відносно порівняння траєкторій).

Для оптимізації параметрів моделі використано останні 10 відліків послідовності. Проведено оптимізація параметра λ для кожного із значень параметра w і визначено залежність СКО від розміру вікна; отримано наступні результати:

Послідовність AIR Глибина прогнозу: 10 відліків, Модифікація: відносне прогнозування				
	Результат оптимі- зації	Результат прогнозування		
w	λ	СКО	Макси-мальне відхилення	Максимальне відхилення, %
3	0,9	32,1931	58	9,3247
4	0,4	32,1931	58	9,3247
5	0,6	32,1931	58	9,3247
6	0,5	32,1931	58	9,3247
7	0,4	27,2598	58	9,3247
8	0,8	28,3813	60	9,6463
9	1	26,7880	58	9,3247
10	0,9	18,3084	36	5,7877
11	0,8	18,3084	36	5,7877

Порівнюючи отримані результати, необхідно відзначити незначну залежність точності прогнозування від порядку моделі, що є безумовною перевагою даного методу.

2. Послідовність Лоренца. Ряд містить дискретизоване рішення рівняння Лоренца:

$$\begin{aligned}
 \dot{x} &= \delta(y - x) \\
 \dot{y} &= rx - y - xz \\
 \dot{z} &= xy - bz
 \end{aligned}
 \tag{7.17}$$

Ряд має наступний характер:

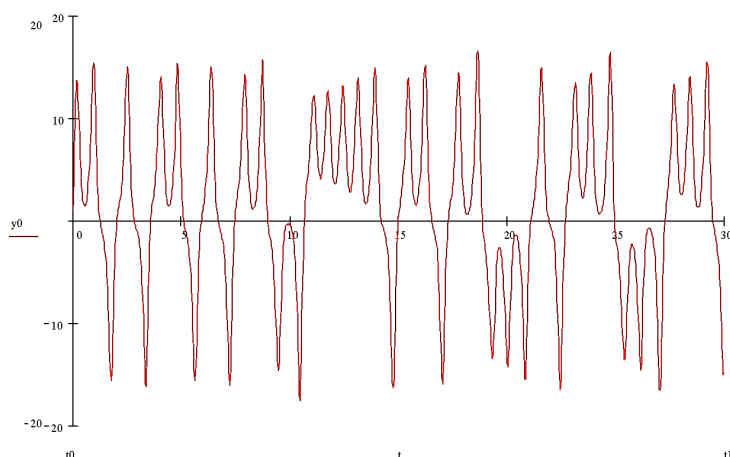


Рис. 7.5. Ряд Лоренца

Результат використання процедури оптимізації параметрів:

Розмір вікна $m = 20$; $\lambda = 0,7$.

Для прогнозування було обрано інтервал прогнозування = [940 ... 1000] (глибина прогнозу 60 відліків – середньострокове прогнозування).

Результат прогнозування наведено на рис. 7.6.

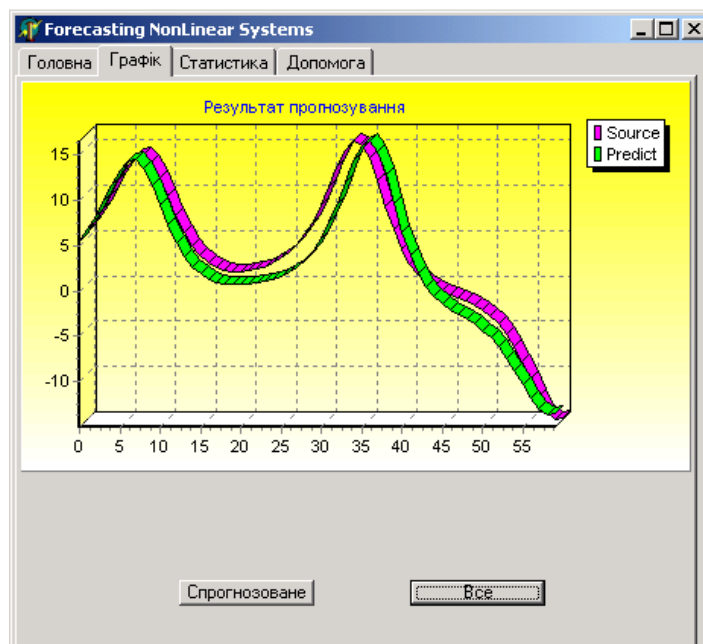


Рис. 7.6. Результати прогнозування ряду Лоренца

Показники точності прогнозування:

– СКО = 2,37588 (15,8392% від максимального значення);

– максимальне відхилення = 6,85 (44,83% максимального значення).

Численні приклади застосування даного методу, які наведені в фаховій літературі, свідчать про його високу ефективність у випадках, коли спостерігаються подібні участки ряду даних. Його безумовними перевагами є такі:

- простота реалізації та подальшої модифікації;
- можливість прогнозування на довільне число кроків без значних втрат точності, що характерно багатьом іншим методам;
- можливість застосування до процесів будь-якого рівня складності (наявність нестационарності та нелінійності).

Запитання і вправи до сьомого розділу

1. Що таке коінтеграція? Чим вона відрізняється від простої кореляції між часовими рядами?
2. Які умови повинні виконуватись, щоб два нестационарні процеси вважалися коінтегрованими?
3. Наведіть приклади макроекономічних змінних, що можуть утворювати коінтегровані пари (або множини).
4. Чому наявність коінтеграційного зв'язку вважається важливою для побудови економетричних моделей?
5. Поясніть суть моделі корегування похибок (ECM). Яке місце вона займає у теорії коінтеграції?
6. Які методи перевірки коінтеграції ви знаєте (тест Енгла–Грейнджера, тест Йогансена тощо)?
7. У чому полягає різниця між одно рівнянням тестом Енгла–Грейнджера та багатовимірним підходом Йогансена?
8. Як впливає на прогнозування наявність або відсутність коінтегрованості між змінними?
9. Чим коінтегровані процеси відрізняються від інтегрованих процесів із випадковим трендом?
10. Чому нестационарні процеси без коінтеграції можуть призводити до «фальшивої регресії»?
11. Візьміть економічні дані (наприклад, рівень інфляції та облікову ставку). Виконайте тест Йогансена і визначте кількість коінтеграційних векторів.
12. Побудуйте регресійну модель між двома нестационарними рядами, які не є коінтегрованими. Проаналізуйте коефіцієнти детермінації (R^2) та зробіть висновки про «фальшиву регресію».

13. Використайте будь-який відкритий фінансовий чи макроекономічний ряд (наприклад, ціни на нафту та курс валюти). Перевірте гіпотезу про коінтеграцію. Опишіть отримані результати та їхнє економічне значення.

РОЗДІЛ 8. НЕПАРАМЕТРИЧНІ МЕТОДИ ПОБУДОВИ МАТЕМАТИЧНИХ МОДЕЛЕЙ НЕЛІНІЙНИХ ЧАСОВИХ РЯДІВ

Вище було розглянуто методику побудови математичних моделей, яка може бути застосована до будь-якого класу процесів. Тепер розглянемо деякі конкретні методи побудови моделей нелінійних фінансово-економічних, соціальних та екологічних процесів.

8.1 Непараметрична регресія

У багатьох випадках побудови математичних (у формі рівнянь різного типу) і статистичних моделей (у формі розподілів випадкових величин) немає достатньої інформації для виявлення та встановлення типу нелінійності процесу, що моделюється. Тобто необхідно побудувати модель нелінійної залежності між змінними $Y(k)$ і $X(k)$, але ми не можемо наперед встановити тип нелінійності і конкретну форму цієї залежності. У таких випадках можна скористатись методами непараметричного моделювання, застосування яких також потребує належної уваги та застосування іноді непростих обчислювальних процедур. До непараметричних методів відносять такі:

- регресію на ядрах;
- локальні лінійні регресійні моделі, які оцінюються за допомогою МНК;
- нейронні мережі;
- метод (машину) опорних векторів;
- метод найближчого сусіда.

Ключовим моментом у застосуванні непараметричних методів є *згладжування* вимірів. Розглянемо дві змінні, зв'язані таким рівнянням:

$$Y(k) = F(X(k)) + \varepsilon(k), \quad (8.1)$$

де $F(X(k))$ – довільна гладка нелінійна функція, яку необхідно оцінити за допомогою наявних даних; $\varepsilon(k)$ – випадковий процес – часто послідовність значень білого шуму, природа якого розглянута вище. Припустимо, що значення

$X(k) = x$ відомі на деякий момент часу; при цьому $Y = [y(1), y(2), \dots, y(N)]^T$, тобто:

$$y(k) = F(x(k)) + \varepsilon(k), \quad k = 1, \dots, N. \quad (8.2)$$

Для усереднених даних можна записати:

$$\frac{1}{N} \sum_{k=1}^N y(k) = F(x(k)) + \frac{1}{N} \sum_{k=1}^N \varepsilon(k) \quad (8.3)$$

За законом великих чисел середнє значення випадкових збурень збігається до нуля при $N \rightarrow \infty$, а тому вибіркове середнє

$$\bar{y} = \frac{1}{N} \sum_{k=1}^N y(k) \quad (8.4)$$

буде консистентною оцінкою функції $F(x)$.

Як правило, при дослідженні даних у формі часових рядів, немає можливості отримати повторні спотереження для змінних, тобто повторити експеримент. Для побудови моделі є такі вибірки: $\{y(k), x(k)\}$, $k=1, \dots, N$. Якщо функція $F(x)$ є достатньо гладкою, вона може забезпечити належну апроксимацію нелінійної залежності, але при обробці фактичних даних це не завжди можливо.

Компромісним варіантом при встановленні залежності може бути застосування зваженого середнього (з різними ваговими коефіцієнтами) замість простого середнього. Вагові коефіцієнти мають бути більшими для $Y(k)$ у тих випадках, коли значення $x(k)$ покращують адекватність функціональної залежності $Y(k)$ та $X(k)$, і меншими для тих $Y(k)$, для яких адекватність погіршується.

Формально оцінку функціональної залежності $F(x)$ для заданих значень x можна записати так:

$$\hat{F}(x) = \frac{1}{N} \sum_{k=1}^N w(k, x) y(k) \quad (8.5)$$

де $w(k, x)$ – вагові коефіцієнти, сума яких може дорівнювати N . У окремому випадку вагові коефіцієнти можуть бути однаковими і дорівнювати $1/N$. Тоді їх сума буде дорівнювати одиниці.

З виразу (8.1.2) випливає, що оцінка $\hat{F}(x)$ є *локальним зваженим середнім*, вагові коефіцієнти якого вибираються на основі двох факторів:

- першим фактором є відстань (міра близькості) між $x(k)$ і x ;
- другим фактором є вибір вагових коефіцієнтів на основі заданої міри близькості.

Різні способи визначення відстані між $x(k)$ і x та оцінювання значень вагових коефіцієнтів приводять до різних методів непараметричного моделювання.

Регресія на ядрах

Регресія на ядрах – це один із самих широко вживаних методів непараметричного згладжування. У цьому випадку вагові коефіцієнти задаються деяким *ядром* $K(x)$, у якості якого вибирають, як правило, функцію щільності розподілу ймовірностей, тобто

$$K(x) \geq 0, \quad \int K(z) dz = 1. \quad (8.6)$$

Для досягнення більшої гнучкості при визначенні міри близькості (відстані) ядро масштабують шляхом введення змінної $h > 0$, яку називають шириною смуги пропускання. Масштабоване ядро має вигляд:

$$\mathbf{K}_h(x) = \frac{1}{h} \mathbf{K}(x/h), \quad \int \mathbf{K}_h(z) dz = 1 \quad (8.7)$$

Тепер вагову функцію можна задати таким чином:

$$w(k, x) = \frac{\mathbf{K}_h(x - x(k))}{\sum_{k=1}^N \mathbf{K}_h(x - x(k))}, \quad (8.8)$$

де знаменник – нормуюча константа, яка дає можливість адаптувати цей метод згладжування до конкретних значень (інтенсивності) змінної X і забезпечує одиничне значення суми вагових коефіцієнтів. Підставляючи (8.1.3) у (8.1.2), отримаємо відомий вираз Надарая-Уотсона для апроксимації функціональної залежності за допомогою ядра [51]:

$$\hat{F}(x) = \sum_{k=1}^N w(k, x) y(k) = \frac{\sum_{k=1}^N \mathbf{K}_h(x - x(k)) y(k)}{\sum_{k=1}^N \mathbf{K}_h(x - x(k))} \quad (8.9)$$

Для вибору типу ядра $\mathbf{K}(x)$ існують різні можливості, але кількість відомих варіантів вибору обмежена, наприклад, досить часто використовують гаусівське ядро:

$$\mathbf{K}_h(x) = \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{x^2}{2h^2}\right), \quad (8.10)$$

а також ядро Єпанечнікова:

$$\mathbf{K}_h(x) = \frac{0,75}{h} \left(1 - \frac{x^2}{h^2}\right) \mathbf{I}\left(\left|\frac{x}{h}\right| \leq 1\right), \quad (8.11)$$

де $\mathbf{I}(A)$ – індикаторна функція, що визначається так:

$$\mathbf{I}(A) = \begin{cases} 1, & \text{якщо виконується умова } A; \\ 0, & \text{якщо не виконується умова } A. \end{cases} \quad (8.12)$$

Для того щоб показати яку роль відіграє ширина смуги пропускання h , розглянемо апроксимацію Надарая-Уотсона у контексті використання ядра Єпанечнікова для аналізу спостережень $\{x(k)\}$. При цьому доцільно розглянути два екстремальні випадки значень смуги пропускання: $h \rightarrow 0$ і $h \rightarrow \infty$. Якщо $h \rightarrow 0$, то

$$\hat{F}(x(k)) \rightarrow \frac{\mathbf{K}_h(0) y(k)}{\mathbf{K}_h(0)} = y(k) \quad (8.13)$$

тобто при малих значеннях смуги пропускання відбувається просте відтворення вихідних даних. Якщо $h \rightarrow \infty$, то

$$\hat{F}(x(k)) \rightarrow \frac{\sum_{k=1}^N \mathbf{K}_h(0) y(k)}{\sum_{k=1}^N \mathbf{K}_h(0)} = \frac{1}{N} \sum_{k=1}^N y(k) = \bar{y} \quad (8.14)$$

тобто використання великих значень смуги пропускання призводить до перегладжування – результатом буде вибіркове середнє. Загалом, можна стверджувати, що при малих значеннях h вагові коефіцієнти фокусуються на декількох спостереженнях, які знаходяться у околі кожного $x(k)$. При великих значеннях h вагові коефіцієнти розповсюджуються на ширший окіл кожного значення $x(k)$. Таким чином, вибір значення h відіграє велике значення при застосуванні методу регресії на ядрах і розв’язанню цієї задачі необхідно

приділяти належну увагу.

Вибір смуги пропускання. Існують кілька підходів до вибору ширини пропускання [51]. Перший підхід ґрунтується на використанні асимптотичного розкладу середньої інтегрованої квадратичної похибки (СІКП або MISE – mean integrated squared error) для методів згладжування на основі ядра:

$$MISE = E \left[\int_{-\infty}^{\infty} [\hat{F}(x) - F(x)]^2 dx \right], \quad (8.15)$$

де $F(x)$ – істинна функція зв'язку. Величина $E[\hat{F}(x) - F(x)]^2$ – це точкова міра середньоквадратичної похибки (СКП) значення функції $\hat{F}(k)$ у точці x . За виконання деяких умов регулярності можна отримати оптимальну ширину смуги пропускання, яка мінімізує похибку MISE. Оптимальне значення смуги пропускання залежить, як правило, від кількох невідомих величин, які необхідно оцінити на основі статистичних даних після деякої попередньої обробки (згладжування). Однак, на практиці вибір процедури такого попереднього згладжування може бути проблематичним, а тому можна скористатись таким емпіричним правилом вибору ширини смуги пропускання:

$$\hat{h}_{opt} = \begin{cases} 1,06 s N^{-1/5} & \text{для гаусового ядра;} \\ 2,34 s N^{-1/5} & \text{для ядра Єпанечнікова,} \end{cases} \quad (8.16)$$

де s – вибіркове стандартне відхилення незалежної змінної, яке має бути стаціонарним за припущенням.

Інший підхід до оцінювання величини смуги пропускання полягає у використанні додаткової перевіркової функції. Спочатку із вибірок спостережень видаляється по одному значенню $(x(j), y(j))$. Потім $N-1$ значень вибірок використовуються для того щоб отримати таку функцію згладжування:

$$\hat{F}_{h,j}(x(j)) = \frac{1}{N-1} \sum_{k \neq j} w(k, x(j)) y(k) \quad (8.17)$$

яка є оцінкою для $y(j)$, а сума вагових коефіцієнтів дорівнює N^{-1} . Далі знаходимо функцію $\hat{F}_{h,j}(x(j))$ для $j = 1, \dots, N$ і визначаємо функціонал:

$$CV(h) = \frac{1}{N} \sum_{j=1}^N [y(j) - \hat{F}_{h,j}(x(j))]^2 W(x(j)) \quad (8.18)$$

де $W(x(j))$ – невід’ємно визначена вагова функція така, що $\sum_{j=1}^N W(x(j)) = N$; $CV(\cdot)$ – функція крос-валідації (cross-validation function). За необхідності, ця функція може бути використана для зменшення впливу граничних (крайніх) вимірів, які мають меншу кількість сусідніх вимірів. Функцію $CV(h)$ називають перехресною перевіркою функцією оскільки вона дає можливість визначити здатність функції згладжування прогнозувати значення $\{y(k)\}_{k=1}^N$. Величину смуги пропускання h вибирають за умови мінімуму функції $CV(\cdot)$, тобто

$$h = \arg \min CV(\cdot). \quad (8.19)$$

Локальна лінійна непараметрична регресія

Припустимо, що існує друга похідна функції $F(x)$ у моделі (8.1.1) і вона є неперервною по x , де x – значення з опорної множини цієї функції. Позначимо наявні для дослідження дані через $\{y(k), x(k)\}_{k=1}^N$. Задача побудови локальної лінійної непараметричної регресії (ЛЛНР) полягає у тому, щоб знайти такі параметри a, b , які мінімізують функціонал:

$$L(a, b) = \sum_{k=1}^N [y(k) - a - b(x - x(k))]^2 \mathbf{K}_h(x - x(k)) \quad (8.20)$$

де $\mathbf{K}_h(\cdot)$ – ядро, визначене вище; h – ширина смуги пропускання; $x(k)$ – незалежна змінна. Надалі оцінка $\hat{a}$ вважається оцінкою для $F(x)$, а оцінка $\hat{b}$ розглядається як значення першої похідної функції $F(x)$ у точці x .

З точки зору теорії оцінювання функціонал (8.1.5) відповідає задачі оцінювання параметрів за допомогою зваженого методу найменших квадратів (МНК), а тому можна отримати вираз для оцінки параметра $\hat{a}$. Для цього застосуємо відому процедуру прирівнювання нулю перших похідних функціонала $L(a, b)$ стосовно параметрів a, b . Таким чином отримаємо систему з двох рівнянь:

$$\begin{aligned} \sum_{k=1}^N \mathbf{K}_h(x - x(k)) y(k) &= a \sum_{k=1}^N \mathbf{K}_h(x - x(k)) + b \sum_{k=1}^N (x - x(k)) \mathbf{K}_h(x - x(k)) \\ \sum_{k=1}^N y(k)(x - x(k)) \mathbf{K}_h(x - x(k)) &= a \sum_{k=1}^N (x - x(k)) \mathbf{K}_h(x - x(k)) + \\ &+ b \sum_{k=1}^N (x - x(k))^2 \mathbf{K}_h(x - x(k)) \end{aligned} \quad (8.21)$$

Введемо позначення:

$$s_{T,l} = \sum_{k=1}^N \mathbf{K}_h(x - x(k)) (x - x(k))^l, \quad l = 0, 1, 2$$

Тепер систему з двох нормальних рівнянь можна записати у матричному вигляді так:

$$\begin{bmatrix} s_{T,0} & s_{T,1} \\ s_{T,1} & s_{T,2} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{k=1}^N \mathbf{K}_h(x - x(k)) y(k) \\ \sum_{k=1}^N (x - x(k)) \mathbf{K}_h(x - x(k)) y(k) \end{bmatrix} \quad (8.22)$$

Звідси знайдемо вираз для оцінки $\hat{a}$:

$$\hat{a} = \frac{s_{T,2} \sum_{k=1}^N \mathbf{K}_h(x-x(k)) y(k) - s_{T,1} \sum_{k=1}^N (x-x(k)) \mathbf{K}_h(x-x(k)) y(k)}{s_{T,0} s_{T,2} - s_{T,1}^2} \quad (8.23)$$

Вирази у чисельнику і знаменнику можна спростити таким чином:

$$\begin{aligned} s_{T,2} \sum_{k=1}^N \mathbf{K}_h(x-x(k)) y(k) - s_{T,1} \sum_{k=1}^N (x-x(k)) \mathbf{K}_h(x-x(k)) y(k) &= \\ &= \sum_{k=1}^N \{ \mathbf{K}_h(x-x(k)) [s_{T,2} - (x-x(k)) s_{T,1}] \} y(k) ; \\ s_{T,0} s_{T,2} - s_{T,1}^2 &= \sum_{k=1}^N \mathbf{K}_h(x-x(k)) s_{T,2} - \sum_{k=1}^N (x-x(k)) \mathbf{K}_h(x-x(k)) s_{T,1} = \\ &= \sum_{k=1}^N \mathbf{K}_h(x-x(k)) [s_{T,2} - (x-x(k)) s_{T,1}] \end{aligned} \quad (8.24)$$

Остаточно оцінка параметра $\hat{a}$ визначається за виразом:

$$\hat{a} = \frac{\sum_{k=1}^N w(k) y(k)}{\sum_{k=1}^N w(k)}, \quad (8.25)$$

де $w(k) = \mathbf{K}_h(x-x(k)) [s_{T,2} - (x-x(k)) s_{T,1}]$.

Для того щоб уникати нульового значення у знаменнику, для обчислення $\hat{F}(x)$ використовують такий вираз:

$$\hat{F}(x) = \frac{\sum_{k=1}^N w(k) y(k)}{\sum_{k=1}^N w(k) + 1/T^2} \quad (8.26)$$

Корисною властивістю виразу (8.1.7) є те, що вагові коефіцієнти задовольняють такій умові:

$$\sum_{k=1}^N (x - x(k)) w(k) = 0 \quad (8.27)$$

Також, якщо припустити, що функція $F(\cdot)$ має першу похідну і мінімізує вираз

$$\sum_{k=1}^N (y(k) - a)^2 \mathbf{K}_h(x - x(k)) \quad (8.28)$$

то результуючою моделлю буде модель Надарая-Уотсона, згадана вище.

Якщо припустити у загальному випадку, що $F(\cdot)$ має обмежену k -ту похідну, то лінійний поліном у рівнянні (8.1.5) можна замінити на поліном $(k-1)$ -го порядку. Вираз (8.1.7) називають моделлю локального лінійного регресійного згладжування.

Застосування до аналізу нелінійних часових рядів. У аналізі часових рядів часто зустрічаються пояснючі змінні у вигляді затриманих у часі значень основної змінної. Розглянемо просту модель з одним лагом:

$$y(k) = F(y(k-1)) + \varepsilon(k), \quad (8.29)$$

до якої застосуємо метод побудови локальної лінійної регресії, розглянутий вище. У випадку існування кількох пояснюючих змінних застосування непараметричного методу потребує деяких модифікацій. Для побудови регресії на основі ядра, можна скористатись багатовимірним нормальним розподілом:

$$\mathbf{K}_h(\mathbf{x}) = \frac{1}{(h\sqrt{2\pi})^p |\Sigma|^{1/2}} \exp\left(-\frac{1}{2h^2} \mathbf{x}^T \Sigma^{-1} \mathbf{x}\right) \quad (8.30)$$

де P – кількість пояснюючих змінних; Σ – додатно визначена матриця. Як альтернативу, у якості багатовимірного ядра можна використати добуток

одновимірних ядерних функцій

$$\mathbf{K}_h(\mathbf{x}) = \prod_{i=1}^p \frac{0,75}{h_i} \left(1 - \frac{x_i^2}{h_i^2} \right) \mathbf{I} \left(\left| \frac{x_i}{h_i} \right| \leq 1 \right) \quad (8.31)$$

Останній підхід простіший, але він не враховує взаємозв'язку між пояснюючими змінними.

Приклад 8.1. Для формального опису процесу ціноутворення вибраного біржового процесу скористаємось рівнянням:

$$y(k) = \mu(x(k-1)) dt + \sigma(x(k-1)) dw(k), \quad (8.32)$$

де $x(k)$ – поточна ціна активу; $y(k) = x(k) - x(k-1)$ – прирости ціни; $w(k)$ – стандартний броунівський процес; $\mu(\cdot)$ і $\sigma(\cdot)$ – гладкі функції змінної $x(k-1)$. Для простоти величина $|y(k)|$ використовується як міра волатильності процесу $x(k)$.

У вибраній моделі $\mu(x(k-1))$ – умовне середнє процесу $y(k)$, тобто умовою є наявність значень $x(k-1)$: $\mu(x(k-1)) = E[y(k) | x(k-1)]$. Оцінка $\hat{\mu}(x(k-1))$, обчислена за допомогою системи статистичного моделювання R (або S-Plus) близька до нуля, але не дорівнює нулю.

Зазначимо, що величина $\hat{\mu}(x(k-1))$ є додатною, коли значення $x(k-1)$ малі, і стає від'ємною, коли значення $x(k-1)$ великі. Це узгоджується з логікою розвитку фінансового процесу: коли доходність велика, то очікується її зниження, і навпаки, коли вона незначна, то очікується її зростання.

Код програми на R і S-plus для прикладу 8.1

```
? z1=read.table('w-3mtbs7097.txt', header=T)
? X=z1[4,1:1460]/100
? Y=(z1[4,2:1461]-z1[4,1:1460])/100
? par(mfcol=c(2,2))
```

```

[?] plot(x,y,pch='*',xlab='x(k-1)',ylab='y(k)')
[?] lines(lowess(x,y))
[?] title(main='(a) y(k) vs x(k-1)')
[?] fit=lowess(x,y)
[?] plot(fit$x, fit$y, xlab='x(k-1)', ylab='mu',type='l',
+ ylim=c(-.002, .002))
[?] title(main='(b) Estimate of mu(.)')
[?] plot(x,abs(y),pch='*',xlab='x(k-1)',ylab='abs(y)')
[?] lines(lowess(x,abs(y)))
[?] title(main='© abs(y) vs x(k-1)')
[?] fit2=lowess(x,abs(y))
[?] plot(fit2$x,fit2$y,type='l',xlab='x(k-1)',ylab='sigma',
+ ylim=c(0,01))
[?] title(main='(d) Estimate of sigma(.)')

```

Дослідження процесу також свідчить, що збільшення доходності приводить до зростання волатильності $|y(k)|$, тобто волатильність є зростаючою функцією від $x(k-1)$. Таким чином, цей приклад дослідження фактичного фінансового процесу демонструє корисність використання простої моделі для аналізу та розуміння динаміки і структури фінансового часового ряду.

8.2 Нелінійна авторегресія, моделі у просторі станів та нейромережі

Нелінійна авторегресія з функціональними коефіцієнтами

Розвиток непараметричних методів дає можливість послабити параметричні обмеження при побудові нелінійних моделей. У деяких дослідженнях непараметричні методи використовують для виконання попереднього аналізу даних з метою отримання додаткової інформації для побудови параметричної нелінійної моделі. Такий підхід використано при побудові авторегресії з функціональними коефіцієнтами (ФАР або FAR – functional coefficient autoregressive model):

$$y(k) = f_1(Y(k-1))y(k-1) + \dots + f_p(Y(k-1))y(k-p) + \varepsilon(k), \quad (8.33)$$

де $Y(k-1) = [y(k-1), \dots, y(k-p)]^T$ – вектор затриманих у часі вимірів основної змінної $y(k)$. За необхідності, вектор $Y(k-1)$ може включати також інші пояснюючі змінні, наявні на момент часу $(k-1)$. За припущення, функції (параметри моделі) $f_i(\cdot)$ у рівнянні (8.2.1) неперервні і двічі диференційовані (майже напевно) стосовно їх аргументів.

Більшість простіших нелінійних моделей, що розглядались раніше, це окремі випадки моделі ФАР. Для оцінювання функціональних коефіцієнтів $f_i(\cdot)$ можна використовувати непараметричні методи – регресію на ядрах або лінійну локальну регресію. Це робиться досить просто особливо у випадку, коли $Y(k-1)$ – має низьку розмірність, наприклад, $Y(k-1)$ – це скаляр.

Нелінійна адитивна авторегресія

Застосування непараметричних методів моделювання до нелінійних вимірів (часових рядів) пов'язане з деякими труднощами, зокрема, це так зване «прокляття розмірності». Розглянемо узагальнену нелінійну авторегресійну модель такого вигляду:

$$y(k) = f(y(k-1), \dots, y(k-p)) + \varepsilon(k). \quad (8.34)$$

Пряме застосування непараметричних методів для оцінювання функції $f(\cdot)$ потребує виконання процедур P -вимірного згладжування, яке вимагає багато обчислювальних витрат при великих значеннях P та великій потужності ряду даних. Простим, але достатньо ефективним методом подолання цих труднощів є використання адитивної моделі, побудова якої пов'язана із згладжуванням низької розмірності.

Часовий ряд $y(k)$ можна описати нелінійною авторегресійною (НААР або NAAR) моделлю

$$y(k) = f_0(k) + \sum_{i=1}^p f_i(y(k-i)) + \varepsilon(k) \quad (8.35)$$

де функції $f_i(\cdot)$ – неперервні майже напевно. Оскільки кожна функція має один аргумент, то її можна оцінити непараметричними методами з використанням алгоритмів одновимірного згладжування і таким чином уникнути високої розмірності. Для оцінювання НААР моделей можна застосовувати ітераційні методи оцінювання, які дають можливість отримати непараметричні оцінки $f_i(\cdot)$ [17, 18]. Необхідно зазначити, що припущення стосовно адитивності накладає суттєві обмеження і його необхідно тестувати перед використанням.

Нелінійні моделі у просторі станів

Загальний вигляд нелінійної моделі у просторі станів такий:

$$\begin{aligned} \mathbf{x}(k) &= f(k, \mathbf{x}(k-1)) + \mathbf{w}(k), \\ \mathbf{z}(k) &= g(k, \mathbf{x}(k)) + \mathbf{v}(k), \end{aligned} \tag{8.36}$$

де $\mathbf{x}(k)$ – вектор стану досліджуваної системи; $f(k, \mathbf{x}(k-1))$ і $g(k, \mathbf{x}(k))$ – відомі функції, які можуть залежати від деяких невідомих параметрів; $\mathbf{w}(k)$ – вектор випадкових зовнішніх збурень з коваріаційною матрицею $\mathbf{Q}(k-l)$; вектор похибок (шумів) вимірів з коваріаційною матрицею $\mathbf{R}(k-l)$; при цьому послідовність значень $\{\mathbf{w}(k)\}$ не залежить від послідовності $\{\mathbf{v}(k)\}$.

Для формального опису еволюції нелінійної динамічної системи (8.2.3), (8.2.4) (тобто переходу із попереднього стану в наступний) необхідно мати повну (для всіх змінних стану) функцію умовного розподілу. Чисельне оцінювання цієї функції виконується за методом Монте-Карло для марковських ланцюгів (МКМЛ). Інші чисельні методи згладжування нелінійних часових рядів розглядаються у багатьох публікаціях, зокрема в роботі [28].

МКМЛ – потужний чисельний метод (інструментарій) для оцінювання параметрів нелінійних математичних моделей. Сьогодні розробляються та використовуються на практиці нові алгоритми на основі теорії МКМЛ, які

також потребують окремого поглибленого дослідження. Припущення стосовно того, що функції $f(k, \mathbf{x}(k-1))$ і $g(k, \mathbf{x}(k))$ мають бути відомими – некритичне, оскільки задача оцінювання структури цих функцій також може бути поставлена та розв’язана. Можливим методом розв’язання цієї задачі може бути попереднє застосування непараметричних методів, наприклад таких, що застосовуються при побудові ФАР і НААР, для ідентифікації функцій $f(k, \mathbf{x}(k-1))$ і $g(k, \mathbf{x}(k))$. Надалі можна зробити перехід до представлення моделі у просторі станів.

Нейронні мережі

Нейронні мережі – один із самих широко використовуваних методів моделювання процесів різної природи. За допомогою нейромережевих моделей найчастіше розв’язуються задачі прогнозування, розпізнавання образів та керування. Розглянемо побудову нейромережі з прямим розповсюдженням, на вхідний шар якої (у вигляді одного або більше нейронів – вузлів) подаються статистичні дані. Виходи вхідного шару подаються на входи наступного шару і так далі відбувається пряма передача даних (сигналів) на вихід мережі.

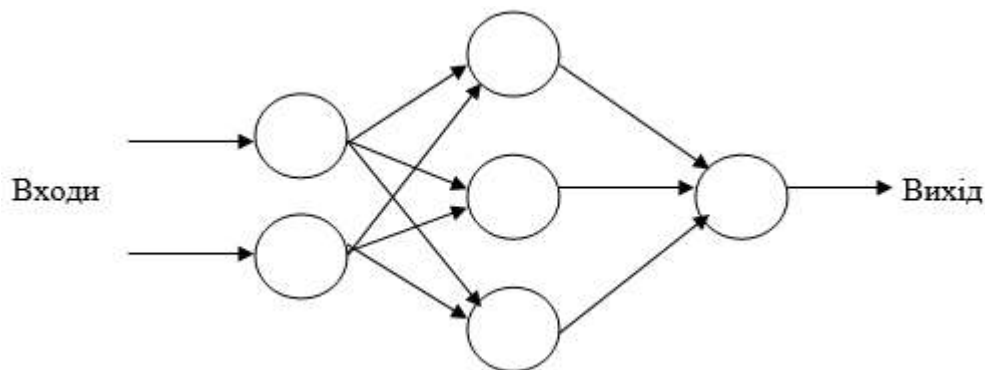


Рис. 8.1 Нейронна мережа з прямим розповсюдженням і одним прихованим шаром для одновимірного часового ряду

На рис. 8.1 подана проста нейронна мережа з прямим розповсюдженням з одним прихованим шаром. Вхідний шар складається з двох вузлів, а прихований – з трьох. Вихідний шар має один вузол. Це мережа 2-3-1 з прямим розповсюдженням. Сьогодні відомі складні каскадні нейромережеві структури із зворотними зв’язками, які описують нелінійні нестационарні багатовимірні процеси (об’єкти) і використовуються для розв’язання задач прогнозування, керування і класифікації.

Обробка даних мережею і передача даних між шарами відбувається з

використанням функцій активації. Розглянемо функціонування мережі з прямим розповсюдженням, яка має один прихований шар. У прихованому шарі j -й вузол формально описується таким виразом:

$$h_j = f_j \left(\alpha_{0j} + \sum_{i \rightarrow j} w_{ij} x_i \right), \quad (8.37)$$

де x_i – значення i -го вхідного вузла; $f_j(\cdot)$ – функція активації, за яку часто вибирають логістичну функцію такого виду:

$$f_j(z) = \frac{\exp(z)}{1 + \exp(z)}; \quad (8.38)$$

α_{0j} – стала величина – зміщення; обчислення суми по значеннях індексів $i \rightarrow j$ означає, що ця сума визначається по всіх вхідних вузлах, які зв'язані з j -м вузлом; w_{ij} – вагові коефіцієнти. Як приклад, запишемо вихідну реакцію j -го вузла мережі 2.3.1, поданої на рис. 8.1:

$$h_j = \frac{\exp(\alpha_{0j} + w_{1j}x_1 + w_{2j}x_2)}{1 + \exp(\alpha_{0j} + w_{1j}x_1 + w_{2j}x_2)}, \quad j = 1, 2, 3 \quad (8.39)$$

Вузол вихідного шару формально описується так:

$$y = f_y \left(\alpha_{0y} + \sum_{j \rightarrow y} w_{jy} h_j \right), \quad (8.40)$$

де $f_0(\cdot)$ – функція активації, яка може бути задана лінійною або функцією

Хевісайда. Якщо $f_0(\cdot)$ лінійна, то реакція на виході визначається так:

$$y = \alpha_{0y} + \sum_{j=1}^m w_{jy} h_j, \quad (8.41)$$

де m – кількість нейронів у прихованому шарі. Функція Хевісайда визначається так:

$$f_y(z) = \begin{cases} 1, & \text{якщо } z > 0, \\ 0, & \text{інакше.} \end{cases} \quad (8.42)$$

Нейрони з функціями Хевісайда називають пороговими. Для нейромережі, поданої на рис. 8.1, можна записати:

$$y = \alpha_{0y} + w_{1y} h_1 + w_{2y} h_2 + w_{3y} h_3, \quad (8.43)$$

і для лінійної функції активації отримаємо:

$$y = \begin{cases} 1, & \text{якщо } \alpha_{0y} + w_{1y} h_1 + w_{2y} h_2 + w_{3y} h_3 > 0, \\ 0, & \text{якщо } \alpha_{0y} + w_{1y} h_1 + w_{2y} h_2 + w_{3y} h_3 \leq 0. \end{cases} \quad (8.44)$$

Об'єднуючи реакції шарів нейромережі, поданої на рис. 8.1, отримаємо такий вираз:

$$y = f_y \left[\alpha_{0y} + \sum_{j \rightarrow y} w_{jy} f_j \left(\alpha_{0j} + \sum_{i \rightarrow j} w_{ij} x_i \right) \right] \quad (8.45)$$

Якщо врахувати можливий зв'язок між вхідним та вихідним шарами, то отримаємо:

Перша сума визначається для вхідних вузлів. Якщо функція активації вихідного вузла лінійна, то прямі зв'язки між вхідними вузлами і вихідним представляють відповідну лінійну функцію. Таким чином, у даному конкретному випадку модель (8.2.8) є узагальненням лінійних моделей.

У випадку використання лінійної функції активації на виході мережі, поданої на рис. 8.1, рівняння (8.2.7) приймає такий вигляд:

$$y = \alpha_{0y} + \sum_{j=1}^3 w_{jy} h_j \quad (8.47)$$

де h_j визначається рівнянням (8.2.6). Ця мережа має 13 параметрів. Якщо ж використовується рівняння (8.2.8), то вихід мережі визначається таким рівнянням:

$$y = \alpha_{0y} + \sum_{i=1}^2 \alpha_{iy} x_i + \sum_{j=1}^3 w_{jy} h_j \quad (8.48)$$

де h_j знову визначається рівнянням (8.2.6), а кількість параметрів збільшується до 15.

Функції, задані рівняннями (8.2.7) і (8.2.8), називають напівпараметричними, оскільки їх функціональна форма (структура) відома, але кількість вузлів, їх зміщення та вагові коефіцієнти невідомі. Наявність прямого зв'язку між вхідним та вихідним шарами у рівнянні (8.2.8) свідчить про те, що мережа може пропустити прихований шар. Таку мережу називають ще мережею прямого розповсюдження з пропущеним шаром.

Мережі з прямим розповсюдженням в літературі з нейронних мереж називають багат шаровими персептронами. Їх універсальна апроксимуюча властивість полягає у тому, що вони можуть апроксимувати будь-яку неперервну нелінійну функцію на компактних множинах шляхом збільшення кількості вузлів у прихованому шарі.

Навчання мережі і прогнозування

Перед навчанням нейромережі вибірка даних у формі часового ряду

розділяється на дві частини: навчальну і перевірочну. На навчальній вибірці оцінюються параметри мережі, а тестова використовується для перевірки якості оцінок прогнозів, обчислених за допомогою мережі.

Нехай для навчання використовується вибірка

$$\{(r(k), x(k)) | k=1, \dots, N\}, \quad (8.49)$$

де $x(k)$ – вектор вхідних значень (наприклад, ціна деякого активу), а $r(k)$ – часовий ряд доходності. Навчання мережі означає, що виконується пошук таких значень змішень (констант) і вагових коефіцієнтів, які мінімізують вибраний критерій якості, наприклад, квадратичний:

$$J = \sum_{k=1}^N [r(k) - y(k)]^2, \quad (8.50)$$

де $y(k)$ – вихід мережі. Це задача нелінійного оцінювання, яку можна розв'язати методами ітераційної оптимізації. Для того щоб гарантувати гладкість побудованої апроксимуючої функції, у постановку задачі оптимізації необхідно ввести додаткові обмеження. Популярним методом навчання є метод зворотного розповсюдження і алгоритм на його основі. Згідно з цим методом навчання починається з вихідного шару з подальшим використанням градієнтного правила, спрямованого на ітераційний пошук оцінок змішень та вагових коефіцієнтів. Після того як ці оцінки знайдені, можна використати нейромережу для обчислення оцінок прогнозів на навчальній та перевірочній вибірках.

Приклад 8.2. Для доходності вибраного активу (на основі 280 вимірів) побудована мережа 3-2-1, що описується виразом (8.2.8). За входи вибрано затримані у часі значення $r(k-1)$, $r(k-2)$ і $r(k-3)$. Використовуючи систему S-plus, обчислені зміщення та вагові коефіцієнти подані нижче:

$$\begin{aligned} r(k) = & 2,76 - 1,57 f_1[\mathbf{r}(k-1)] - 2,16 f_2[\mathbf{r}(k-1)] - \\ & - 0,085 r(k-1) - 0,047 r(k-2) - 0,113 r(k-3), \end{aligned} \quad (8.51)$$

де $\mathbf{r}(k-1) = [r(k-1), r(k-2), r(k-3)]$; дві логістичні функції визначаються формулами:

$$\begin{aligned} f_1[\mathbf{r}(k-1)] &= \frac{\exp[-7,32 - 15,94r(k-1) + 3,08r(k-2) - 16,79r(k-3)]}{1 + \exp[-7,32 - 15,94r(k-1) + 3,08r(k-2) - 16,79r(k-3)]}, \\ f_2[\mathbf{r}(k-1)] &= \frac{\exp[34,85 - 18,36r(k-1) - 13,83r(k-2) - 6,92r(k-3)]}{1 + \exp[34,85 - 18,36r(k-1) - 13,83r(k-2) - 6,92r(k-3)]}. \end{aligned} \quad (8.52)$$

Стандартне відхилення залишків моделі (8.2.9) склала 4,58. Для порівняння на основі тих же даних побудована модель авторегресії першого порядку:

$$r(k) = 1,13 + 0,152r(k-1) + \hat{\varepsilon}(k). \quad (8.53)$$

Стандартна похибка залишків цієї моделі склала 5,31. Порівняння одно крокових прогнозів на навчальній вибірці показало, середня абсолютна похибка у процентах склала 3,8% для нейромережі (8.2.9) і 4,58% для авто регресійної моделі (8.2.10), тобто нейромережа з прямим розповсюдженням дала можливість покращити оцінки однокрокових прогнозів на 0,78%.

Запитання і вправи до восьмого розділу

1. Що таке умовна гетероскедастичність? Чим вона відрізняється від звичайної зміни дисперсії?
2. Поясніть ідею моделі ARCH. У чому полягає її головна відмінність від класичної AR-моделі?
3. Які обмеження існують для параметрів моделі ARCH(q), щоб дисперсія залишалася додатною?
4. Як модель GARCH узагальнює ARCH-модель? Які додаткові можливості вона дає?
5. Які умови необхідні для стаціонарності процесу в моделі GARCH(1,1)?
6. Які методи оцінювання параметрів застосовуються для ARCH та GARCH моделей?

7. Які економічні та фінансові процеси найчастіше описуються за допомогою GARCH-моделей?
8. Що таке ефект «волатильність кластеризується» і як він відображається у часових рядах?
9. У чому полягає суть моделей EGARCH та TGARCH? Які явища вони дозволяють врахувати, на відміну від GARCH?
10. Як можна перевірити наявність ARCH-ефекту у часовому ряді? Які статистичні тести використовують?
11. Для фінансового ряду (наприклад, щоденні зміни курсу валют або індексу S&P 500) перевірте наявність ARCH-ефекту за допомогою тесту Лагранжа (ARCH LM-test). Зробіть висновки.
12. Оцініть параметри моделі GARCH(1,1) для вибраного фінансового ряду. Проінтерпретуйте отримані значення коефіцієнтів.
13. Побудуйте прогноз дисперсії для ряду з використанням ARCH(1) та GARCH(1,1). Порівняйте результати і зробіть висновки щодо точності та поведінки моделей.
14. Використайте модель EGARCH для ряду фондових індексів. Перевірте, чи відображає вона асиметрію у реакції дисперсії на позитивні та негативні шоки.

РОЗДІЛ 9. АНАЛІЗ НЕСТАЦІОНАРНИХ ТА НЕЛІНІЙНИХ ПРОЦЕСІВ БАЙЄСІВСЬКИМИ МЕТОДАМИ

9.1 Побудова адаптивних байєсівських моделей

Оскільки фінансові процеси відносяться до класу нелінійних та нестационарних, то структура і параметри статистичних даних, які використовуються для формалізованого опису та оцінювання ризиків можливих втрат, як правило, змінюються у часі. Це підтверджується відповідним статистичним тестуванням наявних даних. Так, наприклад, наявність нелінійного тренду процесу (тренд другого або вищого порядку) або наявність гетероскедастичності переводять досліджуваний процес до класу нелінійних та нестационарних. Тому у процесі використання побудованих математичних моделей для формального опису та оцінювання ризиків можливих втрат виникає завдання адаптації створених моделей до нових даних. Це стосується всіх типів моделей, що використовуються у роботі: узагальнених лінійних моделей, байєсівських мереж, звичайної регресії та інших можливих типів моделей, що використовуються для математичного опису досліджуваних фінансових процесів [2, 12, 51].

Задача адаптації відноситься до структури і параметрів моделі, що вже використовується, тобто, як правило, це структурно-параметрична адаптація, яка може виконуватись за допомогою різних обчислювальних процедур та критеріїв якості для моделей різних типів. У загальному випадку це одна із задач системного підходу до побудови систем підтримки прийняття рішень (СППР) для моделювання, прогнозування розвитку фінансових процесів та оцінювання ризиків можливих втрат. Серед інших задач системного підходу можна вказати на такі: створення ієрархічної багаторівневої СППР; підготовка даних до моделювання із врахуванням можливих невизначеностей даних структурного, параметричного і статистичного характеру; оцінювання та оптимізація структури і параметрів моделей, що будуються у системі; використання окремих множин критеріїв якості на кожному етапі виконання аналізу даних (критерії якості даних, критерії адекватності моделей, якості оцінок прогнозів та якості альтернативних рішень, що приймаються/підтримуються); функціональна повнота системи підтримки прийняття рішень та інші. Використання окремих множин критеріїв якості на кожному етапі виконання обчислень стосовно аналізу даних дає можливість підвищити якість

проміжних та остаточних обчислень, що програмуються у системі. В цілому, це гарантує високу якість остаточних результатів, які можна отримати за допомогою створеної системи [2, 51].

Незважаючи на велику кількість систем підтримки прийняття рішень, які створені і пропонуються на ринку комерційними підприємствами з розвитку інформаційних технологій для аналізу ризиків, як правило, далеко не всі, необхідні для високоякісного менеджменту ризиків, задачі розв'язуються завдяки використанню цих систем. Одна з причин виникнення такої ситуації полягає у відсутності фахівців необхідної кваліфікації, що особливо стосується аналізу складних сучасних фінансових процесів, які потребують створення інтелектуальних СППР та нових адекватних моделей за їх допомогою, ефективних моделей і методів оцінювання можливих фінансових втрат і генерування альтернативних рішень (альтернатив). Оскільки аналіз фінансових ризиків – це задача високої актуальності, то такі задачі розв'язуються численними фахівцями з теорії оцінювання, математичного моделювання, прогнозування та інформаційних технологій практично у всьому світі.

Далі розглянемо можливий варіант розв'язання задачі адаптації формалізованого опису процесів у формі узагальнених лінійних моделей (УЛМ) і байєсівської мережі до нових даних на прикладі динамічної мережі, яка є прийнятною для аналізу динамічних фінансових даних. Структурна адаптація УЛМ (лінійних та нелінійних) передбачає аналіз та врахування таких елементів моделей: кількість рівнянь, що використовуються для створення моделі; наявність нелінійності у процесі (що підтверджується відповідними тестами); кількість незалежних змінних (регресорів); випадкове збурення та його тип (тип розподілу ймовірностей для цього випадкового процесу); порядок авто регресійної частини моделі (за умови її наявності); порядок ковзного середнього (за умови його наявності); порядок інтегрованості процесу (за умови наявності тренду); можливі обмеження на значення змінних та параметрів. Всі вказані елементи структури моделей досить просто виявляються за допомогою статистичного тестування даних з використанням відомих обчислювальних процедур.

Після корегування структури моделі необхідно виконувати повторне оцінювання параметрів з коректним використанням відповідних методів параметричної ідентифікації (оцінювання параметрів). До цих методів можна віднести такі: звичайний метод найменших квадратів (МНК); нелінійний МНК (НМНК); метод максимальної правдоподібності (ММП); метод допоміжної змінної (МДП); метод Монте-Карло для марківських ланцюгів (МКМЛ), який

здобув популярність завдяки відносній універсальності його застосування до оцінювання параметрів лінійних та нелінійних процесів. Підбір/модифікація методу для оцінювання параметрів моделі ґрунтується на аналізі типу розподілу даних та виявлення наявності не лінійності та встановлення її типу.

Структурна адаптація байєсівських мереж (БМ) передбачає аналіз та корегування таких елементів цих ймовірнісно-статистичних моделей: додавання/скорочення кількості змінних (вузлів); додавання/видалення кількості дуг, що з'єднують вузли мережі; корегування типів локальних розподілів ймовірностей, що відносяться до окремих вузлів [2, 3]. Адаптація структури мережі відбувається з використанням необхідних критеріїв якості мережі (моделі), які надають інформацію про те, що відбувається з моделлю при виконанні тієї чи іншої дії, згаданої вище. Наприклад, при додаванні/видаленні змінних або дуг може відбуватись збільшення/зменшення загальної ймовірності мережі (як спільного багатовимірного розподілу використаних змінних), що підтверджується відповідним вибраним критерієм якості моделі, який використовується у методі адаптації.

Параметрична адаптація байєсівської мережі передбачає корегування значень умовних ймовірностей у відповідних таблицях (таблиці умовних ймовірностей або ТУЙ), значення яких можуть змінюватись в результаті надходження нових даних та/або попереднього корегування структури моделі [3, 30]. Значення у таблицях безумовних ймовірностей, що визначаються батьківськими вузлами, також можуть корегуватись з появою нових батьківських вузлів та зміни границь їх допустимих значень. Необхідно зазначити, що загалом структурно-параметрична адаптація байєсівських мереж – це досить складна задача, яка потребує використання спеціальних обчислювальних процедур і контролю коректності їх виконання. Всі ці процедури спрямовуються на досягнення необхідної адекватності моделей (це стосується також УЛМ та інших типів моделей), що будуються для формального опису вибраних фінансових процесів. Висока ступінь адекватності моделей забезпечує високу якість оцінювання необхідних прогнозів та ризиків можливих фінансових втрат.

9.2 Адаптація байєсівської мережі до даних

У подальшому аналізі будемо розглядати дискретні змінні, які є зручнішими з точки зору простоти представлення розподілів та менших обчислювальних витрат. Такий підхід дає можливість спростити, також,

процедури обчислення ймовірнісного висновку – отримання остаточного результату за мережею. Розглянемо розподіли дискретних випадкових величин, у яких кожна змінна X_i може приймати значення із скінченної множини спостережень $Val(X_i)$. Позначимо розмірність $Val(X_i)$ через $\|X_i\|$. Надалі будемо використовувати заголовні літери, наприклад, X, Y, Z для позначення імен змінних і малі літери x, y, z для позначення конкретних значень, що приймаються цими змінними. Множини змінних будемо позначати напівжирними заголовними літерами $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, з відповідними значеннями, позначеними напівжирними малими літерами $\mathbf{x}, \mathbf{y}, \mathbf{z}$. Змінні мережі і спільні розподіли будемо позначати так: $\mathbf{X} = \{X_1, \dots, X_n\}$.

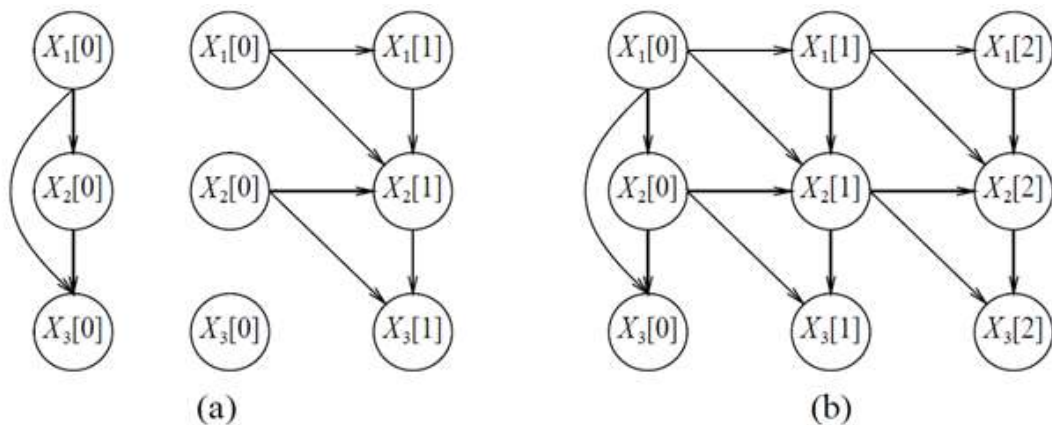


Рис. 9.1. Приклад створення ймовірнісної мережі

Ймовірнісна (байєсівська) мережа (ЙМ або БМ) – це орієнтований ациклічний граф, який відображає спільний розподіл ймовірностей для змінних $\mathbf{X}$. Формально ЙМ для $\mathbf{X}$ може бути представлена парою $B = (G, \Theta)$. Перший компонент цієї пари, G , представляє орієнтований ациклічний граф, вершини якого відповідають випадковим змінним $X_1, \dots, X_n$. Цей компонент кодує таку множину припущень стосовно умовної незалежності: кожна змінна X_i не залежить від його нащадків, враховуючи його батьків $\mathbf{Pa}(X_i)$ в графі G . Другий компонент, Θ , являє собою множину параметрів змінних і визначає

кількісно мережу. У простому випадку він містить параметри $\theta_{i,j_i,k_i} = \Pr(X_i = k_i | \mathbf{Pa}(X_i) = j_i)$ для кожного можливого значення k_i від X_i і кожного можливого набору значень j_i від $\mathbf{Pa}(X_i)$. Кожен умовний розподіл ймовірностей можна представити у вигляді таблиці, яка називається таблицею умовних ймовірностей (ТУЙ). Уявлення, які вимагають менше параметрів, наприклад таких як дерева, також можуть бути використані. Ми використовуємо їх надалі в розділі аналізу результатів експерименту, але для простоти позначень будемо дотримуватися згаданим позначенням ТУЙ.

Враховуючи G і Θ , ймовірнісна байєсівська мережа (ЙМ або БМ) B визначає унікальний спільний розподіл ймовірності для змінних $\mathbf{X}$, що визначається за виразом [19]:

$$P_B(x_1, \dots, x_n) = \prod_{i=1}^n P_B(x_i | \mathbf{pa}(X_i)). \quad (9.1)$$

ЙМ описує розподіл ймовірностей за допомогою фіксованої множини змінних. Динамічна ймовірнісна мережа (ДЙМ) розширює це подання (мережі) для формального опису (моделювання) досліджуваних процесів у часі. Для простоти будемо вважати, що зміни мають місце між дискретними моментами часу, які проіндексовані невід'ємними цілими числами. Тут ми припускаємо, що $\mathbf{X} = \{X_1, \dots, X_n\}$ являє собою множину атрибутів, що змінюють процес. При цьому $X_i[t]$ є випадковою величиною, що визначає значення атрибута X_i у момент часу t , а $\mathbf{X}[t]$ є множиною випадкових величин $X_i[t]$.

Для отримання моделі можливих траєкторій розвитку досліджуваного процесу, необхідно знайти розподіл ймовірностей для випадкових величин $\mathbf{X}[0] \cup \mathbf{X}[1] \cup \mathbf{X}[2] \cup \dots$. Зазвичай, такий спільний розподіл може бути досить складним. У поданій роботі припускаємо, що процес є *марківським* стосовно $\mathbf{X}$, тобто $P(\mathbf{X}[t+1] | \mathbf{X}[0], \dots, \mathbf{X}[t]) = P(\mathbf{X}[t+1] | \mathbf{X}[t])$. Припускаємо також, що процес, який моделюється, є стаціонарним стосовно ймовірностей переходів, тобто у даному випадку стаціонарність означає, що ймовірність переходу $P(\mathbf{X}[t+1] | \mathbf{X}[t])$ не залежить від часу t .

Виходячи з поданих припущень, ДЙМ представляє собою спільний розподіл для всіх можливих траєкторій еволюції процесу і складається з двох частин:

- апіорна структура мережі (моделі) B_0 , що визначає розподіл за відомими початковими станами $\mathbf{X}[0]$;
- мережа переходів $B_{\rightarrow}$ для значень змінних $\mathbf{X}[0] \cup \mathbf{X}[1]$, з метою визначення ймовірності переходу $P(\mathbf{X}[t+1]|\mathbf{X}[t])$ для всіх t .

Рисунок 9.1 (а) ілюструє простий приклад мережі. В мережах переходів (але не в апіорних мережах), змінні в $\mathbf{X}[0]$ не мають батьків. Ймовірність переходу формально можна представити таким виразом:

$$P_{B_{\rightarrow}}(\mathbf{x}[1]|\mathbf{x}[0]) = \prod_{i=1}^n P_{B_{\rightarrow}}(x_i|\mathbf{pa}(X_i[1])) \quad (9.2)$$

Зазначимо, що ДЙМ, яка визначається парою $(B_0, B_{\rightarrow})$, відповідає напів-нескінченній мережі для змінних $\mathbf{X}[0], \dots, \mathbf{X}[\infty]$. На практиці робимо висновок (знаходимо остаточний результат) тільки стосовно скінченного інтервалу $0, \dots, T$. Для того щоб мати можливість отримати цей результат, можемо умовно "розгорнути" структуру ДЙМ в ЙМ над змінними $\mathbf{X}[0], \dots, \mathbf{X}[T]$. Для секції даних 0, батьками $X_i[0]$ будуть ті, які вказані в попередній мережі B_0 ; в секції $t+1$, батьками $X_i[t+1]$ є ті ж вузли, в секціях t і $t+1$ відповідно до батьків $X_i[1]$ в $B_{\rightarrow}$. Тобто умовні розподіли можна скопіювати для цих змінних аналогічним чином. Рисунок 9.1 (б) показує результат розгортання мережі на рис. 9.1 (а) протягом трьох часових інтервалів. Враховуючи, що модель ДЙМ це спільний розподіл по $\mathbf{X}[0], \dots, \mathbf{X}[T]$, процес розгортання можна формально представити у такому вигляді [6, 22]:

$$P_B(\mathbf{x}[0], \dots, \mathbf{x}[T]) = P_{B_0}(\mathbf{x}[0]) \prod_{t=0}^{T-1} P_{B_{\rightarrow}}(\mathbf{x}[t+1]|\mathbf{x}[t]) \quad (9.3)$$

де $P_{B \rightarrow}(\mathbf{x}[t+1]|\mathbf{x}[T])$ можна отримати очевидним чином з моделі переходу.

9.3 Навчання мережі у випадку повних вибірок даних

У цьому підрозділі розглядаємо задачу навчання ДЙМ для випадку наявності *повних вибірок даних*. Результатом цього розділу є те, що навчання ДЙМ для випадку повних вибірок даних використовує ті ж методи, що і навчання ЙМ для випадку повних вибірок даних. Але фактично навчання ДЙМ не є повністю таким самим, що і застосування методів навчання статичної ЙМ у формі розгорнутої мережі у зв'язку з обмеженням на можливість створення повторної структури моделі і обчисленням відповідних параметрів цієї структури.

9.3.1 Навчання ймовірнісної мережі

Задача навчання ймовірнісної мережі формулюється таким чином. Використовуючи навчальну вибірку з D зразків (значень) $\mathbf{X}$, знайти мережу $B = (G, \Theta)$, яка найкращим чином відповідає D . Поняття "найкраща відповідність" визначається за допомогою функції підрахунку відповідних функцій оцінок. В літературі запропоновано кілька різних функцій оцінювання. Найбільш часто використовуються байєсівський інформаційний критерій оцінки і Bayesian-Dirichlet (БДе) критерій [31, 32]. Обидві оцінки поєднують ймовірність даних із структурою мережі, $L(B : D) = \log \Pr(D|B)$.

БІК і оцінки БДе є похідними від ймовірності отримання належної структури мережі. Нехай випадкова величина G змінюється над можливими варіантами структури мережі, які можна отримати за допомогою фактичних даних. Використовуючи правило Байєса, можна отримати такий розподіл для G :

$$\Pr(G|D) \propto \Pr(D|G) \Pr(G), \quad (9.4)$$

де ймовірність даних для цієї структури мережі може бути обчислена за відповідними параметрами розподілу змінних мережі:

$$\Pr(D|G) = \int \Pr(D|G, \Theta) \Pr(\Theta|G) d\Theta. \quad (9.5)$$

Очевидно, що оцінювання апіорних параметрів і обчислення цього інтегралу може бути у загальному випадку складною задачею. Одним з підходів до запобігання повного обчислення інтегралу (9.3.1) є аналіз асимптотичної поведінки цього виразу. Враховуючи можливу наявність великої кількості значень даних, можна показати, що наступна ймовірність є нечутливою до вибору апіорної ймовірності (за умови, що апіорна ймовірність є ненульовою в будь-якому випадку). У роботі [35, 36] отримано асимптотичну оцінку, зручну для аналізу апіорної ймовірності:

$$\log \Pr(D|G) = \log \Pr(D|G, \hat{\Theta}_G) - \frac{\log N}{2} \#G + O(1), \quad (9.6)$$

де $\hat{\Theta}_G$ – параметр налаштування для G , який максимізує ймовірність даних; N є кількістю навчальних зразків (значень); $\#G$ – розмірність G (яка в разі повної інформації є тільки кількістю параметрів); $O(1)$ – постійний член, який не залежить від N і G . Оцінка БІК використовує рівняння (3) для ранжирування кандидатів на структуру мережі, що усуває необхідність пошуку апіорних параметрів.

Альтернативний підхід до аналізу даних і параметрів полягає в оцінюванні виразу (9.3.1) в замкнутій формі. Наближено, апіорні ймовірності параметрів для цієї структури G передбачається розкласти на множники з використанням незалежних апіорних ймовірностей для кожного умовного розподілу $P(X_i | \mathbf{pa}(X_i))$. У випадку повних вибірок даних це означає, що апостеріорні фактори діють так само незалежно. Таким чином, кожен умовний розподіл може бути створено і оновлено/модифіковано окремо, що спрощує аналіз даних. Оцінка цих розподілів може бути обчислена в замкнутій формі, якщо додатково припустити, що апіорні ймовірності кожного умовного розподілу належать відомому спряженому сімейству, наприклад, широко використовуваному розподілу Діріхле. Деякі подробиці такого підходу подано нижче.

Оскільки існує можливість дослідити велику кількість можливих

структур мережі, а також уникнути необхідності призначати попередньо оцінений розподіл параметрів для кожної можливої структури, то необхідно забезпечити множину припущень, які дозволять визначати апіорні ймовірності для всіх структур за допомогою єдиної мережі на основі розподілу Діріхле, який дає впевненість стосовно коректного встановлення значень апіорних ймовірностей [38, 39]. Цей підхід має ту бажану властивість, що оцінки деяких двох мереж будуть еквівалентними, тобто однаковими стосовно формального опису з використанням однакових незалежних припущень.

9.3.2 Оцінки Байєса-Діріхле (БДе) для ймовірнісної мережі

Опишемо можливе оцінювання за критерієм БДе для динамічної ймовірнісної мережі. У цьому підрозділі припускається, що відома навчальна вибірка D , яка складається з N_{seq} повних (необхідних) послідовностей спостережень. Така l послідовність має довжину N_l і визначає значення для змінних $\mathbf{x}^l[0], \dots, \mathbf{x}^l[N_l]$. Така множина даних дає N_{seq} початкових зразків (вимірів), на яких можна навчити структуру B_0 , а також $N = \sum_l N_l$ випадків можливих переходів (структурних змін), на яких можна навчити модифіковану структуру моделі $B_{\rightarrow}$.

Введемо позначення:

$$\theta_{i,j_i,k_i}^{(0)} = \Pr(X_i[0] = k_i' \mathbf{Pa}(X_i[0]) = j_i')$$

$$\text{і так само для моментів часу } t=1, \dots, T; \quad (9.7)$$

$$\theta_{i,j_i,k_i}^{\rightarrow} = \Pr(X_t[t] = k_i' \mathbf{Pa}(X_i[t]) = j_i')$$

Також необхідно ввести позначення для достатньої статистики для кожної наявної множини даних:

$$N_{i,j_i,k_i}^{(0)} = \sum_l I(X_i[0] = k_i' \mathbf{Pa}(X_i[0]) = j_i', \mathbf{x}^l)$$

(9.8)

де $I(\cdot; \mathbf{x}^{\mathbb{N}})$ – індикаторна функція, яка приймає значення 1, якщо подія відбувається (має місце) в послідовності $\mathbf{x}^{\mathbb{N}}$, і дорівнює 0 – у протилежному випадку.

Використовуючи вираз (9.3.1) і перегруповання термів, знайдемо, що функція правдоподібності розкладається, згідно зі структурою ДЙМ, так само, як і у випадку статистичної ймовірнісної мережі [42]:

$$\Pr(D|G, \Theta_G) = \prod_i \prod_{j_i} \prod_{k_i} (\theta_{i,j_i,k_i}^{(0)})^{N_{i,j_i,k_i}^{(0)}} \times \prod_i \prod_{j_i} \prod_{k_i} (\theta_{i,j_i,k_i}^{\rightarrow})^{N_{i,j_i,k_i}^{\rightarrow}}, \quad (9.9)$$

а, отже, логарифмічну функцію правдоподібності можна задати таким чином [10]:

$$L(B:D) = \sum_i \sum_{j_i} \sum_{k_i} N_{i,j_i,k_i}^{(0)} \log \theta_{i,j_i,k_i}^{(0)} + \sum_i \sum_{j_i} \sum_{k_i} N_{i,j_i,k_i}^{\rightarrow} \log \theta_{i,j_i,k_i}^{\rightarrow} \quad (9.10)$$

Таке розкладання полегшує обчислення критерію БІК і оцінок БДє різними способами. По-перше, ймовірність виражається у вигляді суми термів, де кожен терм залежить тільки від умовної ймовірності змінної, враховуючи конкретні значення її батьківських змінних. Таким чином, якщо необхідно знайти для розподілу параметри максимальної правдоподібності, то їх можна максимізувати в межах кожної множини окремо. По-друге, розкладання означає, що можна навчити структуру B_0 незалежно від $B_{\rightarrow}$. Нарешті, можна навчити структуру $B_{\rightarrow}$, таким же чином, як і виконати навчання звичайної мережі на основі наявної множини вимірів.

Уточнимо ці аргументи. Використовуючи звичайну оцінку максимальної правдоподібності для поліноміальних розподілів, можна отримати такий вираз для кількості параметрів $\hat{\Theta}_G$ [41 – 43]:

$$\hat{\theta}_{i,j_i,k_i}^{(0)} = \frac{N_{i,j_i,k_i}^{(0)}}{\sum_{k_i} N_{i,j_i,k_i}^{(0)}}.$$

За аналогією можна записати вираз для випадку перехідної структури.

У випадку побудови таблиці умовних ймовірностей кількість параметрів мережі можна оцінити за виразом:

$$\begin{aligned} \#G &= \#G_0 + \#G_{\rightarrow}, \\ \text{де} \\ \#G_0 &= \sum_i \sum_{p \in \mathbf{Pa}(X_i[0])} \|X_p\| \times (\|X_i[0]\| - 1). \end{aligned} \quad (9.12)$$

За аналогією обчислюється кількість параметрів для перехідної структури.

Нарешті, підставляючи знайдені оцінки у вираз (9.3.2), знаходимо, що критерій БІК буде задаватися у даному випадку так:

$$\begin{aligned} BIC(G : D) &= BIC_0 + BIC_{\rightarrow}, \\ \text{де} \\ BIC_0 &= \sum_i \sum_{j_i'} \sum_{k_i'} N_{i,j_i',k_i'}^{(0)} \log \hat{\theta}_{i,j_i',k_i'}^{(0)} - \frac{\log N_{seq}}{2} \#G_0; \\ BIC_{\rightarrow} &= \sum_i \sum_{j_i} \sum_{k_i} N_{i,j_i,k_i}^{\rightarrow} \log \hat{\theta}_{i,j_i,k_i}^{\rightarrow} - \frac{\log N}{2} \#G_{\rightarrow}. \end{aligned} \quad (9.13)$$

Звернемо увагу, що погіршення значення критерію для множин у вихідний часовий інтервал є функцією від кількості послідовностей N_{seq} , оскільки погіршення оцінки для перехідної моделі є функцією загальної кількості вимірів N .

9.3.3 Оцінки Байєса-Діріхле (БДе) для динамічної ймовірнісної мережі

Нагадаємо, що для обчислення оцінки БДе, необхідно оцінити такий інтеграл:

$$\Pr(D|G) = \int \Pr(D|G, \Theta_G) \Pr(\Theta_G|G) d\Theta_G. \quad (9.14)$$

Перший член всередині інтеграла розкладається, як в рівнянні (9.3.3). Якщо припустити, що апостеріорна ймовірність по кожній умовній ймовірності не залежить від інших, то такий терм також має розклад у вигляді [41]:

$$\Pr(\Theta_G|G) = \prod_i \prod_{j_i'} \Pr(\theta_{i,j_i',k_i'}^{(0)}) \times \prod_i \prod_{j_i} \Pr(\theta_{i,j_i,k_i}^{\rightarrow}) \quad (9.15)$$

Підставляючи цей вираз у попереднє рівняння, бачимо, що цілісний вираз складається з інтеграла по добутку незалежних термів. Тому вираз $\Pr(D|G)$ можна записати у вигляді добутку двох інтегралів таким чином [41]:

$$\prod_i \prod_{j_i'} \int \prod_{k_i'} \theta_{i,j_i',k_i'}^{(0) N_{i,j_i',k_i'}^{(0)}} \times \Pr(\theta_{i,j_i',k_i'}^{(0)}) \times d\theta_{i,j_i',k_i'}^{(0)}, \quad (9.16)$$

і аналогічно можна поступити для випадку перехідної структури.

Для того щоб отримати рішення в замкнутій формі, приймаємо апіорні ймовірності по Діріхле. Для поліноміального розподілу Діріхле змінна X задається множиною гіперпараметрів $\{N'_x : x \in Val(X)\}$ таким чином:

$$\Pr(\Theta_X) = \text{Dirichlet}(\{N'_x : x \in Val(X)\}) \propto \prod_x \theta_x^{N'_x-1} \quad (9.17)$$

Інтуїтивно, гіперпараметри можна розглядати як «псевдо-відліки» (значення), оскільки вони відіграють аналогічну роль з фактичними відліками і впливають з наявних даних. При використанні апіорних ймовірностей по Діріхле, ймовірність спостереження послідовності значень X з кількістю $N(x)$ є такою [41]:

$$\int \prod_x \theta_x^{N(x)} \Pr(\Theta_X|G) d\Theta_X = \frac{\Gamma(\sum_x N'_x)}{\Gamma(\sum_x (N'_x + N(x)))} \times \prod_x \frac{\Gamma(N'_x + N(x))}{\Gamma(N'_x)}, \quad (9.18)$$

де $\Gamma(x) \int_0^\infty t^{x-1} e^{-t} dt$ – функція Гамма, що задовольняє властивостям $\Gamma(1)=1$ і $\Gamma(x+1)=x\Gamma(x)$.

Повертаючись до оцінювання структури за критерієм БДе, припустимо, що для кожної структури G , вибрано гіперпараметри $N_{i,j_i,k_i}' \rightarrow$. Тепер можна переписати $\Pr(D|G)$ у вигляді добутку двох термів таким чином [41]:

$$\prod_i \prod_{j_i} \frac{\Gamma(\sum_{k_i} N_{i,j_i,k_i}'^{(0)})}{\Gamma(\sum_{k_i} N_{i,j_i,k_i}'^{(0)} + N_{i,j_i,k_i}' \rightarrow)} \times \prod_{k_i} \frac{\Gamma(N_{i,j_i,k_i}'^{(0)} + N_{i,j_i,k_i}'^{(0)})}{\Gamma(N_{i,j_i,k_i}'^{(0)})} \quad (9.19)$$

Можна поступити аналогічно для випадку перехідної структури.

Цей підхід, як і раніше, вимагає підстановки гіперпараметрів Діріхле для кожного кандидата у модифіковану структуру ДЙМ. Оскільки фактична кількість можливих структур ДЙМ є досить великою, то практично отримати оцінки апіорних ймовірностей може виявитись важко визначити. Тому можна призначити всі апіорні ймовірності для ДЙМ $B' = (B'_{(0)}, B'_{\rightarrow})$ у вигляді двох еквівалентних вибірок $\hat{N}^{(0)}$ і $\hat{N}^{\rightarrow}$. Враховуючи ці компоненти, можна призначити вагові коефіцієнти для розподілу Діріхле, $G = (G_{(0)}, G_{\rightarrow})$, таким чином:

$$\begin{aligned} N_{i,j_i,k_i}'^{(0)} &= \hat{N}^{(0)} \times P_{B'_{(0)}}(X_i[0] = k_i' \mathbf{Pa}(X_i[0] = j_i')) \\ N_{i,j_i,k_i}'^{\rightarrow} &= \hat{N}^{\rightarrow} \times P_{B'_{\rightarrow}}(X_i[1] = k_i' \mathbf{Pa}(X_i[1] = j_i')) \end{aligned} \quad (9.20)$$

Вибір батьківських вузлів тут ґрунтується на структурі G , і може відрізнятися від структури B' . Інтуїтивно, можна розглянути міру довіри в B' як еквівалентну тому, що раніше використовувалась в послідовності $\hat{N}^{(0)}$ з $\hat{N}^{\rightarrow}$ переходами.

Досить просто можна перевірити, що при виборі апіорних ймовірностей

вказаним чином зберігається основна властивість моделі, яка вимагається у даному випадку: дві структури ДЙМ, які мають ідентичні припущення стосовно незалежності змінних, отримають однакові кількісні оцінки адекватності. Таким чином, можна стверджувати, що розглянуте визначення критерію є природним продовженням процесу оцінювання уточненої структури за допомогою критерію БДе у випадку аналізу статичних і динамічних ймовірнісних мереж.

9.4 Навчання на практиці

В обох видах оцінок структури мережі, які розглянуто вище, є дві важливі властивості. По-перше, оцінку ДЙМ можна записати у вигляді суми термів, (у випадку аналізу БДе розглядаємо логарифм $\Pr(D|G)$), де кожен член визначає вклад певного вибору (множини) батьків для конкретної змінної. Таким чином, локальна зміна однієї множини (наприклад, додавання або видалення дуги) впливає тільки на один з цих термів. По-друге, член, який оцінює $X_i[t]$ із врахуванням його батьків, є функцією відповідних відгуків (або $N^{(0)}(\cdot)$ або $N^{\rightarrow}(\cdot)$) для цієї множини. Таким чином, за рахунок кешування цих відгуків можна ефективно оцінювати вплив багатьох множин даних.

Ці дві властивості можуть бути використані у процедурах пошуку екстремуму використаного функціоналу, що здатні поступово покращити потенційну структуру, застосовуючи інформативно найкращі дуги додавання, видалення або повернення видалених елементів. У випадку аналізу динамічних ймовірнісних мереж, на відміну від статичних ймовірнісних мереж, потрібно враховувати додаткове обмеження, яке структура мережі повинна повторити протягом відносно тривалого часу (можливе часткове або повне повторення структури). Це дає можливість зменшити кількість варіантів структур у кожній точці простору структур конкретного типу. Крім того, можна знайти кращі інформативно початкові структури B_0 незалежно від пошуку можливого кращого варіанту для модифікованої структури $B_{\rightarrow}$.

9.4.1 Навчання для випадку наявності неповних вибірок даних

Розглянемо навчання динамічної ймовірнісної мережі у випадку

наявності неповних вибірок даних. Неповні вибірки даних мають вирішальне значення, оскільки в більшості практичних задач немає повних спостережень процесу, що моделюється. Це означає, що якщо досліджуваний процес – це стаціонарний марківський процес, то часткові спостереження, які є у наявності, можуть не бути марківськими. Припустимо, що відстежується рух автомобіля по шосе, і що спостерігається швидкість автомобіля, відносна відстань до інших автомобілів та інші параметри. Враховуючи вже наявні спостереження, можна стверджувати, що цей приклад не марківський. З іншого боку, марківська модель може бути коректна за умови, якщо включити інформацію про відомі цілі водія, наприклад, їхати по лівій смузі, виїхати на сусідню автомагістраль і т. ін. Вивчаючи приховані змінні, можна перехопити необхідні відомості про стан процесу. Це дозволяє моделі «пам'ятати» додаткову інформацію про минуле і краще передбачати майбутні стани.

Основними труднощами з навчанням від часткових спостережень є те, що більше немає властивостей розкладу оцінки (9.3.3). Це означає, що оптимальний вибір параметрів в одній частині мережі залежить від вибору параметрів в іншій частині мережі. Ця проблема стає зрозумілою краще, якщо відомо, що, як тільки у нас є випадок часткових спостережень, то більше не можна говорити про оптимальне оцінювання на навчальних даних. Для більшості подій, що представляють інтерес для навчання моделі, оцінки є достатньо невизначеними, оскільки не відомі точні значення змінних у наявних вибірках.

Для розв'язання цієї задачі досить часто використовується *ЕМ*-алгоритм [27, 31]. На *Е*-кроці *ЕМ*-алгоритму оцінюються і використовуються наближені параметри для повних вибірок даних, які отримують шляхом обчислення їх очікуваної кількості. На *М*-кроці знову оцінюються значення параметрів максимальної правдоподібності за припущення, що очікувана кількість була до істинних оцінок спостережень. Основне припущення, що лежить в основі поведінки *ЕМ*-алгоритму, є таким, що кожен цикл *ЕМ*-алгоритму гарантовано підвищить ймовірність появи даних, поданих (використаних) у моделі, поки не буде досягнє локального максимуму.

ЕМ-алгоритм традиційно розглядається як метод налаштування параметрів фіксованої структури моделі. Однак, основне припущення можна узагальнити для застосування до довільних структурних, а також параметричних модифікацій моделей (тобто до їх структурної і параметричної адаптації). Так, структурний *ЕМ*-алгоритм (СЕМ) Фрідмана [24] містить той же *Е*-крок, що і класичний *ЕМ*-алгоритм, заповнює дані шляхом обчислення

очікуваної кількості вимірів на основі поточної структури і параметрів. Крім того, для того щоб повторно оцінити параметри, на M -кроці структурний EM -алгоритм може використовувати очікувану кількість параметрів відповідно до наявної поточної структури для того щоб оцінити будь-яку іншу структуру-кандидата, при цьому, по суті, виконуючи структурний пошук повних вибірок даних у внутрішньому циклі виконання обчислень. Фрідман показує, що в результаті модифікована мережа повинна мати вищу оцінку, ніж в оригіналі. Це вірно, навіть у тому випадку, коли очікувана кількість параметрів, що використовується для оцінювання нової структури, обчислюється з використанням інформації про попередню структуру.

Результати Фрідмана можна розширити на випадок ДЙМ таким чином. Нехай $E[BIC((B'_0, B'_{\rightarrow}): D^+): D, (B_0, B_{\rightarrow})]$ є очікуваною оцінкою БІК для ДЙМ $(B'_0, B'_{\rightarrow})$ на основі всіх можливих доповнень D^+ у вигляді нових даних, тобто можна використати також значення можливих прихованих вимірів. Очікування береться по відношенню до $\Pr(D^+ | D, (B_{(0)}, B_{\rightarrow}))$, тобто ймовірність відноситься до цього поповнення на основі попередньої структури ДЙМ. Використовуючи [24], можна довести, що:

$$\begin{aligned} & BIC((B'_0, B'_{\rightarrow}): D) - BIC((B_0, B_{\rightarrow}): D) > \\ & E[BIC((B'_0, B'_{\rightarrow}): D^+): D, (B_0, B_{\rightarrow})] - \\ & E[BIC((B_0, B_{\rightarrow}): D^+): D, (B_0, B_{\rightarrow})] \end{aligned} \quad (9.21)$$

Тобто, якщо вибирається модифікована (нова) структура ДЙМ, яка має більш високі очікувані оцінки, ніж очікувані оцінки старої ДЙМ, то істинна оцінка нової (модифікованої) ДЙМ також буде вищою, ніж істинна оцінка старої ДЙМ. Крім того, різниця в очікуваних оцінках є тільки нижньою межею поліпшення в термінах фактичних оцінок, які оптимізуються; тобто залишається простір для подальшого покращення оцінок нової структури мережі завдяки наявним новим та прихованим вимірам.

Вирішальною властивістю очікуваних оцінок, отриманих за допомогою БІК є те, що вони теж розкладаються в суму місцевих термів таким чином, що в силу лінійності оператора очікування, можемо "спрямувати" оператор E через знаки обчислення суми неявних членів в (9.3.5) таким чином, щоб отримати

рівняння, яке включає в себе терми в такому вигляді:

$$E[N_{i,j_i,k_i}^{\rightarrow}] = \sum_{\mathbf{x}} \sum_t E[I(X_i[t] = k_i, \mathbf{Pa}(X_i[t]) = j_i; \mathbf{x})] = \sum_{\mathbf{x}} \sum \Pr(X_i[t] = k_i, \mathbf{Pa}(X_i[t]) = j_i | \mathbf{x}) \quad (9.22)$$

і можна за аналогією зробити це для членів $E[N_{i,j_i,k_i}^{(0)}]$. Вони називаються очікуваною достатньою статистикою для модифікації структури моделі. Тому ключовою вимогою для застосування структурного ЕМ-алгоритму (СЕМ), у випадку існування неповних даних, є можливість обчислити ймовірність всіх допустимих сімейств ймовірностей, $\Pr(X_i[t] = k_i, \mathbf{Pa}(X_i[t]) = j_i | \mathbf{x})$, для всіх мереж, які можна оцінити, тобто саму поточну мережну структуру і «сусідні» мережі у допустимому полі пошуку структур ДЙМ.

Для ефективного обчислення загальної ймовірності для конкретних сімейств ймовірностей, можна перетворити ДЙМ в об'єднані дерева і використати динамічний двопрхідний алгоритм [42] аналізу структур, схожий на алгоритм прямого-зворотнього ходу, який використовується при оцінюванні прихованих марківських моделей (ПММ). Для того щоб ефективно обчислити повну ймовірність множини вузлів, які ще не містяться у жодному вузлі об'єднаних дерев, необхідно використовувати складніші методи. Простіший підхід, який використовується сьогодні, повинен об'єднати разом всі змінні, що містяться в одному інтервалі, в одному вузлі (тобто для того щоб перетворити ДЙМ у ланцюг Маркова), обчислити очікувану кількість переходів між кожною парою послідовних станів в цьому ланцюгу, а потім ізолювати ці кількості станів для того щоб отримати кількості переходів для кожного сімейства змінних.

Також можна виконувати кешування (збереження) очікуваних кількостей структур, обчислених на основі поточної моделі поповнення, і використати надалі цей кеш для того щоб уникнути повторного обчислення очікуваних кількостей. Методика, яка використовується сьогодні, для обчислення кількості і значень очікуваних достатніх статистик, залишається чисельно складною. Таким чином, процедура застосування алгоритму СЕМ з використанням згенерованих або фактичних даних полягає у вигляді, поданому нижче.

Вибрати початкову структури $(B_0^0, B_{\rightarrow}^0)$ (можливо випадково).

Проаналізувати значення вимірів для $n = 0, 1, \dots$ на збіжність.

Відслідкувати поліпшення параметрів $(B_0^n, B_{\rightarrow}^n)$ за допомогою ЕМ.

Реалізувати пошук по можливих структурах ДЙМ (з використанням очікуваних оцінок, які обчислені за допомогою структур $(B_0^n, B_{\rightarrow}^n)$). Нехай кращою оцінкою для ДЙМ, при цьому, буде варіант $(B_0^{n+1}, B_{\rightarrow}^{n+1})$: якщо якість $(B_0^{n+1}, B_{\rightarrow}^{n+1}) = (B_0^n, B_{\rightarrow}^n)$, то необхідно повернутись до структури $(B_0^{n+1}, B_{\rightarrow}^{n+1})$.

Таблиця 9.1. Значення вибраних критеріїв для одного набору даних

Розмір даних		Критерій БІК			Критерій BDe		
N_{seq}	N	0	1	2	0	1	2
300	11387	-1.4527	-1.5198	-1.9842	-1.5588	-1.5138	-1.359
600	28043	-1.1638	-1.0443	-1.0369	-1.2762	-1.0235	-1.0458
1200	61864	-0.9883	-0.9977	-0.9885	-0.9672	-0.9587	-0.9869
1800	102906	-0.9623	-0.9746	-0.9889	-0.9637	-0.9618	-0.9913

Поданий вище результат отримано для випадку застосування алгоритму СЕМ з метою отримання оцінок за байєсівським інформаційним критерієм (БІК). Процедура навчання на основі алгоритму СЕМ також розширена на використання оцінок БДе, але у цьому випадку потрібно пам'ятати про нелінійність цих оцінок. Однак, це не відіграє критичної ролі при їх практичному використанні, тобто їх можна використовувати, пам'ятаючи, що виконується дослідження нелінійних нестационарних процесів, які мають свої особливості моделювання.

9.5 Байєсівська фільтрація нелінійних даних

9.5.1 Модель системи даних

Ще однією прикладною задачею, яка відноситься до підсистеми адаптації моделі, є фільтрація даних, тобто покращення їх якості перед моделюванням і прогнозуванням. Існують різні підходи та алгоритми фільтрації даних і серед

них байєсівські методи, які активно використовуються на практиці.

Прикладна нелінійна байєсівська фільтрація базується на нелінійній моделі в просторі станів у дискретному часі, які пов'язує прихований стан x_k з спостереженням z_k [44]:

$$\begin{aligned} x_k &= f(x_{k-1}, v_{k-1}) \\ x_k &= f(x_{k-1}, v_{k-1}), \quad v_{k-1} \sim p_{v_{k-1}}, \quad v_{k-1} \sim p_{v_{k-1}}, \quad x_1 \sim p_{x_1}, \quad x_1 \sim p_{x_1} \\ z_k &= h(x_k) + n_k \\ z_k &= h(x_k) + n_k, \quad n_k \sim p_{n_k} \end{aligned} \quad (9.23)$$

Тут k означає номер дискретного моменту часу, v_k – процес випадкового шуму з відомою щільністю розподілу p_{v_k} ; n_k – адитивний шум вимірів з відомою щільністю розподілу p_{n_k} . Перший вимір позначається як z_1 , тому перший невідомий стан це x_1 , він має щільність розподілу p_{x_1} . Модель може також залежати від відомого керуючого впливу u_k , що в даному випадку опущено. Запис $s_{1:k}$ означає послідовність $s_1, s_2, \dots, s_k$ (s є одним із сигналів x, v, z, n).

В статистичній літературі модель часто записується в термінах умовної щільності розподілу

$$\begin{aligned} x_k &\sim p(x_k | x_{k-1}) \\ z_k &\sim p(z_k | x_k) \end{aligned} \quad (9.24)$$

У деякому сенсі така модель є більш загальною стосовно досліджуваного процесу. Завданням методів гранулярної байєсівської фільтрації є апроксимація апостеріорної щільності розподілу невідомого стану системи (процесу) за умови наявних вимірів, яка позначається так: $p(x_k | z_{1:k})$.

2. Методи гранулярної фільтрації

Алгоритм послідовної вибірки за значимістю

Алгоритм послідовної вибірки за значимістю (Sequential Importance Sampling (SIS) Algorithm) є методом реалізації рекурсивного байєсівського фільтра з використанням симуляцій Монте-Карло [50]. Це базовий алгоритм гранулярної фільтрації. Основною ідеєю є представлення необхідної апостеріорної щільності розподілу як множини випадкових елементів з асоційованими вагами та обчислення оцінок на основі цих елементів та вагів. При збільшенні кількості елементів вибірки така Монте-Карло характеристика стає еквівалентною до звичайного функціонального опису апостеріорної щільності та SIS фільтр наближається до оптимальної байєсівської оцінки.

Нехай $\{x_{1:k}^i, w_k^i\}_{i=1}^{N_s}$ позначає випадкову міру, яка характеризує апостеріорну щільність $p(x_{1:k}|z_{1:k})$, де $\{x_{1:k}^i, i = 0, \dots, N_s\}$ – множина k -крокових траєкторій з опорних точок (гранул, частинок) з асоційованими нормованими вагами $\{w_k^i, i = 0, \dots, N_s\}$, $\sum_{i=1}^{N_s} w_k^i = 1$, N_s – задана кількість гранул. Тоді апостеріорна щільність в момент часу k може бути апроксимована так :

$$p(x_{1:k}|z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_{1:k} - x_{1:k}^i) \quad (9.25)$$

де $\delta(x)$ – δ -функція Дірака. Тобто $p(x_{1:k}^i|z_{1:k}) \approx w_k^i$.

Ваги обираються за принципом *вибірки за значимістю*. Він полягає в наступному. Припустимо, що $p(x) \propto \pi(x)$ ($\propto$ означає пропорційність) – це щільність розподілу випадкової величини, реалізацію якої з її справжнього розподілу складно згенерувати, але для якої $\pi(x)$ можна

обчислити. Нехай також $x^i \sim q(x), i = 0, \dots, N_s$ $x^i \sim q(x), i = 0, \dots, N_s$ – реалізації випадкової величини, які легко генеруються з розподілу зі щільністю $q(\cdot)q(\cdot)$, яка називається *пропонууючою щільністю* або значимою щільністю (*proposal density, importance density*). Тоді зважена апроксимація щільності $p(\cdot)p(\cdot)$ виглядає таким чином:

$$w^i \propto \frac{\pi(x^i)}{q(x^i)} \quad (9.26)$$

є нормованою вагою i -ої гранули (після обчислення відношення проводиться нормування вагових коефіцієнтів: $\sum_{i=1}^{N_s} w_k^i = 1, \sum_{i=1}^{N_s} w_k^i = 1$).

Тоді, якщо реалізації $x_{1:k}^i, x_{1:k}^i$ були згенеровані з розподілу з пропонууючою щільністю $q(x_{1:k} | z_{1:k})q(x_{1:k} | z_{1:k})$, то ваги в (9.3.2) відповідно до (9.3.3) будуть такими:

$$w^i \propto \frac{p(x_{1:k} | z_{1:k})}{q(x_{1:k} | z_{1:k})} \quad (9.27)$$

У випадку послідовних обчислень, на кожній ітерації маючи зважену вибірку, яка апроксимує $p(x_{1:k-1} | z_{1:k-1})p(x_{1:k-1} | z_{1:k-1})$, можна було б апроксимувати $p(x_{1:k} | z_{1:k})p(x_{1:k} | z_{1:k})$ новою вибіркою. Але якщо пропонууюча щільність обрана так, що вона розкладається на множники:

$$q(x_{1:k} | z_{1:k}) = q(x_k | x_{1:k-1}, z_{1:k})q(x_{1:k-1} | z_{1:k-1}) \quad (9.28)$$

то можна отримати елементи $x_{1:k}^i \sim q(x_{1:k} | z_{1:k})$ $x_{1:k}^i \sim q(x_{1:k} | z_{1:k})$ доповнюючи кожен з існуючих елементів $x_{1:k-1}^i \sim q(x_{1:k-1} | z_{1:k-1})$

$$x_{1:k-1}^i \sim q(x_{1:k-1}|z_{1:k-1}) \quad \text{новим станом} \quad x_k^i \sim q(x_k|x_{1:k-1}, z_{1:k})$$

$$x_k^i \sim q(x_k|x_{1:k-1}, z_{1:k}).$$

У практичних застосуваннях найчастіше вимагається лише фільтрована оцінка розподілу $p(x_k|z_{1:k})p(x_k|z_{1:k})$ на кожному кроці, а не умовний розподіл одразу всієї траєкторії $p(x_{1:k}|z_{1:k})p(x_{1:k}|z_{1:k})$. Тому надалі розглядати саме цей випадок.

Використовуючи прямий наслідок з теореми Байєса, отримаємо:

$$p(x_k|z_{1:k}) = \frac{p(z_k|x_k)p(x_k|z_{1:k-1})}{p(z_k|z_{1:k-1})} \quad (9.29)$$

за виконання певних умов, зокрема, якщо $q(x_k|x_{1:k-1}, z_{1:k})$ $q(x_k|x_{1:k-1}, z_{1:k}) = q(x_k|x_{k-1}, z_k)q(x_k|x_{k-1}, z_k)$, тобто пропонуюча щільність залежить лише від x_{k-1} та z_k , можна показати [32], що для оновлення вагових коефіцієнтів використовується таке співвідношення:

$$w_k^i \propto w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, z_k)} \quad (9.30)$$

і фільтрований апостеріорний розподіл може бути апроксимований так:

$$p(x_k|z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(x_k - x_k^i) \quad (9.31)$$

Ще раз необхідно наголосити, що вагові коефіцієнти w_k^i мають бути нормовані щоб забезпечити виконання умови: $\sum_{i=1}^{N_s} w_k^i = 1$. Це гарантує коректність виконання обчислень. Вибір пропонуючого розподілу є одним із найбільш важливих етапів проектування гранулярного фільтра.

Можливі варіанти вибору, їх переваги та недоліки розглянуті в літературі. Зокрема важливо, щоб дисперсія вагів $w_k^i w_k^i$ була невеликою.

Найчастіше в якості пропонуемого розподілу використовують апіорний з (9.5.2a):

$$q(x_k | x_{k-1}^i, z_k) = p(x_k | x_{k-1}^i), \quad (9.32)$$

що є зручним. В такому разі (6) спрощується до вигляду:

$$w_k^i \propto w_{k-1}^i p(z_k | x_k^i). \quad (9.33)$$

Але *не в усіх задачах* такий вибір пропонуемого розподілу однозначно оптимальний.

9.5.3 Основний алгоритм фільтрації

Підсумуємо алгоритм послідовної вибірки за значимістю. Елементи $x_{1:1}^i$ у зваженій вибірці на 1-му кроці $\left\{x_{1:1}^i, \frac{1}{N_s}\right\}_{i=1}^{N_s}$ генеруються з початкового розподілу $p_{x_1} p_{x_1}$ (9.5.1a). Оскільки цей розподіл істинний, то корекції не потрібно і всі ваги рівні $w_1^i = \frac{1}{N_s} w_1^i = \frac{1}{N_s}$.

Процедуру генерації зваженої вибірки на k -му кроці, маючи зважену вибірку на $(k-1)$ -му кроці, можна представити наступним псевдокодом:

Алгоритм 9.1: SIS Particle Filter

$$\left[\{x_k^i, w_k^i\}_{i=1}^{N_s}\right] \left[\{x_k^i, w_k^i\}_{i=1}^{N_s}\right] = \text{SIS}\left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k\right] \left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k\right]$$

FOR

— Згенерувати $x_k^i \sim q(x_k | x_{k-1}^i, z_k) x_k^i \sim q(x_k | x_{k-1}^i, z_k)$

— Присвоїти гранулі x_k^i вагу w_k^i відповідно до (9.5.6)

END FOR

Для цього та всіх подальших алгоритмів фільтрації на кожному кроці апостеріорний розподіл можна апроксимувати за формулою (9.3.5.). Оцінкою умовного математичного сподівання стану x_k буде:

$$\widehat{x}_k = \sum_{i=1}^{N_s} w_k^i x_k^i \quad (9.34)$$

Відсів або виключення гранул

Частою проблемою SIS гранулярного фільтра є виродження вагових коефіцієнтів, коли після певної кількості ітерацій усі гранули, крім однієї, мають незначну вагу. Було показано, що дисперсія вагів може лише збільшуватись з часом, тому феномену виродження уникнути неможливо. Виродження спричиняє те, що багато обчислень витрачається на оновлення гранул, внесок яких у апроксимацію $p(x_k|z_{1:k})p(x_k|z_{1:k})$ майже нульовий.

Одним із способів зменшення ефекту виродження є проведення *відсіву* (resampling). Основною ідеєю відсіву є виключення гранул, що мають малу вагу та концентрація на гранулах з великою вагою. На кроці відсіву відбувається генерація нової множини $\{x_k^{i*}\}_{i=1}^{N_s}$ вибором (з повтореннями) N_s реалізацій випадкової величини x_k з наближеного дискретного представлення розподілу $p(x_k|z_{1:k})p(x_k|z_{1:k})$, заданого формулою (9.5.7), тобто таким чином, що $P\{x_k^{i*} = x_k^j\} = w_k^j P\{x_k^{i*} = x_k^j\} = w_k^j$. Отримана вибірка є насправді вибіркою незалежних однаково розподілених (i.i.d.) величин з дискретного розподілу (9.5.7), тому ваги тепер встановлюються рівними, $w_k^i = \frac{1}{N_s} w_k^i = \frac{1}{N_s}$.

Наведемо псевдокод процедури відсіву. Такий варіант відсіву, *систематичний відсів* (systematic resampling), обраний через простоту реалізації та оцінку складності алгоритму $O(N_s)$. Для кожної частинки нової вибірки зберігається також її індекс у попередній вибірці, i^j , що знадобиться в наступних алгоритмах.

Алгоритм 9.2: Resampling Algorithm

$$\left[\{x_k^{j*}, w_k^j, i^j\}_{j=1}^{N_s} \right] \left[\{x_k^{j*}, w_k^j, i^j\}_{j=1}^{N_s} \right] = \text{RESAMPLE} \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right]$$

Ініціалізувати функцію розподілу (ФР): $c_1 = 0, c_1 = 0$

FOR $i = \overline{2, N_s} i = \overline{2, N_s}$

— Побудувати ФР: $c_i = c_{i-1} + w_k^i c_i = c_{i-1} + w_k^i$

END FOR

Почати з початку ФР: $i = 1, i = 1$

Згенерувати початкову точку: $u_1 \sim U[0, N_s^{-1}] u_1 \sim U[0, N_s^{-1}]$

FOR $j = \overline{1, N_s} j = \overline{1, N_s}$

— Рухатись уздовж ФР: $u_j = u_1 + N_s^{-1}(j-1) u_j = u_1 + N_s^{-1}(j-1)$

— WHILE $u_j > c_i u_j > c_i$

— — $i = i + 1, i = i + 1$

— END WHILE

— Присвоїти елемент: $x_k^{j*} = x_k^i x_k^{j*} = x_k^i$

— Присвоїти вагу: $w_k^j = N_s^{-1} w_k^j = N_s^{-1}$

— Присвоїти батька: $i^j = i, i^j = i$

END FOR

Процедура відсіву має деякі недоліки. По-перше, вона звужує можливості паралелізму, оскільки на кроці відсіву всі гранули мають комбінуватися. По-друге, гранули з великою вагою будуть обиратися багаторазово, що може призвести до втрати різноманітності вибірки, результуюча вибірка міститиме багато повторів. Ця проблема, відома як *збідніння вибірки* (sample impoverishment) є особливо гострою у випадку малого шуму процесу. Тоді всі гранули можуть зліпитися в одну через кілька ітерацій.

9.5.4 Фільтр вибірки за значимістю з відсівом

Фільтр вибірки за значимістю з відсівом (Sampling Importance Resampling

Filter (SIR)) – це метод Монте-Карло, який може бути застосованим до задач рекурсивної байєсівської фільтрації. Обмеження, накладені на його використання, є дуже слабкими. Функції $f(\cdot, \cdot)$ та $h(\cdot, \cdot)$ в (9.5.1a) і (9.5.1b) мають бути відомими, необхідно мати змогу генерувати реалізації з розподілу шуму $p_{v_{k-1}}$ та апіорного розподілу $p(x_k|x_{k-1})$ і визначати значення щільності розподілу $p(z_k|x_k)$ в потрібних точках, принаймні з точністю до спільної константи.

SIR алгоритм може бути легко виведений з алгоритму SIS за рахунок відповідного вибору:

- 1) пропонуемого розподілу $q(x_k|x_{k-1}^i, z_k)$ – в якості нього обирається апіорна щільність $p(x_k|x_{k-1}^i)$;
- 2) кроку відсіву, який здійснюється у кожний момент часу.

Такий вибір пропонуемого розподілу обумовлює необхідність генерування реалізацій з $p(x_k|x_{k-1}^i)$. Реалізацію $x_k^i \sim p(x_k|x_{k-1}^i)$ можна отримати, якщо згенерувати реалізацію шуму $v_{k-1}^i \sim p_{v_{k-1}}$ та покласти $x_k^i = f(x_{k-1}^i, v_{k-1}^i)$.

Для такого спеціального вибору пропонуємої щільності формула оновлення вагів набуває вигляду (9.5.9). Але, взявши до уваги, що відсів відбувається в кожний момент часу, отримуємо, що $w_{k-1}^i = 1/N_s \forall i$, а тому:

$$w_k^i \propto p(z_k|x_k^i) \quad (9.35)$$

Ваги задані в (3.5.11) нормуються перед фазою відсіву. Подамо псевдокод ітерації цього алгоритму:

Алгоритм 9.3. Particle Filter

$$\left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] = \text{SIR} \left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k \right] \left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k \right]$$

FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— Згенерувати $x_k^i \sim p(x_k | x_{k-1}^i) x_k^i \sim p(x_k | x_{k-1}^i)$

— Обчислити $w_k^i = p(z_k | x_k^i) w_k^i = p(z_k | x_k^i)$

END FOR

Обчислити загальну вагу: $t = \sum_{i=1}^{N_s} w_k^i t = \sum_{i=1}^{N_s} w_k^i$

FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— Нормувати i -ту вагу: $w_k^i = t^{-1} w_k^i w_k^i = t^{-1} w_k^i$

END FOR

Провести відсів, використовуючи алгоритм 9.2 (Resampling Algorithm):

$$\left[\{x_k^i, w_k^i, -\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i, -\}_{i=1}^{N_s} \right] = \text{RESAMPLE} \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right]$$

5. Допоміжний фільтр вибірки за значимістю з відсівом

Допоміжний фільтр вибірки за значимістю з відсівом (Auxiliary Sampling Importance Resampling Filter (ASIR)) є варіантом стандартного фільтра вибірки за значимістю з відсівом. Він може бути отриманий з SIR фільтра введенням пропонууючої щільності $q(x_k, i | z_{1:k}) q(x_k, i | z_{1:k})$, з якої генеруються реалізації пар $\{x_k^j, i^j\}_{j=1}^{N_s} \{x_k^j, i^j\}_{j=1}^{N_s}$, де $i^j i^j$ позначає номер гранули в $(k-1)(k-1)$ -ий момент. $i^j i^j$ в даному фільтрі називають допоміжною змінною, оскільки її єдина мета допомогти в задачі симуляції. Звідси назва фільтра.

Використовуючи теорему Байєса, можна показати, що

$$p(x_k, i | z_{1:k}) \propto p(z_k | x_k) p(x_k | x_{k-1}^i) w_{k-1}^i. \quad (9.36)$$

ASIR фільтр генерує реалізацію зі спільного розподілу $p(x_k, i | z_{1:k})$ $p(x_k, i | z_{1:k})$, а потім опускає номер ii в парі $(x_k, i)(x_k, i)$, щоб отримати вибірку

$\{x_k^j\}_{j=1}^{N_s} \{x_k^j\}_{j=1}^{N_s}$ з маргінального розподілу $p(x_k|z_{1:k})p(x_k|z_{1:k})$. Пропонуюча щільність для $\{x_k^j, i^j\}_{j=1}^{N_s} \{x_k^j, i^j\}_{j=1}^{N_s}$ має задовольняти пропорційність:

$$q(x_k, i|z_{1:k}) \propto p(z_k|\mu_k^i)p(x_k|x_{k-1}^i)w_{k-1}^i, \quad (9.37)$$

де $\mu_k^i \mu_k^i$ – певна характеристика $x_k x_k$ за умови $x_{k-1}^i x_{k-1}^i$. Наприклад, це може бути умовне математичне сподівання $\mu_k^i = E[x_k|x_{k-1}^i] \mu_k^i = E[x_k|x_{k-1}^i]$, мода або реалізація $\mu_k^i \sim p(x_k|x_{k-1}^i) \mu_k^i \sim p(x_k|x_{k-1}^i)$.

Якщо записати

$$q(x_k, i|z_{1:k}) = q(x_k|i, z_{1:k})q(i|z_{1:k}),$$

та покласти

$$q(x_k|i, z_{1:k}) = p(x_k|x_{k-1}^i),$$

то з (9.37) випливає, що:

$$q(i|z_{1:k}) \propto p(z_k|\mu_k^i)w_{k-1}^i. \quad (9.38)$$

Парі $\{x_k^j, i^j\}_{j=1}^{N_s} \{x_k^j, i^j\}_{j=1}^{N_s}$ присвоюється вага, пропорційна відношенню (9.5.12) і (9.5.13):

$$w_k^j \propto \frac{p(z_k|x_k^j)}{p(z_k|\mu_k^{i^j})}. \quad (9.39)$$

Наведемо псевдокод ітерації цього алгоритму:

Алгоритм 9.4: Auxiliary Particle Filter

$$\left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] = \text{APF} \left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k \right] \left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k \right]$$

FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— Обчислити $\mu_k^i \mu_k^i$

— Обчислити $w_k^i = q(i|z_{1:k}) \propto p(z_k|\mu_k^i) w_{k-1}^i w_k^i = q(i|z_{1:k}) \propto p(z_k|\mu_k^i) w_{k-1}^i$.

END FOR

Обчислити загальну вагу: $t = \sum_{i=1}^{N_s} w_k^i t = \sum_{i=1}^{N_s} w_k^i$

FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— Нормувати i-ту вагу: $w_k^i = t^{-1} w_k^i w_k^i = t^{-1} w_k^i$

END FOR

Провести відсів, використовуючи алгоритм 2 (Resampling Algorithm):

— $\left[\{-, -, i^j\}_{j=1}^{N_s} \right] \left[\{-, -, i^j\}_{j=1}^{N_s} \right] = \text{RESAMPLE} \left[\{x_{k-1}^i, w_k^i\}_{i=1}^{N_s} \right] \left[\{x_{k-1}^i, w_k^i\}_{i=1}^{N_s} \right]$

FOR $j = \overline{1, N_s} j = \overline{1, N_s}$

— Згенерувати $x_k^j \sim q(x_k|i^j, z_{1:k}) = p(x_k|x_{k-1}^{i^j}) x_k^j \sim q(x_k|i^j, z_{1:k}) = p(x_k|x_{k-1}^{i^j})$, як в

SIR фільтрі

— Присвоїти вагу $w_k^j w_k^j$, використовуючи (9.5.15)

END FOR

Обчислити загальну вагу: $t = \sum_{i=1}^{N_s} w_k^i t = \sum_{i=1}^{N_s} w_k^i$

FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— Нормувати i-ту вагу: $w_k^i = t^{-1} w_k^i w_k^i = t^{-1} w_k^i$

END FOR

Таким чином, ідейно, щоб згенерувати реалізацію з розподілу $p(x_k, i|z_{1:k})$ спочатку генерується номер i з ймовірністю $w_k^i \propto q(i|z_{1:k})$ (ці ймовірності називаються ймовірностями першого етапу), а потім x_k з розподілу $p(x_k|x_{k-1}^i) p(x_k|x_{k-1}^i)$. Після цього індекс i відкидається, а x_k присвоюються ваги відповідно до (9.5.15) – ваги другого етапу. Є надія, що отримані ваги другого рівня мають меншу дисперсію, ніж ваги в оригінальному SIR алгоритмі.

Регуляризований гранулярний фільтр

Метод відсіву введено для зменшення ефекту виродження, яким страждають гранулярні фільтри. Але, як було зазначено при його обговоренні, відсів призводить до інших проблем, зокрема проблеми втрати різноманітності гранул. Вона виникає через те, що на кроці відсіву реалізації генеруються з дискретного, а не неперервного розподілу. Якщо не приділяти уваги цій проблемі, вона може призвести до екстремального випадку «зліплення гранул», коли всі $N_s N_s$ гранул знаходяться в одному й тому самому місці простору станів і погано відображають апостеріорний розподіл.

Для боротьби з цією проблемою був запропонований регуляризований гранулярний фільтр (Regularized Particle Filter (RPF)). В ньому при проведенні процедури відсіву генеруються гранули з неперервної апроксимації $p(x_k | z_{1:k})$, а не з дискретної, як в SIR фільтрі. А саме апроксимації вигляду:

$$p(x_k | z_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i K_h(x_k - x_k^i) \quad (9.40)$$

де

$$K_h(x) = \frac{1}{h^{n_x}} K\left(\frac{x}{h}\right) \quad (9.41)$$

це масштабоване статистичне ядро (kernel density) $K(\cdot)K(\cdot)$, $h > 0$ – це параметр згладжування (скалярний), n_x – розмірність вектора стану x , $w_k^i, i = 1, \dots, N_s$ – нормовані ваги. Статистичне ядро $K(\cdot)K(\cdot)$ – це симетрична функція щільності розподілу така, що:

$$\int x K(x) dx = 0 \quad \int \|x\|^2 K(x) dx < \infty \quad (9.42)$$

Тобто випадковий вектор з щільністю $K(\cdot)K(\cdot)$ має нульове математичне

сподівання та скінченні моменти другого порядку.

Статистичне ядро $K(\cdot)$ та параметр згладжування $h > 0$ обираються так, щоб мінімізувати середнє інтегроване квадратичне відхилення (mean integrated square error (MISE)) між справжньою апостеріорною щільністю та відповідним регуляризованим емпіричним представленням в (9.5.16), яке визначається так:

$$\int x K(x) dx = 0 \quad \int \|x\|^2 K(x) dx < \infty$$

$$MISE(\hat{p}) = E \left[\int (\hat{p}(x_k | z_{1:k}) - p(x_k | z_{1:k}))^2 dx_k \right] \quad (9.43)$$

де $\hat{p}(x_k | z_{1:k})$ позначає апроксимацію, визначену в (9.5.16). У спеціальному випадку, коли всі елементи мають однакову вагу $\frac{1}{N_s}$, оптимальним вибором ядра є ядро Єпанечнікова [19, 20]:

$$K_e(x) = \begin{cases} \frac{n_x + 2}{2c_{n_x}} (1 - \|x\|^2), & \text{якщо } \|x\| < 1 \\ 0, & \text{інакше} \end{cases} \quad (9.44)$$

$$\int x K(x) dx = 0$$

$$\int \|x\|^2 K(x) dx < \infty$$

де c_{n_x} – об'єм одиничної гіперсфери в просторі R^{n_x} .

Часто зручно використовувати гаусівське ядро:

$$K_n(x) = (2\pi)^{-n_x/2} e^{-\frac{1}{2}\|x\|^2} \quad (9.45)$$

Крім того, якщо щільність, що апроксимується, є гаусівською з одиничною коваріаційною матрицею, оптимальним вибором параметра згладжування є такий:

$$h_{opt} = A(K) N_s^{1/(n_x+4)} \quad (9.46)$$

де

$$A(K_e) = [8c_{n_x}^{-1}(n_x + 4)(2\sqrt{\pi})^{n_x}]^{1/(n_x+4)}$$

$$A(K_e) = [8c_{n_x}^{-1}(n_x + 4)(2\sqrt{\pi})^{n_x}]^{1/(n_x+4)} \text{ для ядра Єпанечнікова та}$$

$$A(K_n) = [4/(n_x + 2)]^{1/(n_x+4)} A(K_n) = [4/(n_x + 2)]^{1/(n_x+4)} \text{ д л я}$$

гаусівського ядра.

В алгоритмі регуляризованого гранулярного фільтра відсів проводиться не на кожному кроці, а тоді, коли значення міри виродження стає нижчим деякого заданого порогу $N_T N_T$. Сама міра виродження обчислюється за формулою:

$$N_{eff} = \frac{1}{\sum_{i=1}^{N_s} (w_k^i)^2} \quad (9.47)$$

Зазначимо, що якщо встановити порогове значення $N_T N_T$ рівним $N_s N_s$, то відсів відбуватиметься на кожній ітерації.

Наведемо псевдокод ітерації алгоритму:

Алгоритм 9.5: Regularized Particle Filter

$$\left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] = \text{RPF} \left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k \right] \left[\{x_{k-1}^i, w_{k-1}^i\}_{i=1}^{N_s}, z_k \right]$$

FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— Згенерувати $x_k^i \sim q(x_k | x_{k-1}^i, z_k) x_k^i \sim q(x_k | x_{k-1}^i, z_k)$

— Присвоїти гранулі x_k^i вагу w_k^i відповідно до (9.5.6)

END FOR

Обчислити загальну вагу: $t = \sum_{i=1}^{N_s} w_k^i t = \sum_{i=1}^{N_s} w_k^i$

FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— Нормувати i -ту вагу: $w_k^i = t^{-1} w_k^i w_k^i = t^{-1} w_k^i$

END FOR

Обчислити $N_{eff} N_{eff}$, використовуючи (9.5.21)

IF $N_{eff} < N_T N_{eff} < N_T$

— Обчислити емпіричну коваріаційну матрицю $S_k S_k$ для $\{x_k^i, w_k^i\}_{i=1}^{N_s}$

— Обчислити матрицю $\Sigma_k \Sigma_k$ таку, що $\Sigma_k \Sigma_k^T = S_k \Sigma_k \Sigma_k^T = S_k$ (наприклад, $\Sigma_k = E_k D_k^{\frac{1}{2}} \Sigma_k = E_k D_k^{\frac{1}{2}}$, де $S_k = E_k D_k E_k^T S_k = E_k D_k E_k^T$ – спектральний розклад матриці $S_k S_k$)

— Провести відсів, використовуючи алгоритм 2 (Resampling Algorithm):

— — $\left[\{x_k^i, w_k^i, -\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i, -\}_{i=1}^{N_s} \right] = \text{RESAMPLE} \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right] \left[\{x_k^i, w_k^i\}_{i=1}^{N_s} \right]$

— FOR $i = \overline{1, N_s} i = \overline{1, N_s}$

— — Згенерувати $\varepsilon^i \sim K \varepsilon^i \sim K$ зі статистичного ядра (використати (9.5.18) або (9.5.19))

— — $x_k^{i*} = x_k^i + h_{opt} \Sigma_k \varepsilon^i x_k^{i*} = x_k^i + h_{opt} \Sigma_k \varepsilon^i$ (використати (9.5.20))

— END FOR

END IF

Хоча результати (9.5.18), (9.5.19), (9.5.20) є оптимальними лише в спеціальних випадках, вони можуть використовуватись в загальному випадку для отримання субоптимального фільтра.

Введення регуляризації покращує результат порівняно зі звичайним SIR фільтром у випадках, коли наявна значна втрата різноманітності вибірки гранул, наприклад, якщо шум у процесі є невеликим. Однією з проблем такого підходу є збільшення дисперсії апостеріорного розподілу.

Запитання і вправи до дев'ятого розділу

1. Що таке порогові моделі у часових рядах? Яку основну ідею вони реалізують?
2. У чому полягає відмінність між лінійними та пороговими моделями?
3. Які типи порогових моделей ви знаєте (TAR, STAR, SETAR)?
4. Як визначається порогове значення? Яку роль воно відіграє у

моделі?

5. Чим порогова модель стохастичної волатильності відрізняється від моделей ARCH/GARCH?
6. Що таке асиметрія у волатильності? Як її враховують порогові моделі?
7. Які методи оцінювання параметрів порогових моделей застосовуються найчастіше?
8. Як можна перевірити гіпотезу про наявність порогового ефекту у часовому ряді?
9. Наведіть приклади економічних і фінансових процесів, де проявляється порогова поведінка.
10. Чому порогові моделі вважаються більш гнучкими порівняно з класичними GARCH-підходами?
11. Візьміть фінансовий ряд (наприклад, дохідність індексу S&P 500). Побудуйте TAR-модель і визначте порогове значення. Проаналізуйте отримані параметри.
12. Побудуйте GARCH(1,1) та порогову GARCH (TGARCH) модель для одного й того ж фінансового ряду. Порівняйте результати оцінки волатильності.
13. Використайте модель EGARCH або TGARCH для перевірки ефекту «важливішого впливу негативних новин». Проінтерпретуйте результати.

Література

1. *Бідюк П.І.* Системний підхід до прогнозування на основі моделей часових рядів // Системні дослідження та інформаційні технології, 2003, № 3, с. 88-110.
2. *Бідюк П.І., Романенко В.Д., Тимошук О.Л.* Аналіз часових рядів. – Київ: НТУУ «КПІ», 2012. – 600 с.
3. *Бідюк П.І., Терентьєв О.М., Гасанов А.С.* Побудова і методи навчання байєсівських мереж // Кібернетика і системний аналіз, 2005, № 4, с. 133-147.
4. *Бідюк П.І.* Моделювання і прогнозування гетероскедастичних процесів // Системні дослідження та інформаційні технології, 2004, № 1, с. 115-134.
5. *Бідюк П.І., Савенков О.І., Баклан І.В.* Часові ряди: моделювання і прогнозування. – Київ: ЕКМО, 2003. – 144 с.
6. *Бідюк П.І.* Оцінювання і прогнозування стану малого підприємства за допомогою мережі Байєса // Збірник наукових праць Миколаївського гуманітарного університету ім. Петра Могили, 2006, т. 57, Вип. 44, с. 8-29.
7. *Бідюк П.І., Литвиненко В.І., Помазан Л.В.* Алгоритм ідентифікації стохастичних нелінійних систем на основі генетичного алгоритму // Адаптивні САУ, 2006, № 9(29), с. 3-15.
8. *Грін В. Г.* Економетричний аналіз. – Київ: Основи, 2005. – 1198 с.
9. *Кроптя А.В., Бідюк П.І.* Аналіз і методи оцінювання екстремальних значень // Наукові вісті НТУУ «КПІ», 2005, № 4, с. 34 – 47.
10. *Лук'яненко І.Г., Городніченко Ю.О.* Сучасні економетричні методи у фінансах. – Київ: Літера, 2002. – 350 с.
11. *Andreassen S., Jensen F.V., Olesen K.G.* Medical expert systems based on causal probabilistic networks // Int. J. of Biomedical Computing, 1991, v. 28, pp. 1-30.
12. *Bernardo J.M., Smith A.F.M.* Bayesian theory. – New York: John Wiley & Sons, Ltd., 2000. – 586 p.
13. *Beveridge S., Nelson C.* A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of business cycle // J. of Monetary Economics, 1981, v. 7, pp. 151-174.
14. *Brooks C.* Testing for nonlinearity in daily pound exchange rates // Applied Financial Economics, 1996, Vol. 6, pp. 307 – 317.
15. *Brooks C.* Linear and nonlinear (non-) forecastability of daily sterling exchange rates // Journal of Forecasting, 1997, Vol. 16, pp. 125 – 145.

16. *Brooks C., Hinich M.J.* Bicorrelations and cross-bicorrelations as non-linearity tests and tools for exchange rate forecasting // *Journal of Forecasting*, 2001, Vol. 20, pp. 181 – 196.
17. *Chen R., Tsay R.S.* Nonlinear additive ARX models // *Journal of the American Statistical Association*, Vol. 88, pp. 955 – 967.
18. *Chen R., Tsay R.S.* Functional coefficient autoregressive models // *Journal of the American Statistical Association*, Vol. 88, pp. 298 – 308.
19. *Cooper J.F.* Probabilistic inference using belief networks is NP-hard / TR-KSL-87-27, Knowledge Systems Lab., Stanford University, 1987. – 65 p.
20. *Dagum P., Luby M.* Appximating probabilistic inference in Bayesian belief networks is NP-hard // *Artificial Intelligence*, 1993, vol. 45, – pp. 141-153.
21. *Dahl G.* Combining disparate sources of information in the safety assessment of software-based systems // *Nuclear Eng. and Design*, 2000, v. 195, No.3, pp. 307-319.
22. *Diez F.J.* Local conditioning in Bayesian networks // *Artificial intelligence*, 1996, v. 87, pp. 1-20.
23. *Enders W.* *Applied Econometric Time Series*. – New York: John Wiley & Sons, Inc., 1995. – 450 p.
24. *Friedman J.H.* Multivariate adaptive regression splines // *Annals of Statistics*, 1991, Vol. 19, pp. 1 – 141.
25. *Gallant A.R.* *Nonlinear statistical models*. – New York: John Wiley and Sons Inc., 1987. – 622 p.
26. *Gilks W.R., Richardson S., Spiegelhalter D.J.* *Markov Chain Monte Carlo in Practice*. – London: Chapman & Hall, 1996. – 486 p.
27. *Grunwald P.* A tutorial introduction to the minimum description length principle / *Advances in minimum description length: Theory and applications* MIT Press, Cambridge, MA, USA, 2005. – 80 p.
28. *Handbook of Econometrics*, Vol. 4 / Eds.: R.F. Engle, D.L. McFadden. – New York: Elsevier Science, 1994. – 1044 p.
29. *Harvey A.C.* *Econometric analysis of time series*. – Cambridge, MA: MIT Press, 1990. – 402 p.
30. *Hautaniemi S.K.* Target identification with Bayesian networks. – Master of Science Thesis / <http://www.cs.tut.fi/~samba/publications>, 2000. – 99 p.
31. *Heckerman D.* An empirical comparison of three inference methods. (In Kanal L.N., Lemmer J.F. (Eds.). – *Uncertainty in artificial intelligence 4*. – Amsterdam: Elsevier Sc. Publishers, 1990, pp. 283 – 302.
32. *Heckerman D., Geiger D., Chickering D.* *Learning Bayesian Networks: The combination of knowledge and statistical data* / Technical report, MSR-TR-94-09, March 1994. – 54 p.
33. *Hinich M.J.* Testing for dependence in the input to a linear time series model // *Journal of Nonparametric Statistics*, 1996, No. 2, pp. 205 – 221.

34. *Hinich M.J., Patterson D.M.* Evidence of nonlinearity in daily stock returns // *Journal of Business and Economic Statistics*, 1985, Vol. 3, No. 1, pp. 69 – 77.
35. *Jensen F.V., Olesen K.G., Andersen S.K.* An algebra of Bayesian belief universes for knowledge-based systems // *Networks*, 1990, v. 20, pp. 637-660.
36. *Jensen F.V.* An Introduction to Bayesian networks. – Berlin: Springer, 1996. – 178 p.
37. *Johnston J., DiNardo J.* Econometric methods. - New York: McGraw-Hill, Inc., 1997. – 525 p.
38. *Kjerulff U.* Constructing Bayesian Networks / Report of Reykjavik University, April, 2005. – 77 p.
39. *Lauritzen S.L., Spiegelhalter D.J.* Local computations with probabilities on graphical structures and their application to expert systems // *J. of the Royal Stat. Society, B*, 1988, v. 50, pp. 157-224.
40. *Lewis P.A.W., Stevens L.G.* Semi-multivariate nonlinear modeling of time series using multivariate adaptive regression splines (MARS) // *Journal of the American Statistical Association*, Vol. 86, pp. 864 – 877.
41. *Neil M., Fenton N., Nielsen L.* Building large-scale Bayesian networks // *The Knowledge Engineering Review*, 2000, V. 15, No.3, pp. 257-284.
42. *Pearl J.* Probabilistic reasoning in intelligent systems. – San Francisco: Morgan Kaufmann Publishers, 1988. – 552 p.
43. *Press S.J.* Subjective and objective Bayesian statistics. – New Jersey: Wiley & Sons, Inc., 2003. – 563 p.
44. *Stensholt B.K., Tjostheim D.* Multiple bilinear time series models // *Journal of Time Series Analysis*, 1987, Vol. 8, pp. 221 – 233.
45. *Suzuki J.* Learning Bayesian belief networks based on the MDL principle: an efficient algorithm using the branch and bound technique. // *IEICE Trans. on Information and Systems*, February, 1999. – pp. 356-367.
46. *Terasvirta T., Tjostheim D., Granger C.W.J.* Aspects of modeling nonlinear time series / in *Handbook of Econometrics*, Vol. 4 / Eds.: R.F. Engle, D.L. McFadden. – New York: Elsevier Science B.V., 1994, pp. 2919 – 2957.
47. *Tjostheim D.* Some doubly stochastic time series models // *Journal of Time Series Analysis*, 1986, Vol. 7, pp. 51 – 72.
48. *Tjostheim D.* Nonlinear time series: a selective review // *Scandinavian Journal of Statistics*, 1994, Vol. 21, No. 2, pp. 97 – 130.
49. *Tsay R.S.* Nonlinearity tests for time series // *Biometrika*, 1986, Vol. 73, pp. 461 – 466.
50. *Tsay R.S.* Testing and modeling threshold autoregressive processes // *Journal of the American Statistical Association*, 1989, Vol. 84, pp. 231 – 240.
51. *Tsay R.S.* Analysis of financial time series. – New York: John Wiley & Sons, Inc., 2010. – 715 p.

