

Міністерство освіти і науки України
Чорноморський державний університет імені Петра Могили

Бідюк П. І., Гожий О. П.

ЙМОВІРНІСНО-СТАТИСТИЧНІ МЕТОДИ МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ

Монографія



Миколаїв – 2014

Ухвалено до видання Вченою Радою Чорноморського державного університету імені Петра Могили (протокол №1, від 11.09.2014 р.).

Рецензенти:

Андрєєв Микола Варфоломійович – докт. фіз.-мат. наук, професор (Інститут прикладного системного аналізу НАН та Міносвіти і науки України);

Снитюк Віталій Євгенович – докт. техн. наук., професор (Київський національний університет імені Тараса Шевченка)

Бідюк П. І. Ймовірно-статистичні методи моделювання і прогнозування : [монографія] / П. І. Бідюк, О. П. Гожий. – Миколаїв : Чорноморський державний університет ім. Петра Могили, 2014. – 440 с.

У монографії представлено сучасні методи ймовірно-статистичного моделювання і прогнозування нелінійних нестационарних процесів різної природи, яка ґрунтується на сучасних принципах системного аналізу. Запропоновано методика побудови математичних моделей у формі авторегресії та авторегресії з інтегрованим ковзним середнім (АРІКС), яка ґрунтується на кореляційному аналізі досліджуваних процесів, статистичних тестах на нелінійність та нестационарність, а також множинах критеріїв якості даних, моделей і прогнозів. Окремий розділ присвячено аналізу стаціонарності процесів АРКС. Наведено послідовність формування функцій прогнозування на основі побудованих моделей та приклади їх застосування для оцінювання прогнозів. Розглянуто методика побудови оптимального фільтра Калмана та його застосування для отримання оптимальних оцінок стану і короткострокових прогнозів за моделями у просторі станів. Докладно розглянуто приклади представлення моделей АРКС у просторі станів. Показано особливості протікання реальних ФЕП, моделювання сезонних ефектів, гетероскедастичних процесів та процесів із детермінованими і стохастичними трендами. Окремий розділ присвячено аналізу нестационарних коінтегрованих процесів та можливостям їх прогнозування. Розглянуто правила обчислення ймовірностей, безумовну незалежність подій, а також поняття правдоподібності і ймовірності. Наведено кілька варіантів формулювання теореми Байєса та приклади їх застосування.

Науково-навчальне видання рекомендується для студентів, аспірантів, викладачів, а також інженерів, що спеціалізуються в галузі математичного моделювання (створення моделей ймовірно-статистичного типу) на основі статистичних даних та коротко- і середньострокового прогнозування за допомогою побудованих моделей. Розглянуті методи і методики можуть бути застосовані для аналізу ФЕП та інших процесів у різних галузях виробництва, бізнесу, науки і техніки.

ЗМІСТ

Стор.

Вступ		7
Розділ 1	Мета моделювання, вимоги до даних і моделей	13
1.1	Мета побудови математичних моделей	13
1.2	Поняття структури математичної моделі	15
1.3	Два підходи до побудови математичних моделей	17
1.4	Узагальнений алгоритм побудови моделі	20
1.5	Вимоги до експериментальних даних, оцінок параметрів та моделі у цілому	22
1.5.1	Вимоги до експериментальних даних	22
1.5.2	Вимоги до оцінок параметрів моделі	24
1.5.3	Вимоги до математичної моделі	26
1.6	Спрощена класифікація математичних моделей	28
1.7	Деякі типи регресійних та різницевих рівнянь (PP)	29
	Запитання і вправи до розділу 1	34
Розділ 2	Аналіз стаціонарності процесів авторегресії з ковзним середнім	36
2.1	Слабка і сильна стаціонарність	36
2.2	Необхідні та достатні умови стаціонарності процесу $AR(1)$	38
2.3	Умови стаціонарності часткового розв'язку рівняння АРКС	41
2.4	Узагальнення умов стаціонарності на процеси $AR(p, q)$ довільного порядку	45
2.5	Теоретична автокореляційна функція	48
2.6	Обчислення кореляційних функцій за допомогою рівнянь Юла-Уокера	50
2.7	Часткова автокореляційна функція	55
	Запитання і вправи до розділу 2	61
Розділ 3	Застосування різницевих рівнянь до опису динаміки процесів	63
3.1	Загальні відомості про різницеві рівняння	63
3.2	Ітераційний метод знаходження розв'язків різницевих рівнянь	68
3.3	Знаходження однорідних та часткових розв'язків	75
3.3.1	Розв'язування однорідних рівнянь	75
3.3.2	Знаходження часткових розв'язків різницевих рівнянь	85
3.4	Приклади знаходження повних розв'язків різницевих рівнянь	98
	Запитання і вправи до розділу 3	109
Розділ 4	Прогнозування динаміки процесів за допомогою різницевих рівнянь	111
4.1	Для чого необхідні прогнози?	111
4.2	Які складові процесу можна прогнозувати?	114
4.3	Умовні та безумовні статистичні характеристики	116
4.4	Прогнозування без знаходження розв'язку рівняння	119
4.5	Побудова функцій прогнозування на основі розв'язку різницевого рівняння	123
4.6	Прогнозування з мінімальною дисперсією	128
4.7	Функції прогнозування, отримані в даному розділі	140
4.8	Оцінювання якості прогнозу	142
4.9	Довірчий інтервал для оцінок прогнозу	149

4.10	Комбінування оцінок прогнозів, отриманих за різними методів	150
4.10.1	Усереднювання прогнозів (вагові коефіцієнти однакові)	150
4.10.2	Зважене комбінування прогнозів	152
4.10.3	Вибір вагових коефіцієнтів за допомогою похибок оцінок прогнозів	153
	Запитання і вправи до розділу 5	154
Розділ 5	Методика побудови математичних моделей часових рядів	156
5.1	Аналіз процесу	157
5.2	Попередня обробка даних	158
5.3	Аналіз наявності нелінійностей	166
5.4	Формування інших елементів структури моделі	168
5.5	Оцінювання параметрів моделей-кандидатів	174
5.6	Діагностика моделей – вибір кращої з множини оцінених кандидатів ...	176
5.7	Приклади побудови математичних моделей за статистичними даними ..	182
5.7.1	Модель індексу споживчих цін	183
	Запитання і вправи до розділу 5	194
Розділ 6	Оптимальне оцінювання й прогнозування стану за допомогою фільтра Калмана	197
6.1	Принцип рекурсивного оцінювання	197
6.2	Дискретний фільтр Калмана для вільної динамічної системи	199
6.3	Дискретний фільтр Калмана для лінійної системи з детермінованими і стохастичними впливами	208
6.4	Причини розбіжності та заходи щодо підвищення обчислювальної стійкості оптимального фільтра	212
6.5	Приклади побудови оптимального фільтра	214
6.6	Оцінювання неспостережуваних компонент вектора стану за допомогою оптимального фільтра	222
6.7	Функція прогнозування на основі оптимального фільтра	224
	Запитання і вправи до розділу 6	230
Розділ 7	Представлення математичних моделей у просторі станів	232
7.1	Вступ	232
7.2	Перехід від лінійних неперервних до лінійних дискретних систем	237
7.3	Дискретні рівняння типу АРКС у просторі станів	239
7.4	Приклади представлення моделей у просторі станів	247
	Запитання і вправи до розділу 7	255
Розділ 8	Моделювання та прогнозування сезонних ефектів	257
8.1	Особливості перебігу реальних фінансово-економічних процесів	257
8.2	Моделювання процесів із сезонними ефектами	264
	Запитання і вправи до розділу 8	268
Розділ 9	Прогнозування гетероскедастичних процесів	270
9.1	Загальні відомості про гетероскедастичні процеси	270
9.2	Узагальнений авторегресійний умовно гетероскедастичний (УАРУГ) процес	278
9.3	Деякі інші моделі гетероскедастичних процесів	282
9.4	Врахування великих імпульсних значень у моделях ГСП	283

9.5	Приклади побудови математичних моделей гетероскедастичних процесів	289
9.6	Оцінювання стохастичної моделі волатильності	298
	Запитання і вправи до розділу 9	309
Розділ 10	Моделювання детермінованих і випадкових трендів	311
10.1	Вступ	311
10.2	Модель випадкового кроку	313
10.3	Методи вилучення тренду	320
10.4	Бізнес-цикли	326
10.5	Стохастичні тренди та декомпозиція ряду	330
10.6	Виділення стохастичного тренду із узагальненої моделі АРІКС	333
10.7	Розкладання тренду для моделі випадкового кроку + шум + дрейф	336
	Запитання і вправи до розділу	341
Розділ 11	Прогнозування коінтегрованих процесів	343
11.1	Задача моделювання нестационарних коінтегрованих процесів	343
11.2	Моделі коінтеграції	350
11.2.1	Коінтеграція і тренди	350
11.2.2	Коінтеграція і корегування похибок	353
11.3	Методика побудови моделі коінтегрованих процесів	362
	Запитання і вправи до розділу 11	373
Розділ 12	Вступ до ймовірнісного аналізу даних	375
12.1	Задачі, що розв'язуються ймовірнісно-статистичними методами	375
12.2	Можливі помилки при використанні ймовірнісно-статистичних методів	377
12.3	Ймовірнісні та детерміновані моделі	379
12.4	Безумовна незалежність подій	384
12.5	Правила обчислення ймовірностей	384
12.6	Умовні ймовірності, теорема Байєса	388
12.7	Розширення правила додавання	395
12.8	Правдоподібність і ймовірність	398
12.9	Обчислення ймовірностей для змінних	399
12.10	Вступ до алгебри потенціалів	410
12.11	Випадкові змінні	412
Розділ 13	Деякі варіанти теореми Байєса	416
13.1	Загальна форма теореми Байєса для довільних подій	416
13.2	Теорема Байєса для дискретних даних і дискретних параметрів	421
13.3	Теорема Байєса для неперервних даних і дискретного параметра	425
13.4	Теорема Байєса для дискретних даних і неперервного параметра	428
Перелік	посилань	432

ПЕРЕЛІК СКОРОЧЕНЬ

АКФ – автокореляційна функція
АР – авторегресія
АРКС – авторегресія з ковзним середнім
АРІКС – авторегресія з інтегрованим ковзним середнім
АРУГ – авторегресія з умовною гетероскедастичністю
АЧХ – амплітудно-частотна характеристика
ВП – випадковий процес
ВВП – валовий внутрішній продукт
ЕПП – економіка перехідного періоду
ЕУАРУГ – експоненціальна узагальнена авторегресія з умовною гетероскедастичністю
ІІ – індекс інфляції
ІМ – імітаційна модель
МАП – максимальна абсолютна похибка
МВК – модель випадкового кроку
МіАП – мінімальна абсолютна похибка
МГВА – метод групового врахування аргументів
МКМЛ – метод Монте-Карло для марковських ланцюгів
МКП – модель корегування похибки
ММП – метод максимальної правдоподібності
МНК – метод найменших квадратів
МР – множинна регресія
НКФ – нелінійна кореляційна функція
НМНК – нелінійний метод найменших квадратів
ПС – простір станів
РМНК – рекурсивний метод найменших квадратів
РМДП – рекурсивний метод допоміжної змінної
РММП – рекурсивний метод максимальної правдоподібності
САП – середнє абсолютне значення похибки
СЕ – сезонні ефекти
САПП – середня абсолютна похибка у відсотках
СКП – середньоквадратична похибка
СППР – система підтримки прийняття рішення
УАРУГ – узагальнена авторегресія з умовною гетероскедастичністю
ФК – фільтр Калмана
ЦФ – цифровий фільтр
ЧАКФ – часткова автокореляційна функція

ВСТУП

Побудова математичних моделей за статистичними (експериментальними) даними і обчислення оцінок прогнозів на їх основі моделей – один із самих популярних підходів до прогнозування динаміки соціально-економічних, фінансових, технічних та інших процесів. Як приклади можна навести коротко- і середньострокове прогнозування об'ємів виробництва та накопичення продукції на складах, оцінювання альтернативних економічних стратегій, формування бюджетів підприємств і держави, прогнозування та менеджмент ризиків різної природи, аналіз демографічних процесів та розв'язання багатьох інших задач. Необхідно зазначити, що неможливо побудувати систему автоматичного керування будь-яким об'єктом без розв'язання задачі короткострокового прогнозування. Це зумовлено тим, що будь-який об'єкт керування має час запізнення по входу (лаг), на який необхідно прогнозувати розвиток основних змінних.

Загалом методи прогнозування можна розділити на три широкі класи: (1) прогнозування на основі суджень, тобто, прогнозування, що ґрунтується на суб'єктивних судженнях (оцінках), інтуїції, поглиблених знаннях конкретної області та іншій інформації, що має відношення до прогнозованого процесу; (2) методи прогнозування на основі використання часового ряду однієї змінної, тобто, на основі авторегресії, авторегресії з інтегрованим ковзним середнім (АРІКС), АРКС плюс тренд і т. ін.; (3) методи прогнозування на основі використання часових рядів кількох змінних (багатовимірні векторні процеси). У останньому випадку ендогенна змінна, що прогнозується, залежить від кількох регресорів або екзогенних змінних у правій частині рівняння. Очевидно, що у загальному випадку процедура прогнозування може поєднувати у собі 2-3 методи.

На сьогоднішній день у спеціальній літературі описано досить широку множину методів прогнозування лінійних та нелінійних процесів з використанням даних у формі часових рядів. Найбільш поширеними серед них є метод групового врахування аргументів (МГВА), авторегресія (АР),

АРКС, авторегресія з інтегрованим ковзним середнім (АРІКС), лінійна та нелінійна множинна регресія, квантильна регресія, регресійні дерева, нейромережі, байєсівські мережі, нечіткі множини, нечіткі нейромережі та інші. Відносно „універсальними” методами моделювання та прогнозування є МГВА і нечіткі нейромережі. Однак, практика показує, що одного, навіть досить універсального методу, недостатньо для досягнення повноти аналізу процесу. Так, коректний аналіз гетероскедастичних процесів вимагає застосування моделей спеціальної структури для опису умовної дисперсії, що не забезпечує МГВА. Кожний метод моделювання має також свої недоліки і переваги стосовно обчислювальних витрат та характеристик точності оцінок прогнозів. Так, висока точність прогнозу за допомогою МГВА або нейромережі іноді досягається за рахунок високих обчислювальних витрат і моделей складної структури. Це особливо стосується застосування моделі у системі керування реальним часом, де модель необхідна і для оцінювання прогнозу і для синтезу керуючого впливу. Суттєвий вииграш щодо обчислювальних витрат можна досягти у такому випадку за допомогою набагато простішої моделі АРКС (або АРІКС), перевагами яких є простота структури та можливість їх оперативної адаптації до характеристик процесу в реальному часі. Цьому сприяє також наявність множини відносно простих методів рекурсивного оцінювання параметрів моделей.

Однак на сьогодні відсутній систематизований підхід до вибору структури математичних моделей та методів прогнозування, а також рекомендацій стосовно їх застосування. Під структурою моделі надалі будемо розуміти сукупність таких елементів: (1) розмірність моделі – кількість рівнянь, що описують процес; (2) порядок рівнянь; (3) нелінійність та її тип (стосовно змінних або параметрів); (4) час запізнення по входу (лаг) та його оцінка; (5) випадкове збурення та його тип.

Метою даної роботи є (1) аналіз та систематизація деяких існуючих методів прогнозування стаціонарних та нестаціонарних нелінійних процесів, представлених часовими рядами і часовими перерізами даних; (2) представлення нових методів прогнозування лінійних та нелінійних нестаціонарних процесів, зокрема введення поняття функцій прогнозування та їх практичне використання; (3) вироблення рекомендацій щодо комплексного застосування методів прогнозування. Розглядаються спеціальні структури моделей гетероскедастичних та коінтегрованих процесів, які забезпечують отримання високоякісних оцінок прогнозів нестаціонарних процесів із змінною дисперсією і трендами різних порядків.

Структура книги ґрунтується на системному підході до аналізу процесів різної природи, який передбачає врахування у моделях можливих невизначеностей описання процесів, зокрема це невизначеності структурного і статистичного характеру. Структурні невизначеності зумовлені значною нестаціонарністю багатьох реальних процесів, яка призводить до швидкої зміни елементів структури (складових) процесу, що вимагає адаптації

структури моделі до змінних параметрів досліджуваного процесу. Невизначеності статистичного характеру зумовлені недостатньою інформацією стосовно випадкових збурень. Як правило, такі невизначеності вимагають застосування до їх опису розподілів ймовірностей різних типів, хоча у більшості випадків ми достатньо обґрунтовано вважаємо збурення нормально розподіленим процесом. Цьому сприяє коректне застосування центральної граничної теореми.

Також відомо, що у більшості випадків моделювання та прогнозування ми будемо оптимальні за параметрами математичні моделі та оцінюємо моделі і прогнози за множиною деяких статистичних критеріїв. При цьому оптимізація стосується не тільки безпосередньо параметрів (коефіцієнтів) моделі, але й статистичних параметрів якості моделі. Тобто виникає необхідність багатокритеріальної оптимізації всього процесу побудови моделі та оцінювання прогнозу. Така задача частково ставиться і розглядається у даній роботі, хоча розв'язання таких задач у явному вигляді на сьогодні є досить проблематичним – потрібно шукати компроміс між витратами на створення таких комп'ютерних систем і отриманими результатами. Такі роботи під силу потужним компаніям, які працюють у співдружності із науково-дослідними лабораторіями.

Статистичне моделювання і прогнозування ґрунтується на двох основних типах експериментальних (статистичних) даних: часові ряди і часові перерізи.

Часовий ряд – це множина рівновіддалених (рідко різновіддалених) у часі спостережень (вимірів), які характеризують поведінку процесу чи об'єкта на вибраному часовому інтервалі.

Іншим різновидом даних є *часовий переріз*.

Часовий переріз – це множина спостережень вибраних змінних досліджуваної системи (процесу) на вибраній момент часу. Наприклад, дані, які характеризують стан всіх галузей української економіки на кінець грудня 2013 року.

Часові ряди – це основа для аналізу динаміки технічних, технологічних, біологічних, соціально-економічних, фінансових та систем інших видів. Якщо протягом місяця (тридцяти днів) ми будемо вимірювати о шостій ранку та о шостій вечора *температуру свого тіла, частоту пульсу, артеріальний тиск та поверхневий опір шкіри*, то отримаємо в результаті чотири часових ряди, які (у деякому наближенні) характеризують наш поточний фізичний стан. Кожний ряд можна описати своєю математичною моделлю, яка може бути використана для прогнозу конкретної змінної. Можна побудувати одну чотиривимірну модель, яка дасть можливість прогнозувати всі змінні одночасно. Іншими прикладами часових рядів є значення індексу інфляції, зареєстровані щомісячно протягом кількох років, або щорічні значення валового внутрішнього продукту (ВВП), зареєстровані протягом 30 років. Проміжок часу між вимірами називають періодом дискретизації вимірів

(часто позначають через T_s , де s – перша буква англійського слова *sampling* – дискретизація).

Часові ряди у фінансово-економічних системах є, як правило, обмеженими за розміром (кілька десятків або сотень значень), а ряди, які характеризують поведінку технічних систем і технологічних процесів, мають практично нескінченну довжину. Наприклад, в сучасних автомобілях встановлюють систему автоматичного керування швидкістю руху, яка підтримує постійну швидкість без втручання водія. Швидкість руху постійно вимірюється тахометром, перетворюється у цифрову форму і поступає на елемент порівняння із заданим значенням швидкості. Відхилення поточного значення швидкості від заданого використовується для зміни кута нахилу заслінки в карбюраторі або регулювання системи впорскування палива, яка регулює подачу горючої суміші в циліндри. Якщо прийняти період дискретизації вимірів змінних у системі керування швидкістю $T_s = 20$ мс, то за 1 секунду отримаємо 50 вимірів, за 1 хвилину – 3000 вимірів, а за 1 годину – 180000. Таким чином, у даному випадку ми маємо справу з надзвичайно довгими часовими рядами значень швидкості руху та відхилень від завдання.

Період дискретизації даних повинен бути строго постійним, оскільки його величина впливає на характеристики часового ряду. Чим більшим є період дискретизації, тим біднішим, з точки зору частотного наповнення, буде ряд експериментальних даних. В техніці період дискретизації вибирають за умови (теорема Котельникова-Шеннона, яка визначає необхідну умову вибору T_s):

$$T_s \leq \frac{1}{2f_{\max}},$$

де f_{\max} – максимальна частота гармонічної складової сигналу, яка необхідна для подальшого дослідження процесу чи об'єкта. Вона вибирається дослідником з фізичних міркувань та особливостей функціонування процесу. Визначити частотний склад сигналу (представленого, наприклад, часовим рядом) можна за допомогою перетворення Фур'є або іншого зручного ортогонального перетворення даних.

Необхідно зазначити, що при невеликих значеннях періоду дискретизації може з'явитись значна лінійна залежність між вимірами, що призводить до появи лінійно залежних стовпчиків вимірів у відповідній матриці, що використовується у методі оцінювання параметрів моделі, наприклад, у методі найменших квадратів (МНК), який реалізується за виразом:

$$\hat{\theta} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y},$$

де $\hat{\theta}$ – вектор оцінок параметрів математичної моделі об'єкта; X – матриця вимірів незалежних змінних (регресорів або екзогенних змінних); y – вектор вимірів залежної (основної або ендогенної) змінної. Це призводить до неможливості оцінювання параметрів (коефіцієнтів) рівнянь на основі рядів даних. Тому вибір значення T_s завжди є компромісним – тобто не занадто великим щоб не втратити інформативність сигналу (його частотне наповнення) і не занадто малим щоб не з'явилась значна лінійна залежність між вимірами. В технічних системах період дискретизації може складати від кількох мікросекунд до десятків хвилин.

При дослідженні соціально-економічних та фінансових процесів період дискретизації може складати кілька хвилин (ціни на біржі), один день, тиждень, місяць, квартал або рік. При цьому можливості змінювати T_s у сторону зменшення в таких системах дуже обмежені внаслідок особливостей протікання цих процесів та методів збору даних. Незважаючи на те, що масштаби часу, у яких протікають процеси в технічних та соціально-економічних системах суттєво відрізняються, методи аналізу відповідних часових рядів практично залишаються майже однаковими. Про це свідчать численні приклади із спеціальної літератури.

Економетричні часові ряди містять ряд компонент, математичний опис яких буде докладно розглянуто у даній роботі, зокрема це такі:

- детерміновані і випадкові коливання;
- детерміновані і випадкові тренди;
- сезонні ефекти (складові);
- нелінійності стосовно змінних і параметрів;
- значні випадкові імпульсні значення (викиди);
- змінна у часі умовна дисперсія процесу.

Перед побудовою математичної моделі процесу дуже важливо визначитись, які компоненти будуть описані та спрогнозовані за моделлю. Для цього потрібно коректно сформулювати постановку задачі. Послідовність побудови функцій прогнозування сформульована у четвертому розділі, а послідовність побудови моделей процесів різної природи розглядається у п'ятому розділі. В подальших розділах розглядаються особливості моделювання нестационарних та нелінійних процесів, а також деякі методи їх оцінювання.

Оптимальне оцінювання стану і обчислення короткострокового прогнозу розглянуто у шостому розділі. Докладно проаналізовано можливі причини розбіжності фільтра, а також заходи стосовно підвищення його обчислювальної стійкості. Наведено приклади його застосування до аналізу реальних процесів, які описано математичними моделями у просторі станів. Представлення моделей типу АРКС у просторі станів розглянуто у сьомому розділі.

Розділи 8 – 10 присвячено аналізу особливостей протікання реальних фінансово-економічних процесів, моделюванню сезонних ефектів,

прогнозуванню гетероскедастичних процесів із змінною у часі дисперсією та побудові детермінованих і випадкових трендів. Викладення матеріалу супроводжується ілюстративними прикладами та прикладами на основі фактичних даних.

Окремий розділ присвячено моделюванню і прогнозуванню нестационарних коінтегрованих процесів, які поширені у перехідних економіках. Отримані результати можна використати при побудові систем підтримки прийняття рішень, оскільки коінтеграційний аналіз надає можливість коректно розділити множину змінних на керуючі та керовані і створити відповідні системи оптимального керування.

У дванадцятому розділі наведено вступ до ймовірнісного аналізу даних. Розглянуто правила обчислення ймовірностей, поняття правдоподібності, обчислення ймовірностей для змінних, а також загальне формулювання правила (теореми) Байєса.

Тринадцятий розділ присвячено деяким альтернативним формулюванням теореми Байєса, які стосуються дискретних даних і дискретних параметрів, неперервних даних і дискретних параметрів і т. ін. Всі формулювання супроводжуються прикладами застосування.

Автори сподіваються, що книга в цілому буде корисною для студентів, аспірантів, інженерів, викладачів та всім іншим, хто займається практичними задачами математичного моделювання та прогнозування на основі експериментальних і статистичних даних в різних галузях техніки, економіки, фінансів, екології, соціальних дослідженнях та ін.

Розділ 1

МЕТА МОДЕЛЮВАННЯ, ВИМОГИ ДО ДАНИХ І МОДЕЛЕЙ

Розділ присвячено аналізу принципів 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. Наведено аналіз структурного і функціонального методів побудови моделей; перший представляє собою теоретичний метод на основі аналізу законів та закономірностей функціонування процесу, а другий орієнтований на збір та використання експериментальних даних. Узагальнений алгоритм побудови моделі за експериментальними даними 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 та скорочувати витрати часу на побудову та аналіз якості моделі. Розділ завершується переліком деяких типів математичних моделей, що описують стаціонарні і нестаціонарні процеси, а також процеси з нелінійностями відносно змінних та параметрів.

1.1 Мета побудови математичних моделей

Можна виділити три цілі побудови математичних моделей процесів та об'єктів:

- *поглиблене вивчення процесів;*

- прогнозування значень змінних за допомогою моделі;
- використання моделі для створення (синтезу) систем керування.

Поглиблене вивчення процесів за допомогою математичних моделей дозволяє дослідити кількісні зв'язки між вхідними та вихідними змінними, дослідити яким чином змінюються вихідні змінні при варіації вхідних в широкому діапазоні та розглянути поведінку процесів на будь-яких часових інтервалах у прийнятному масштабі часу. 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 реального часу, тобто в даному випадку часто використовують імітаційне моделювання. За допомогою математичних 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 системи завдяки можливості її швидкого дослідження на моделі, виявити вплив початкових умов та обмежень на ключові змінні.

Прогнозування значень змінних виконується, як правило, на основі набагато простіших моделей ніж поглиблене вивчення процесів. Для цієї мети дуже зручними є дискретні моделі у вигляді авторегресійних рівнянь (АР), авторегресії з ковзним середнім (АРКС) та авторегресії з інтегрованим ковзним середнім (АРІКС). Остання модель враховує нестационарність у формі тренду. Часто використовують також стандартизоване представлення в просторі станів неперервних чи дискретних моделей. Тип моделі залежить також від того, який прогноз необхідний – короткостроковий, середньостроковий чи довгостроковий. Задача прогнозування може бути дуже складною, якщо стохастичний процес нелінійний та нестационарний, наприклад, у випадку прогнозування такого параметра часового ряду як дисперсія.

Для розв'язання задачі *синтезу систем керування 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 функціонують у замкненому контурі з від'ємним зворотним зв'язком. Це сприяє асимптотичному зменшенню похибок керування до мінімально можливих значень навіть при використанні досить грубих моделей.

Моделювання системи керування на комп'ютері дає можливість за мінімальний час дослідити можливість застосування різних варіантів законів керування, комбінацій керуючих дій та визначити ефективність системи керування при зміні вхідних керуючих сигналів у широкому діапазоні їх значень. Комп'ютерна система керування може працювати 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 процесами. Як правило, такий тип керування реалізується за допомогою систем підтримки прийняття рішень, які дають можливість генерувати множину альтернатив з подальшим вибором кращої з них за допомогою множини критеріїв якості. При використанні комп'ютерної системи як порадики, керуючі дії, згенеровані комп'ютером, порівнюються з експертними оцінками оцінками входів і на основі деякого критерію якості вибираються найбільш прийнятні варіанти керуючих впливів.

1.2 Поняття структури математичної моделі

Введемо поняття структури математичної моделі, яким будемо користуватись у подальшому викладенні матеріалу. Поняття структури моделі містить у собі такі елементи (параметри):

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

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

де $y(k)$ – основна (залежна) змінна; a_0, a_1, a_2 – параметри (коефіцієнти) моделі; $\varepsilon(k)$ – випадкова змінна.

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

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

3. Наявність нелінійностей і їх характер. Визначення наявності нелінійностей – не завжди проста задача. Так, для механічних і деяких інших систем наявність нелінійностей можна визначити шляхом вивчення законів, закономірностей і особливостей їхнього функціонування. Наприклад, відомо, що для механічних систем характерною є наявність нелінійностей типу «люфт», «тертя», гістерезис.

При побудові регресійних моделей частіше зустрічаються нелінійності стосовно змінних і нелінійності стосовно параметрів. Прикладом нелінійності стосовно змінних може бути поліноміальна регресія виду

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

Коефіцієнти цього рівняння можна оцінювати за допомогою звичайного методу найменших квадратів (МНК) при належній побудові матриці вимірів (її структура буде розглянута нижче). Нелінійність стосовно параметрів зумовлена наявністю в моделі добутків коефіцієнтів, наприклад, у вигляді

$$y(k) = a_0 + a_1 a_2 x(k) + a_2 \exp(-bx(k)) + \varepsilon(k).$$

Коефіцієнти (параметри) такої моделі неможливо оцінити за допомогою звичайного МНК, тому для розв'язання цієї задачі використовують нелінійний МНК, метод максимальної правдоподібності або інші методи нелінійного оцінювання, наприклад відповідно модифікований метод Монте-Карло.

4. Час запізнення реакції на виході об'єкта стосовно вхідного сигналу. Запізнювання по входу (лаг) досить легко враховується як у неперервних, так і в дискретних моделях. Для моделі з дискретними змінними у вигляді різницевого рівняння

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

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

Розглянемо модель неперервного процесу у вигляді передатної функції із запізненням:

$$W(p) = \frac{K e^{-p\tau}}{1 + Tp},$$

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 в неперервній формі τ зв'язані між собою таким чином:

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

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

5. Тип збурень, що діють на процес, 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 в адитивній чи мультиплікативній формах. Вище ми привели різницеві рівняння, у яких збурення $\varepsilon(k)$ входить в адитивній формі. Прикладом мультиплікативної форми може бути така:

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

де $v(k)$ – мультиплікативне збурення для процесу $h(k)$.

Частіше всього збурення описують розподілами випадкових величин (статистичні моделі), але в окремих випадках його 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 структури моделі, що адекватна процесу, – задача не завжди проста і вирішується, як правило, ітераційно. Спочатку структуру моделі оцінюють наближено на підставі дослідження закономірностей протікання процесу, аналізу кореляційних функцій, візуального аналізу даних. При цьому вибирають кілька найбільш ймовірних структур (кандидатів). Потім обчислюють оцінки параметрів моделей-кандидатів і вибирають кращу з них, використовуючи відповідні статистичні характеристики якості моделей.

Остаточним критерієм придатності моделі є можливість її застосування для розв'язання конкретної задачі, наприклад, прогнозування або керування. Якщо жодна з моделей-кандидатів не може вважатися адекватною для конкретного застосування, то необхідно досліджувати на інформативність експериментальні дані, які можуть бути недостатньо інформативними для оцінювання моделі. У такому випадку може знадобитися повторний чи додатковий збір експериментальних даних і корегування структури моделі. Іноді взагалі приходиться вибирати не один, а більше класів моделей для опису процесу або ж для досягнення належної якості прогнозування використовувати схеми комбінування оцінок прогнозів з коректно вибраними (обчисленими) ваговими коефіцієнтами.

1.3 Два підходи до побудови математичних моделей

Основними відомими підходами до побудови математичних моделей є такі:

- структурний;
- функціональний.

Структурний підхід передбачає математичний опис внутрішнього механізму взаємодії змінних, відображення їх фактичних взаємозв'язків за допомогою рівнянь та співвідношень.

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

Розглянемо, наприклад, зростання інфляції внаслідок випуску додаткової грошової маси. Оскільки логіка цього процесу досить проста і існують експериментальні (статистичні) дані, які ілюструють зростання інфляції, то можна постулювати, що інфляція описується диференціальним або різницеvim рівнянням першого/другого порядку (рис.1.1).

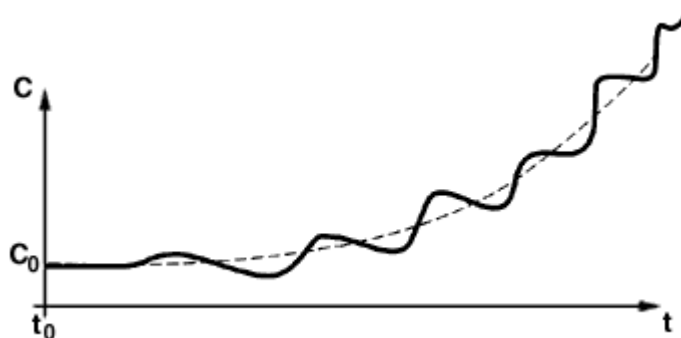


Рис. 1.1 Можлива крива зростання інфляції

Структурним підходом можна скористатись, наприклад, для побудови математичної моделі процесу трансформування власності (модель процесу приватизації, який є невід'ємною складовою перехідної економіки) або ж для макроекономіки в цілому. Для цього необхідно визначити вхідні керуючі змінні, збурення та вихідні змінні, а також визначити якого типу зв'язки існують між ними (рис. 1.2). Крім того, завжди існує задача виявлення збурень та встановлення типів розподілів, до яких вони відносяться.



Рис. 1.2 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 – рівень виробництва валового внутрішнього продукту (ВВП), індекси інфляції, індекс людського розвитку, середня заробітна плата і таке інше.

Як правило, в такі моделі вводять у явному вигляді *збурення* – випадкові змінні, які негативно впливають на протікання процесу. Так, у випадку створення моделі макроекономіки у цілому можуть мати місце такі збурення:

- помилкові рішення уряду;
- затримки платежів між підприємствами та державами;
- значні коливання цін на енергоносії;
- хронічна технологічна відсталість;
- швидкі зміни податкового законодавства;
- неконтрольований вплив капіталу за кордон;
- використання недостовірної статистичної інформації.

Очевидно, що врахувати такі впливи в моделі надзвичайно складно, а тому випадкові змінні агрегують (об'єднують) і представляють в моделі однією чи двома випадковими змінними (процесами), які представляють всі збурення в інтегрованому вигляді.

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 взаємодії його змінних.

В даному випадку процес = “чорна скринька”.

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

$$C = f(I, D, P).$$

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

Очевидно, що функціональний підхід є простішим від структурного і саме він найчастіше використовується на практиці. Гнучкість даного підходу дає можливість відносно швидко будувати високоякісні моделі для прогнозування та синтезу систем керування. Крім того, як правило, дискретні моделі такого типу досить легко адаптуються до часових змін, що мають місце у модельованих процесах та об'єктах.

1.4 Узагальнений алгоритм побудови моделі

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

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 *кількості входів і виходів, логіки взаємодії* складових частин процесу, визначення можливих *зовнішніх збурень* та їх типу: детерміновані чи випадкові. За можливості необхідно встановити розподіл ймовірностей для випадкових збурень або функціональний опис 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 встановити типи структур моделей-кандидатів. Їх може бути кілька залежно від того наскільки невизначеною є інформація стосовно характеристик процесу. Чим більшою є 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 з метою отримання експериментальних даних:

- визначаються діапазони зміни вхідних та вихідних величин і збурень;
- встановлюється дискретність зміни вхідних величин, період дискретизації вимірів (якщо змінні неперервні), визначаються типи вимірювальних та реєструючих приладів;
- плануються режими роботи процесу, для яких необхідно зібрати експериментальні дані;
- якщо дані носять статистичний характер, то визначається періодичність їх збору та занесення в базу даних;

– за необхідності задається тип, об'єм і якість продукції, що буде вироблена на протязі 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 вибраних структур 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 залежно від її структури;
- зробити попередню обробку даних; залежно від конкретної задачі це може бути масштабування, логарифмування, цифрова фільтрація, видалення недостовірних даних і т. і.;
- обчислити оцінки (векторів) параметрів моделей.

6. *Визначити ступінь адекватності кожної моделі-кандидата процесу за допомогою множини статистичних критеріїв (статистик) якості.* Визначити кращу модель з множини оцінених кандидатів.

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

8. *Уточнити структуру моделі,* зібрати, за необхідності, додаткові експериментальні дані 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 технології обробки даних і методи прогнозування розвитку процесів на виробництві. Наприклад, цілком можливо спланувати та провести інвестиційний експеримент, експерименти із впровадженням нових технологій, новими типами продукції.

1.5 Вимоги до експериментальних даних, оцінок параметрів та моделі в цілому

1.5.1 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 (послідовності вимірів), що недопустимо, оскільки при цьому змінюється інформативність сигналу. Крім того, вимірювання вхідних та вихідних сигналів необхідно робити синхронно, тобто в одні й ті ж моменти часу. Інакше вони будуть непридатні для побудови рівнянь та передатних функцій, оскільки порушуються причинно-наслідкові зв'язки між входами та виходами. Як правило, в системах керування реального часу задача збору вимірювальних даних має найвищий пріоритет серед множини всіх задач, які необхідно розв'язувати.

2. *Вибірка даних повинна бути представницькою*. Це означає, що вона повинна охоплювати досить довгий період часу, щоб включити в розгляд всі режими роботи, які передбачається описати математичною моделлю. Розрізняють два основних режими роботи об'єктів (процесів): *перехідний та усталений*. В перехідному режимі система керування переводить процес з деякого початкового стану у заданий. Перебування процесу у заданому (номінальному) стані на протязі 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 у 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. Цей процес розпочався у 1991 році і, за оцінками фахівців з економіки, триватиме 30-40 років.

3. *Вибірка вимірювальних даних повинна бути інформативною.* Частіше всього інформативність пов'язують з *числом похідних*, що їх містить вибірка даних. Чим більше число похідних можна отримати з вимірів, тим інформативнішою є часова вибірка. Наприклад, припустимо, що процес описується диференціальним рівнянням другого порядку:

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

де $y(t)$ – вихідний сигнал процесу; $u(t)$ – вхідний сигнал; $\theta^T = [a_0 \ a_1 \ a_2]$ – вектор коефіцієнтів рівняння, які необхідно оцінити за допомогою експериментальних даних. Очевидно, що оцінки коефіцієнтів a_2 і a_1 можна обчислити тільки у тому випадку, якщо виміри $y(t)$ містять другу і першу похідну стосовно часу.

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

Вимога інформативності виконується у тому випадку, коли вхідний сигнал задовольняє умові *достатнього збудження* процесу. Основна ідея достатнього рівня збудження полягає у тому, щоб смуга частот вхідного сигналу перекривала амплітудно-частотну характеристику процесу. Тобто, вихідний сигнал $y(t)$ буде інформативним у тому випадку, якщо достатньо інформативним буде вхідний сигнал $u(t)$. Умові інформативності (достатнього збудження) задовольняють такі основні типи вхідних сигналів: *білий шум*, *псевдовипадковий двійковий сигнал* та *одиничний імпульс*. Білий шум (гаусів процес) теоретично має нескінченний частотний спектр; достатньо широкі 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 фізики функціонування об'єкта чи процесу (подача на вхід подібного сигналу може призвести до руйнування об'єкта чи до створення аварійної ситуації). Тому в таких системах часто використовують як збуджуючий сигнал завдання регулятора, якщо він має форму сходинок, тобто має фронт прямокутного імпульсу. У багатьох випадках можливе використання гармонічних сигналів, які сприймаються

"легше" більшістю фізичних 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 і забезпечує достатній рівень „збудження” процесу.

1.5.2 Вимоги до оцінок параметрів моделі

Точність оцінок параметрів моделі залежить від якості вимірювальних даних, коректності попередньої обробки даних та від того, наскільки правильно вибрано метод оцінювання. Так, для оцінювання параметрів лінійних та псевдолінійних (нелінійних відносно змінних) моделей можна застосовувати звичайний МНК та його модифікації, а для оцінювання моделей, нелінійних відносно параметрів, необхідно застосовувати нелінійний МНК, метод максимальної правдоподібності та інші методи, розроблені для оцінювання параметрів нелінійних моделей.

Існують такі усталені вимоги до оцінок параметрів математичних моделей:

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 або на різних відрізках однієї вибірки. Формально незміщеність оцінок параметрів записують так:

$$E[\hat{\theta}] = \theta,$$

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

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

$$p(|\hat{\theta}_k - \theta| < \varepsilon) \rightarrow 1 \text{ при } k \rightarrow \infty,$$

де $\varepsilon > 0$ – мале число, яке обмежує допустиму похибку оцінки параметра; p – символ ймовірності; $\hat{\theta}_k$ – 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 стосовно мінімальності дисперсії оцінки, яка формально записується так:

$$Var(\hat{\theta}) \rightarrow \min.$$

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 випадковою величиною; де $y(k)$ – значення ряду, отримане експериментально (або статистичні дані); $\hat{y}(k)$ – оцінка змінної, отримана за допомогою побудованої моделі;

– *похибки моделі* – це некорельований процес, тобто відсутня автокореляція похибок:

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

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

Корельованість значень похибки між собою означає, що вона містить інформацію про основний процес. Тобто необхідно корегувати структуру моделі таким чином, щоб похибка стала некорельованою.

1.5.3 Вимоги до математичної моделі

1. Модель повинна бути *адекватною* процесу чи об'єкту.

Адекватність означає, що модель повинна:

- а) відобразити найбільш характерні зв'язки та взаємодію між змінними процесу;
- б) враховувати можливі керуючі дії (сигнали);
- в) враховувати вплив зовнішніх збурень та шуми вимірів;
- г) враховувати початкові значення змінних та обмеження на них.

Формально адекватність визначають за допомогою ряду статистичних параметрів або статистик. Наприклад, дуже часто використовують середньо-квадратичну похибку моделі (СКП), яка обчислюється за формулою:

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

де $x_s(k)$ – вимір вихідного сигналу об'єкта в момент k ; $x_m(k)$ – оцінка вихідного сигналу об'єкта, отримана за оціненою моделлю. Для лінійних моделей запропоновано кілька 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. Розв'язок необхідний для асимптотичного аналізу поведінки процесу, а також для обчислення оцінок прогнозів.

Одним із принципів, яких необхідно дотримуватись при побудові моделі є наступний: “*в моделі не повинно бути нічого зайвого крім*

необхідного”. 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 відповідають статистичним характеристикам часового ряду, на основі якого вона оцінюється. Такі спрощені моделі виявляються цілком придатними для прогнозування та автоматичного керування процесами. Загалом питання складності моделі вирішується в кожному випадку окремо.

3. Модель повинна буди достатньо *універсальною*, щоб її можна було застосувати до опису вибраного класу однотипних процесів або до опису 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:

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

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 передатна 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, наведена модель описує широкий клас біологічних систем і цілком відповідає умові універсальності.

При моделюванні технічних систем широко застосовують ланки першого і другого порядку, що відповідають звичайним диференціальним рівнянням таких же порядків. На основі таких простих ланок можна будувати моделі будь-якої складності. Дуже поширений в техніці та екології клас систем з розподіленими параметрами. Наприклад, процес розповсюдження домішок в атмосфері та водному середовищі, механічні коливання сонячних батарей та антен супутників, крила літака, коливання локомотива з вагонами на залізниці, автомобіля з причепом і багато інших. Динаміку таких систем описують диференціальними рівняннями з частинними похідними.

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 досліджуваного об'єкта. 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 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 яких є функціями часу. Системи керування, побудовані для нестационарних процесів, називають адаптивними. Такі системи є досить складними з точки зору аналізу збіжності оцінок параметрів та похибок керування, а тому при проектуванні адаптивних систем необхідно особливу увагу приділяти питанням достатнього збудження процесу, вибору методу оцінювання параметрів та контролю обчислювального процесу в системі у цілому.

1.6 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 прямолінійний рух автомобіля, потяга або літака;

б) *імітаційні* – моделі, які докладно відтворюють поточне функціонування процесу у вибраному масштабі часу.

Імітаційні (ІМ) моделі нагадують активний фізичний експеримент з використанням фактичних даних, отриманих безпосередньо з процесу. Характерною особливістю ІМ є генерування та використання псевдовипадкових послідовностей (процесів) із заданими характеристиками.

Переваги імітаційних моделей:

- наглядність результатів (проміжних та остаточних);
- динамічний характер відображення протікання процесу;
- можливість врахування детермінованих і випадкових факторів, а також складних залежностей від них;
- простота корекції моделі (ввести додаткове рівняння, правило і т. ін.);
- можливість дослідити процес на множині його реалізацій (тобто, провести статистичний експеримент Монте-Карло);
- практично необмежені можливості введення в модель таких елементів, як
 - рівняння будь-якого типу;
 - логічних правил;
 - нечіткої логіки;
 - статистичних розрахунків;
 - оптимізаційних процедур;
 - процедур прийняття рішень;
 - об'єднання чітких алгоритмів з нечіткими (це непросто, але корисно при аналізі функціонування процесів з невизначеностями та прийнятті рішень).

Недоліки імітаційних моделей:

- висока вартість побудови та використання;
- вимагають багато часу для розробки.

1.7 Деякі типи регресійних та різницевих рівнянь (РР)

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

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

де $a_i, i = 1, \dots, p$ – коефіцієнти моделі, які оцінюються на основі значень часового ряду; p – порядок авторегресії, який визначається кількістю затриманих у

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 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 рівняння.

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

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

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

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

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

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

Авторегресія + множинна регресія = змішана регресія:

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

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

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

Регресія, нелінійна стосовно змінних (псевдолінійна регресія):

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

Тобто у 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:

$$y(k) = a_0 + a_1 e^{b \cdot x(k)} + \varepsilon(k).$$

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. Рівняння для умовної дисперсії (авторегресія першого порядку):

$$h(k) = \beta_0 + \beta_1 \varepsilon^2(k-1) + \varepsilon_1(k),$$

де $h(k)$ – умовна дисперсія процесу в момент k ; $\varepsilon^2(k)$ – квадрат залишків авторегресії низького порядку (1-го або 2-го), побудованої для основної змінної; $\varepsilon_1(k)$ – похибка моделі в момент k . Докладно моделі гетероскедастичних процесів будуть розглянуті нижче.

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

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

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

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

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):

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

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

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

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

Модель для прогнозування волатильності за допомогою УАРУГ:

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

Рівняння для умовної дисперсії та коваріації:

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

Коефіцієнт хеджування при використанні двофакторної моделі:

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

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

Концепція коінтегрованості змінних передбачає існування довгострокового зв'язку між значеннями 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.

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 наступний вигляд:

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

$$\Delta y(k) = a_{20} + \sum_{i=1}^p b_{2i} \Delta y(k-i) + \sum_{i=1}^p c_{2i} \Delta x(k-i) + \lambda_2 e_x(k-1) + \varepsilon_2(k).$$

Коефіцієнти λ_1, λ_2 в наведених рівняннях називають швидкістю пристосування (корегування).

Вище ми навели тільки деякі структури (типи) математичних моделей стаціонарних та нелінійних нестаціонарних процесів, які широко використовуються для опису динаміки процесів різної природи. На сьогодні існує дуже багато інших різновидів дискретних моделей, які описують різні складові реальних процесів. В подальшому ми будемо намагатися виконати аналіз по можливості більшого числа типів існуючих моделей, але необхідно зазначити, що зробити це в межах однієї публікації досить складно.

Запитання і вправи до першого розділу

1. Назвіть три напрями застосування математичних моделей? Наведіть приклади практичного застосування.
2. Чому математична модель, призначена для застосування в замкненому контурі керування процесом (об'єктом) може мати нижчу ступінь адекватності об'єкту ніж модель, побудована для прогнозування?
3. Які вимоги висуваються до математичних моделей, призначених для застосування у реальному часі?
4. Сформулюйте необхідну умову для визначення оцінки періоду дискретизації вимірів (теорема Котельникова-Шеннона). Чому ця теорема представляє тільки необхідну умову?
5. Назвіть 5 елементів структури математичної моделі, поясніть їх на прикладі.
6. Поясніть фізичну суть запізнення (лагу) по входу, наведіть приклади? Чи існують у природі процеси без запізнення і що необхідно для їх існування? Який середній час запізнення (реакції) для людини?
7. Яким чином враховується запізнення по входу в дискретних та неперервних моделях?
8. На чому ґрунтується існування авторегресії?
9. Назвіть основні типи нелінійностей процесів та які методи застосовують для оцінювання параметрів нелінійних моделей?
10. Сформулюйте узагальнений алгоритм побудови математичної моделі на основі експериментальних даних. Поясніть необхідність оцінювання декількох моделей-кандидатів.
11. Поясніть вимогу неперервності та синхронності до експериментальних даних?
12. Що означають вимоги представництва та інформативності до експериментальних даних?
13. Сформулюйте три основних вимоги до оцінок параметрів математичної моделі. При яких умовах оцінки, отримані за методом найменших квадратів, задовольняють трьом загальним вимогам до оцінок моделей?
14. Яким вимогам повинна задовольняти модель процесу?

15. Що означає *адекватність* моделі процесу? Яким чином можна її визначити?
16. Для чого необхідно знаходити розв'язки рівнянь, що описують динаміку процесу?
17. Що означають *універсальність* та *робастність* моделі?
18. Яким чином досягається адаптивність моделі процесу?
19. Сформулюйте переваги та недоліки імітаційного моделювання?
20. В чому полягає відмінність між парною та множинною регресіями?
21. Чому регресійні рівняння називають стохастичними? Яку ще складову містять, як правило, часові ряди окрім випадкової?
22. Назвіть причини введення випадкової змінної $\varepsilon(k)$ у праву частину рівняння?
23. Якими методами можна коректно оцінювати параметри моделей, нелінійних стосовно параметрів?
24. Який процес називають гетероскедастичним? Яка іншу назву мають такі процеси?
25. В чому полягає суть інтегрованості та коінтегрованості випадкових процесів?

Розділ 2

АНАЛІЗ СТАЦІОНАРНОСТІ ПРОЦЕСІВ АВТОРЕГРЕСІЇ З КОВЗНИМ СЕРЕДНІМ

2.1 Слабка і сильна стаціонарність

Процеси з одиничними коренями

Розглянемо докладніше властивості рівнянь, які застосовують для опису динаміки процесів різної природи, представлених даними у формі часових рядів. Спочатку уточнимо, яке ж рівняння коректно називати авторегресією з ковзним середнім?

Рівняння виду

$$\begin{aligned} y(k) &= a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j \varepsilon(k-j) + \varepsilon(k) = \\ &= a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j), \end{aligned} \quad (2.1.1)$$

$$b_0 = 1, \quad \sum_{j=1}^q b_j = 1,$$

формально називають авторегресією з ковзним середнім, якщо корені його характеристичного рівняння лежать всередині одиничного кола на комплексній площині (тобто їх значення по модулю не перевищують одиницю), а $\{\varepsilon(k)\}$ – випадковий процес, причини появи якого розглянуто у попередньому розділі.

Якщо один або більше коренів характеристичного рівняння, записаного для (2.1.1) (для його авторегресійної частини), дорівнюють одиниці, то послідовність $\{y(k)\}$ називають *інтегрованою* або *процесом з одиничними коренями* (інша назва – процес із трендом), а рівняння (2.1.1) називають авторегресією з інтегрованим ковзним середнім (АРІКС). Процеси з одиничними коренями відносять до *класу нестационарних процесів*. Оскільки такі процеси є досить характерними для виробничих технологій, економіки, фінансів, екології та інших галузей діяльності, то їм необхідно приділити значну увагу. Нестационарні процеси такого типу особливо часто зустрічаються в перехідній економіці (практично всі процеси перехідної економіки нестационарні) та відповідній фінансовій діяльності, для яких властива висока нестационарна динаміка розвитку. Так, інфляція в Україні зросла за три роки (1991 – 1993 роки, тобто перші роки перехідного періоду) від 15-20% до 10250%. Таке катастрофічне зростання рівня інфляції супроводжувалось її повільним спадом до рівня 15-20% на протязі наступних 8-10 років. Це приклад процесу з надзвичайно високою динамікою (швидкістю зміни), а також позитивним і негативним трендами на різних етапах його розвитку. У подальшому розгляді трендом вважатимемо поточне середнє значення процесу, яке вказує на його довгострокові зміни.

Сильна і слабка стаціонарність

Розрізняють стаціонарність *слабку* і *сильну*. Слабку стаціонарність називають ще так:

- стаціонарність *по коваріації*;
- стаціонарність *другого порядку*;
- стаціонарність *у широкому смислі*.

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:

- *математичне сподівання* не залежить від часу:

$$E[y(k)] = E[y(k - s)] = \mu = \text{const}; \quad (2.1.2)$$

- *дисперсія* залишається постійною для всього часового інтервалу, на якому розглядається процес:

$$E\{[y(k) - \mu]^2\} = E\{[y(k - s) - \mu]^2\} = \sigma_y^2 = \text{const} \quad (2.1.3)$$

або $\text{var}[y(k)] = \text{var}[y(k - s)] = \sigma_y^2 = \text{const};$

- *автоковаріація* залишається незмінною в часі для всього часового інтервалу, на якому досліджується процес:

$$\begin{aligned}
 E\{[y(k) - \mu][y(k-s) - \mu]\} &= \\
 &= E\{[y(k-j) - \mu][y(k-j-s) - \mu]\} = \gamma(s) = \text{const},
 \end{aligned}$$

або

$$(2.1.4)$$

$$\text{cov}[y(k), y(k-s)] = \text{cov}[y(k-j), y(k-j-s)] = \gamma(s) = \text{const}.$$

Сильна стаціонарність не вимагає щоб математичне сподівання та дисперсія були постійними величинами. Поняття сильної стаціонарності використовують, в основному, у теоретичному аналізі випадкових процесів. При виконанні практичних розрахунків частіше використовують поняття слабкої стаціонарності.

Таким чином, слабка стаціонарність накладає строгіші обмеження на поведінку випадкового процесу ніж сильна. У подальшому викладенні матеріалу будемо розглядати ті процеси, які є стаціонарними по коваріації, тобто стаціонарні у слабкому сенсі. Відповідно, процес будемо вважати *нестационарним*, якщо не виконується хоча б одна з умов (2.1.2) – (2.1.4).

Зазначимо також, що оцінки основних статистичних характеристик 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 визначати конкретні параметри та їх комбінації, які впливають на стаціонарність, а також коректно вибирати метод оцінювання параметрів моделі. Такі знання необхідні для дослідження характеристик нових (побудованих) моделей різноманітних структур, визначення можливостей їх практичного застосування, встановлення областей стійкості розв'язків і т. ін.

2.2 Необхідні та достатні умови стаціонарності процесу AP(1)

Розглянемо стохастичний процес AP(1), що описується рівнянням:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad (2.2.1)$$

де $\{\varepsilon(k)\}$ – послідовність білого шуму з нульовим середнім, $E[\varepsilon(k)] = 0$, та скінченною постійною дисперсією $E[\varepsilon^2(k)] = \sigma_\varepsilon^2$; $y(0) = y_0$ – початкова умова, яка може бути детермінованою.

Повний розв'язок цього рівняння має вигляд (докладно методика знаходження повних розв'язків різницевого рівнянь розглядається нижче у наступному розділі):

$$y(k) = a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_0 + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i). \quad (2.2.2)$$

Запишемо математичне сподівання для цього розв'язку:

$$E[y(k)] = a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_0, \quad (2.2.3)$$

а також математичне сподівання для (2.2.2), але при зміні моменту часу на s кроків вперед:

$$E[y(k+s)] = a_0 \sum_{i=0}^{k+s-1} a_1^i + a_1^{k+s} y_0. \quad (2.2.4)$$

Вирази (2.2.3) і (2.2.4) 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

$$\begin{aligned} \lim_{k \rightarrow \infty} [y(k)] &= \lim_{k \rightarrow \infty} \left[a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_0 + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) \right] = \\ &= \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i). \end{aligned} \quad (2.2.5)$$

Таким чином, безумовне математичне сподівання

$$\mu = E[y(k)] = \frac{a_0}{1-a_1}, \quad k \rightarrow \infty, \quad (2.2.6)$$

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 (2.2.5):

$$\begin{aligned} \text{var}[y(k)] &= E[(y(k) - \mu)^2] = E \left[\left(\sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) \right)^2 \right] = \\ &= E \left\{ \left[\varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots \right] \times \left[\varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots \right] \right\} = \end{aligned}$$

$$\sigma_{\varepsilon}^2 (1 + a_1^2 + a_1^4 + \dots) = \frac{\sigma_{\varepsilon}^2}{1 - a_1^2} = \text{const},$$

оскільки $E[\varepsilon(k-i)\varepsilon(k-j)] = 0$ при $i \neq j$.

Запишемо вираз для автоковаріації:

$$\begin{aligned} \text{cov}[y(k) - \mu] &= E[(y(k) - \mu)(y(k-s) - \mu)] = \\ &= E\left\{[\varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots + a_1^s \varepsilon(k-s) + a_1^{s+1} \varepsilon(k-s-1) + \dots] \times \right. \\ &\quad \left. \times [\varepsilon(k-s) + a_1 \varepsilon(k-1-s) + a_1^2 \varepsilon(k-2-s) + \dots]\right\} = \\ &= \sigma_{\varepsilon}^2 a_1^s [1 + a_1^2 + a_1^4 + \dots] = \frac{\sigma_{\varepsilon}^2 a_1^s}{1 - a_1^2} = \text{const} \end{aligned}$$

для заданого значення параметра s .

Таким чином, стохастичний процес $AR(1)$ стаціонарний при $|a_1| < 1$ і $k \rightarrow \infty$, тобто стаціонарність проявляється на довгих часових інтервалах.

Якщо врахувати однорідний розв'язок (тобто розв'язок, зумовлений ненульовою початковою умовою для основної змінної), то отримаємо:

$$y(k) = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) + A a_1^k.$$

Тепер математичне сподівання

$$E[y(k)] = \frac{a_0}{1 - a_1} + A \cdot E[a_1^k]$$

залежить від часу і послідовність $\{y(k)\}$ буде стаціонарною тільки у тому випадку, коли $A a_1^k = 0$. Тобто для досягнення стаціонарності необхідно вибрати з початкових умов довільну константу $A = 0$. Величину $A a_1^k$ можна інтерпретувати як величину відхилення від довгострокової рівноваги.

Сформулюємо остаточно умови стаціонарності стохастичного процесу $AR(1)$:

а) модуль коефіцієнта $|a_1| < 1$ (корінь характеристичного рівняння);

б) однорідний розв'язок повинен дорівнювати нулю, тобто $A = 0$, або процес починається досить давно у минулому, що забезпечує збіжність $a_1^k \rightarrow 0$.

Умови стаціонарності процесу авторегресії довільного порядку

Отримані умови стаціонарності процесу $AR(1)$ легко розповсюджуються на процес авторегресії довільного порядку $AR(p)$. Однорідний розв'язок рівняння $AR(p)$ має вигляд:

$$y^h(k) = \sum_{i=1}^p A_i \alpha_i^k,$$

а якщо є кратні корені, то

$$y^h(k) = \alpha^k \sum_{i=1}^m A_i k^{i-1} + \sum_{i=m+1}^p A_i \alpha_i^k, \quad (2.2.7)$$

де α – корінь характеристичного рівняння кратності m ; $\alpha_i, i = m+1, \dots, p$ – різні дійсні корені. Очевидно, що середнє, дисперсія та коваріація, обчислені для (2.2.7), будуть залежати від часу.

Таким чином, умовою стаціонарності процесу $AR(p)$ є рівність нулю його однорідного розв'язку.

2.3 Умови стаціонарності часткового розв'язку рівняння АРКС

Розглянемо задачу визначення умов стаціонарності на прикладі. На практиці досить часто використовують модель АРКС(2, 1), яка виявляється адекватною для багатьох процесів:

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) + b_1 \varepsilon(k-1), \quad (2.3.1)$$

де $a_0 = 0$, оскільки 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:

$$y_{np}^p(k) = \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i), \quad (2.3.2)$$

і підставимо його у (2.3.1):

$$\begin{aligned} \alpha_0 \varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots = \\ = a_1 [\alpha_0 \varepsilon(k-1) + \alpha_1 \varepsilon(k-2) + \dots] + a_2 [\alpha_0 \varepsilon(k-2) + \alpha_1 \varepsilon(k-3) + \dots] + \\ + \varepsilon(k) + b_1 \varepsilon(k-1). \end{aligned}$$

Для того щоб отримати тотожність, прирівняємо коефіцієнти при однакових змінних:

$$\begin{aligned} \alpha_0 &= 1; \\ \alpha_1 &= a_1 \alpha_0 + b_1 \quad \text{або} \quad \alpha_1 = a_1 + b_1; \\ \alpha_2 &= a_1 \alpha_1 + a_2 \alpha_0; \\ &\vdots \\ \alpha_i &= a_1 \alpha_{i-1} + a_2 \alpha_{i-2}, \quad \forall i \geq 2. \end{aligned}$$

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 (2.3.1):

$$\lambda^2 - a_1 \lambda - a_2 = 0.$$

Таким чином, якщо корені цього характеристичного рівняння знаходяться всередині одиничного кола на комплексній площині, то $\{\alpha_i\}$ – збіжний процес. Це і є умовою стаціонарності неоднорідного розв'язку рівняння АРКС(2, 1).

Статистичні характеристики послідовності, згенерованої за допомогою пробної функції $\{y_{np}^p(k) = \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i)\}$

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:

$$E[y^p(k)] = E[\alpha_0 \varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots] = 0,$$

а також

$$E[y^p(k)] = E[y^p(k-s)] = 0, \quad \forall k, s. \quad (2.3.3)$$

б) дисперсія:

$$\begin{aligned} \text{var}[y^p(k)] &= E[\alpha_0\varepsilon(k) + \alpha_1\varepsilon(k-1) + \alpha_2\varepsilon(k-2) + \dots]^2 = \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha_i^2, \end{aligned} \quad (2.3.4)$$

тобто, $\text{var}[y^p(k)] = \text{var}[y^p(k-s)], \forall k, s$.

в) автоковаріація:

$s = 1$:

$$\begin{aligned} \text{cov}[y^p(k), y^p(k-1)] &= E\{[\varepsilon(k) + \alpha_1\varepsilon(k-1) + \alpha_2\varepsilon(k-2) + \dots] \times \\ &\quad \times [\varepsilon(k-1) + \alpha_1\varepsilon(k-2) + \alpha_2\varepsilon(k-3) + \dots]\} = \\ &= \sigma_\varepsilon^2 (\alpha_1 + \alpha_2\alpha_1 + \alpha_3\alpha_2 + \dots). \end{aligned}$$

$s = 2$:

$$\begin{aligned} \text{cov}[y^p(k), y^p(k-2)] &= E\{[\varepsilon(k) + \alpha_1\varepsilon(k-1) + \alpha_2\varepsilon(k-2) + \dots] \times \\ &\quad \times [\varepsilon(k-2) + \alpha_1\varepsilon(k-3) + \alpha_2\varepsilon(k-4) + \dots]\} = \\ &= \sigma_\varepsilon^2 (\alpha_2 + \alpha_3\alpha_1 + \alpha_4\alpha_2 + \dots) \end{aligned}$$

і для довільного значення параметра s можна записати:

$$\text{cov}[y(k), y(k-s)] = \sigma_\varepsilon^2 (\alpha_s + \alpha_{s+1}\alpha_1 + \alpha_{s+2}\alpha_2 + \dots) = \text{const}, \quad (2.3.5)$$

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 у такому випадку послідовність $\{y^p(k)\}$ також буде розбіжною.

Приклад 2.1. Визначимо умови стаціонарності для процесу АРКС(2,1)

$$y(k) = 1,6y(k-1) - 0,9y(k-2) + \varepsilon(k) + 0,5\varepsilon(k-1). \quad (2.3.6)$$

За допомогою методу невизначених коефіцієнтів знайдемо коефіцієнти неоднорідного розв'язку, зумовленого наявністю $\varepsilon(k)$:

$$\begin{aligned} \alpha_0 &= 1; \\ \alpha_1 &= 1,6 + 0,5 = 2,1; \\ &\vdots \\ \alpha_i &= 1,6\alpha_{i-1} - 0,9\alpha_{i-2}, \quad \forall i \geq 2, \end{aligned} \quad (2.3.7)$$

тобто, (2.3.7) – це рівняння авторегресії другого порядку для коефіцієнтів неоднорідного розв'язку. Знайдемо розв'язок характеристичного рівняння для (2.3.6) та (2.3.7):

$$\lambda^2 - 1,6\lambda + 0,9 = 0;$$

$$\lambda_{1,2} = \frac{1}{2}(1,6 \pm \sqrt{2,56 - 3,6}) = \frac{1}{2}(1,6 \pm j1,02); \quad j = \sqrt{-1}.$$

Розв'язок однорідного рівняння (2.3.7) у полярних координатах має вигляд:

$$\alpha^h(i) = \beta_1 r^i \cos(\theta i + \beta_2),$$

де $r = \sqrt{-a_2} = \sqrt{0,9} = 0,949$; $\cos(\theta) = \frac{a_1}{2(a_2)^{1/2}} = \frac{1,6}{2 \cdot 0,949} = 0,843$, а звідси значення кута в радіанах: $\theta = 0,567$.

Таким чином, можна конкретизувати однорідний розв'язок

$$\alpha^h(i) = \beta_1 (0,949)^i \cos(0,567i + \beta_2). \quad (2.3.8)$$

Константи β_1 і β_2 знайдемо за допомогою початкових умов: $\alpha_0 = 1, \alpha_1 = 2,1$:

$$\begin{cases} 1 = \beta_1 \cos(\beta_2) \\ 2,1 = 0,949\beta_1 \cos(0,567 + \beta_2) \end{cases}$$

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

$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta),$$

знайдемо, що $\beta_2 = -1,197$. Таким чином, розв'язок, із врахуванням початкових умов, приймає вигляд:

$$\alpha(i) = -\frac{1}{\cos(1,197)} \cdot 0,949^i \cos(0,567i - 1,197). \quad (2.3.9)$$

З (2.3.9) випливає, що загальна тенденція зміни коефіцієнтів α_i у часі буде в сторону зменшення їх значень. Деякі значення цих коефіцієнтів наведено в таблиці 6.1.

Таблиця 2.1

i	0	1	2	3	4	5	6	7	8	9	10
α_i	1,0	2,1	2,46	2,046	1,06	-0,146	1,187	-1,786	-1,761	-1,22	-0,37

Якщо продовжити обчислення коефіцієнтів α_i , то побачимо, що цей процес збіжний – згасаючі гармонічні коливання. Звідси випливає, що процес $\{y(k)\}$ також повинен бути збіжним.

2.4 Узагальнення умов стаціонарності на процеси АРКС(p, q) довільного порядку

У даному параграфі розглянемо задачу визначення умов стаціонарності для процесів ковзного середнього та авторегресії довільного порядку.

Стаціонарність процесу КС(∞)

Процес КС(∞) описується рівнянням:

$$x(k) = \sum_{i=0}^{\infty} b_i \varepsilon(k-i), \quad (2.4.1)$$

де $\{\varepsilon(k)\}$ – процес білого шуму із скінченною постійною дисперсією σ_ε^2 . Необхідно встановити чи буде процес $x(k)$ стаціонарним за коваріацією. Запишемо вирази для математичного сподівання, дисперсії та коваріації процесу (2.4.1):

а) математичне сподівання:

$$\begin{aligned} \mu_x &= E[x(k)] = E[\varepsilon(k) + b_1\varepsilon(k-1) + b_2\varepsilon(k-2) + \dots] = 0; \\ \mu_x &= E[x(k)] = E[x(k-s)] = 0; \end{aligned}$$

б) дисперсія:

$$\begin{aligned} \text{var}[x(k)] &= E\{[\varepsilon(k) + b_1\varepsilon(k-1) + b_2\varepsilon(k-2) + \dots]^2\} = \\ &= E[\varepsilon(k)]^2 + b_1^2 E[\varepsilon(k-1)]^2 + b_2^2 E[\varepsilon(k-2)]^2 + \dots = \\ &= \sigma_\varepsilon^2 [1 + b_1^2 + b_2^2 + \dots]. \end{aligned} \quad (2.4.2)$$

Очевидно, що $\text{var}[x(k-s)] = \sigma_\varepsilon^2 [1 + b_1^2 + b_2^2 + \dots]$ і

$$\text{var}[x(k)] = \text{var}[x(k-s)] = \text{const}, \quad \forall k, s.$$

Звідси випливає, що умовою стаціонарності процесу $\{x(k)\}$ є скінченність суми $\sum b_i^2$.

в) автоковаріація:

$$\begin{aligned} \text{cov}[x(k), x(k-s)] &= E\{[\varepsilon(k) + b_1\varepsilon(k-1) + b_2\varepsilon(k-2) + \dots] \times \\ &\quad \times [\varepsilon(k-s) + b_1\varepsilon(k-s-1) + b_2\varepsilon(k-s-2) + \dots]\} = \\ &= \sigma_\varepsilon^2 (b_s + b_1 b_{s+1} + b_2 b_{s+2} + \dots). \end{aligned} \quad (2.4.3)$$

Припустимо, що сума $(b_s + b_1 b_{s+1} + b_2 b_{s+2} + \dots)$ скінченна. У 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 коефіцієнтів:

$$\sum b_i^2 \quad \text{і} \quad (b_s + b_1 b_{s+1} + b_2 b_{s+2} + \dots). \quad (2.4.4)$$

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 $AR(p)$

Визначимо обмеження стаціонарності на коефіцієнти авторегресії довільного порядку p . Частковий розв'язок рівняння

$$y(k) = a_0 + \sum_{i=0}^p a_i y(k-i) + \varepsilon(k) \quad (2.4.5)$$

має вигляд:

$$y^p(k) = \frac{a_0}{1 - \sum_{i=1}^p a_i} + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i). \quad (2.4.6)$$

За аналогією з рівнянням другого порядку у даному випадку також можна знайти різницеве рівняння для обчислення коефіцієнтів α_i :

$$\alpha_i = a_1\alpha_{i-1} + a_2\alpha_{i-2} + \dots + a_p\alpha_{i-p}. \quad (2.4.7)$$

Якщо корені характеристичного рівняння, записаного для (2.4.7), 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 (2.4.6). Для цього знайдемо вирази для математичного сподівання, дисперсії та коваріації:

а) математичне сподівання:

$$E[y(k)] = E[y(k-s)] = \frac{a_0}{1 - \sum a_i} = \text{const}.$$

Ця величина буде скінченною, якщо $1 - \sum a_i \neq 0$, тобто це умова того, що корені відповідного характеристичного рівняння знаходяться всередині одиничного кола на комплексній площині.

б) Дисперсія:

$$\begin{aligned} \text{var}[y(k)] &= E\{[\varepsilon(k) + \alpha_1\varepsilon(k-1) + \alpha_2\varepsilon(k-2) + \dots]^2\} = \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha_i^2. \\ \text{var}[y(k-s)] &= E\{[\varepsilon(k-s) + \alpha_1\varepsilon(k-s-1) + \alpha_2\varepsilon(k-s-2) + \dots]^2\} = \\ &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \alpha_i^2. \end{aligned}$$

Таким чином, дисперсія буде скінченною, якщо сума $\sum \alpha_i^2$ скінченна.

в) Автоковаріація:

$$\begin{aligned} \gamma_s = \text{cov}[y(k), y(k-s)] &= E\{[\varepsilon(k) + \alpha_1\varepsilon(k-1) + \alpha_2\varepsilon(k-2) + \dots] \times \\ &\times [\varepsilon(k-s) + \alpha_1\varepsilon(k-s-1) + \alpha_2\varepsilon(k-s-2) + \dots]\} = \\ &= \sigma_\varepsilon^2 (\alpha_s + \alpha_1\alpha_{s+1} + \alpha_2\alpha_{s+2} + \dots) = \text{const} \end{aligned}$$

і не залежить від часу. Таким чином, процес $AR(p)$ буде стаціонарним, якщо відповідні суми коефіцієнтів будуть збіжними.

Процес АРКС(p, q) в цілому

Розглянемо тепер процес АРКС(p, q) в цілому:

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j), \quad b_0 = 0. \quad (2.4.8)$$

Запишемо неоднорідний розв'язок за допомогою лагового оператора:

$$y(k) = \frac{a_0}{1 - \sum_{i=1}^p a_i} + \frac{\varepsilon(k)}{1 - \sum_{i=1}^p a_i L^i} + \frac{b_1 \varepsilon(k-1)}{1 - \sum_{i=1}^p a_i L^i} + \dots \quad (2.4.9)$$

Перший член у правій частині (2.4.9) є константою, а всі інші будуть стаціонарними, якщо корені поліномів $(1 - \sum_{i=1}^p a_i L^i)$ лежать всередині одиничного кола.

Таким чином, якщо послідовність $\{x(k)\}$, яка описується рівнянням (2.4.1), стаціонарна, 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 (2.4.1), стаціонарна, то і послідовність $\{y(k)\}$ буде стаціонарною за умови, що корені характеристичного рівняння для авторегресійної частини знаходяться всередині одиничного кола на комплексній площині.

2.5 Теоретична автокореляційна функція

У попередньому параграфі ми встановили, що для визначення стаціонарності процесів необхідно обчислювати три основні статистичні характеристики: *математичне сподівання, дисперсію і коваріацію*. Ці параметри можна обчислювати безпосередньо за вибіркою, тобто знаходити *вибіркові характеристики*, або за допомогою математичної моделі – *теоретичні характеристики*.

Розглянемо розрахунок теоретичних статистичних характеристик для моделі АР(1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad E[\varepsilon(k)] = 0,$$

$$E[\varepsilon(k)\varepsilon(k-j)] = \begin{cases} \sigma_\varepsilon^2, & j=0, \\ 0, & j \geq 1; \end{cases} \quad E[\varepsilon(k)\varepsilon(y-j)] = \begin{cases} \sigma_\varepsilon^2, & j=0, \\ 0, & j \geq 1. \end{cases}$$

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:

$$\sigma_y^2 = \text{var}[y(k)] = \frac{\sigma_\varepsilon^2}{1 - a_1^2} = \gamma_0, \quad |a_1| < 1. \quad (2.5.1)$$

Нагадаємо, що вибірка дисперсія часового ряду $\{y(k)\}$ обчислюється за виразом:

$$\sigma_y^2 = \frac{1}{N-1} \sum_{k=1}^N [y(k) - \bar{y}]^2,$$

де \bar{y} – вибіркове середнє значення ряду.

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 АР(1) для моменту $(k+1)$:

$$y(k+1) = a_1 y(k) + \varepsilon(k+1) = a_1 [a_1 y(k-1) + \varepsilon(k)] + \varepsilon(k+1),$$

$$\begin{aligned} \gamma(1) &= \text{cov}[y(k), y(k+1)] = \\ &= E\{[a_1 y(k-1) + \varepsilon(k)][a_1^2 y(k-1) + a_1 \varepsilon(k) + \varepsilon(k+1)]\} = \\ &= a_1^3 \text{var}[y(k-1)] + a_1 \text{var}[\varepsilon(k)] = a_1^3 \frac{\sigma_y^2}{1 - a_1^2} + a_1 \sigma_\varepsilon^2. \end{aligned}$$

Покладаючи $\sigma_y^2 \approx \sigma_\varepsilon^2$ (що не входить в протиріччя із означенням випадкових процесів), отримаємо:

$$\gamma(1) = \text{cov}[y(k), y(k+1)] = \frac{a_1^3 \sigma_\varepsilon^2 + a_1 \sigma_\varepsilon^2 - a_1^3 \sigma_\varepsilon^2}{1 - a_1^2} = \frac{a_1 \sigma_\varepsilon^2}{1 - a_1^2}. \quad (2.5.2)$$

Для довільного значення s тепер можна записати:

$$\gamma(s) = E\{[y(k) - \mu_y], [y(k-s) - \mu_y]\} = \frac{\sigma_\varepsilon^2 a_1^s}{1 - a_1^2}. \quad (2.5.3)$$

Автокореляційну функцію (АКФ) можна знайти шляхом нормування коваріаційної функції діленням на $\gamma(0) = \gamma_0$:

$$\rho(s) = \frac{\gamma(s)}{\gamma_0}; \quad \rho_0 = 1, \rho(1) = a_1, \rho(2) = a_1^2, \dots, \rho(s) = a_1^s.$$

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 встановити адекватність побудованої моделі процесу. Крім того, теоретичну АКФ використовують для оцінювання стаціонарності процесів.

2.6 Обчислення кореляційних функцій за допомогою рівнянь 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

$$\begin{aligned} y(k) &= a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k), \\ \{\varepsilon(k)\} &\sim N(0, \sigma_\varepsilon^2), \end{aligned} \quad (2.6.1)$$

на $y(k-s)$, $s=0,1,2,\dots$ і знайдемо математичне сподівання обох частин. Зазначимо, що перетин a_0 не впливає на кореляційну функцію, а тому його виключено з рівняння (2.6.1).

Система рівнянь Юла-Уокера у даному випадку приймає вигляд:

$$s=0: E[y(k)y(k)] = a_1 E[y(k-1)y(k)] + a_2 E[y(k-2)y(k)] + E[\varepsilon(k)y(k)],$$

$$s=1: E[y(k)y(k-1)] = a_1 E[y(k-1)y(k-1)] + a_2 E[y(k-2)y(k-1)] + E[\varepsilon(k)y(k-1)],$$

$$\begin{aligned}
 s = 2: \quad E[y(k) y(k-2)] &= a_1 E[y(k-1) y(k-2)] + a_2 E[y(k-2) y(k-2)] + \\
 &\quad + E[\varepsilon(k) y(k-2)] \\
 &\quad \vdots \qquad \qquad \qquad (2.6.2)
 \end{aligned}$$

$$\exists s: \quad E[y(k) y(k-s)] = a_1 E[y(k-1) y(k-s)] + a_2 E[y(k-2) y(k-s)] + \\
 + E[\varepsilon(k) y(k-s)]$$

За означенням коваріація обчислюється так:

$$E[y(k) y(k-s)] = E[y(k-s) y(k)] = \gamma_s,$$

а дисперсія ряду $\{y(k)\}$ визначається дисперсією випадкової змінної $\varepsilon(k)$, тобто, $E[\varepsilon(k) y(k)] \approx \sigma_\varepsilon^2$. Крім того, прийнемо, що $b_0 = 1$ при $\varepsilon(k)$, і $\varepsilon(k)$ не повинно бути корельованим з $y(k-s)$, $s \geq 1$ або $E[\varepsilon(k) y(k-s)] = 0$ при $s \geq 1$.

Запишемо систему (2.6.2) через коваріації $\gamma_0, \gamma_1, \dots, \gamma_s$:

$$\begin{cases}
 \gamma_0 = a_1 \gamma_1 + a_2 \gamma_2 + \sigma_\varepsilon^2 \\
 \gamma_1 = a_1 \gamma_0 + a_2 \gamma_1 \\
 \quad \quad \quad \vdots \\
 \gamma_s = a_1 \gamma_{s-1} + a_2 \gamma_{s-2}
 \end{cases} \qquad (2.6.3)$$

і поділимо два останніх рівняння на γ_0 :

$$\begin{cases}
 \rho_1 = a_1 \rho_0 + a_2 \rho_1, \\
 \rho_s = a_1 \rho_{s-1} + a_2 \rho_{s-2}.
 \end{cases} \qquad (2.6.4)$$

Оскільки $\rho_0 = 1$, то $\rho_1 = a_1 / (1 - a_2)$. Всі інші значення дискретної кореляційної функції можна знайти за допомогою рекурсивної формули:

$$\begin{aligned}
 \rho_s &= a_1 \rho_{s-1} + a_2 \rho_{s-2}, \\
 \text{або} \quad \rho(s) &= a_1 \rho(s-1) + a_2 \rho(s-2).
 \end{aligned} \qquad (2.6.5)$$

Так,

$$\rho(2) = \frac{a_1}{1 - a_2} + a_2, \quad \rho(3) = a_1 \left(\frac{a_1^2}{1 - a_2} + a_2 \right) + \frac{a_2 a_1}{1 - a_2}.$$

АКФ процесу ковзного середнього 1-го порядку КС(1)

Для процесу КС(1), $y(k) = \varepsilon(k) + b_1 \varepsilon(k-1)$, рівняння Юла-Уокера мають вигляд:

$$\begin{aligned}\gamma_0 &= E[y(k)y(k)] = E\{[\varepsilon(k) + b_1\varepsilon(k-1)][\varepsilon(k) + b_1\varepsilon(k-1)]\} = \\ &= \sigma_\varepsilon^2 + b_1^2 \sigma_\varepsilon^2 = (1 + b_1^2) \sigma_\varepsilon^2, \\ \gamma_1 &= E[y(k)y(k-1)] = E\{[\varepsilon(k) + b_1\varepsilon(k-1)][\varepsilon(k-1) + b_1\varepsilon(k-2)]\} = \\ &= b_1 \sigma_\varepsilon^2, \\ \gamma_2 &= E[y(k)y(k-2)] = E\{[\varepsilon(k) + b_1\varepsilon(k-1)][\varepsilon(k-2) + b_1\varepsilon(k-3)]\} = 0 \\ \text{і} \quad \gamma_s &= 0.\end{aligned}$$

Поділимо γ_0, γ_1 на γ_0 і отримаємо:

$$\rho_0 = 1, \quad \rho_1 = \frac{b_1}{1 + b_1^2}, \quad \rho_s = 0, \quad \forall s > 1,$$

тобто, АКФ процесу КС(1) має тільки два значення, відмінних від нуля.

АКФ процесу АРКС(1,1)

Для рівняння

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) + b_1 \varepsilon(k-1)$$

необхідно знайти коваріаційну функцію γ_s і АКФ ρ_s . За означенням можна записати:

$$\gamma_s = E\{[y(k) - \mu_y][y(k-s) - \mu_y]\},$$

$$E[\varepsilon(k)y(k-s)] = 0, \quad s \geq 1,$$

$$E[\varepsilon(k)y(k)] = \sigma_\varepsilon^2,$$

$$\begin{aligned}E[\varepsilon(k-1)y(k)] &= E\{\varepsilon(k-1)[a_1 y(k-1) + \varepsilon(k) + b_1 \varepsilon(k-1)]\} = \\ &= a_1 \sigma_\varepsilon^2 + b_1 \sigma_\varepsilon^2 = \sigma_\varepsilon^2 (a_1 + b_1).\end{aligned}$$

Рівняння Юла-Уокера у даному випадку мають вигляд:

$$\begin{aligned}\gamma_0 &= E[y(k)y(k)] = a_1 E[y(k)y(k-1)] + \\ &+ E[y(k)\varepsilon(k)] + b_1 E[y(k)\varepsilon(k-1)] =\end{aligned}$$

$$= a_1 \gamma_1 + \sigma_\varepsilon^2 + b_1(a_1 + b_1) \sigma_\varepsilon^2,$$

$$\begin{aligned} \gamma_1 &= E[y(k) y(k-1)] = a_1 E[y(k-1) y(k-1)] + \\ &+ E[y(k-1) \varepsilon(k)] + b_1 E[y(k-1) \varepsilon(k-1)] = \\ &= a_1 \gamma_0 + b_1 \sigma_\varepsilon^2. \end{aligned}$$

$$\begin{aligned} \gamma_2 &= E[y(k) y(k-2)] = a_1 E[y(k-1) y(k-2)] + \\ &+ E[y(k-2) \varepsilon(k)] + b_1 E[y(k-2) \varepsilon(k-1)] = \\ &= a_1 \gamma_1. \end{aligned}$$

⋮

$$\begin{aligned} \gamma_s &= E[y(k) y(k-s)] = a_1 E[y(k-1) y(k-s)] + \\ &+ E[y(k-s) \varepsilon(k)] + b_1 E[y(k-s) \varepsilon(k-1)] = \\ &= a_1 \gamma_{s-1}. \end{aligned}$$

Запишемо отримані рівняння у скороченій формі:

$$\begin{aligned} \gamma_0 &= a_1 \gamma_1 + \sigma_\varepsilon^2 + b_1(a_1 + b_1) \sigma_\varepsilon^2, \\ \gamma_1 &= a_1 \gamma_0 + b_1 \sigma_\varepsilon^2, \\ \gamma_2 &= a_1 \gamma_1, \\ \gamma_3 &= a_1 \gamma_2, \\ &\vdots \\ \gamma_s &= a_1 \gamma_{s-1}. \end{aligned}$$

Розв'язуючи цю систему рівнянь, отримаємо вирази для коваріацій:

$$\gamma_0 = \frac{1 + b_1^2 + 2a_1 b_1}{1 - a_1^2} \cdot \sigma_\varepsilon^2; \quad \gamma_1 = \frac{(1 + a_1 b_1)(a_1 + b_1)}{1 + b_1^2 + 2a_1 b_1};$$

а звідси знайдемо коефіцієнти кореляції:

$$\rho_0 = 1, \quad \rho_1 = \frac{(1 + a_1 b_1)(a_1 + b_1)}{1 + b_1^2 + 2a_1 b_1}, \quad \rho_s = a_1 \rho_{s-1}, \quad \forall s > 1.$$

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

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) + b_1 \varepsilon(k-1)$$

необхідно знайти коваріаційну функцію γ_s і АКФ ρ_s . За означенням,

$$\gamma_s = E\{[y(k) - \mu_y][y(k-s) - \mu_y]\},$$

$$E[\varepsilon(k) y(k-s)] = 0, \quad s \geq 1,$$

$$E[\varepsilon(k) y(k)] = \sigma_\varepsilon^2,$$

$$\begin{aligned} E[\varepsilon(k-1) y(k)] &= E\{\varepsilon(k-1)[a_1 y(k-1) + \varepsilon(k) + b_1 \varepsilon(k-1)]\} = \\ &= a_1 \sigma_\varepsilon^2 + b_1 \sigma_\varepsilon^2 = \sigma_\varepsilon^2 (a_1 + b_1). \end{aligned}$$

Рівняння Юла-Уокера:

$$\begin{aligned} \gamma_0 &= E[y(k) y(k)] = a_1 E[y(k) y(k-1)] + a_2 E[y(k) y(k-2)] + \\ &\quad + E[y(k) \varepsilon(k)] + b_1 E[y(k) \varepsilon(k-1)] = \\ &= a_1 \gamma_1 + a_2 \gamma_2 + \sigma_\varepsilon^2 + b_1 (a_1 + b_1) \sigma_\varepsilon^2, \end{aligned}$$

$$\begin{aligned} \gamma_1 &= E[y(k) y(k-1)] = a_1 E[y(k-1) y(k-1)] + a_2 E[y(k-1) y(k-2)] + \\ &\quad + E[y(k-1) \varepsilon(k)] + b_1 E[y(k-1) \varepsilon(k-1)] = \\ &= a_1 \gamma_0 + a_2 \gamma_1 + b_1 \sigma_\varepsilon^2. \end{aligned}$$

$$\begin{aligned} \gamma_2 &= E[y(k) y(k-2)] = a_1 E[y(k-1) y(k-2)] + a_2 E[y(k-2) y(k-2)] + \\ &\quad + E[y(k-2) \varepsilon(k)] + b_1 E[y(k-2) \varepsilon(k-1)] = \\ &= a_1 \gamma_1 + a_2 \gamma_0. \end{aligned}$$

$$\begin{aligned} \gamma_3 &= E[y(k) y(k-3)] = a_1 E[y(k-1) y(k-3)] + a_2 E[y(k-2) y(k-3)] + \\ &\quad + E[y(k-3) \varepsilon(k)] + b_1 E[y(k-3) \varepsilon(k-1)] = \\ &= a_1 \gamma_2 + a_2 \gamma_1. \end{aligned}$$

⋮

$$\gamma_s = a_1 \gamma_{s-1} + a_2 \gamma_{s-2}.$$

Перейдемо до скороченої форми запису:

$$\begin{aligned}\gamma_0 &= a_1\gamma_1 + a_2\gamma_2 + \sigma_\varepsilon^2 [1 + b_1(a_1 + b_1)], \\ \gamma_1 &= a_1\gamma_0 + a_2\gamma_1 + b_1\sigma_\varepsilon^2, \\ \gamma_2 &= a_1\gamma_1 + a_2\gamma_0, \\ \gamma_3 &= a_1\gamma_2 + a_2\gamma_1, \\ \gamma_4 &= a_1\gamma_3 + a_2\gamma_2, \\ &\vdots \\ \gamma_s &= a_1\gamma_{s-1} + a_2\gamma_{s-2}.\end{aligned}$$

Розв'язуючи цю систему рівнянь, знайдемо вирази для коваріацій:

$$\gamma_0 = \frac{a_1(1+a_2)c_3 + c}{1 - [a_1c_2(1+a_2) + a_2^2]}, \quad \gamma_1 = \frac{a_1\gamma_0 + c_1}{1 - a_2},$$

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:

$$\rho_i = a_1\rho_{i-1} + a_2\rho_{i-2} + \dots + a_p\rho_{i-p}.$$

Тобто декілька перших значень коефіцієнтів кореляції дають можливість обчислити оцінки всіх інших, необхідних для аналізу, значень.

2.7 Часткова автокореляційна функція

Перед тим як розглядати часткову автокореляційну функцію визначимо деякі особливості звичайної АКФ. Так, АКФ процесу АР(1) обчислюється за виразами:

$$\rho_0 = 1, \quad \rho(1) = a_1, \quad \rho(2) = a_1^2, \dots, \quad \rho(s) = a_1^s,$$

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 $y(k-2)$:

$$\rho_2 = \rho_1 \cdot \rho_1 = \rho_1^2.$$

Очевидно, що такі “непрямі” кореляції наявні в АКФ будь-якого процесу авторегресії. В цілому це призводить до того, що АКФ недостатньо 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. 2.1).

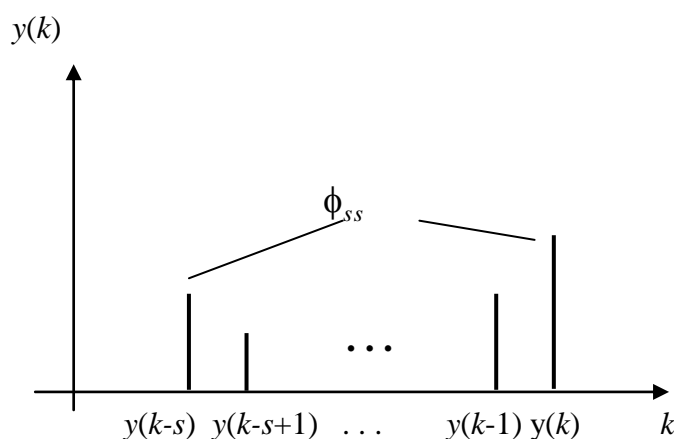


Рис. 2.1 Коефіцієнт часткової кореляції між $y(k)$ і $y(k-s)$

Для процесу $AR(1)$ значення ЧАКФ між $y(k)$ і $y(k-2)$ дорівнює нулю за означенням. Звідси випливає, що у 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 дуже корисною при розв’язанні задач ідентифікації моделей процесів $AR(p)$.

Однак це не відноситься до процесів ковзного середнього. Так, для процесу $КС(1)$

$$y(k) = \varepsilon(k) + b_1 \varepsilon(k-1)$$

можна записати $y(k) = (1 + b_1 L) \varepsilon(k)$ або, скориставшись властивістю операторів зсуву, отримаємо:

$$\varepsilon(k) = \frac{y(k)}{1 + b_1 L} = y(k) - b_1 y(k-1) + b_1^2 y(k-2) - b_1^3 y(k-3) + \dots, \quad |b_1| < 1.$$

З цього виразу випливає, що значення $y(k)$ буде корельованим із всіма іншими, затриманими у часі значеннями, і ЧАКФ цього процесу не спадає різко до нуля.

Коефіцієнти ЧАКФ можна обчислити безпосередньо за допомогою значень часового ряду. Процедуру обчислень можна представити у вигляді двох кроків:

1. Сформуванати вибірку з відхилень від середнього:

$$\{y'(k)\} = \{y(k)\} - \mu_y.$$

2. Оцінити рівняння необхідного порядку:

$$y'(k) = \phi_{11} y'(k-1) + e(k), \quad (2.7.1)$$

де $e(k)$ – похибка моделі, яка може і не бути білим шумом. Оскільки рівняння (2.7.1) не містить інших значень крім $y'(k)$ та $y'(k-1)$, то ϕ_{11} є коефіцієнтом часткової кореляції між цими значеннями.

За аналогією можна сформуванати та оцінити параметри рівнянь другого порядку

$$y'(k) = \phi_{21} y'(k-1) + \phi_{22} y'(k-2) + e(k),$$

де ϕ_{22} – коефіцієнт часткової кореляції між $y'(k)$ і $y'(k-2)$, який не враховує впливу значення $y'(k-1)$.

В результаті прийдемо до обчислення ЧАКФ за такими виразами:

$$\phi_{11} = \rho_1, \quad \phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}; \quad \phi_{ss} = \frac{\rho_s - \sum_{j=1}^{s-1} \phi_{s-1,j} \rho_{s-j}}{1 - \sum_{j=1}^{s-1} \phi_{s-1,j} \rho_j}, \quad (2.7.2)$$

де $\phi_{sj} = \phi_{s-1,j} - \phi_{ss} \phi_{s-1,s-j}$, $j = 1, 2, 3, \dots, s-1$.

Приклад обчислення ЧАКФ

Обчислити АКФ і ЧАКФ для процесу АРКС(1,1):

$$y(k) = -0,7y(k-1) + \varepsilon(k) - 0,7\varepsilon(k-1).$$

- а) Коефіцієнти АКФ обчислимо за допомогою отриманих вище виразів: $\rho_0 = 1$;

$$\rho_1 = \frac{(1 + a_1 b_1)(a_1 + b_1)}{1 + b_1^2 + 2a_1 b_1} = \frac{(1 + 0,49)(-0,7 - 0,7)}{1 + 0,49 + 2 \cdot 0,49} = -0,8445.$$

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:

$$\phi_{11} = \rho_1 = -0,8445;$$

$$\phi_{22} = (\rho_2 - \rho_1^2)/(1 - \rho_1^2) = \frac{0,591 - (-0,8445)^2}{1 - (-0,8445)^2} = -0,425;$$

$$\phi_{21} = \phi_{11} - \phi_{22}\phi_{11} = -1,204;$$

$$\phi_{33} = \frac{\rho_3 - \sum_{j=1}^2 \phi_{2,j} \rho_{3-j}}{1 - \sum_{j=1}^2 \phi_{2,j} \rho_j} = -0,262; \quad \phi_{44} = \frac{\rho_4 - \sum_{j=1}^3 \phi_{3,j} \rho_{4-j}}{1 - \sum_{j=1}^3 \phi_{3,j} \rho_j} = -0,173,$$

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:

$$\phi_{55} = -0,117; \quad \phi_{66} = -0,0681; \quad \phi_{77} = -0,056; \quad \phi_{88} = -0,039.$$

У загальному випадку ЧАКФ стаціонарного процесу $АРКС(p,q)$ повинна збігатись до нуля, починаючи з p -го значення, а швидкість збіжності залежить від значень коефіцієнтів полінома

$$1 + b_1 L + b_2 L^2 + \dots + b_q L^q = 0.$$

Висновки стосовно стаціонарних процесів

1. АКФ процесу $АРКС(p,q)$ починає збігатись до нуля при значеннях зсуву $s = q$. Коефіцієнти АКФ задовольняють різницевому рівнянню:

$$\rho_i = a_1 \rho_{i-1} + a_2 \rho_{i-2} + \dots + a_p \rho_{i-p}.$$

Оскільки корені характеристичного рівняння знаходяться всередині одиничного кола на комплексній площині, то значення АКФ будуть збігатись до нуля при значеннях зсуву $s \geq q$.

2. ЧАКФ процесу $АРКС(p,q)$ починає збігатись до нуля при $s \geq p$. Починаючи із значення $s = p$ значення ЧАКФ за своїм характером відповідають значенням АКФ.

Ці властивості кореляційних функцій дають можливість вибрати порядок моделі $АРКС(p,q)$. Так, для деяких конкретних стаціонарних процесів АКФ і ЧАКФ мають такі значення [...]:

- *білий шум*: всі $\rho_s = 0$, $\phi_{ss=0}$;
- *процес $АР(1)$* при $-1 < a_1 < 0$: $\rho_s = a_1^s$ – експоненційно спадають; $\phi_{11} = \rho_1$; $\phi_{ss} = 0$, $s > 1$;
- *процес $АР(p)$* : 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 $КС(1)$* 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 $КС(1)$* при $-1 < b_1 < 0$: ρ_1 має додатне значення і $\rho_s = 0$, $s > 1$; $\phi_{ss} > 0$ і мають коливальну збіжність;
- *процес $АРКС(1,1)$* 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 $АРКС(1,1)$* при $-1 < a_1 < 0$: ρ_s мають коливальну збіжність і $sign(\rho_1) = sign(a_1 + b_1)$; $\phi_{11} = \rho_1$, далі настає гладка збіжність; $sign(\phi_{ss}) = sign(\phi_{11})$;
- *процес $АРКС(p,q)$* : значення ρ_s мають гладку або коливальну збіжність, починаючи із значення $s > q$; ϕ_{ss} мають гладку або коливальну збіжність, починаючи із $s > p$.

Означення статистичної значущості оцінок коефіцієнтів АКФ

Для того щоб коректно визначити порядок моделі $АРКС(p,q)$, необхідно користуватись відповідними статистиками, які дають можливість формально встановити статистичну значущість отриманих оцінок. Для стаціонарного процесу обчислені значення r_s мають нормальний розподіл (за умови, що дані мають такий же розподіл). Оскільки при визначенні порядку ми користуємось вибірковими статистичними характеристиками, то для визначення статистичної значущості r_s (тобто суттєвої відмінності від нуля значень АКФ) можна скористатись статистикою Бокса-Пірса або Лунга-Бокса.

Статистика Бокса-Пірса або Q – статистика розраховується за виразом:

$$Q = N \sum_{i=1}^s r_i^2 .$$

Якщо досліджуваний процес АРКС стаціонарний, то статистика Q має розподіл χ^2 з s ступенями свободи. Це досить наближена статистика, але її використовують для визначення стаціонарності *довгих вибірок* даних.

Для *коротких вибірок* запропонована статистика Люнга-Бокса:

$$Q_{LB} = N(N + 2) \sum_{i=1}^s r_i^2 / (N - i),$$

яку використовують також при визначенні порядку процесу (він є також порядком відповідної моделі). Оскільки значення Q пропорціональне $\sum r_i^2$, то менші значення Q відповідають меншій ступені корельованості досліджуваного процесу. Порядок прийняття рішень стосовно значущості обчислених оцінок такий же, як і у випадку аналізу інших статистик.

Запитання і вправи до другого розділу

1. Дайте означення процесу авторегресії з ковзним середнім? Поясніть можливу природу випадкового процесу $\{\varepsilon(k)\}$ у правій частині рівняння. Як можна обчислити оцінки $\hat{\varepsilon}(k)$?
2. Які процеси називають інтегрованими? Наведіть приклади інтегрованих процесів. Які інші назви використовують для інтегрованих процесів?
3. Дайте означення слабкої стаціонарності? Чим відрізняється слабка стаціонарність від сильної? Які інші назви слабкої стаціонарності ви знаєте? Яке поняття стаціонарності використовують на практиці частіше?
4. У чому полягає різниця між теоретичними і вибірковими статистичними характеристиками? У яких випадках використовують теоретичні статистичні характеристики?
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:

$$i. \quad x(k) = 0,15\gamma(k-1) + 0,25\gamma(k-2) + 0,35\gamma(k-3) + 0,25\gamma(k-4),$$

$$\text{де } \gamma(k) \text{ – випадковий процес білого шуму: } E[\gamma(k)\gamma(l)] = \begin{cases} \sigma_\gamma^2, & k=l; \\ 0, & k \neq l. \end{cases}$$

7. Для авторегресійного процесу другого порядку

$$a. \quad y(k) = a_0 + a_2 y(k-2) + \varepsilon(k),$$

визначте такі статистичні характеристики:

(а) умовні математичні сподівання:

$$E_{k-2}[y(k)]; \quad E_{k-1}[y(k)]; \quad E_k[y(k+2)];$$

(б) коваріації: $\text{cov}[y(k), y(k-1)]; \quad \text{cov}[y(k), y(k-2)];$

(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 вираз для похибок оцінок прогнозів $f(s) = y(k+s) - \hat{y}(k+s)$ і теоретичну автокореляційну функцію для послідовності $\{f(s)\}$. Які статистичні характеристики мають похибки оцінок прогнозів?
9. Поясніть умови стаціонарності процесу АР(1)? Наведіть приклад фінансового чи економічного процесу, для якого доцільно виконати умови стаціонарності, тобто бажано щоб цей процес розвивався як стаціонарний.
10. Яка на вашу думку загальна мета аналізу стаціонарності процесів? Чи можна без нього обійтись при побудові математичних моделей процесів та об'єктів?
11. Для чого будують систему рівнянь Юла-Уокера? Які величини ми отримуємо в результаті знаходження розв'язку цієї системи рівнянь?
12. Як можна інтерпретувати однорідний розв'язок $A\lambda_i^k$ (λ_i – це i -й корінь характеристичного рівняння; k – дискретний час) при аналізі стаціонарності процесів АРКС?
13. Сформулюйте умови стаціонарності процесу ковзного середнього довільного порядку?
14. Сформулюйте умови стаціонарності для процесу АРКС довільного порядку?
15. При якому значенні параметра часового 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 оцінок АКФ?

Р о з д і л 3

ЗАСТОСУВАННЯ РІЗНИЦЕВИХ РІВНЯНЬ ДО ОПИСУ ДИНАМІКИ ПРОЦЕСІВ

3.1 Загальні відомості про різницеві рівняння

При використанні дискретних рівнянь (моделей) незалежну змінну, час t , замінюють дискретним часом, тобто, покладають $t = kT_s$, де T_s – період дискретизації вимірів, який у технічних системах приймає значення від десятків мікросекунд до десятків секунд і навіть хвилин, а при моделюванні фінансово-економічних процесів – від декількох хвилин до одного року, в залежності від того, які статистичні дані можна отримати. У моделях, які описують процеси у дискретному часі, період дискретизації вимірів, як правило, нормують до одиниці і незалежною змінною залишається k (дискретний час), яка приймає цілі значення від 0 до ∞ . При цьому для кожної прикладної задачі дискретна одиниця часу має відповідне фактичне значення.

Завдяки простоті структури та наявності надійних методів оцінювання параметрів різницеві рівняння (РР) знайшли надзвичайно широке застосування при моделюванні процесів у технічних системах, економіці та фінансах, екології, біології та інших прикладних і наукових галузях. Простим прикладом різницевого рівняння є стохастичне рівняння авторегресії першого порядку з одиничним коефіцієнтом (окремий випадок) при затриманому в часі значенні основної змінної:

$$y(k) = y(k-1) + \varepsilon(k), \quad (3.1.1)$$

де $y(k)$ – основна змінна; $\varepsilon(k)$ – випадкова величина, яка відображає вплив різноманітних невимірюваних факторів на основну змінну, неточність структури моделі та похибки обчислень її параметрів. В першу чергу це випадкові збурення, що діють на процес. Частіше всього припускають, що випадкова величина має нормальний розподіл: $\{\varepsilon(k)\} \sim N_n(0, \sigma_\varepsilon^2)$, тобто це некорельований процес із нульовим середнім та скінченною дисперсією σ_ε^2 . Наприклад, у випадку коли $y(k)$ – ціни на деякі біржові акції в k – й день, то $\varepsilon(k)$ характеризує коливання ціни під впливом багатьох випадкових факторів, які неможливо ввести в модель, оскільки неможливо отримати їх виміри, або ж ці впливи носять якісний характер. Як правило, у наведеній моделі процес $\{\varepsilon(k)\}$ має такі обмеження: $E[\varepsilon(k)] = 0$, $E[\varepsilon(k)\varepsilon(l)] = \begin{cases} \sigma_\varepsilon^2, & k = l, \\ 0, & k \neq l. \end{cases}$ Докладніше

роль випадкової змінної буде розглядатись окремо у кожному конкретному випадку моделювання.

За допомогою рівняння (3.1.1) описують, наприклад, ціну акції на біржі в момент часу, що відповідає аргументу k . Його називають ще рівнянням, яке описує *процес випадкового кроку* (випадкове блукання або *random walk*). Таку назву воно отримало з тієї причини, що приріст значення основної змінної визначається, фактично, випадковим процесом. Воно може бути записане також у формі першої різниці

$$\Delta y(k) = \varepsilon(k),$$

де $\Delta y(k) = y(k) - y(k-1)$. Саме можливість застосування різниць перших і вищих порядків до опису часових рядів зумовило використання назви *різницеві рівняння*.

Для рівняння випадкового кроку можна записати однорідне у вигляді:

$$y(k) - y(k-1) = 0,$$

яке має характеристичне рівняння: $\alpha - 1 = 0$ (як отримати характеристичне рівняння в загальному випадку, ми розглянемо нижче у даному розділі). Таким чином, дане характеристичне рівняння має один корінь $\alpha = 1$. Якщо характеристичне рівняння має хоча б один одиничний корінь, то кажуть, що відповідне йому різницеве рівняння описує *процес з одиничним коренем* (нестационарний процес з трендом). Нижче покажемо, що процеси з одиничними коренями – це *процеси з трендами* або *інтегровані процеси*. У подальших викладках під трендом будемо розуміти загальний довгостроковий напрям розвитку процесу; фактично він узгоджується з поточним середнім значенням.

Більш загальною формою різницевого рівняння (3.1.1) є така (процес авторегресії першого порядку – АР(1)):

$$\Delta y(k) = \alpha_0 + \alpha_1 y(k-1) + \varepsilon(k), \quad (3.1.2)$$

але для того, щоб воно відповідало процесу випадкового кроку необхідно покласти: $\alpha_0 = 0$, $\alpha_1 = 1$; інакше це рівняння вже не буде відповідати своєму означенню. Зазначимо, що порядок авторегресійної частини різницевого рівняння визначається кількістю минулих (затриманих) вимірів залежної змінної, які використовуються у правій частині для пояснення її зміни в часі.

Різницеві рівняння, у правій частині яких наявні минулі (затримані) виміри основної змінної, називають *авторегресійними* (АР), тобто регресія змінної на саму себе. Рівняння авторегресії n -го порядку має вигляд:

$$y(k) = a_0 + \sum_{i=1}^n a_i y(k-i) + \varepsilon(k). \quad (3.1.3)$$

Якщо на процес, що моделюється, впливає деяка вхідна змінна, то вона вона записується у правій частині:

$$y(k) = a_0 + \sum_{i=1}^n a_i y(k-i) + \sum_{j=1}^q b_j x(k-j) + \varepsilon(k). \quad (3.1.3)$$

Якщо $x(k)$ – випадковий процес, то таке рівняння називають авторегресією із ковзним середнім (АРКС). Формально, у випадку ковзного середнього також необхідно виконати умову: $\sum_{j=1}^q b_j = 1$.

Характеристичне рівняння, записане для (3.1.3), може мати одиничні корені, тобто один або більше коренів характеристичного рівняння можуть приймати значення «1». Такі процеси називають процесами авторегресії з інтегрованим ковзним середнім – АРІКС(p, d, q), де d – кількість одиничних коренів характеристичного рівняння. Процеси цього класу *нестационарні* – вони мають тренд, порядок якого визначається кількістю одиничних коренів. Якщо $d = 1$, то тренд лінійний; якщо $d = 2$, то тренд квадратичний і т. д. Таким чином, *процеси з трендом, інтегровані процеси і процеси з одиничними коренями* – це три різні назви нестационарних процесів, що містять тренд.

Застосування перших та різниць вищих порядків

Разом із основною змінною для опису процесів використовують *перші та другі різниці*, наприклад:

$$\Delta y(k) = y(k) - y(k-1),$$

$$\begin{aligned}\Delta y(k+1) &= y(k+1) - y(k), \\ \Delta y(k+2) &= y(k+2) - y(k+1).\end{aligned}$$

Наведені перші різниці відображають швидкість зміни основної змінної, що відповідає першій похідній для рівнянь, які описують модельований процес у неперервному часі, тобто для дифференціальних рівнянь.

Обчислення перших різниць приводить до видалення лінійного тренду з процесу. Наприклад, нехай тренд описується лінійними рівняннями:

$$y(k) = a_0 + a_1 k.$$

Перша різниця для цього процесу:

$$\Delta y(k) = y(k) - y(k-1) = a_0 + a_1 k - a_0 - a_1 \cdot (k-1) = a_1,$$

тобто, після дискретного диференціювання залишилась константа.

Другі різниці відображають швидкість зміни у часі перших різниць (тобто прискорення) і записуються так:

$$\begin{aligned}\Delta^2 y(k) &= \Delta(\Delta y(k)) = \Delta[y(k) - y(k-1)] = [y(k) - y(k-1)] - [y(k-1) - y(k-2)] = \\ &= y(k) - 2y(k-1) + y(k-2),\end{aligned}$$

$$\Delta^2 y(k+1) = \Delta(\Delta y(k+1)) = y(k+1) - 2y(k) + y(k-1).$$

Останній вираз називають різницевою схемою обчислення другої похідної. На практиці другі різниці використовують досить рідко, а різниці вищого порядку не використовуються. Застосування других різниць до процесу приводить до видалення квадратичного тренду, що можна легко проілюструвати на прикладі описання тренду поліномом другого порядку від k . Наприклад, нехай:

$$y(k) = a_0 + a_1 k + a_2 k^2.$$

Перша різниця:

$$\begin{aligned}\Delta y(k) &= a_0 + a_1 k + a_2 k^2 - [a_0 + a_1 \cdot (k-1) + a_2 \cdot (k-1)^2] = \\ &= a_0 + a_1 k + a_2 k^2 - a_0 - a_1 k + a_1 - a_2 k^2 + 2a_2 k - a_2 = \\ &= a_1 + 2a_2 k - a_2.\end{aligned}$$

Друга різниця:

$$\begin{aligned}\Delta^2 y(k) &= a_1 + 2a_2 k - a_2 - [a_1 + 2a_2 \cdot (k-1) - a_2] = \\ &= 2a_2.\end{aligned}$$

Редукована та структурована форми різницевого рівняння. Для ілюстрації рівнянь такого типу розглянемо модель Самуельсона для валового внутрішнього продукту (ВВП), споживання та інвестицій:

$$y(k) = c(k) + I(k), \quad (3.1.4)$$

$$c(k) = \alpha y(k-1) + \varepsilon_c(k), \quad 0 < \alpha < 1, \quad (3.1.5)$$

$$I(k) = \beta[c(k) - c(k-1)] + \varepsilon_I(k), \quad \beta > 0, \quad (3.1.6)$$

де $y(k), c(k), I(k)$ – ВВП, споживання та інвестиції в момент k відповідно. У цій моделі $y(k), c(k), I(k)$ – ендогенні змінні, а $\varepsilon_c(k), \varepsilon_I(k)$ – стохастичні змінні з нульовим середнім, які характеризують вплив збурень на споживання та інвестиції, тобто вплив тих змінних, які не введені у модель. Третє рівняння ілюструє принцип прискорення (акселерації), оскільки в нього введена швидкість зміни споживання $[c(k) - c(k-1)]$.

Рівняння (3.1.6) називають **структурованим рівнянням**, оскільки в ньому ендогенна змінна $I(k)$ залежить від поточного (а не минулого) значення іншої ендогенної змінної $c(k)$.

Редукованою формою різницевого рівняння називають таку, в якій основна змінна (у лівій частині) залежить від минулих вимірів цієї ж змінної, минулих вимірів інших ендогенних змінних, поточних та минулих вимірів екзогенних змінних та збурення.

Таким чином, рівняння для споживання вже представлене у редукованій формі. Рівняння для інвестицій не відповідає редукованій формі, оскільки у правій частині наявне поточне значення споживання. Для того щоб отримати редуковану форму для інвестицій, підставимо (3.1.5) у рівняння для інвестицій і отримаємо:

$$\begin{aligned} I(k) &= \beta[\alpha y(k-1) + \varepsilon_c(k) - c(k-1)] + \varepsilon_I(k) = \\ &= \alpha\beta y(k-1) - \beta c(k-1) + \beta\varepsilon_c(k) + \varepsilon_I(k). \end{aligned}$$

Необхідно зазначити, що отримана редукована форма для інвестицій не єдина. Якщо дискретний час k зменшити на одиницю, то рівняння (3.1.3) можна переписати у вигляді: $c(k-1) = \alpha y(k-2) + \varepsilon_c(k-1)$. За допомогою отриманого рівняння запишемо редуковану форму для інвестицій у вигляді:

$$\begin{aligned} I(k) &= \alpha\beta y(k-1) - \beta[\alpha y(k-2) + \varepsilon_c(k-1)] + \beta\varepsilon_c(k) + \varepsilon_I(k) = \\ &= \alpha\beta[y(k-1) - y(k-2)] + \beta[\varepsilon_c(k) - \varepsilon_c(k-1)] + \varepsilon_I(k). \quad (3.1.7) \end{aligned}$$

За аналогією можна отримати редуковану форму для ВВП, якщо підставити (3.1.5) і (3.1.7) у рівняння (3.1.4):

$$\begin{aligned}
 y(k) &= \alpha y(k-1) + \varepsilon_c(k) + \alpha\beta[y(k-1) - y(k-2)] + \beta[\varepsilon_c(k) - \varepsilon_c(k-1)] + \varepsilon_l(k) = \\
 &= \alpha(1+\beta)y(k-1) - \alpha\beta y(k-2) + (1+\beta)\varepsilon_c(k) + \varepsilon_l(k) - \beta\varepsilon_c(k-1).
 \end{aligned}
 \tag{3.1.8}$$

Рівняння (3.1.8) називають редукованою формою стосовно однієї змінної, оскільки у правій частині наявні тільки минулі виміри основної змінної та збурення. Така форма зручна для прогнозування, оскільки для цього необхідно мати виміри тільки однієї змінної. Нижче будуть розглянуті підходи до прогнозування, які ґрунтуються на розв'язках різницевих рівнянь. Вони дають можливість обчислювати оцінки прогнозів на довільну кількість кроків вперед.

Необхідно також зазначити, що різницеві рівняння, які описують фінансово-економічні процеси перехідного періоду, можна використати для створення систем адаптивного оптимального керування цими процесами шляхом використання популярних сьогодні інформаційних систем підтримки прийняття оптимальних (чи субоптимальних) рішень, при створенні експертних систем і т. ін.

Позитивною стороною використання РР є те, що їх параметри легко обновляються при поступанні нових вимірів (за допомогою методів рекурсивного оцінювання – рекурсивного МНК, ММП і т. ін.). Також можна досить просто змінити порядок і структуру математичної моделі у цілому. Таким чином досить просто виконується вимога структурно-параметричної адаптації математичної моделі, яка будується на основі статистичних даних.

При цьому під структурою моделі будемо розуміти *кількість рівнянь* (розмірність) моделі, *їх порядок, наявність нелінійностей* та їхній тип, *наявність запізнення (лагу)* стосовно входу та його оцінка, а також *збурення процесу* і його тип. У поняття структури також можна включити обмеження на параметри та основні змінні моделі. Оцінювання структури моделі на основі статистичних або експериментальних даних буде розглянуто нижче.

3.2 Ітераційний метод знаходження розв'язків різницевих рівнянь

Поняття розв'язку різницевого рівняння

Розв'язки різницевих рівнянь (так само як і диференціальних) необхідні для аналізу асимптотичної поведінки процесу (на нескінченних часових інтервалах), для формування на їх основі функцій прогнозування, а також для порівняння характеристик різних процесів між собою. Вони використовуються, також, при проектуванні систем автоматичного керування технічними системами і технологічними процесами.

Розв'язок різницевого рівняння представляє собою *функцію часу, вхідної змінної, початкових умов, параметрів та збурення*. Існують різні підходи до знаходження розв'язків РР.

Як простий приклад знаходження розв'язку розглянемо процес, який описується першою різницею

$$\Delta y(k) = 2,5, \quad (3.2.1)$$

тобто, приріст змінної за один період дискретизації складає 2,5 одиниці у вибраному масштабі. Замінімо першу різницю відповідними значеннями основної змінної:

$$y(k) = y(k-1) + 2,5,$$

і задамо початкову умову $y(0) = y_0 = 0,25$. Тепер запишемо значення $y(k)$ для кількох моментів часу, починаючи з $k = 1$:

$$\begin{aligned} y(1) &= y_0 + 2,5, \\ y(2) &= y(1) + 2,5 = y_0 + 2,5 + 2,5, \\ y(3) &= y(2) + 2,5 = y_0 + 2,5 + 2,5 + 2,5 = y_0 + 3 \cdot 2,5 \end{aligned}$$

і для довільного k :

$$y(k) = y_0 + 2,5 \cdot k = 2,5k + 0,25. \quad (3.2.2)$$

Отриманий розв'язок для (3.2.1) справедливий для будь-якого значення k . Для того щоб це довести, підставимо отриманий розв'язок у рівняння (3.2.1):

$$2,5k + 0,25 = 2,5(k-1) + 0,25 + 2,5$$

або

$$0,25 = 0,25.$$

Таким чином, знайдений розв'язок справедливий для $\forall k, k = 0, \dots, \infty$. Він представляє собою рівняння прямої, яка перетинає вісь ординат в точці $y_0 = 0,25$. В зв'язку з цим значення $y_0 = a_0$ називають ще *перетином*.

Ітераційний підхід до знаходження розв'язку рівняння першого порядку

Розглянемо знаходження розв'язку РР першого порядку ітераційним методом. Таким же методом знайдено розв'язок рівняння (3.2.1). Рівняння першого порядку з ненульовим перетином має вигляд:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad (3.2.3)$$

з початковою умовою y_0 . Для моментів часу $k = 1, 2, 3$ можна записати, що

$$y(1) = a_0 + a_1 y(0) + \varepsilon(1),$$

$$\begin{aligned} y(2) &= a_0 + a_1 y(1) + \varepsilon(2) = a_0 + a_1 [a_0 + a_1 y(0) + \varepsilon(1)] + \varepsilon(2) = \\ &= a_0 + a_0 a_1 + (a_1)^2 y(0) + a_1 \varepsilon(1) + \varepsilon(2). \end{aligned}$$

$$\begin{aligned} y(3) &= a_0 + a_1 y(2) + \varepsilon(3) = \\ &= a_0 [1 + a_1 + (a_1)^2] + (a_1)^3 y(0) + a_1^2 \varepsilon(1) + a_1 \varepsilon(2) + \varepsilon(3). \end{aligned}$$

Узагальнюючи запис, можна записати розв'язок рівняння (3.2.3) для довільного моменту k :

$$y(k) = a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k y_0 + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i). \quad (3.2.4)$$

Розв'язок (3.2.4) також можна отримати, якщо розпочати ітерації в зворотному часі, тобто

$$\begin{aligned} y(k) &= a_0 + a_1 [a_0 + a_1 y(k-2) + \varepsilon(k-1)] + \varepsilon(k) = \\ &= a_0 + (1 + a_1) + a_1^2 [a_0 + a_1 y(k-3) + \varepsilon(k-2)] + \varepsilon(k) + a_1 \varepsilon(k-1). \end{aligned}$$

Продовження ітераційного процесу приведе до розв'язку у формі (3.2.4).

Повний розв'язок (3.2.4) складається з двох частин. Перша частина залежить від початкової умови $a_1^k y_0$, а друга від наявності в рівнянні перетину a_0 і випадкової змінної $\varepsilon(k)$. Відповідно, першу частину називають однорідним розв'язком або розв'язком однорідного рівняння і позначають так:

$$y^h(k) = a_1^k y_0 = A \alpha^k,$$

де індекс h – *homogeneous* (означає однорідний). У більш загальному вигляді однорідний розв'язок записують так: $y^h(k) = A \alpha^k$, де A – довільна константа, яка визначається за допомогою початкових умов; α – корінь характеристичного рівняння, який дорівнює параметру a_1 у випадку рівняння $AP(1)$.

Неоднорідний або частковий розв'язок позначають так:

$$y^p(k) = a_0 \sum_{i=0}^{k-1} a_1^i + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i),$$

де верхній індекс p – *partial* (частковий). Повний розв'язок представляє собою суму однорідного і часткового розв'язків, тобто,

$$y(k) = y^h(k) + y^p(k).$$

Якщо початкова умова не задана, то замість y_0 в (3.2.4) можна підставити $y(0) = a_0 + a_1 y(-1) + \varepsilon(0)$ (випадок можливий, якщо існують дані про процес до моменту $k=0$):

$$\begin{aligned} y(k) &= a_0 \sum_{i=0}^{k-1} a_1^i + a_1^k [a_0 + a_1 y(-1) + \varepsilon(0)] + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) = \\ &= a_0 \sum_{i=0}^k a_1^i + \sum_{i=0}^k a_1^i \varepsilon(k-i) + a_1^{k+1} y(-1). \end{aligned} \quad (3.2.5)$$

Продовжуючи ітерації у зворотному напрямі на m кроків, отримаємо:

$$y(k) = a_0 \sum_{i=0}^{k+m} a_1^i + \sum_{i=0}^{k+m} a_1^i \varepsilon(k-i) + a_1^{k+m+1} y(-m-1). \quad (3.2.6)$$

При $|a_1| < 1$ член a_1^{k+m+1} збігається до нуля, якщо $m \rightarrow \infty$, а нескінченна сума $[1 + a_1 + a_1^2 + \dots]$ збігається до виразу $1/(1 - a_1)$ як сума членів геометричної прогресії. Таким чином, розв'язок (3.2.6) на нескінченності має вигляд:

$$y(k) = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i), \quad |a_1| < 0, \quad k \rightarrow \infty. \quad (3.2.7)$$

Для того щоб перевірити правильність розв'язку, необхідно підставити (3.2.7) у (3.2.3). В результаті отримаємо тотожність. Однак, (3.2.7) – це не єдиний розв'язок для (3.2.3). Розв'язком є також таке рівняння:

$$y(k) = \frac{a_0}{1 - a_1} + A a_1^k + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i), \quad |a_1| < 0, \quad (3.2.8)$$

де A – довільна константа. Підставимо (3.2.8) в (3.2.3) і отримаємо тотожність, що свідчить про правильність розв'язку:

$$\frac{a_0}{1-a_1} + Aa_1^k + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) = a_0 + a_1 \left[\frac{a_0}{1-a_1} + Aa_1^{k-1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-1-i) \right] + \varepsilon(k).$$

З наведеного вище розв'язку рівняння першого порядку видно, що коефіцієнт a_1 відіграє важливу роль у поведінці розв'язку. Якщо $|a_1| > 1$, то розв'язок буде розбіжним. Збіжність чи розбіжність буде мати коливальний характер, якщо $a_1 < 0$; це зумовлено тим, що $a_1^k > 0$ при парному k і $a_1^k < 0$ при непарному k .

Альтернативний підхід до знаходження розв'язків

Існує **альтернативний підхід** до знаходження розв'язків різницевого рівнянь, який полягає у знаходженні окремо розв'язку однорідного рівняння та часткового розв'язку. Наприклад, розв'язок однорідного рівняння виду

$$y(k) = a_1 y(k-1) \quad \text{або} \quad y(k) = a_1 y(k-1) + a_2 y(k-2)$$

називають *однорідним розв'язком*. Тривіальним розв'язком однорідного рівняння є $y(1) = y(2) = \dots = 0$. Якщо покласти $a_0 = 0$ і всі значення випадкової змінної $\{\varepsilon(k)\} = 0$, то із (3.2.4) випливає, що розв'язком рівняння буде також $y(k) = a_1^k y_0$. Можна показати, що $y(k) = Aa_1^k$, де A – довільна константа, також буде розв'язком однорідного рівняння. Підставимо останнє значення в однорідне рівняння і отримаємо тотожність:

$$A(a_1)^k = a_1 A(a_1)^{k-1}. \quad (3.2.9)$$

Із наведеного вище аналізу розв'язку рівняння першого порядку, а також з (3.2.9) випливає, що однорідний розв'язок має такі властивості:

1. Якщо $|a_1| < 1$, то вираз $(a_1)^k$ збігається до нуля при $k \rightarrow \infty$. Збіжність має гладкий характер, якщо $0 < a_1 < 1$, і коливальний характер при $-1 < a_1 < 0$.

2. Якщо $|a_1| > 1$, то однорідний розв'язок буде нестійким (тобто, спостерігається розбіжність). Якщо $a_1 > 1$, то однорідний розв'язок наближається до нескінченності при $k \rightarrow \infty$ без коливань, а якщо $a_1 < -1$, то розбіжність носить коливальний характер.

3. Якщо $a_1 = 1$, то однорідному рівнянню $y(k) = a_1 y(k-1)$ задовольняє будь-яка довільна константа A . Якщо ж $a_1 = -1$, то розв'язок називають *метастійким* – $(a_1)^k = 1$ для парних значень k , і $(a_1)^k = -1$ при непарних значеннях k .

Рівняння (3.2.7) називають *частковим* (або *неоднорідним*) *розв'язком* різницевого рівняння. Ця частина загального розв'язку зумовлена наявністю в правій частині рівняння членів a_0 та $\varepsilon(k)$. Процес знаходження часткового розв'язку буде розглянуто нижче, а зараз наведемо загальну методику розв'язання різницевого рівняння, яка логічно впливає з отриманого ітераційним методом повного розв'язку рівняння $AP(1)$.

Загальна методика знаходження розв'язку різницевого рівняння

Користуючись результатами знайденого розв'язку рівняння першого порядку, отриманими вище, можна сформулювати загальну методику знаходження розв'язків різницевого рівняння типу AP та $APKC$:

Крок 1: Для моделі $APKC(p, q)$ сформулювати однорідне рівняння та знайти всі його однорідні розв'язки (їх кількість дорівнює p).

Крок 2: Знайти всі складові часткового розв'язку, зумовлені константами, детермінованими та випадковими функціями у правій частині рівняння.

Крок 3: Записати загальний розв'язок у вигляді суми однорідного та часткового розв'язків.

Крок 4: За допомогою початкових умов обчислити значення довільних констант.

Розглянемо приклад знаходження повного розв'язку рівняння авторегресії другого порядку.

Приклад 3.1. Знайдемо загальний розв'язок PP другого порядку

$$y(k) = 2,1 + 0,8y(k-1) - 0,15y(k-2).$$

Сформуємо однорідне рівняння:

$$y(k) - 0,8y(k-1) + 0,15y(k-2) = 0.$$

Оскільки це рівняння другого порядку, то воно має два однорідних розв'язки, які знаходять за допомогою розв'язку характеристичного рівняння,

записаного для однорідного (докладніше ця методика буде розглянута нижче). Характеристичне рівняння має вигляд:

$$\lambda^2 - 0,8\lambda + 0,15 = 0.$$

Це квадратне рівняння має два корені: $\lambda_1 = 0,5$; $\lambda_2 = 0,3$. Таким чином, два однорідних розв'язки можна записати так: $y_1^h(k) = A_1(0,5)^k$; $y_2^h(k) = A_2(0,3)^k$.

Правильність знаходження однорідних розв'язків можна перевірити шляхом підстановки в однорідне рівняння. Наприклад:

$$A_1(0,5)^k - 0,8A_1(0,5)^{k-1} + 0,15A_1(0,3)^{k-2} = 0.$$

Якщо розділити всі члени на $(0,3)^{k-2}$, то отримаємо

$$(0,3)^2 - 0,9(0,3) + 0,15 = 0,09 - 0,27 + 0,15 = 0.$$

Другий однорідний розв'язок перевіряється аналогічно.

Частковий розв'язок для детермінованого збурення знайдемо так:

$$y^p = 2,1 + 0,8y^p - 0,15y^p,$$

тобто під впливом константи на виході також буде константа y^p для всіх моментів часу. Звідси $y^p = 6,0$, де верхній індекс $p = \text{partial}$ (частковий). Загальна методика визначення часткового розв'язку буде наведена нижче. Тепер об'єднаємо однорідний і частковий розв'язки:

$$y(k) = 6,0 + A_1(0,5)^k + A_2(0,3)^k,$$

де A_1, A_2 – довільні константи, які знайдемо за допомогою початкових умов. Оскільки маємо дві невідомі константи, то необхідно мати два значення для початкових умов. Нехай $y(0) = 1,0$; $y(1) = 2,0$. Тепер можемо записати систему із двох рівнянь для знаходження констант:

$$\begin{aligned} k = 0: & 1,0 = 6,0 + A_1 + A_2, \\ k = 1: & 2,0 = 6,0 + 0,5A_1 + 0,3A_2 \end{aligned}$$

Розв'язуючи систему, знайдемо, що $A_1 = -12,5$; $A_2 = 7,5$, і розв'язок приймає вигляд

$$y(k) = 6,0 - 12,5(0,5)^k + 7,5(0,4)^k.$$

Таким чином, знайдений розв'язок має середнє значення (його називають ще зміщенням) 6,5, а два інші члени збігаються до нуля при $k \rightarrow \infty$.

3.3 Знаходження однорідних та часткових розв'язків

3.3.1 Розв'язування однорідних рівнянь

Однорідне рівняння другого порядку

При моделюванні економічних процесів найчастіше використовують різницеві рівняння 1-3 порядку, але зовсім не виключена можливість використання рівнянь вищих порядків. У зв'язку з цим виникає необхідність знаходження розв'язків однорідних рівнянь різних порядків. Оскільки розв'язок рівняння другого порядку дає можливість зробити деякі узагальнення на рівняння вищих порядків, то розглянемо спочатку рівняння другого порядку:

$$y(k) - a_1 y(k-1) - a_2 y(k-2) = 0. \quad (3.3.1)$$

По аналогії з рівнянням першого порядку виберемо пробний однорідний розв'язок у загальному вигляді $y^h(k) = A\alpha^k$. Підставимо пробний розв'язок в рівняння (3.3.1):

$$A\alpha^k - a_1 A\alpha^{k-1} - a_2 A\alpha^{k-2} = 0 \quad (3.3.2)$$

і поділимо обидві частини отриманого рівняння на $A\alpha^{k-2}$. В результаті отримаємо так зване *характеристичне рівняння* відносно α :

$$\alpha^2 - a_1 \alpha - a_2 = 0. \quad (3.3.3)$$

Запишемо корені цього квадратного рівняння:

$$\alpha_{1,2} = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2} = \frac{a_1 \pm \sqrt{d}}{2}, \quad (3.3.4)$$

де дискримінант $d = \sqrt{a_1^2 + 4a_2}$. Кожний корінь характеристичного рівняння дає один однорідний розв'язок. Розв'язком є також лінійна комбінація

$$A_1 \alpha_1^k + A_2 \alpha_2^k, \quad (3.3.5)$$

що легко перевірити підстановкою (3.3.5) в (3.3.1):

$$A_1\alpha_1^k + A_2\alpha_2^k = a_1(A_1\alpha_1^{k-1} + A_2\alpha_2^{k-1}) + a_2(A_1\alpha_1^{k-2} + A_2\alpha_2^{k-2}).$$

Перенесемо праву частину вліво і згрупуємо члени:

$$A_1(\alpha_1^k - a_1\alpha_1^{k-1} - a_2\alpha_1^{k-2}) + A_2(\alpha_2^k - a_1\alpha_2^{k-1} - a_2\alpha_2^{k-2}) = 0.$$

Оскільки α_1, α_2 – це корені характеристичного рівняння (3.3.3), то обидва члени в дужках дорівнюють нулю. Таким чином, повний однорідний розв'язок РР другого порядку має вигляд:

$$y^h(k) = A_1\alpha_1^k + A_2\alpha_2^k. \quad (3.3.6)$$

Оскільки *рівняння другого порядку* – це основа для аналізу розв'язків рівнянь вищих порядків, в літературі розглядають три можливих випадки знаходження їх розв'язку в залежності від значення дискримінанта: $d > 0$, $d = 0$ і $d < 0$. Крім того, рівняння другого порядку – досить поширена модель економічних, фінансових, технічних та інших процесів. Так, в технічних системах рівняннями другого порядку можна описати гармонічні коливання елементів конструкцій.

А. Дискримінант додатний: $a_1^2 + 4a_2 > 0$. У такому випадку корені будуть різними дійсними числами. Однорідний розв'язок має вигляд:

$$y^h(k) = A_1\alpha_1^k + A_2\alpha_2^k.$$

Очевидно, що при $|\alpha_1| > 1$ або $|\alpha_2| > 1$ розв'язок буде розбіжним.

Приклад 3.2. Розглянемо процес другого порядку

$$y(k) = 0,95y(k-1) + 0,35y(k-2),$$

тобто, $a_1 = 0,95$; $a_2 = 0,35$. Для однорідного рівняння $y(k) - a_1 y(k-1) - a_2 y(k-2) = 0$ характеристичним є наступне: $\alpha^2 - 0,95\alpha - 0,35 = 0$. Дискримінант $d = a_1^2 + 4a_2 = 0,95^2 + 4 \cdot 0,35 = 2,3025$. Корені характеристичного рівняння:

$$\alpha_1 = \frac{1}{2}(a_1 + \sqrt{d}) = \frac{1}{2}(0,95 + 1,517) = 1,233; \quad \alpha_2 = \frac{1}{2}(0,95 - 1,517) = -0,284.$$

Таким чином, однорідний розв'язок має вигляд:

$$y^h(k) = A_1 \cdot 1,233^k + A_2 \cdot (-0,284)^k,$$

що свідчить про розбіжність розв'язку.

Б. Дискримінант $a_1^2 + 4a_2 = 0$ (випадок кратних коренів). При $d = 0$ отримаємо два однакових корені: $\alpha_1 = \alpha_2 = a_1/2$. Таким чином, однорідним розв'язком є $A \cdot (a_1/2)^k$. Однак, при $d = 0$ існує також однорідний розв'язок такого вигляду: $A \cdot k(a_1/2)^k$ (тобто дискретний час k виступає множником, що приводить до нелінійної поведінки однорідного розв'язку). Покажемо, що це так шляхом підстановки другого розв'язку в рівняння:

$$k(a_1/2)^k - a_1[(k-1)(a_1/2)^{k-1}] - a_2[(k-2)(a_1/2)^{k-2}] = 0.$$

Для зручності аналізу поділимо обидві частини цього рівняння на $(a_1/2)^{k-2}$ і отримаємо:

$$-[(a_1^2/4) + a_2]k + [(a_1^2/2) + 2a_2] = 0.$$

Оскільки $a_1^2 + 4a_2 = 0$, то обидва вирази в дужках будуть дорівнювати нулю і, таким чином, $k(a_1/2)^k$ є розв'язком рівняння другого порядку при $d = 0$. Тепер можемо записати однорідний розв'язок для даного випадку у вигляді лінійної комбінації двох розв'язків:

$$y^h(k) = A_1 \left(\frac{a_1}{2}\right)^k + A_2 k \left(\frac{a_1}{2}\right)^k.$$

Цей розв'язок буде розбіжним, якщо $|a_1| > 2$. Якщо $|a_1| < 2$, то складова розв'язку $A_1(a_1/2)^k$ буде збіжною, але поведінка складової $A_2 k(a_1/2)^k$ буде невизначеною, тому що в ній наявний множник k . Знайдемо границю, до якої прямує другий однорідний розв'язок при $k \rightarrow \infty$. Введемо позначення:

$$\beta = \frac{1}{a_1/2} = \frac{1}{\alpha}.$$

Тепер скористаємось правилом Лопіталя:

$$\lim_{k \rightarrow \infty} k \alpha^k = \lim_{k \rightarrow \infty} \frac{k}{\beta^k} = \lim_{k \rightarrow \infty} \frac{1}{\beta^k \ln \beta} = 0,$$

оскільки $|\beta| > 1$ при $|\alpha| < 1$. Якщо виконати аналіз розв'язку в даному випадку, то виявиться, що він має немонотонний характер – деякий сплеск на перших 20-30 кроках з подальшою збіжністю до нуля при $k \rightarrow \infty$.

При $-2 < a_1 < 0$ розв'язок має коливальний характер з досить великою амплітудою на перших 20-30 кроках (додатна амплітуда при парних k і від'ємна при непарних). Однак він також збігається до нуля при $k \rightarrow \infty$, тобто в даному випадку процес збіжний, незважаючи на наявність в розв'язку дискретного часу k в явній формі.

У випадку, якщо характеристичне рівняння має m -кратний корінь, однорідний розв'язок буде мати такий вигляд:

$$y^h(k) = A_1 \alpha^k + A_2 k \alpha^k + A_3 k^2 \alpha^k + \dots + A_m k^{m-1} \alpha^k.$$

В. Дискримінант $a_1^2 + 4a_2 < 0$. У даному випадку характеристичне рівняння має два комплексно спряжених корені. Оскільки $a_1^2 \geq 0$ завжди, то комплексні корені можуть з'явитись тільки у випадку, коли $a_2 < 0$.

Таким чином, корені характеристичного рівняння можна записати у вигляді:

$$\alpha_1 = \frac{1}{2}(a_1 + i\sqrt{d}), \quad \alpha_2 = \frac{1}{2}(a_1 - i\sqrt{d}), \quad \text{де } i = \sqrt{-1}.$$

При знаходженні розв'язку однорідного рівняння скористаємось такими тригонометричними тотожностями:

$$\begin{aligned} \sin(\theta_1 + \theta_2) &= \sin(\theta_1)\cos(\theta_2) + \cos(\theta_1)\sin(\theta_2); \\ \cos(\theta_1 + \theta_2) &= \cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2), \end{aligned}$$

які при $\theta_1 = \theta_2$ спрощуються до вигляду:

$$\begin{aligned} \sin(2\theta) &= 2\sin(\theta)\cos(\theta); \\ \cos(2\theta) &= \cos^2(\theta) - \sin^2(\theta). \end{aligned}$$

Нагадаємо, що комплексне число $a + ib$ можна представити точкою на комплексній площині (рис. 3.1) і перейти до представлення коренів характеристичного рівняння в полярних координатах.

Запишемо комплексне число через тригонометричні функції кута θ і отримаємо його представлення в полярних координатах:

$$a = r \cos(\theta), \quad b = r \sin(\theta).$$

По відношенню до виразів для коренів однорідного рівняння α_1, α_2 можна записати, що $a = a_1/2$, $b = \sqrt{d}/2$.

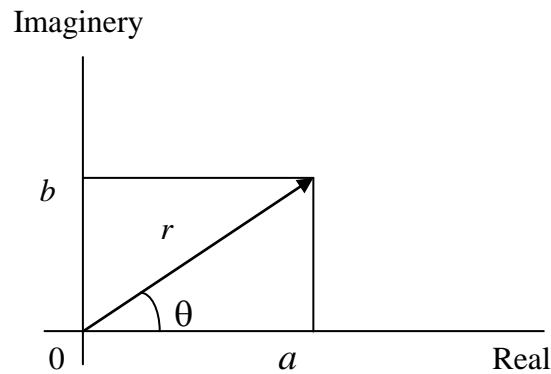


Рис. 3.1 Представлення комплексного числа на площині

Таким чином, для коренів α_1, α_2 можна записати вирази в полярних координатах:

$$\begin{aligned}\alpha_1 &= a + ib = r [\cos(\theta) + i \sin(\theta)], \\ \alpha_2 &= a - ib = r [\cos(\theta) - i \sin(\theta)].\end{aligned}$$

Для того щоб сформулювати розв'язок, необхідно отримати вирази для α_1^k і α_2^k . Спочатку запишемо вираз для α_1^2 із врахуванням того, що $i^2 = -1$:

$$\begin{aligned}\alpha_1^2 &= \{r[\cos(\theta) + i \sin(\theta)]\} \{r[\cos(\theta) + i \sin(\theta)]\} = \\ &= r^2 [\cos(\theta)\cos(\theta) - \sin(\theta)\sin(\theta) + 2i \sin(\theta)\cos(\theta)] = \\ &= r^2 [\cos(2\theta) + i \sin(2\theta)].\end{aligned}$$

Якщо продовжити процес піднесення до степеня, то отримаємо такі вирази для коренів:

$$\alpha_1^k = r^k [\cos(k\theta) + i \sin(k\theta)], \quad \alpha_2^k = r^k [\cos(k\theta) - i \sin(k\theta)].$$

Оскільки однорідний розв'язок $y^h(k)$ – це дійсне число, а α_1, α_2 – комплексні числа, то довільні константи A_1, A_2 – повинні бути спряженими комплексними числами у наступній формі:

$$A_1 = B [\cos(\varphi) + i \sin(\varphi)], \quad A_2 = B [\cos(\varphi) - i \sin(\varphi)],$$

де B, φ – довільні дійсні числа.

Тепер можна обчислити добутки $A_1\alpha_1^k$ і $A_2\alpha_2^k$. Так,

$$A_1\alpha_1^k = B[\cos(\varphi) + i \sin(\varphi)]r^k [\cos(k\theta) + i \sin(k\theta)] =$$

$$= Br^k [\cos(\varphi)\cos(k\theta) - \sin(\varphi)\sin(k\theta) + i\cos(k\theta)\sin(\varphi) + i\sin(k\theta)\cos(\varphi)].$$

Використовуючи наведені вище тригонометричні тотожності, отримаємо:

$$A_1\alpha_1^k = Br^k [\cos(k\theta + \varphi) + i\sin(k\theta + \varphi)].$$

За аналогією можна отримати вираз для другого добутку:

$$A_2\alpha_2^k = Br^k [\cos(k\theta + \varphi) - i\sin(k\theta + \varphi)].$$

Тепер запишемо однорідний розв'язок як суму двох розв'язків:

$$\begin{aligned} y^h(k) &= Br^k [\cos(k\theta + \varphi) + i\sin(k\theta + \varphi)] + Br^k [\cos(k\theta + \varphi) - i\sin(k\theta + \varphi)] = \\ &= 2Br^k \cos(k\theta + \varphi). \end{aligned}$$

Якщо покласти $2B = \beta_1$ і $\theta = \omega$, то розв'язок прийме вигляд:

$$y^h(k) = \beta_1 r^k \cos(\omega k + \varphi),$$

де β_1 , φ – довільні константи; $r = \sqrt{-a_2}$; значення ω вибирається із умови:

$$\cos(\omega) = \frac{a_1}{2(-a_2)^{1/2}} = \frac{a_1}{2r}.$$

З фізичної точки зору величина ω має смисл кругової частоти, тобто $\omega = 2\pi f$ (f – лінійна частота в герцах), а φ – початкова фаза коливань розв'язку. Тобто однорідний розв'язок має, в даному випадку, форму затухаючих або розбіжних коливань в залежності від значення r .

Наведений вище вираз для r можна уточнити так:

$$r = \sqrt{a^2 + b^2} = \sqrt{\left(\frac{a_1}{2}\right)^2 + \left(\frac{\sqrt{|d|}}{2}\right)^2} = \sqrt{\frac{a_1^2 + [-(a_1^2 + 4a_2)]}{4}} = \sqrt{-a_2}.$$

Розглянемо *приклад знаходження однорідного розв'язку у випадку комплексних коренів.*

Приклад 3.3. Нехай рівняння другого порядку має вигляд:

$$y(k) = 1,6y(k-1) - 0,8y(k-2).$$

Характеристичне рівняння для нього:

$$\alpha^2 - 1,6\alpha + 0,8 = 0.$$

Дискримінант $d = a_1^2 + 4a_2 = -0,64$. Корені характеристичного рівняння і параметри однорідного розв'язку:

$$\alpha_1 = 0,5(1,6 + i0,8), \quad \alpha_2 = 0,5(1,6 - i0,8).$$

$$r = (-a_2)^{1/2} = (0,80)^{1/2} = 0,894.$$

$$\cos(\omega) = \frac{a_1}{2(-a_2)^{1/2}} = \frac{1,6}{2 \cdot 0,894} = 0,894.$$

Умови стійкості розв'язку однорідного рівняння другого порядку

У першому випадку (**A** – різні дійсні корені) для забезпечення стійкості розв'язку необхідно виконати умову:

$$-1 < \alpha_i < 1, \quad i = 1, 2 \quad \text{або} \quad -1 < \frac{1}{2} \left(a_1 \pm \sqrt{a_1^2 + 4a_2} \right) < 1.$$

Більший корінь характеристичного рівняння буде меншим одиниці, якщо

$$a_1 + (a_1^2 + 4a_2)^{1/2} < 2 \quad \text{або} \quad (a_1^2 + 4a_2)^{1/2} < 2 - a_1.$$

Таким чином, можна записати, що $a_1^2 + 4a_2 < 4 - 4a_1 + a_1^2$, або умовою стійкості є така:

$$a_1 + a_2 < 1.$$

Менший корінь $\alpha_2 = 0,5(a_1 - \sqrt{d})$ буде більшим -1 , якщо

$$a_1 - (a_1^2 + 4a_2)^{1/2} > -2 \quad \text{або} \quad 2 + a_1 > (a_1^2 + 4a_2)^{1/2}.$$

Звідси отримаємо: $4 + 4a_1 + a_1^2 > a_1^2 + 4a_2$, а умова стійкості матиме вигляд:

$$a_2 - a_1 < 1.$$

У другому випадку (**B** – кратні корені) $a_1^2 + 4a_2 = 0$. Умовою стійкості розв'язку є $|a_1| < 2$.

У третьому випадку (**B** – комплексні корені) $a_1^2 + 4a_2 < 0$. Умова стійкості має вигляд: $r = (-a_2)^{1/2} < 1$ або $-a_2 < 1$, при $a_2 < 0$.

Таким чином, загальною умовою стійкості розв'язку для всіх трьох випадків є така: корені характеристичного рівняння повинні лежати всередині одиничного кола на комплексній площині.

У випадку різних та кратних дійсних коренів вони будуть лежати на горизонтальній осі комплексної площини, але всередині кола одиничного радіуса. У третьому випадку корені розташовуються в довільних точках площини, але для забезпечення стійкості розв'язку вони також повинні лежати всередині одиничного кола.

Структура розв'язку однорідного рівняння вищого порядку

Розв'язки однорідних рівнянь вищих порядків знаходять по аналогії з рівняннями другого порядку. Розглянемо однорідне рівняння n -го порядку:

$$y(k) - \sum_{i=1}^n a_i y(k-i) = 0.$$

Підставимо в нього однорідний розв'язок загального виду $y^h(k) = A\alpha^k$:

$$A\alpha^k - a_1 A\alpha^{k-1} - a_2 A\alpha^{k-2} - \dots - a_n A\alpha^{k-n} = 0$$

і розділимо отримане рівняння на $A\alpha^{k-n}$. В результаті отримаємо характеристичне рівняння n -го порядку:

$$\alpha^n - a_1 \alpha^{n-1} - a_2 \alpha^{n-2} - \dots - a_n = 0.$$

Воно має n розв'язків, тобто n коренів $\alpha_i, i=1, \dots, n$. Згідно з отриманими вище результатами лінійна комбінація n коренів також буде розв'язком. Якщо серед знайдених коренів буде корінь кратності m , то розв'язок матиме вигляд:

$$y^h(k) = A_1 \alpha^k + A_2 k \alpha^k + A_3 k^2 \alpha^k + \dots + A_m k^{m-1} \alpha^k + A_{m+1} \alpha_{m+1}^k + \dots + A_n \alpha_n^k.$$

Очевидно, що у випадку наявності комплексно-спряженого кореня розв'язок буде мати відповідну гармонічну складову вигляду $y^h(k) = \beta_1 r^k \cos(\omega k + \varphi)$. Очевидно, що для рівнянь вищих порядків також необхідно аналізувати стійкість розв'язків, тому розглянемо спочатку стійкість (збіжність) розв'язку рівняння другого порядку.

Умови стійкості розв'язку однорідного рівняння другого порядку

На основі умов стійкості розв'язку однорідного рівняння другого порядку можна сформулювати умови стійкості розв'язку однорідного рівняння n -го порядку:

1. Необхідною умовою того, щоб корені характеристичного рівняння лежали в одиничному колі є така:

$$\sum_{i=1}^n a_i < 1.$$

2. Оскільки значення коефіцієнтів a_i можуть бути додатними і від'ємними, то достатньою умовою того, що корені характеристичного рівняння будуть лежати всередині одиничного кола є:

$$\sum_{i=1}^n |a_i| < 1.$$

3. Щонайменше один корінь характеристичного рівняння дорівнює одиниці, якщо виконується наступна рівність:

$$\sum_{i=1}^n a_i = 1.$$

Будь-яку послідовність (часовий ряд), яка містить один або більше одиничних коренів відповідного характеристичного рівняння, називають *процесом з одиничним коренем* (або процес із трендом).

4. Умови стійкості розв'язку рівняння третього порядку мають вигляд:

$$\begin{aligned} 1 - a_1 - a_2 - a_3 &> 0, \\ 1 + a_1 - a_2 + a_3 &> 0, \\ 1 - a_1 a_3 + a_2 - a_3^2 &> 0, \\ 3 + a_1 + a_2 - 3a_3 &> 0 \quad \text{або} \quad 3 - a_1 + a_2 + 3a_3 > 0. \end{aligned}$$

При цьому, якщо перші три умови виконуються, то остання є надлишковою.

На практиці досить часто зустрічаються випадки, коли виникають труднощі з визначенням точних значень коренів характеристичного рівняння.

Однак, у більшості випадків при дослідженні розв'язків достатньо знати якісну картину їх поведінки, для чого необхідно знати чи знаходяться корені характеристичного рівняння всередині одиничного кола. Тому для дослідження розв'язків можна застосувати *теорему Шура*, яка визначає необхідні та достатні умови стійкості [...].

Теорема Шура формулюється так: якщо для характеристичного рівняння n -го порядку всі n детермінантів (які наведені нижче) є додатними, то дійсні частини всіх характеристичних коренів будуть меншими одиниці по модулю, тобто:

$$|\operatorname{Re}(\alpha_i)| < 1, \quad i = 1, \dots, n.$$

$$\Delta_1 = \begin{vmatrix} 1 & -a_n \\ -a_n & 1 \end{vmatrix}; \quad \Delta_2 = \begin{vmatrix} 1 & 0 & -a_n & -a_{n-1} \\ -a_1 & 1 & 0 & -a_n \\ -a_n & 0 & 1 & -a_1 \\ -a_{n-1} & -a_n & 0 & 1 \end{vmatrix};$$

$$\Delta_3 = \begin{vmatrix} 1 & 0 & 0 & -a_n & -a_{n-1} & -a_{n-2} \\ -a_1 & 1 & 0 & 0 & -a_n & -a_{n-1} \\ -a_2 & -a_1 & 1 & 0 & 0 & -a_n \\ -a_n & 0 & 0 & 1 & -a_1 & -a_2 \\ -a_{n-1} & -a_n & 0 & 0 & 1 & -a_1 \\ -a_{n-2} & -a_{n-1} & -a_n & 0 & 0 & 1 \end{vmatrix};$$

$$\Delta_n = \begin{vmatrix} 1 & 0 & 0 & \cdot & \cdot & 0 & -a_n & -a_{n-1} & \cdot & \cdot & \cdot & -a_1 \\ -a_1 & 1 & 0 & \cdot & \cdot & 0 & 0 & -a_n & \cdot & \cdot & \cdot & -a_2 \\ -a_2 & -a_1 & 1 & \cdot & \cdot & 0 & 0 & 0 & -a_n & \cdot & \cdot & -a_3 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_{n-1} & -a_{n-2} & -a_{n-3} & \cdot & \cdot & 1 & 0 & 0 & 0 & \cdot & \cdot & -a_n \\ -a_n & 0 & 0 & \cdot & \cdot & 0 & 1 & -a_1 & -a_2 & \cdot & \cdot & -a_{n-1} \\ -a_{n-1} & -a_n & 0 & \cdot & \cdot & 0 & 0 & 1 & \cdot & \cdot & \cdot & -a_{n-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_2 & -a_3 & -a_4 & \cdot & \cdot & 0 & 0 & 0 & \cdot & \cdot & 1 & -a_1 \\ -a_1 & -a_2 & -a_3 & \cdot & \cdot & -a_n & 0 & 0 & \cdot & \cdot & \cdot & 1 \end{vmatrix}.$$

Зазначимо, що кожний з наведених детермінантів можна розглядати як такий, що складається з чотирьох блоків. Кожний блок Δ_i представляє собою трикутну матрицю розмірності $i \times i$. Лівий верхній блок має одиничну головну діагональ і всі нулі над нею. В цьому блоці індекс збільшується на одиницю по мірі руху вниз по стовпчику від діагоналі. Лівий нижній блок представляє собою транспонований верхній правий.

На головній діагоналі верхнього правого блоку знаходяться коефіцієнти a_n , а під цією діагоналлю – нулі. Якщо рухатись вгору по стовчиках від діагоналі, то індекс зменшується на одиницю. Нижній лівий блок представляє собою транспонований верхній правий. Значення коефіцієнта a_0 прийнято за одиницю. Все ці особливості полегшують побудову наведених вище детермінантів.

Якщо прийняти до уваги те, що на практиці досить рідко використовуються рівняння, порядок яких складає більше 3, то обчислення детермінантів реалізується досить просто. Крім того, можна легко побудувати універсальну програму для обчислення детермінантів довільного порядку. Разом з тим необхідно зауважити, що надзвичайно простими у користуванні є наведені вище чотири умови визначення стійкості розв'язку однорідного рівняння n -го порядку, які дозволяють оперативно встановити якісний характер розв'язку.

3.3.2 Знаходження часткових розв'язків різницевого рівнянь

Частковий (неоднорідний) розв'язок зумовлений наявністю у рівнянні констант, а також детермінованих і випадкових змінних і функцій у правій частині. Вигляд часткового розв'язку залежить від конкретного типу збурюючої функції у правій частині різницевого рівняння. Методика знаходження часткових розв'язків складається з таких кроків:

- вибір (формування) пробної функції;
- підстановка пробної функції у рівняння і визначення невідомих коефіцієнтів пробної функції (наприклад, за методом невизначених коефіцієнтів);
- аналіз характеру та збіжності отриманого розв'язку.

Розглянемо декілька характерних випадків.

Випадок 1

У правій частині різницевого рівняння наявні тільки постійна складова (зміщення) a_0 і авторегресійна (АР) частина. Тобто рівняння має вигляд:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \dots + a_n y(k-n). \quad (3.3.7)$$

При подачі константи на вхід лінійного процесу на виході також повинна бути константа. Тобто логічно припустити, що частковим розв'язком цього рівняння повинна бути деяка константа, що можна записати так: $y(k) = y(k-1) = \dots = c$. Підставимо цей пробний розв'язок в (3.3.7) і отримаємо:

$$c = a_0 + a_1 c + a_2 c + \dots + a_n c;$$

звідси

$$c = \frac{a_0}{1 - a_1 - a_2 - \dots - a_n}. \quad (3.3.8)$$

Оскільки знаменник не повинен дорівнювати нулю, то константа c є частковим розв'язком і можна записати:

$$y^p = \frac{a_0}{1 - a_1 - a_2 - \dots - a_n} = \frac{a_0}{1 - \sum_{i=1}^n a_i}. \quad (3.3.9)$$

Якщо $\sum_{i=0}^n a_i = 1$, то маємо окремий випадок, коли процес $\{y(k)\}$ називають *процесом з одиничним коренем*. Оскільки у цьому випадку процес $\{y(k)\}$ розбіжний, то частковий пробний розв'язок у вигляді константи не підходить для використання. Очевидно, що цю розбіжність (*лінійний* або *квазілінійний тренд*) необхідно описати функцією часу. У простому випадку (наявний один одиничний корінь характеристичного рівняння) можна скористатися пробним розв'язком $y^p(k) = ck$. Підставимо цей пробний розв'язок в рівняння (3.3.7) і отримаємо:

$$ck = a_0 + (k-1)a_1 c + (k-2)a_2 c + \dots + (k-n)a_n c$$

або

$$(1 - a_1 - a_2 - \dots - a_n)ck = a_0 - c(a_1 + 2a_2 + 3a_3 + \dots + na_n).$$

Оскільки $1 - a_1 - a_2 - \dots - a_n = 0$, то

$$c = \frac{a_0}{a_1 + 2a_2 + 3a_3 + \dots + na_n}.$$

Наприклад, для рівняння

$$y(k) = 2,5 + 0,75 y(k-1) + 0,25 y(k-2),$$

$a_1 + a_2 = 0,75 + 0,25 = 1$, і частковий розв'язок має вигляд ck , де $c = 2,5 / (0,75 + 2 \cdot 0,25) = 2,0$. Якщо пробна функція ck не приводить до успішного знаходження розв'язку, то необхідно взяти іншу пробну функцію. Наприклад, можна скористатися нелінійними функціями виду $y^p(k) = ck^2, ck^3, \dots, ck^n$. Для рівняння довільного порядку одна з цих функцій обов'язково буде частковим розв'язком.

Випадок 2

Лінійний процес збурюється експонентою $x(k) = bd^{rk}$, де b, d, r – константи. Наприклад, для економічних процесів коефіцієнт r можна розглядати як фактор зростання. Розглянемо послідовність знаходження часткового розв'язку на прикладі рівняння першого порядку:

$$y(k) = a_0 + a_1 y(k-1) + bd^{rk}. \quad (3.3.10)$$

У правій частині цього рівняння є константа a_0 . Тому можна очікувати, що в частковому розв'язку також буде наявна константа. Для лінійного процесу можна також передбачити, що при подачі на його вхід сигналу bd^{rk} на виході з'явиться реакція такої ж форми, але з деяким коефіцієнтом передачі. Тому логічно записати пробний розв'язок у формі:

$$y^p(k) = c_0 + c_1 d^{rk},$$

де c_0, c_1 – невідомі константи, які знайдемо за методом невизначених коефіцієнтів. Підставимо пробний розв'язок в рівняння (3.3.10) і в результаті отримаємо:

$$c_0 + c_1 d^{rk} = a_0 + a_1 (c_0 + c_1 d^{r(k-1)}) + bd^{rk}. \quad (3.3.11)$$

Перенесемо в ліву частину члени, які містять невідомі константи c_0, c_1 і об'єднаємо їх таким чином:

$$c_0 + c_1 \frac{d^{rk}}{1 - a_1} \left(\frac{d^r - a_1}{d^r} \right) = \frac{a_0}{1 - a_1} + \frac{b}{1 - a_1} d^{rk}.$$

Звідси знайдемо, що

$$c_0 = \frac{a_0}{1 - a_1}; \quad c_1 \frac{d^{rk}}{1 - a_1} \frac{d^r - a_1}{d^r} = \frac{b}{1 - a_1} d^{rk},$$

або

$$c_1 = \frac{bd^r}{d^r - a_1}.$$

Тепер запишемо частковий розв'язок рівняння (3.3.10):

$$y^p(k) = \frac{a_0}{1 - a_1} + \frac{bd^r}{d^r - a_1} d^{rk}. \quad (3.3.12)$$

Якщо $|d^r| < 1$, то (3.3.12) збігається до значення $a_0/(1 - a_1)$ при $k \rightarrow \infty$. Якщо $a_1 = 1$, то при знаходженні часткового розв'язку необхідно взяти пробну функцію $c_0 = ck$. А якщо $a_1 = d^r$, то необхідно сформулювати пробну функцію у вигляді:

$$c_1 = \frac{kbd^r}{d^r - a_1}.$$

Методика знаходження часткового розв'язку для рівнянь вищих порядків залишається такою самою. Спочатку необхідно коректно вибрати пробну функцію у відповідність із структурою рівняння, а потім застосувати метод невизначених коефіцієнтів для знаходження невідомих параметрів.

Випадок 3

На вхід процесу подається *функція* у формі $x(k) = bk^d$, де b – константа; d – ціле додатне число. Різницеве рівняння має вигляд:

$$y(k) = a_0 + \sum_{i=1}^n a_i y(k-i) + bk^d. \quad (3.3.13)$$

У даному випадку $y(k)$ залежить від k^d , значення $y(k-1)$ залежить від $(k-1)^d$, значення $y(k-2)$ залежить від $(k-2)^d$ і так далі. Тому пробну функцію для часткового розв'язку можна вибрати у формі:

$$y^p(k) = c_0 + c_1 k + c_2 k^2 + \dots + c_d k^d.$$

Для того щоб знайти невідомі коефіцієнти c_i , підставимо пробний розв'язок в рівняння другого порядку:

$$c_0 + c_1 k = a_0 + a_1 [c_0 + (k-1)c_1] + a_2 [c_0 + (k-2)c_1] + bk. \quad (3.3.14)$$

Значення коефіцієнтів c_0, c_1 виберемо такими, щоб (3.3.14) залишалось тотожністю для всіх можливих значень k . Згрупуємо всі константи і всі члени, які мають множник k :

$$c_0 - a_0 - a_1 c_0 + a_1 c_1 - a_2 c_0 + 2a_2 c_1 = a_1 k c_1 + a_2 k c_1 + bk - c_1 k.$$

Звідси отримаємо значення невідомих констант за *методом невизначених коефіцієнтів*:

$$\begin{aligned} c_1 &= b / (1 - a_1 - a_2), \\ c_0 &= [a_0 - (2a_2 + a_1)c_1] / (1 - a_1 - a_2) \end{aligned}$$

або

$$c_0 = \frac{a_0}{1 - a_1 - a_2} - \frac{b(2a_2 + a_1)}{(1 - a_1 - a_2)^2}.$$

Таким чином, частковий розв'язок включає в себе лінійний часовий тренд. Якщо $a_1 + a_2 = 1$, то пробний розв'язок необхідно домножити на k . Методика залишається незмінною для рівнянь вищих порядків.

Приклад 3.4. Розглянемо детермінований процес авторегресії другого порядку:

$$\begin{aligned} y(k) &= 2,5 + a_1(k-1) + a_2 y(k-2), \\ a_1 &= 0,75; \quad a_2 = 0,25; \quad y(0) = 1; \quad y(1) = 2. \end{aligned}$$

Оскільки сума коефіцієнтів моделі $\sum_{i=1}^2 a_i = 0,75 + 0,25 = 1$, то за пробну функцію часткового розв'язку вибираємо $y_{проб}^p(k) = ck$, де

$$c = \frac{a_0}{\sum_{i=1}^2 a_i} = \frac{2,5}{0,75 + 2 \cdot 0,25} = \frac{2,5}{1,25} = 2,0.$$

Таким чином, частковий розв'язок має вигляд: $y^p(k) = 2,0k$.

Характеристичне рівняння $\alpha^2 - 0,75\alpha - 0,25 = 0$ має корені: $\alpha_1 = 1$, $\alpha_2 = -0,25$. Запишемо повний розв'язок

$$y(k) = A_1 + A_2(-0,25)^k + 2,0k.$$

Використаємо початкові умови для знаходження довільних констант:

$$\begin{aligned} k=0: \quad 1 &= A_1 + A_2, \\ k=1: \quad 2 &= A_1 - 0,25A_2 + 2,0. \end{aligned}$$

Розв'язуючи цю систему двох рівнянь, знайдемо: $A_1 = 0,2$; $A_2 = 0,8$.

Приклад 3.5. Необхідно знайти повний розв'язок рівняння:

$$y(k) = a_0 + 2y(k-1) - y(k-2).$$

Характеристичне рівняння $\alpha^2 - 2\alpha + 1 = 0$ має корені: $\alpha_1 = \alpha_2 = 1$, тобто маємо випадок кратних коренів. Оскільки пробна функція $y_{проб.}^p(k) = ck$ не приводить до успіху, то вибираємо за пробну $y_{проб.}^p(k) = ck^2$. В результаті підстановки пробного розв'язку у вихідне рівняння знайдемо:

$$c = \frac{a_0}{k^2 - 2(k-1)^2 + (k-2)^2} = \frac{a_0}{2}.$$

Тепер запишемо повний розв'язок:

$$y(k) = A_1 + A_2k + \frac{a_0}{2}k^2.$$

Випадок 4

Застосування методу невизначених коефіцієнтів для знаходження часткового розв'язку *стохастичного* різницевого рівняння (у правій частині наявний випадковий процес).

Процес авторегресії першого порядку

Як уже зазначалось вище, стохастична авторегресія першого порядку описується рівнянням

$$y(k) = a_0 + a_1y(k-1) + \varepsilon(k).$$

Природа процесу $\{y(k)\}$ в даному випадку така, що частковий розв'язок залежить тільки від постійного члена a_0 , часу k та послідовності $\{\varepsilon(k)\}$. Враховуючи, що часова змінна у збурюючій функції задана неявно, то змінна k може входити в частковий розв'язок тільки у тому випадку, коли корінь характеристичного рівняння дорівнює 1.

Для ілюстрації методу скористаємось такою пробною функцією:

$$y(k) = b_0 + b_1k + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i), \quad (3.3.15)$$

де b_0, b_1, α_i – невідомі коефіцієнти. Для того щоб знайти значення цих коефіцієнтів, підставимо пробний розв'язок (3.3.15) в РР першого порядку:

$$\begin{aligned} b_0 + b_1k + \alpha_0\varepsilon(k) + \alpha_1\varepsilon(k-1) + \alpha_2\varepsilon(k-2) + \dots = \\ = a_0 + a_1[b_0 + b_1(k-1) + \alpha_0\varepsilon(k-1) + \alpha_1\varepsilon(k-2) + \dots] + \varepsilon(k). \end{aligned}$$

Після групування членів отримаємо:

$$(b_0 - a_0 - a_1 b_0 + a_1 b_1) + b_1(1 - a_1)k + (\alpha_0 - 1)\varepsilon(k) + (\alpha_1 - a_1 \alpha_0)\varepsilon(k - 1) + (\alpha_2 - a_1 \alpha_1)\varepsilon(k - 2) + (\alpha_3 - a_1 \alpha_2)\varepsilon(k - 3) + \dots = 0. \quad (3.3.16)$$

Отримане рівняння (3.3.16) повинне бути справедливим при всіх можливих значеннях k і всіх можливих значеннях $\varepsilon(k - i)$, $i = 0, 1, 2, \dots$; тобто повинні виконуватись такі умови:

$$\begin{aligned} \alpha_0 - 1 &= 0, \\ \alpha_1 - a_1 \alpha_0 &= 0, \\ \alpha_2 - a_1 \alpha_1 &= 0, \\ &\vdots \\ b_0 - a_0 - a_1 b_0 + a_1 b_1 &= 0, \\ b_1 - a_1 b_1 &= 0. \end{aligned} \quad (3.3.17)$$

Зазначимо, що першою групою умов (перші три рядки) можна скористатись для знаходження розв'язку відносно α_i в рекурсивній формі. Використовуючи умови (3.3.17), знайдемо:

$$\alpha_0 = 1, \quad \alpha_1 = a_1, \quad \alpha_2 = a_1 \alpha_1 = a_1^2, \quad \alpha_i = a_1^i.$$

Якщо $a_1 \neq 1$ (немає одиничного кореня характеристичного рівняння), то $b_1 = 0$, $b_0 = a_0 / (1 - a_1)$, і частковий розв'язок має вигляд:

$$y(k) = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k - i).$$

Такий же розв'язок був отриманий вище для РР першого порядку за методом ітерацій. Загальний розв'язок отримаємо в результаті об'єднання часткового та однорідного розв'язків:

$$y(k) = \frac{a_0}{1 - a_1} + A a_1^k + \sum_{i=0}^{\infty} a_1^i \varepsilon(k - i). \quad (3.3.18)$$

Якщо відома початкова умова $y(0) = y_0$, то можна знайти значення довільної константи із рівняння:

$$y_0 = \frac{a_0}{1 - a_1} + A + \sum_{i=0}^{\infty} a_1^i \varepsilon(-i).$$

Таким чином загальний розв'язок можна записати так:

$$y(k) = \frac{a_0}{1-a_1} + a_1^k \left[y_0 - \frac{a_0}{1-a_1} - \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) \right] + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i),$$

або

$$y(k) = \frac{a_0}{1-a_1} + a_1^k \left(y_0 - \frac{a_0}{1-a_1} \right) + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i). \quad (3.3.19)$$

Якщо $a_1 = 1$ (корінь характеристичного рівняння також дорівнює одиниці), то b_0 може бути довільною константою, а $b_1 = a_0$. У цьому випадку можливою формою часткового розв'язку може бути функція, яка включає добуток $a_0 k$, оскільки з'являється одиничний корінь. Така функція має вигляд:

$$y(k) = b_0 + a_0 k + \sum_{i=0}^{\infty} \varepsilon(k-i). \quad (3.3.20)$$

Таку функцію називають «неналежною», тому що сума елементів послідовності $\{\varepsilon(k)\}$ може бути нескінченною. Для початкової умови y_0 запишемо

$$y_0 = b_0 + \sum_{i=0}^{\infty} \varepsilon(-i) \quad \text{або} \quad b_0 = y_0 - \sum_{i=0}^{\infty} \varepsilon(-i).$$

Із врахуванням цього результату (3.3.20) приймає вигляд:

$$y(k) = y_0 + a_0 k + \sum_{i=0}^k \varepsilon(k-i).$$

Стохастичний процес АРКС(1,1)

Як розширення попереднього випадку, розглянемо тепер стохастичне рівняння АРКС(1,1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-1). \quad (3.3.21)$$

Розв'язок цього рівняння буде включати константу, зважену послідовність $\{\varepsilon(k)\}$ та час k у першій степені (тому що рівняння першого порядку). Як і у попередньому прикладі, k потрібно включати у пробний розв'язок, якщо корінь характеристичного рівняння дорівнює одиниці. Таким чином, скористаємось таким пробним розв'язком:

$$y_{np}(k) = b_0 + b_1 k + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i).$$

Підставимо пробний розв'язок в (3.3.21) і отримаємо:

$$b_0 + b_1 k + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i) = a_0 + a_1 \left[b_0 + b_1(k-1) + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i-1) \right] + \varepsilon(k) + \beta_1 \varepsilon(k-1)$$

або

$$b_0 + b_1 k + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i) = a_0 + a_1 b_0 + a_1 b_1 k - a_1 b_1 + a_1 \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i-1) + \varepsilon(k) + \beta_1 \varepsilon(k-1).$$

Прирівнюючи коефіцієнти при однакових змінних $\varepsilon(k)$, $\varepsilon(k-1)$, $\varepsilon(k-2)$, ..., отримаємо такі значення невідомих коефіцієнтів розв'язку:

$$\begin{aligned} \alpha_0 &= 1; \\ \alpha_1 &= a_1 \alpha_0 + \beta_1 = a_1 + \beta_1; \\ \alpha_2 &= a_1 \alpha_1 = a_1(a_1 + \beta_1); \\ \alpha_3 &= a_1 \alpha_2 = a_1^2(a_1 + \beta_1); \\ &\vdots \\ \alpha_i &= a_1 \alpha_{i-1} \quad \text{або} \quad \alpha_i = a_1^{i-1} (a_1 + \beta_1). \end{aligned}$$

Прирівнюючи константи та члени, які містять час k , отримаємо:

$$\begin{aligned} b_0 &= a_0 + a_1 b_0 - a_1 b_1; \\ b_1 &= a_1 b_1. \end{aligned}$$

Для даного прикладу також необхідно розглянути два випадки: $a_1 \neq 1$ і $a_1 = 1$. Якщо $a_1 \neq 1$, то це означає, що тренду немає і $b_1 = 0$. Звідси випливає, що

$$b_0 = \frac{a_0}{1 - a_1},$$

а частковий розв'язок має вигляд:

$$y(k) = \frac{a_0}{1 - a_1} + \varepsilon(k) + (a_1 + \beta_1) \sum_{i=1}^{\infty} a_1^{i-1} \varepsilon(k-i).$$

Загальний розв'язок включає ще однорідний:

$$y(k) = \frac{a_0}{1 - a_1} + A a_1^k + \varepsilon(k) + (a_1 + \beta_1) \sum_{i=1}^{\infty} a_1^{i-1} \varepsilon(k-i).$$

Якщо $a_1 = 1$, то з виразу $b_0 = a_0 + a_1 b_0 - a_1 b_1$ випливає, що $b_1 = a_0$, а коефіцієнт b_0 є довільною константою, яка може бути визначена за допомогою початкових умов. Таким чином, «неналежна» форма розв'язку для даного випадку має вигляд:

$$y(k) = b_0 + a_0 k + \varepsilon(k) + (1 + \beta_1) \sum_{i=1}^{\infty} \varepsilon(k-i). \quad (3.3.22)$$

Покладаючи початкову умову $y(0) = y_0$ при $k = 0$, отримаємо:

$$y_0 = b_0 + \varepsilon_0 + (1 + \beta_1) \sum_{i=1}^{\infty} \varepsilon(-i),$$

що дає змогу визначити b_0 за виразом: $b_0 = y_0 - \varepsilon(0) - (1 + \beta_1) \sum_{i=1}^{\infty} \varepsilon(-i)$.

Підставимо цей вираз в (3.3.22):

$$\begin{aligned} y(k) &= y_0 - \varepsilon(0) - (1 + \beta_1) \sum_{i=1}^{\infty} \varepsilon(-i) + a_0 k + \varepsilon(k) + (1 + \beta_1) \sum_{i=1}^{\infty} \varepsilon(k-i) = \\ &= y_0 + a_0 k - \varepsilon(0) - (1 + \beta_1) \varepsilon(-1) - (1 + \beta_1) \varepsilon(-2) - (1 + \beta_1) \varepsilon(-3) - \dots + \varepsilon(k) + \\ &+ (1 + \beta_1) \varepsilon(k-1) + (1 + \beta_1) \varepsilon(k-2) + \dots + (1 + \beta_1) \varepsilon(0) + (1 + \beta_1) \varepsilon(-1) + \dots = \\ &= y_0 + a_0 k + \varepsilon(k) + (1 + \beta_1) \sum_{i=1}^{k-1} \varepsilon(k-i). \end{aligned}$$

Стохастичний процес $AR(2)$

Оскільки на практиці досить часто зустрічаються моделі другого порядку, то розглянемо знаходження розв'язку для стохастичної авторегресії $AR(2)$. Це рівняння має вигляд:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k). \quad (3.3.23)$$

Оскільки це рівняння другого порядку, то його характеристичне рівняння може мати два одиничних корені, а тому максимальний порядок полінома для опису можливого тренду повинен складати 2. Тобто пробний розв'язок сформуємо так:

$$y_{np}(k) = b_0 + b_1 k + b_2 k^2 + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i),$$

де b_0, b_1, b_2, α_i – невідомі коефіцієнти розв’язку рівняння (3.3.23). Підстановка пробного розв’язку в рівняння (3.3.23) приводить до такого виразу:

$$\begin{aligned} & b_0 + b_1 k + b_2 k^2 + \alpha_0 \varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \dots = \\ & = a_0 + a_1 [b_0 + b_1(k-1) + b_2(k-1)^2 + \alpha_0 \varepsilon(k-1) + \alpha_1 \varepsilon(k-2) + \alpha_2 \varepsilon(k-3) + \dots] + \\ & + a_2 [b_0 + b_1(k-2) + b_2(k-2)^2 + \alpha_0 \varepsilon(k-2) + \alpha_1 \varepsilon(k-3) + \alpha_2 \varepsilon(k-4) + \dots] + \varepsilon(k). \end{aligned} \quad (3.3.24)$$

Із умов, які приводять наведений вираз до тотожності при довільних реалізаціях процесу $\{\varepsilon(k)\}$, отримуємо такі значення невизначених коефіцієнтів:

$$\begin{aligned} \alpha_0 &= 1; \\ \alpha_1 &= a_1 \alpha_0 = a_1; \\ \alpha_2 &= a_1 \alpha_1 + a_2 \alpha_0 = a_1^2 + a_2; \\ \alpha_3 &= a_1 \alpha_2 + a_2 \alpha_1 = a_1^3 + 2a_1 a_2; \\ &\vdots \end{aligned}$$

З наведених виразів для обчислення коефіцієнтів α_i видно, що починаючи з α_2 , всі коефіцієнти можна визначити за допомогою різницевого рівняння другого порядку:

$$\alpha_j = a_1 \alpha_{j-1} + a_2 \alpha_{j-2}.$$

Оскільки α_0 і α_1 відомі, то всі наступні коефіцієнти можна визначити ітераційно.

Властивості збіжності невідомих коефіцієнтів можна визначити так само, як і властивості збіжності розв’язку однорідного рівняння другого порядку, тобто:

1. Для збіжності розв’язку однорідного рівняння другого порядку необхідно виконати такі умови: $|a_2| < 1$, $a_1 + a_2 < 1$ і $a_2 - a_1 < 1$. Збіжність також передбачає, що вплив минулих значень послідовності $\{\varepsilon(k)\}$ на поточне значення $y(k)$ зменшується із плином часу.

2. Якщо коефіцієнти збігаються до деяких скінченних усталених значень, то ця збіжність має безпосередній прямий характер при $(a_1^2 + 4a_2) > 0$. Якщо $(a_1^2 + 4a_2) < 0$, то збіжність буде мати гармонічну форму, а при $(a_1^2 + 4a_2) = 0$ спочатку спостерігається швидке збільшення цих значень з подальшою збіжністю.

Розглянемо члени тотожності (3.3.24), які не зв'язані з випадковою послідовністю $\{\varepsilon(k)\}$:

$$b_2(1-a_1-a_2)k^2 + [b_1(1-a_1-a_2) + 2b_2(a_1+2a_2)]k + [b_0(1-a_1-a_2) - a_0 + a_1(b_1-b_2) + 2a_2(b_1-2b_2)] = 0. \quad (3.3.25)$$

Рівність лівої частини цього виразу нулю повинна виконуватись для $\forall k$.

Спочатку розглянемо випадок, коли $a_1 + a_2 \neq 1$. Оскільки в такому випадку $1 - a_1 - a_2 \neq 0$, то для збіжності необхідно покласти $b_2 = 0$. Оскільки для забезпечення збіжності коефіцієнт при k повинен дорівнювати нулю, то необхідно також покласти $b_1 = 0$. Якщо $b_1 = b_2 = 0$, то можна знайти, що

$$b_0 = \frac{a_0}{1 - a_1 - a_2}.$$

Якщо ж $a_1 + a_2 = 1$, то значення коефіцієнтів b_0, b_1, b_2 будуть залежати від конкретних значень коефіцієнтів a_0, a_1, a_2 . Такі випадки потрібно розглядати окремо.

Таким чином, стосовно збіжності часткового розв'язку можна зробити наступний висновок.

Умовами збіжності часткового розв'язку є умови стійкості розв'язку однорідного рівняння. Якщо корені характеристичного рівняння одиничні, то в частковому розв'язку з'являється поліноміальний тренд, а порядок полінома дорівнює кількості одиничних коренів.

Примітка: Якщо складові збурюючої функції у правій частині рівняння, тобто $x(k)$, мають такий же вигляд як і складові $y^h(k)$, то пробна функція часткового розв'язку змінюється. Всі складові $y^p(k)$, які відповідають таким же складовим $y^h(k)$, необхідно помножити на k в такій найменшій степені, щоб їх тотожність не порушувалась.

Відповідний приклад буде розглянуто в наступному параграфі.

Процес $AP(2)$ з випадковою та детермінованою функціями на вході

Розглянемо процес з випадковою та детермінованою функціями на вході:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + x(k) + \varepsilon(k), \quad (3.3.26)$$

де $a_0 = 0,3$; $a_1 = -0,5$; $a_2 = 1,5$; $x(k) = c_1 e^{c_2(k+1)}$; $\varepsilon(k)$ – центрований випадковий процес типу білого шуму; c_1, c_2 – константи.

Характеристичне рівняння $\alpha^2 + 0,5\alpha - 1,5 = 0$ має корені: $\alpha_1 = 1$ і $\alpha_2 = -1,5$. Тобто, це процес з одиничним коренем, який має тренд. Таким чином, однорідний розв'язок має вигляд:

$$y^h(k) = A_1 + A_2(-1,5)^k.$$

Підставимо пробну функцію

$$y_{np}^p(k) = b_0 + b_1 k + b_2 k^2 + b_3 e^{c_2(k+1)} + \sum_{i=1}^{\infty} \alpha_i \varepsilon(k-i)$$

в рівняння (3.3.26):

$$\begin{aligned} b_0 + b_1 k + b_2 k^2 + b_3 e^{c_2(k+1)} + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i) = \\ = a_0 + a_1 [b_0 + b_1(k-1) + b_2(k-1)^2 + b_3 e^{c_2 k} + \alpha_0 \varepsilon(k-1) + \alpha_1 \varepsilon(k-2) + \dots] + \\ + a_2 [b_0 + b_1(k-2) + b_2(k-2)^2 + b_3 e^{c_2(k-1)} + \alpha_0 \varepsilon(k-2) + \alpha_2 \varepsilon(k-3) + \dots] + \\ + c_1 e^{c_2(k+1)}. \end{aligned}$$

Прирівняємо коефіцієнти в лівій і правій частині при однакових змінних:

$$b_0 = a_0 + a_1 b_0 - a_1 b_1 + a_1 b_2 + a_2 b_0 - a_2 b_1 + 4a_2 b_2; \quad (3.3.27)$$

Коефіцієнти при k :

$$b_1 = a_1 b_1 - 2a_1 b_2 + a_2 b_1 - 4a_2 b_2 \quad \text{або} \quad b_1(1 - a_1 - a_2) = -2b_2(a_1 + 2a_2).$$

Оскільки $1 - a_1 - a_2 = 1 + 0,5 - 1,5 = 0$, а $a_1 + 2a_2 = -0,5 + 5 = 2,5$, то необхідно покласти $b_2 = 0$.

При $b_2 = 0$ із (3.3.27) можна записати, що

$$b_0 = a_0 + a_1 b_0 - a_1 b_1 + a_2 b_0 - a_2 b_1 \quad \text{або} \quad b_0(1 - a_1 - a_2) = a_0 - b_1(a_1 + a_2).$$

$$\text{Звідси} \quad b_1 = \frac{a_0}{a_1 + a_2} = \frac{0,3}{-0,5 + 1,5} = 0,3.$$

Прирівняємо ще коефіцієнти при $e^{c_2 k}$:

$$b_3 e^{c_2} = a_1 b_3 + a_2 b_3 e^{-c_2} + c_1 e^{c_2},$$

і визначимо константу b_3 :

$$b_3 = \frac{c_1 e^{c_2}}{e^{c_2} - a_1 - a_2 e^{-c_2}} \quad \text{або} \quad b_3 = \frac{c_1}{1 - a_1 e^{-c_2} - a_2 e^{-2c_2}} = \frac{c_1}{1 + 0,5 e^{-c_2} - 1,5 e^{-2c_2}}.$$

Залишився невизначеним коефіцієнт b_0 , який можна знайти за допомогою початкових умов.

Запишемо повний розв'язок:

$$y(k) = b_0 + A_1 + A_2 (-1,5)^k + 0,3k + \frac{c_1}{1 + 0,5 e^{-c_2} - 1,5 e^{-2c_2}} e^{c_2(k+1)} + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i),$$

де $\alpha_i = -0,5\alpha_{i-1} + 1,5\alpha_{i-2}$. Очевидно, що цей процес має тренд, на який накладаються коливання у вигляді адитивних складових $A_2(-1,5)^k$ та $\sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i)$.

3.4 Приклади знаходження повних розв'язків різницевого рівнянь

Приклад 3.4.1. Модель процесу інфляції.

Для моделювання процесу інфляції запропоновано різницеве рівняння такого вигляду:

$$m(k) = \alpha + p(k) - \beta[p^e(k+1) - p(k)], \quad \beta > 0,$$

де $m(k)$ – логарифм номінального забезпечення грошовою масою в момент k ; $p(k)$ – логарифм поточного рівня цін; $p^e(k+1)$ – логарифм рівня цін, які

очікуються в момент часу $k + 1$. Очевидно, що наведене рівняння може бути використане тільки для нормованих безрозмірних змінних.

Перепишемо наведену вище модель у вигляді:

$$m(k) - p(k) = \alpha - \beta[p^e(k+1) - p(k)].$$

Тепер можна стверджувати, що потреба в грошах $[m(k) - p(k)]$ знаходиться у від'ємному взаємозв'язку із приростом цін $[p^e(k) - p(k)]$, тобто відносне перевищення цін над об'ємом грошової маси в лівій частині рівняння вимагає знаку мінус перед приростом цін у правій частині.

Звичайно, що наведена вище модель є дуже спрощеною для економіки перехідного періоду, тому що для її побудови використано тільки дві змінні. Фактично, на інфляцію в перехідний період впливає ряд інших факторів, такі як значне падіння рівня виробництва, перекачування валюти за кордон (в той час як вона повинна використовуватися для збільшення внутрішніх інвестицій), рівень тіньової економіки і т. ін. Така модель буде набагато складнішою за своєю структурою, але адекватнішою процесу інфляції.

Для визначення впливу на процес інфляції майбутніх збурень отримаємо розв'язок рівняння відносно цих значень. Нехай забезпечення грошовою масою має такий простий вигляд:

$$m(k) = m_a + \varepsilon(k),$$

де m_a – середнє значення випуску грошової маси; $\varepsilon(k)$ – некорельований випадковий процес з нульовим середнім, $E[\varepsilon(k)] = 0$. Покладемо також, що

$$p^e(k+1) = p(k+1),$$

тобто прогнозовані ціни відповідають дійсним значенням цін в момент $k + 1$.

Таким чином, рівняння рівноваги грошової маси та рівня цін приймає вигляд:

$$m_a + \varepsilon(k) - p(k) = \alpha - \beta[p(k+1) - p(k)]$$

або

$$p(k+1) - \left(1 + \frac{1}{\beta}\right)p(k) = -(m_a - \alpha)/\beta - \varepsilon(k)/\beta.$$

Знайдемо розв'язок цього рівняння. Для знаходження часткового розв'язку скористаємось методом невизначених коефіцієнтів. Оскільки перед змінною $p(k)$ стоїть коефіцієнт $a_1 = 1 + 1/\beta > 1$, то будемо шукати розв'язок відносно майбутніх відліків збурень. Пробний розв'язок запишено у вигляді:

$$p_{проб}(k) = b_0 + \sum_{i=0}^{\infty} \gamma_i \varepsilon(k+i).$$

Підставимо його у наведене вище рівняння інфляції для того щоб знайти невідомі коефіцієнти b_0, γ_i :

$$b_0 + \sum_{i=0}^{\infty} \gamma_i \varepsilon(k+1+i) - (1+1/\beta) \left[b_0 + \sum_{i=0}^{\infty} \gamma_i \varepsilon(k+i) \right] = [\alpha - m_a - \varepsilon(k)]/\beta.$$

Для того щоб останнє рівняння було тотожністю для всіх можливих ненульових значень $\{\varepsilon(k)\}$ необхідно прирівняти коефіцієнти при однакових значеннях $\varepsilon(k+i), i=0, 1, 2, \dots$:

– константи: $b_0 - b_0(1+\beta)/\beta = (\alpha - m_a)/\beta \Rightarrow b_0 = m_a - \alpha,$

– коефіцієнти при $\varepsilon(k)$: $-\gamma_0(1+\beta)/\beta = -1/\beta \Rightarrow \gamma_0 = 1/(1+\beta),$

– коефіцієнти при $\varepsilon(k+1)$: $\gamma_0 - \gamma_1(1+\beta)/\beta = 0 \Rightarrow \gamma_1 = \beta/(1+\beta)^2,$

•
•
•

– для довільного моменту часу: $\gamma(i) - \gamma(i-1)(1+\beta)/\beta \Rightarrow \gamma_i = \beta^i / (1+\beta)^{i+1}.$

Частковий розв'язок у компактній формі матиме вигляд:

$$p^p(k) = m_a - \alpha + \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(k+i).$$

Для рівняння першого порядку,

$$p(k+1) - \left(1 + \frac{1}{\beta}\right) p(k) = 0,$$

однорідний розв'язок запишемо так:

$$p^h(k) = A \left(1 + \frac{1}{\beta}\right)^k.$$

Таким чином, загальний розв'язок, як сума часткового та однорідного, приймає такий вигляд:

$$p(k) = m_a - \alpha + \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(k+i) + A \left(1 + \frac{1}{\beta}\right)^k.$$

Очевидно, що $\lim_{i \rightarrow \infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} = 0,$ тобто частковий розв'язок не

розбігається, але однорідний буде зростати при збільшенні k . Для того щоб уникнути розбіжності однорідного розв'язку, необхідно покласти $A = 0$.

Скористаємось початковими умовами для визначення умов збіжності однорідного розв'язку:

$$p_0 = m_a - \alpha + \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(i) + A,$$

або

$$A = p_0 - m_a + \alpha - \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(i).$$

Таким чином, для того щоб $A = 0$ необхідно мати таку початкову умову:

$$p_0 = m_a - \alpha + \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(i).$$

У цьому рівнянні $(m_a - \alpha)$ – детермінована умова довгострокової “рівноваги”, яка зустрічається також в інших моделях макроекономіки. Якщо розв'язок початкового рівняння (моделі інфляції) збігається, то він повинен збігатися до цієї детермінованої частини. Складова розв'язку

$$\frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(k+1)$$

ілюструє вплив збурень на величину грошової маси в обороті. Якщо прийняти до уваги те, що

$$\left| \frac{\beta}{1+\beta} \right| < 1,$$

то вплив цієї складової на поведінку розв'язку в цілому буде помітним тільки на короткому початковому проміжку часу (при відносно невеликих значеннях k), оскільки коефіцієнти зменшуватимуться з ростом значення степені $i+1$.

Із сказаного можна зробити висновок, що частковий розв'язок характеризує поведінку (рівновагу) головної змінної $p(k)$ на короткому та більш значному інтервалах часу. Однорідний розв'язок можна розглядати як порушення рівноваги в початковий період часу.

Оскільки частковий розв'язок

$$p^p(k) = m_a - \alpha + \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(k+i)$$

представляє собою рівняння рівноваги між ціною та грошовою масою для k – го періоду часу, то початкове значення

$$p_0 = m_a - \alpha + \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(i)$$

є рівноважним значенням ціни для нульового періоду. Таким чином, умова

$$A(1+1/\beta)^k = 0$$

є умовою нульового відхилення від рівноваги (ціни \leftrightarrow гроші) в початковий період часу.

Якщо на послідовність $\{p(k)\}$ накласти обмеження, що вона не повинна розбігатися, то загальний розв'язок (його називають у такому вигляді ще імпульсною характеристикою – частковий розв'язок без константи) матиме вигляд:

$$p(k) = m_a - \alpha + \frac{1}{\beta} \sum_{i=0}^{\infty} \left(\frac{\beta}{1+\beta} \right)^{i+1} \varepsilon(k+i).$$

Відзначимо, що у кожний момент часу k ціна пропорційна середньому значенню грошової маси в обороті. Це можна легко показати, оскільки всі змінні задані в логарифмічному масштабі і часткова похідна

$$\frac{\partial p(k)}{\partial m} = 1.$$

Короткочасні зміни поставок грошової маси в оборот поводять себе таким чином. Наслідком останнього розв'язку є те, що зростання грошової маси в майбутньому, яке представлене значеннями $\varepsilon(k+1)$, призводять до зростання цін в поточний період часу. Але основна ідея полягає в тому, що збільшення випуску грошової маси в майбутньому буде сприяти також збільшенню цін в майбутньому, а звідси впливає, що гроші не вигідно тримати без руху, вони повинні бути в обороті.

Приклад 3.4.2. Знайти загальний розв'язок рівняння:

$$y(k) = -2y(k-1) - y(k-2) + k(-1)^k. \quad (3.4.1)$$

Запишемо характеристичне рівняння:

$$\lambda^2 + 2\lambda + 1 = 0$$

і знайдемо його корені: $\lambda_{1,2} = \frac{1}{2}(-2 \pm \sqrt{4-4})$, тобто, $\lambda_1 = \lambda_2 = -1$. Запишемо однорідний розв'язок із врахуванням наявності кратного кореня

$$y^h(k) = A_1(-1)^k + A_2 k(-1)^k.$$

Якщо складові збурюючої функції у правій частині рівняння (3.4.1), тобто, $x(k) = k(-1)^k$ мають такий же вигляд як і складові $y^h(k)$, то пробна функція часткового розв'язку змінюється. Всі складові $y^p(k)$, які відповідають таким же складовим $y^h(k)$, необхідно помножити на k в такій найменшій степені, щоб їх тотожність порушувалась. Таким чином, пробна функція для часткового розв'язку:

$$y_{проб}^p(k) = [c_2 k + c_3 k^2] k(-1)^k = [c_2 k^2 + c_3 k^3](-1)^k. \quad (3.4.2)$$

Підставимо (3.4.2) в (3.4.1) і знайдемо невідомі коефіцієнти c_2, c_3 за допомогою методу невизначених коефіцієнтів. Таким чином, отримаємо:

$$[c_2 k^2 + c_3 k^3](-1)^k = -2[c_2(k-1)^2 + c_3(k-1)^3](-1)^{(k-1)} - \\ - [c_2(k-2)^2 + c_3(k-2)^3](-1)^{(k-2)} + k(-1)^k$$

або

$$[c_2 k^2 + c_3 k^3](-1)^k = 2[c_2(k-1)^2 + c_3(k-1)^3](-1)^k - \\ - [c_2(k-2)^2 + c_3(k-2)^3](-1)^k + k(-1)^k.$$

Звідси знайдемо значення невідомих коефіцієнтів $c_2 = 1/2$; $c_3 = 1/6$ і повний розв'язок

$$y(k) = \left[c_0 + c_1 k + \frac{1}{2} k^2 + \frac{1}{6} k^3 \right] (-1)^k.$$

Приклад 3.4.3. Для порівняння дискретної моделі з неперервною розглянемо розв'язок диференціального рівняння другого порядку, яке описує вимушені коливання

$$\ddot{y}(t) + 2h\dot{y}(t) + r^2 y(t) = x(t), \quad (3.4.3)$$

де h – коефіцієнт, що характеризує властивості опору середовища (демпфірування); r – коефіцієнт жорсткості. Нехай збурююча сила змінюється за гармонічним законом $x(t) = M \sin(\omega t)$, а опір середовища покладемо для простоти нульовим, $h = 0$. Тепер рівняння (3.4.3) приймає вигляд:

$$\ddot{y}(t) + r^2 y(t) = M \sin(\omega t). \quad (3.4.4)$$

Власні коливання тіла визначаються однорідним диференціальним рівнянням

$$\ddot{y}(t) + r^2 y(t) = 0, \quad (3.4.5)$$

і описуються рівнянням гармонічних коливань

$$y(t) = A \sin(rt + \varphi).$$

Необхідно знайти частковий розв'язок рівняння (3.4.3). Вигляд часткового розв'язку залежить від того, чи буде число $a + ib = i\omega$, $i = \sqrt{-1}$, коренем характеристичного рівняння для рівняння (3.4.5). Оскільки характеристичне рівняння

$$\lambda^2 + r^2 = 0$$

має корені $\lambda_{1,2} = \pm i r$, то аналіз зводиться до випадку, коли частота впливу збурюючої сили співпадає з частотою власних коливань тіла (*резонанс*) і коли не співпадає (*відсутність резонансу*).

Розглянемо нерезонансний випадок ($\omega \neq r$). Згідно теорії диференціальних рівнянь частковий розв'язок рівняння (3.4.4) необхідно шукати за допомогою пробної функції:

$$y^p(t) = A \cos(\omega t) + B \sin(\omega t), \quad (3.4.6)$$

де A і B – коефіцієнти, які необхідно визначити.

З останнього рівняння отримаємо:

$$y^p(t) = A \cos(\omega t) + B \sin(\omega t),$$

$$\dot{y}^p(t) = -A\omega \sin(\omega t) + B\omega \cos(\omega t),$$

$$\ddot{y}^p(t) = -A\omega^2 \cos(\omega t) - B\omega^2 \sin(\omega t).$$

Підставляючи другу похідну в (3.4.4), отримаємо:

$$A(r^2 - \omega^2) \cos(\omega t) + B(r^2 - \omega^2) \sin(\omega t) = M \sin(\omega t).$$

З початкових умов знайдемо невідомі коефіцієнти:

$$\left. \begin{array}{l} A(r^2 - \omega^2) = 0 \\ B(r^2 - \omega^2) = M \end{array} \right\} \Rightarrow A = 0, B = \frac{M}{r^2 - \omega^2}.$$

Частковий розв'язок:

$$y^p(t) = \frac{M}{r^2 - \omega^2} \sin(\omega t). \quad (3.4.7)$$

Загальний розв'язок приймає вигляд:

$$y^p(t) = \frac{M}{r^2 - \omega^2} \sin(\omega t) + A \sin(r t + \varphi). \quad (3.4.8)$$

Таким чином, коливання, що описуються рівнянням (3.4.8), визначаються лінійною комбінацією гармонічних коливань різної частоти.

Резонансні коливання. Явище резонансу виникає у випадку, коли частота зміни збурюючої сили співпадає з частотою власних коливань тіла ($\omega = r$). Рівняння коливань має вигляд:

$$\ddot{y}(t) + r^2 y(t) = M \sin(rt). \quad (3.4.9)$$

У даному випадку частковий розв'язок буде містити множник t :

$$y^p(t) = t[A \cos(rt) + B \sin(rt)]. \quad (3.4.10)$$

Отримуємо такий результат:

$$y^p(t) = t[A \cos(rt) + B \sin(rt)],$$

$$\dot{y}^p(t) = A \cos(rt) + B \sin(rt) + t[-Ar \sin(rt) + Br \cos(rt)],$$

$$\ddot{y}^p(t) = 2[-Ar \sin(rt) + Br \cos(rt)] + t[-Ar^2 \cos(rt) - Br^2 \sin(rt)]$$

а звідси

$$-2Ar \sin(rt) + 2Br \cos(rt) = M \sin(rt);$$

$$\left. \begin{array}{l} -2Ar = M \\ 2Br = 0 \end{array} \right\} \Rightarrow A = -\frac{M}{2r}, B = 0,$$

а тому неоднорідний розв'язок має вигляд:

$$y^p(t) = -\frac{M}{2r}t \cos(rt). \quad (3.4.11)$$

Частковий розв'язок (3.4.11) представляє собою вимушені коливання з необмеженою (розбіжною) амплітудою. Гармонічні коливання знаходяться між прямими:

$$y(t) = \frac{M}{2r}t \quad \text{і} \quad y(t) = -\frac{M}{2r}t.$$

Загальний розв'язок

$$y(t) = -\frac{M}{2r}t \cos(rt) + A \sin(rt + \varphi)$$

представляє собою лінійну комбінацію коливань з необмеженою амплітудою та гармонічних коливань з постійною амплітудою тієї ж частоти.

Приклад 3.4.4. Розглянемо дискретне рівняння з гармонічним збуренням

$$y(k) = a_0 + a_1 y(k-1) + A \sin(\omega k),$$

де A і ω – відомі значення амплітуди і кругової частоти вхідного сигналу.

Виберемо пробну функцію у вигляді:

$$y_{\text{проб}}^p(k) = c_0 + c_1 k + c_2 \sin(\omega k + \varphi),$$

де φ – зсув фази сигналу при проходженні через лінійну систему.

Скористаємось тригонометричною тотожністю

$$\sin(\alpha + \beta) = \sin(\alpha) \cos(\beta) + \cos(\alpha) \sin(\beta)$$

і підставимо пробну функцію в рівняння процесу:

$$c_0 + c_1 k + c_2 [\sin(\omega k) \cos(\varphi) + \cos(\omega k) \sin(\varphi)] = a_0 + a_1 \{c_0 + c_1(k-1) + c_2 [\sin(\omega(k-1)) \cos(\varphi) + \cos(\omega(k-1)) \sin(\varphi)]\} + A \sin(\omega k).$$

Введемо позначення: $\cos(\varphi) = b_1$, $\sin(\varphi) = b_2$, $\sin(\omega) = b_3$, $\cos(\omega) = b_4$ і в результаті отримаємо тотожності:

$$c_0 = a_0 + a_1 c_0 - a_1 c_1,$$

$$c_2 b_1 = c_2 a_1 b_1 b_4 + c_2 a_1 b_2 b_3 + A,$$

$$c_2 b_2 = -c_2 a_1 b_1 b_3 + c_2 a_1 b_2 b_4.$$

Якщо $a_1 \neq 1$, то $c_1 = 0$ і з першої тотожності знайдемо, що $c_0 = a_0 / (1 - a_1)$. За допомогою другої можна визначити коефіцієнт c_2 :

$$c_2 = \frac{A}{b_1 - a_1 b_1 b_4 - a_1 b_2 b_3}.$$

Оскільки коефіцієнти b_3 і b_4 відомі, а $b_1^2 + b_2^2 = 1$, то з третьої тотожності можна знайти b_1 і b_2 і фазовий кут φ .

Таким чином, при $a_1 \neq 1$ розв'язок має вигляд:

$$y(k) = \frac{a_0}{1 - a_1} + B a_1^k + \frac{A}{b_1 - a_1 b_1 b_4 - a_1 b_2 b_3} \sin(\omega k + \varphi).$$

Приклад 3.4.5. Знайдемо повний розв'язок рівняння третього порядку АРКС(3, 2):

$$y(k) = 0,5 + y(k-1) + 0,25y(k-2) - 0,25y(k-3) + \varepsilon(k) - \\ - 0,125\varepsilon(k-1) + 0,125\varepsilon(k-2).$$

Запишемо характеристичне рівняння

$$\lambda^3 - \lambda^2 - 0,25\lambda + 0,25 = 0$$

і знайдемо його корені: $\lambda_1 = 1$; $\lambda_2 = -0,5$; $\lambda_3 = 0,5$ (процес з одиничним коренем). Однорідний розв'язок має вигляд:

$$y^h(k) = A_1 + A_2(-0,5)^k + A_3(0,5)^k.$$

Для знаходження часткового розв'язку виберемо пробну функцію у вигляді:

$$y_{проб}^p(k) = b_0 + b_1 k + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i)$$

і підставимо її в рівняння процесу, скориставшись загальними позначеннями для коефіцієнтів:

$$b_0 + b_1 k + \alpha_0 \varepsilon(k) + \alpha_1 \varepsilon(k-1) + \alpha_2 \varepsilon(k-2) + \alpha_3 \varepsilon(k-3) + \dots = \\ = a_0 + a_1 [b_0 + b_1(k-1) + \alpha_0 \varepsilon(k-1) +$$

$$\begin{aligned}
& + \alpha_1 \varepsilon(k-2) + \alpha_2 \varepsilon(k-3) + \alpha_3 \varepsilon(k-4) + \dots] + \\
& + a_2 [b_0 + b_1(k-2) + \alpha_0 \varepsilon(k-2) + \alpha_1 \varepsilon(k-3) + \\
& \quad + \alpha_2 \varepsilon(k-4) + \alpha_3 \varepsilon(k-5) + \dots] + a_3 [b_0 + b_1(k-3) + \\
& + \alpha_0 \varepsilon(k-3) + \alpha_1 \varepsilon(k-4) + \alpha_2 \varepsilon(k-5) + \alpha_3 \varepsilon(k-6) + \dots] + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \\
& \quad + \beta_2 \varepsilon(k-2).
\end{aligned}$$

Прирівняємо константи в лівій і правій частині, а також коефіцієнти при однакових змінних:

$$b_0 = a_0 + a_1 b_0 - a_1 b_1 + a_2 b_0 - 2a_2 b_1 + a_3 b_0 - 3a_3 b_1;$$

$$\text{при } k: \quad b_1 = a_1 b_1 + a_2 b_1 + a_3 b_1;$$

$$\text{при } \varepsilon(k): \quad \alpha_0 = 1;$$

$$\text{при } \varepsilon(k-1): \quad \alpha_1 = a_1 \alpha_0 + \beta_1;$$

$$\text{при } \varepsilon(k-2): \quad \alpha_2 = a_1 \alpha_1 + a_2 \alpha_0;$$

$$\text{при } \varepsilon(k-3): \quad \alpha_3 = a_1 \alpha_2 + a_2 \alpha_1 + a_3 \alpha_0.$$

Таким чином, $b_1 \approx 0,67$, а коефіцієнт b_0 необхідно визначити за допомогою початкових умов. Повний розв'язок:

$$y(k) = A + A_2 (-0,5)^k + A_3 (0,5)^k + 0,67k + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i),$$

$$\text{де } A = A_1 + b_0, \quad \alpha_i = \alpha_{i-1} + 0,25\alpha_{i-2} - 0,25\alpha_{i-3}.$$

Запитання і вправи до третього розділу

1. Поясніть, для чого необхідно знаходити розв'язки різницевого рівнянь? Чи можна скористатись моделлю АРКС(p, q) для асимптотичного аналізу поведінки процесу?
2. Який процес називають процесом з одиничними коренями? Які синоніми цієї назви? Дайте означення тренду? Що означає не-стаціонарність процесу?
3. На яку частину розв'язку впливають початкові умови? В чому проявляється вплив початкових умов? Яким чином можна задати початкові умови?
4. Що характеризують перші і другі різниці? Запишіть рівняння АР(2) через перші різниці; для розв'язання якої задачі можна скористатись цією моделлю?
5. Який процес називають метастійним? Чи зустрічаються такі процеси на практиці?
6. Яким чином можна звести однорідний розв'язок до нуля? Що дасть на практиці зведення однорідного розв'язку до нуля?
7. У якому випадку неоднорідний розв'язок дорівнює нулю?
8. Знайдіть однорідні розв'язки рівнянь:

$$y(k) = 0,35 - 0,75y(k-1) + 0,25y(k-2) + \varepsilon(k);$$

$$y(k) = a_0 + 1,5y(k-1) - 0,5y(k-2) + \varepsilon(k);$$

$$y(k) = a_0 + y(k-2) + \varepsilon(k);$$

$$y(k) = a_0 + 2y(k-1) - y(k-2) + \varepsilon(k);$$

$$y(k) = a_0 + y(k-1) + 0,25y(k-2) - 0,25y(k-3) + \varepsilon(k).$$

9. Запишіть загальний вигляд однорідного рівняння для випадку, коли характеристичне рівняння містить трикратний корінь? Чи буде такий розв'язок стійким (збіжним), дайте пояснення?

10. У яких випадках однорідний розв'язок містить гармонічну складову? Чим визначається амплітуда, частота і фаза коливань в даному випадку?
11. Що таке кругова частота? Чим зручне використання кругової частоти у порівнянні з лінійною при аналізі поведінки процесів?
12. Сформулюйте умови збіжності однорідного розв'язку через значення коренів характеристичного рівняння; чи є ці умови достатніми?
13. Сформулюйте необхідні і достатні умови збіжності однорідного розв'язку через коефіцієнти рівняння авторегресії; яким чином можна отримати ці умови?
14. Чи досить знати необхідні і достатні умови збіжності однорідного розв'язку рівняння AP(2) для того щоб узагальнити ці умови на рівняння вищих порядків?
15. Чи відомі вам інші методи аналізу стійкості розв'язків крім методів, які ґрунтуються на коренях характеристичного рівняння та коефіцієнтах математичної моделі?
16. Які переваги має графічний аналіз збіжності розв'язків, чи можна його застосувати на практиці?
17. У чому полягає подібність та відмінності розв'язків однорідних різницевих та диференціальних рівнянь? Покажіть це на прикладі рівняння другого порядку.
18. Запишіть у загальному вигляді розв'язок наведеного нижче рівняння:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + a_3 y(k-3) + b_1 \varepsilon(k-1) + \varepsilon(k).$$

- a. Нехай характеристичне рівняння має, у даному випадку, один дійсний
- b. і два кратних корені. Виконайте аналіз збіжності отриманого розв'язку.
19. Нарисуйте графік однорідного розв'язку для випадку двократного кореня характеристичного рівняння?
20. Запишіть у загальному вигляді розв'язок однорідного рівняння довільного порядку з m -кратним коренем характеристичного рівняння; який вигляд матиме графік цього розв'язку?
21. Запишіть у загальному вигляді розв'язок однорідного рівняння, якщо його характеристичне має два дійсних різних корені і комплексно спряжену пару коренів.

Р о з д і л 4

ПРОГНОЗУВАННЯ ДИНАМІКИ ПРОЦЕСІВ ЗА ДОПОМОГОЮ РІЗНИЦЕВИХ РІВНЯНЬ

4.1 Для чого необхідні прогнози?

Прогнозування розвитку подій у процесі приватної та ділової активності – це суттєва складова повсякденного життя, яка стосується кожного із нас. Навряд чи знайдеться такий напрям людської діяльності, де не потрібно робити прогнози розвитку подій, ситуацій, показників і т. ін. Для того щоб продемонструвати важливість коректного розв’язання задачі прогнозування розглянемо ієрархію процесу прийняття рішень на рівні виробничого, торговельного або іншого підприємства. В управлінському процесі, що стосується діяльності підприємств різних видів діяльності, можна виділити наведені нижче *рівні ієрархії* планування, які ґрунтуються на прогнозуванні розвитку подій.

1. Стратегічне бізнес-планування. На цьому рівні визначають стратегічні цілі функціонування організації (фірми, підприємства) на довгому часовому інтервалі: від 3-х до 20-и років. План діяльності складається *щомісячний, щоквартальний* або *щорічний*. Основними цілями є, як правило, такі: отримання визначеного рівня прибутку; досягнення необхідного рівня якості продукції; визначення своєї участі (частки) на ринку конкретних товарів; задоволення вимог покупців; досягнення визначеного рівня автоматизації виробництва; досягнення необхідного рівня відносин між роботодавцем і найманими робітниками.

2. Середньострокове планування управлінської діяльності. На даному рівні визначаються управлінські дії, необхідні для досягнення стратегічних цілей на часовому інтервалі від 3-х місяців до 3-х років. План складається *щотижневий* або *щомісячний*. Планування стосується необхідних об'ємів фінансів; маркетингу; об'ємів продаж; перевезення (розповсюдження) продукції; кількості, місцезнаходження та характеру необхідного устаткування; агрегування (концентрації) виробництва – тобто того, яка продукція буде вироблятися на конкретних відділеннях підприємства; формування та обґрунтування основних календарних планів.
3. Оперативне планування та управління. Головна мета цього рівня управління полягає в тому, щоб ефективно розподілити ресурси для досягнення наведених вище головних цілей. Оперативне планування та управління реалізується на часових інтервалах від 1 місяця до 2-х років; план складається по *тижнях* або по *місяцях*. Воно стосується управління розподілом матеріальних ресурсів, виробництвом та розподілом продукції, регулювання необхідних виробничих потужностей (приміщення, устаткування, робоча сила, витрати енергії та сировини).
4. Щоденне управління (обробка транзакцій). На цьому рівні реалізуються управлінські операції на часових інтервалах від 1-го дня до 1-го місяця. При цьому контролюються, як правило, такі операції: закупівля необхідних об'ємів та номенклатури матеріалів; формування замовлень; виставлення та приймання рахунків; оперативний контроль якості продукції на всьому протязі її виробництва; управління транспортним господарством підприємства.

Докладний аналіз кожного рівня прийняття рішень на підприємстві свідчить, що управлінський персонал не може ефективно розв'язувати задачі планування та контролю без належного високоякісного *прогнозування* показників розвитку підприємства та інших процесів, пов'язаних з діловою активністю. При цьому горизонт прогнозу (проміжок часу, на який робиться прогноз) залежить від рівня ієрархії і може бути досить довгим (до 20-и років) або коротким (один день). Таким чином, ієрархія процесу прийняття рішень потребує створення та використання ієрархічної системи прогнозування, яка забезпечить обчислення високоякісних оцінок коротко-, середньо- та довгострокових прогнозів.

Необхідно також зазначити, що рух вниз по управлінській ієрархії супроводжується суттєвими змінами об'єму, характеру інформації та її докладності. Так, на вершині піраміди необхідно приймати *неструктуровані* та *погано структуровані* рішення при неповній інформації (або її відсутності). Чим нижчим є рівень управління, тим більше програмованими

(структурованими, упорядкованими, прозорими) є рішення, для яких існує інформація у все повнішому об'ємі.

За наявності на підприємстві комп'ютеризованої системи підтримки прийняття рішень з'являється можливість автоматизації процесів управління та контролю. Наприклад, автоматизуються операції обліку на складах, процеси, пов'язані з контролем якості продукції, аналіз номенклатури обладнання, формування технологічних карт, необхідних для реалізації технологічних процесів, а також процесів накопичення і підтримки необхідних запасів сировини і т. ін. В цілому, рухаючись вниз по ієрархії прийняття рішень можна сформулювати такі висновки:

- рішення та прогнози стають менш стратегічними, конкретнішими і точнішими;
- зменшується участь у прийнятті рішень управлінського персоналу верхнього рівня;
- скорочується горизонт прогнозування від років до місяців, тижнів та днів;
- рішення та прогнози стають більш програмованими (стандартизованими, структурованими, чітко визначеними);
- значно зростає докладність опису ситуацій та задач, стосовно яких приймаються рішення;
- зростає ступінь автоматизації процесу прийняття рішень;
- для прийняття рішень, а особливо тих, що повторюються, доцільно застосовувати відповідні комп'ютерні системи підтримки прийняття рішень при прогнозуванні та плануванні діяльності галузі, підприємства або індивідуума.

Із сказаного випливає, що прогноуюча система також повинна бути побудована за ієрархічною архітектурою (структурою), що є цілком природним для систем подібного типу. Такий підхід дає можливість структурувати (упорядкувати) задачі прогнозування, створити множину необхідних для прогнозування спеціалізованих моделей, ситуацій та методів прийняття рішень на кожному рівні, а також наблизити комп'ютеризований процес прийняття рішень до того аналітичного процесу, який є звичним для нас у повсякденному житті. Все це сприяє прискоренню процесу аналізу проблеми, прийняття та реалізації рішення.

Крім бізнес-процесів досить часто ставиться задача оцінювання прогнозу при прийнятті приватних рішень. Наприклад, необхідно спрогнозувати сімейний бюджет на 4-5 років із врахуванням необхідності оплати навчання у навчальному закладі та виплат за житловими кредитами, або спрогнозувати витрати часу на виконання різних робіт в залежності від їх складності. Прийняття приватних рішень також включає необхідність довгострокового прогнозування (оцінювання) можливостей стосовно рівня заробітку, зростання цін на нерухомість і товари повсякденного вжитку,

вартості споживчого кошика, витрат на транспортні послуги та багатьох інших процесів.

В наступному параграфі розглянемо можливості прогнозування розвитку процесів довільної природи, представлених часовими рядами статистичних або експериментальних даних.

4.2 Які складові процесу можна прогнозувати?

У загальному випадку прогноз може бути представлений одним (точковим) значенням змінної, інтервалом, у який попадає значення випадкової змінної, а також ймовірністю прийняття змінною (чи подією) деякого значення у конкретному інтервалі допустимих значень. Якщо для опису процесу застосовують лінгвістичні змінні, то прогнозом буде нечітке значення, але його також можна перетворити у чітке число (дефазифікувати).

Можна по-різному ставити задачу прогнозування в залежності від рівня прийняття рішення та конкретної поставленої задачі управління або контролю. Прогнозування може стосуватись таких складових процесу:

- детермінованого тренду (поточного середнього значення, яке можна описати деякою вибраною функцією), як індикатора довгострокових змін процесу;
- випадкового (нерегулярного) тренду, як показника коротко- та середньострокових змін;
- короткострокових змін, тобто прогнозування коливань (відхилень), що накладаються на тренд;
- сезонних ефектів;
- приростів (швидкості) зміни процесу, які визначаються першими різницями;
- дисперсії або стандартного відхилення, як міри розсіювання процесу (наприклад, волатильність, яку часто використовують за міру ризику у інвестуванні або міру якості продукції на виробництві);
- якісних змінних процесів довільної природи (за допомогою нечіткої логіки, мереж Байєса і т. ін.);
- комбінацій вказаних елементів процесів.

Відповідно до того, які складові процесу необхідно прогнозувати, ставиться задача побудови математичної, ймовірнісної (статистичної) або логічної моделі, що має меті забезпечити високу якість оцінок прогнозу на заданий часовий горизонт. Розглянемо деякі можливості створення математичного опису складових процесів різної природи.

Детермінований тренд

Якщо описати детермінований тренд за допомогою полінома від часу довільного порядку, p :

$$y(k) = a_0 + a_1 \cdot k + a_2 \cdot k^2 + \dots + a_p \cdot k^p + \varepsilon(k), \quad E[\varepsilon(k)] = 0,$$

то визначення прогнозу тренду зводиться до підстановки в це рівняння бажаного значення часу k і застосування до нього оператора безумовного математичного сподівання (для того щоб позбутися впливу випадкового процесу $\varepsilon(k)$, що, звичайно, призводить до внесення додаткової похибки). Оцінювання майбутніх значень тренду вважається довгостроковим прогнозом, оскільки детермінований тренд вказує на довгострокові зміни процесів. Обмеження центрованості на випадковий процес, $E[\varepsilon(k)] = 0$, необхідне для коректного застосування методів оцінювання параметрів моделей, а також для виконання подальшого аналізу якості оцінок прогнозів.

Стохастичний тренд

Для опису і прогнозування стохастичного тренду можна скористатись, наприклад, відомим рівнянням випадкового кроку із перетином (константою):

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

розв'язок якого має вигляд:

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

Сума $\sum_{i=1}^k \varepsilon(i)$ у правій частині останнього рівняння описує випадкову складову тренду. Цю складову називають ще *нерегулярною*. Ускладнені моделі стохастичних трендів включають комбінації декількох випадкових процесів, що дає можливість описувати складні випадки швидких трудно прогнозованих змін досліджуваних процесів.

Прогнозування коливань, що накладаються на тренд

Колівання, що накладаються на тренд, або короткострокові зміни можна описати рівняннями авторегресії з ковзним середнім:

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

Техніку обчислення оцінок прогнозу, $\hat{y}(k+s)$, де s – кількість кроків прогнозування, за допомогою умовного та безумовного математичного сподівання розглянемо нижче у цьому розділі.

При переході до перших та різних вищих порядків з процесу автоматично вилучається тренд відповідного порядку. Наприклад, якщо процес містить лінійний тренд, то перші різниці вилучають його і після переходу до різниць ми маємо справу з коливаннями, що накладаються на тренд.

Можливість прогнозування *сезонних ефектів* досягається за рахунок введення у модель процесу відповідних значень основної змінної із затримками (лагами), що відповідають періодичності ефекту. Як буде показано нижче, сезонний ефект можна врахувати як за допомогою основної змінної, так і за допомогою процесу ковзного середнього.

Прогнозування дисперсії

Якщо дисперсія процесу змінюється із плином часу, то для її опису можна вибрати рівняння для формування відповідної функції прогнозування та обчислення у подальшому оцінки прогнозованого значення. Процеси із змінною дисперсією отримали назву гетероскедастичних (від англійського *heteroscedastic*), вони будуть розглянуті нижче. Методика побудови моделей гетероскедастичних процесів передбачає виконання таких кроків:

- математичний опис основної змінної рівнянням авторегресії невисокого порядку (наприклад, першого або другого) з метою вилучення з процесу детермінованої складової;
- математичний опис умовної дисперсії, як динамічної змінної, за допомогою рівняння прийнятної (за якістю прогнозу) структури.

Докладний опис даної методики і приклади її застосування наведено нижче у іншому розділі.

4.3 Умовні та безумовні статистичні характеристики

При виконанні статистичного аналізу випадкових процесів використовують *умовні та безумовні статистичні характеристики*. Зокрема, для знаходження короткострокових та довгострокових прогнозів розвитку процесів застосовують умовне, E_k , та безумовне, E , математичне сподівання, відповідно. Розглянемо у чому полягає різниця визначення цих характеристик.

Безумовні статистичні характеристики – це константи, які оцінюють і аналізують на довільних (вибраних дослідником) часових інтервалах, не накладаючи умов на змінні, функції та об'єми інформації, необхідної для обчислення цих характеристик. Тобто інформація для обчислення дисперсії вважається відомою на всьому часовому інтервалі, що розглядається у процесі аналізу.

Так, безумовне математичне сподівання використовують для знаходження довгострокових прогнозів або умов економічної рівноваги, його позначають символом E . Такі характеристики, як безумовне вибіркоче середнє і дисперсія обчислюють за відомими формулами:

$$\bar{x} = \frac{1}{N} \sum_{k=1}^N y(k), \quad \text{Var}(x) = \frac{1}{N-1} \sum_{k=1}^N [x(k) - \bar{x}]^2.$$

У виразі для дисперсії немає необхідності зменшувати значення N на одиницю, якщо потужність ряду досить велика, наприклад, декілька десятків.

Умовні статистичні характеристики в аналізі динаміки процесів, представлених часовими рядами, – це *функції часу*, які визначають на кожний конкретний момент часу k . При цьому для їх обчислення необхідно, щоб на вибраній (заданий) момент була наявна інформація щодо значень змінних та функцій, яка необхідна для виконання відповідних обчислень.

Умовне математичне сподівання застосовують для обчислення короткострокових та середньострокових прогнозів. Умовну дисперсію процесу і стандартне відхилення також часто використовують як міру ризику. Наприклад, в аналізі фінансових процесів, при формуванні правил торгівлі на біржі, аналізі банківських і економічних ризиків. У технічних системах дисперсія також відіграє значну роль при визначенні ступеня відхилення вузла (наприклад, підшипника, корпусного елемента, крила) від нормального (заданого) стану. Тому вміння правильно аналітично описати дисперсію дає можливість з високою точністю описати і спрогнозувати значення відхилень від норми.

Так, умовне математичне сподівання стохастичного процесу AP(1) визначається за виразом:

$$\begin{aligned} E_k[y(k+1) | y(k), y(k-1), \dots, y(0), \varepsilon(k), \varepsilon(k-1), \dots, \varepsilon(0)] = \\ = E_k[a_0 + a_1 y(k) + \varepsilon(k+1)] = a_0 + a_1 y(k), \end{aligned}$$

за умови, що $E_k[\varepsilon(k+l)] = 0, l \geq 1$. Умовне вибіркоче середнє та умовну вибіркочу дисперсію можна наближено обчислити за виразами:

$$\bar{x}(k) = \frac{1}{k} \sum_{i=1}^k x(i); \quad \text{Var}_k(x) = \frac{1}{k-1} \sum_{i=1}^k [x(i) - \bar{x}(k)]^2, \quad k = 2, \dots, N.$$

В результаті застосування цих виразів отримаємо ряд значень умовного середнього та умовної дисперсії, тобто ще дві характеристики процесу, якими можна скористатись при побудові математичних і статистичних моделей.

Наближені значення умовних вибірових статистичних характеристик можна визначити, також, за аналогією з обчисленням ковзного середнього. При такому підході необхідно вибрати ширину ковзного вікна і обчислити значення статистичного параметра рухаючись крок за кроком від початку до кінця часового ряду. Наприклад, якщо вибрати вікно шириною п'ять значень ряду, то умовна дисперсія буде обчислюватись за виразом:

$$\text{Var}_k(x) = \frac{1}{4} \sum_{i=k-2}^{k+2} [y(i) - \bar{y}]^2, \quad k = 3, \dots, N-2,$$

або у загальному вигляді так:

$$\text{Var}_k(x) = \frac{1}{w-1} \sum_{i=k-(w-1)/2}^{k+(w-1)/2} [y(i) - \bar{y}]^2, \quad k = \frac{d-1}{2}, \dots, N - \frac{d-1}{2},$$

де w – ширина ковзного вікна. Ширина вікна залежить від того, наскільки швидко змінюється у часі дисперсія. Якщо вона має високу динаміку, то для розміру вікна вибирають значення 5-9. Незважаючи на наближеність таких розрахунків, практика моделювання свідчить, що обчислені значення виявляються, як правило, досить корисними (інформативними) при побудові моделей умовної дисперсії.

У складніших випадках обчислюють функцію умовної дисперсії, яка точніше відображає характер її зміни у часі. Наприклад, можна розглянути таку модель процесу:

$$y(k) = \sqrt{f[x(k)]} \varepsilon(k), \quad (4.3.1)$$

$$\varepsilon(k) = \beta \varepsilon(k-1) + v(k), \quad (4.3.2)$$

де $y(k) \in \mathfrak{R}$, $k = 1, 2, \dots, N$; $\{v(k)\}$ – множина незалежних однаково розподілених (НОР) величин, що мають нормальний розподіл з параметрами $N(0, 1)$; $\beta \in \Theta = (-1; 1)$; $f[x(k)] \in C^p[0, 1]$. Припущення про те, що $v(k)$ – гаусів процес, зроблено для зручності викладок. Змінна $x(k) \in [0, 1]$ повинна бути упорядкована за значеннями, тобто:

$$x(1) \leq x(2) \leq \dots \leq x(N),$$

де $x(k) = \frac{k}{N}$, $k = 1, 2, \dots, N$. Функцію $f[x(k)]$ називають функцією дисперсії,

хоча вона не повністю описує дисперсійно-коваріаційну структуру процесу (4.3.1)–(4.3.2). Якщо $f[x(k)]$ має p неперервних похідних, то безумовну дисперсію процесу $y(k)$ можна визначити таким чином:

$$\begin{aligned}
\text{Var}[y(k)] &= f[x(k)] E[\varepsilon^2(k)] = f[x(k)] E\{\beta \varepsilon(k-1) + v(k)\}^2 = \\
&= \{\beta^2 E[\varepsilon^2(k-1)] + E[v^2(k)]\} f[x(k)] = \\
&= \beta^2 f[x(k)] E[\varepsilon^2(k-1)] + f[x(k)] = \\
&= \beta^2 \text{var}[y(k)] + f[x(k)].
\end{aligned}$$

Звідси отримаємо:

$$\text{Var}[y(k)] = \frac{f[x(k)]}{1 - \beta^2}.$$

Таким чином, процес (4.3.1)–(4.3.2) є умовно і безумовно гетероскедастичним. Задача полягає у тому, щоб оцінити функцію $f[x(k)]$. Така задача потребує окремого розгляду.

4.4 Прогнозування без знаходження розв'язку рівняння

Обчислення оцінки прогнозу без розв'язку рівняння

Структура різницевого рівняння така, що воно дозволяє виконувати прогнозування на один крок (один період дискретизації вимірів) без додаткових перетворень. Тобто у праву частину необхідно підставити минулі значення змінних і обчислити оцінку прогнозу головної змінної у лівій частині. Але для того щоб знайти оцінку прогнозу на більшу кількість кроків, необхідно застосувати деякі попередні перетворення різницевих рівнянь. Розглянемо деякі можливі підходи до формування функцій прогнозування та обчислення оцінок прогнозованих значень.

Як приклад, розглянемо рівняння AP(1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad E[\varepsilon(k)] = 0. \quad (4.4.1)$$

Збільшимо незалежну змінну, час, на одиницю і запишемо рівняння знову:

$$y(k+1) = a_0 + a_1 y(k) + \varepsilon(k+1). \quad (4.4.2)$$

Якщо коефіцієнти a_0, a_1 відомі, то можна знайти умовне математичне сподівання на основі відомої інформації до моменту k включно:

$$\begin{aligned}
E_k[y(k+1)] &= E_k[y(k+1) | y(k), y(k-1), \dots, \varepsilon(k), \varepsilon(k-1), \dots] = \\
&= a_0 + a_1 E_k[y(k)] = a_0 + a_1 y(k),
\end{aligned} \quad (4.4.3)$$

оскільки $y(k)$ в момент k є відомою константою.

За аналогією запишемо рівняння (4.4.1) для моменту $k + 2$:

$$y(k + 2) = a_0 + a_1 y(k + 1) + \varepsilon(k + 2), \quad (4.4.4)$$

і знайдемо умовне математичне сподівання:

$$\begin{aligned} E_k[y(k + 2)] &= a_0 + a_1 E_k[y(k + 1)] = a_0 + a_1 E_k[a_0 + a_1 y(k)] = \\ &= a_0 + a_0 a_1 + a_1^2 y(k). \end{aligned}$$

Для наступного моменту часу маємо:

$$E_k[y(k + 3)] = a_0 + a_0 a_1 + a_0 a_1^2 + a_1^3 y(k).$$

Таким чином, для загального випадку прогнозування на s кроків можна записати:

$$\hat{y}(k + s) = E_s[y(k + s)] = a_0 \left(\sum_{i=0}^{s-1} a_1^i \right) + a_1^s y(k) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k). \quad (4.4.5)$$

Отримане рівняння називають функцією прогнозування на довільне число кроків s . Прогноз представляє собою збіжний процес, якщо $|a_1| < 1$, тобто

$$\lim_{s \rightarrow \infty} E_k[y(k + s)] = \frac{a_0}{1 - a_1}, \quad |a_1| < 1, \quad (4.4.6)$$

де a_1 – знаменник геометричної прогресії. Вираз (4.4.6) свідчить про те, що для будь-якого стаціонарного процесу АР чи АРКС оцінка умовного прогнозу асимптотично, при $s \rightarrow \infty$, збігається до безумовного середнього.

Похибка оцінки прогнозу на довільне число кроків

Знайдемо похибку оцінок прогнозів за умови, що $E[\varepsilon(k)] = 0$:

$$f_k(s) = y(k + s) - E_k[y(k + s)]. \quad (4.4.7)$$

Похибка прогнозу на один крок:

$$f_k(1) = y(k + 1) - E_k[y(k + 1)] = a_0 + a_1 y(k) + \varepsilon(k + 1) - a_0 - a_1 y(k) = \varepsilon(k + 1)$$

Похибку прогнозу на два кроки можна визначити так:

$$\begin{aligned} f_k(2) &= y(k+2) - E_k[y(k+2)] = \\ &= a_0 + a_1[a_0 + a_1 y(k) + \varepsilon(k+1)] + \varepsilon(k+2) - E_k[y(k+2)] = \\ &= a_0 + a_0 a_1 + a_1^2 y(k) + a_1 \varepsilon(k+1) + \varepsilon(k+2) - a_0 - a_0 a_1 - a_1^2 y(k) = \\ &= \varepsilon(k+2) + a_1 \varepsilon(k+1). \end{aligned}$$

Таким чином, можемо записати вираз для похибки оцінки прогнозування на довільну кількість кроків:

$$f_k(s) = \varepsilon(k+s) + a_1 \varepsilon(k+s-1) + a_1^2 \varepsilon(k+s-2) + \dots + a_1^{s-1} \varepsilon(k+1). \quad (4.4.8)$$

Враховуючи те, що $E[f_k(s)] = 0$, оцінка прогнозу, яка обчислюється за виразом (4.4.5), є незміщеною. Дисперсія похибки прогнозування:

$$\text{Var}[f_k(s)] = \sigma^2 [1 + a_1^2 + a_1^4 + a_1^6 + \dots + a_1^{2(s-1)}],$$

тобто дисперсія є функцією s . Асимптотичне значення дисперсії похибки прогнозу для стаціонарного процесу

$$\lim_{s \rightarrow \infty} \text{Var}[f_k(s)] = \frac{\sigma^2}{1 - a_1^2}, \quad (4.4.9)$$

де a_1^2 – знаменник геометричної прогресії.

Узагальнення функції прогнозування (на процес АРКС(p,q))

Знайдемо спочатку функцію прогнозування для процесу АРКС(2,1), який описується рівнянням:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \varepsilon(k) + \beta_1 \varepsilon(k-1). \quad (4.4.10)$$

Для наступного моменту часу можна записати:

$$y(k+1) = a_0 + a_1 y(k) + a_2 y(k-1) + \varepsilon(k+1) + \beta_1 \varepsilon(k).$$

Умовне математичне сподівання для $y(k+1)$ має вигляд:

$$E_k[y(k+1)] = a_0 + a_1 y(k) + a_2 y(k-1) + \beta_1 \varepsilon(k),$$

де $\varepsilon(k)$ розглядається як відома величина на момент k включно. При цьому

$$E_k[\varepsilon(k+j)] = 0, \quad \forall j > 0.$$

Для моменту часу $k+2$ маємо

$$y(k+2) = a_0 + a_1 y(k+1) + a_2 y(k) + \varepsilon(k+2) + \beta_1 \varepsilon(k+1)$$

і умовне математичне сподівання

$$\begin{aligned} E_k[y(k+2)] &= a_0 + a_1 E_k[y(k+1)] + a_2 E_k[y(k)] = \\ &= a_0 + a_1 [a_0 + a_1 y(k) + a_2 y(k-1) + \beta_1 \varepsilon(k)] + a_2 y(k) = \\ &= a_0 + a_0 a_1 + a_1^2 y(k) + a_1 a_2 y(k-1) + a_1 \beta_1 \varepsilon(k) + a_2 y(k) = \\ &= a_0 (1 + a_1) + (a_1^2 + a_2) y(k) + a_1 a_2 y(k-1) + a_1 \beta_1 \varepsilon(k). \end{aligned}$$

Можна знайти також умовне математичне сподівання для оцінки прогнозу на три кроки:

$$\begin{aligned} E_k[y(k+3)] &= a_0 + a_1 E_k[y(k+2)] + a_2 E_k[y(k+1)] = \\ &= a_0 (1 + a_1 + a_1^2 + a_2) + (a_1^3 + 2a_1 a_2) y(k) + (a_1^2 a_2 + a_2^2) y(k-1) + \\ &\quad + \beta_1 (a_1^2 + a_2) \varepsilon(k). \end{aligned}$$

З отриманих виразів для умовного математичного сподівання видно, що рекурсивна формула для обчислення оцінки прогнозу на довільну кількість кроків s може бути записана у вигляді:

$$E_k[y(k+s)] = a_0 + a_1 E_k[y(k+s-1)] + a_2 E_k[y(k+s-2)]. \quad (4.4.11)$$

Якщо корені характеристичного рівняння, записаного для (4.4.11), знаходяться всередині одиничного кола, то оцінка прогнозу асимптотично збігається до безумовного середнього значення:

$$\lim_{s \rightarrow \infty} E[y(k+s)] = \frac{a_0}{1 - a_1 - a_2},$$

а для процесу АРКС(p, q) довільного порядку оцінку умовного прогнозу можна записати так:

$$E_k[y(k+s)] = a_0 + \sum_{i=1}^p a_i E_k[y(k+s-i)].$$

4.5 Побудова функції прогнозування на основі розв'язку різницевого рівняння

Розглянемо як приклад рівняння АРКС(1,1):

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-1), \quad |a_1| < 1, \quad (4.5.1)$$

де $\varepsilon(k)$ – білий шум з нульовим середнім; $y(0) = y_0$ – відома початкова умова. Для однорідного рівняння $y(k) - a_1 y(k-1) = 0$ розв'язком є $A a_1^k$, де A – довільна константа.

Частковий розв'язок можна знайти за допомогою лагового оператора L у вигляді:

$$y(k) = \frac{a_0}{1-a_1} + \frac{\varepsilon(k)}{1-a_1 L} + \frac{\beta_1 \varepsilon(k-1)}{1-a_1 L}. \quad (4.5.2)$$

Використовуючи властивості лагового оператора, запишемо загальний розв'язок так:

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i-1) + A a_1^k. \quad (4.5.3)$$

Для того щоб знайти значення довільної константи скористаємось початковою умовою:

$$k=0: \quad y_0 = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(-i-1) + A.$$

Запишемо розв'язок із врахуванням отриманого значення довільної константи:

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i-1) + \left[y_0 - \frac{a_0}{1-a_1} - \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) + \beta_1 \sum_{i=0}^{\infty} a_1^i \varepsilon(-i-1) \right] a_1^k.$$

Оскільки

$$\sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) = \varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots + a_1^k \varepsilon(0) + a_1^{k+1} \varepsilon(-1) + \dots,$$

$$- a_1^k \sum_{i=0}^{\infty} a_1^i \varepsilon(-i) = -a_1^k \varepsilon(0) - a_1^{k+1} \varepsilon(-1) - a_1^{k+2} \varepsilon(-2) - \dots,$$

то

$$y(k) = \frac{a_0}{1-a_1} + \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{k-1} a_1^i \varepsilon(k-i-1) + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k. \quad (4.5.4)$$

Знайдемо рівняння для оцінювання прогнозу із врахуванням того, що на момент $k=0$ відоме значення збурення $E_0[\varepsilon(0)] = \varepsilon_0$, тобто необхідно знайти $E_0[y(k)]$. Таким чином, функція прогнозу приймає такий вигляд:

$$E_0[y(0+k) | y(0), \varepsilon(0)] = \frac{a_0}{1-a_1} + E_0 \left[\sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) + \beta_1 \sum_{i=0}^{k-1} a_1^i \varepsilon(k-1-i) \right] + E_0 \left(\left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k \right).$$

Враховуючи, що

$$\sum_{i=0}^{k-1} a_1^i \varepsilon(k-i) = \varepsilon(k) + a_1 \varepsilon(k-1) + a_1^2 \varepsilon(k-2) + \dots + a_1^{k-1} \varepsilon(-1),$$

і

$$\beta_1 \sum_{i=0}^{k-1} a_1^i \varepsilon(k-1-i) = \beta_1 \varepsilon(k-1) + \beta_1 a_1 \varepsilon(k-2) + \beta_1 a_1^2 \varepsilon(k-3) + \dots + \beta_1 a_1^{k-1} \varepsilon(0),$$

отримаємо:

$$\hat{y}(0+k) = E_0[y(k)] = \frac{a_0}{1-a_1} + \beta_1 a_1^{k-1} \varepsilon_0 + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^k. \quad (4.5.5)$$

Рівняння (4.5.5) можна розглядати як функцію прогнозування на k кроків наперед на основі інформації, наявної на момент $k=0$.

Знайдемо функцію прогнозування на s кроків вперед на основі інформації, наявної на момент k . Спочатку зробимо заміну індексів у рівнянні (4.5.5):

$$\hat{y}(0+s) = E_0[y(s)] = \frac{a_0}{1-a_1} + \beta_1 a_1^{s-1} \varepsilon_0 + \left[y_0 - \frac{a_0}{1-a_1} \right] a_1^s =$$

$$= \left(\frac{a_0}{1 - a_1} \right) (1 - a_1^s) + \beta_1 a_1^{s-1} \varepsilon_0 + y_0 a_1^s.$$

Тепер виконаємо оновлення часового індексу для змінних y і ε на k одиниць вперед:

$$\hat{y}(k + s) = E_k[y(k + s)] = \left(\frac{a_0}{1 - a_1} \right) (1 - a_1^s) + \beta_1 a_1^{s-1} \varepsilon(k) + y(k) a_1^s. \quad (4.5.6)$$

Отримане рівняння представляє собою функцію прогнозування на основі відомої інформації про процес на момент k , включно. Використовуючи наведені вище викладки, можна записати функції прогнозування для довільної кількості кроків таким чином:

$$s = 1: \quad E_k[y(k + 1)] = a_0 + \beta_1 \varepsilon(k) + a_1 y(k);$$

$$s = 2: \quad E_k[y(k + 2)] = \left(\frac{a_0}{1 - a_1} \right) (1 - a_1^2) + \beta_1 a_1 \varepsilon(k) + a_1^2 y(k);$$

$$s = 3: \quad E_k[y(k + 3)] = \left(\frac{a_0}{1 - a_1} \right) (1 - a_1^3) + \beta_1 a_1^2 \varepsilon(k) + a_1^3 y(k);$$

⋮
⋮
⋮

$$\text{При цьому} \quad \lim_{s \rightarrow \infty} E_k[y(k + s)] = \frac{a_0}{1 - a_1}.$$

Можна досить просто перейти від моделі АРКС(1,1) до моделі АР(1), якщо покласти $\beta_1 = 0$. Для АР(1) отримаємо функцію прогнозування у вигляді:

$$E_k[y(k + s)] = \left(\frac{a_0}{1 - a_1} \right) (1 - a_1^s) + a_1^s y(k). \quad (4.5.7)$$

По аналогії можна знайти функції прогнозування для моделей іншої структури, табулювати отримані функції і використовувати їх при прогнозуванні процесів різної природи.

Примітка: Експериментально встановлено, що модель, яка може бути використана для прогнозування, повинна бути отримана (оцінена) на основі не менше ніж 50 спостережень (вимірів) основної змінної.

Якщо модель побудована на основі вибірки даних довжиною N , то для рівняння АРКС(2,1),

$$y(k) = \hat{a}_0 + \hat{a}_1 y(k-1) + \hat{a}_2 y(k-2) + \hat{\varepsilon}(k) + \hat{\beta}_1 \hat{\varepsilon}(k-1),$$

функцію прогнозу можна записати так:

$$E_N[y(N+1)] = \hat{a}_0 + \hat{a}_1 E_N[y(N)] + \hat{a}_2 E_N[y(N-1)] + \hat{\beta}_1 \hat{\varepsilon}(k);$$

$$E_N[y(N+2)] = \hat{a}_0 + \hat{a}_1 E_N[y(N+1)] + \hat{a}_2 E_N[y(N)];$$

.

.

$$E_N[y(N+s)] = \hat{a}_0 + \hat{a}_1 E_N[y(N+s-1)] + \hat{a}_2 E_N[y(N+s-2)],$$

$$s \geq 2.$$

Приклад 4.1. Побудувати функцію прогнозування для рівняння АР(2)

$$y(k) = 3,0 + 0,9 y(k-1) - 0,2 y(k-2) + \varepsilon(k)$$

З початковими умовами: $y(0) = y_0$; $y(1) = y_1$.

а) Знайдемо однорідний розв'язок:

$$\alpha^2 - 0,9\alpha + 0,2 = 0;$$

$$\alpha_1 = 0,5; \alpha_2 = 0,4.$$

$$y^h(k) = A_1 0,5^k + A_2 0,4^k.$$

б) Частковий розв'язок для детермінованої частини:

$$d = 3,0 + 0,9d - 0,2d, \quad \text{або} \quad d = \frac{3,0}{0,3} = 10.$$

в) Частковий розв'язок для стохастичної частини:

$$y^{ps}(k) = \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i),$$

де $\alpha_0 = 1$; $\alpha_1 = 0,9$; $\alpha_i = 0,9\alpha_{i-1} - 0,2\alpha_{i-2}$, $\forall i > 1$.

г) Повний розв'язок:

$$y(k) = 10,0 + A_1 0,5^k + A_2 0,4^k + \sum_{i=0}^{\infty} \alpha_i \varepsilon(k-i),$$

$$\alpha_i = 0,9\alpha_{i-1} - 0,2\alpha_{i-2}.$$

Для того щоб знайти значення невідомих довільних констант, використаємо початкові умови y_0, y_1 . В результаті отримаємо два рівняння:

$$y_0 = 10,0 + A_1 + A_2 + \sum_{i=0}^{\infty} \alpha_i \varepsilon(-i),$$

$$y_1 = 10,0 + 0,5A_1 + 0,4A_2 + \sum_{i=0}^{\infty} \alpha_i \varepsilon(1-i),$$

з яких можна визначити невідомі константи. Із врахуванням початкових умов розв'язок прийме такий вигляд:

$$y(k) = 10,0 + 0,4^k [5,0(y_0 - 10) - 10(y_1 - 10)] + \\ + 0,5^k [10(y_1 - 10) - 4(y_0 - 10)] + \sum_{i=0}^{k-1} \alpha_i \varepsilon(k-i).$$

Значення залежної змінної в момент $k + s$ при початкових умовах $y(k)$ і $y(k-1)$ можна записати так:

$$y(k+s) = 10,0 + 0,4^s \{5,0[y(k-1) - 10] - 10[(y(k) - 10)]\} + \quad (4.5.8) \\ + 0,5^s \{[10y(k) - 10] - 4[y(k-1) - 10]\} + \sum_{i=0}^{s-1} \alpha_i \varepsilon(k+s-i),$$

де $y(k-1), y(k)$ – початкові умови відносно k -го моменту часу. У рівнянні (4.5.8) виконано заміну часової змінної k на $k+s$, але із врахуванням того, що прогнозоване значення визначається відносно k -го моменту часу. Тому верхнє граничне значення над знаком суми визначається так: $k-2 = k+s-2$, а із врахуванням того, що початковим моментом часу (на який відома необхідна інформація) є $k=1$, отримаємо $1+s-2 = s-1$. Перевірка нижнього індекса свідчить, що $i=0 \Rightarrow \varepsilon(k+s)$ і $i=s-1 \Rightarrow \varepsilon(k+1)$, тобто границі зміни часового індекса для збурення визначені правильно.

Запишемо умовне математичне сподівання або прогноз на короткий проміжок часу:

$$E_k[y(k+s)] = 10,0 + 0,4^s \{5,0[y(k-1) - 10] - 10[(y(k) - 10)]\} + \\ + 0,5^s \{[10y(k) - 10] - 4[y(k-1) - 10]\}. \quad (4.5.9)$$

Безумовне середнє або асимптотичний прогноз приймає значення:

$$\lim_{S \rightarrow \infty} E_k[y(k+s)] = 10,$$

тобто прогноз на нескінченності дорівнює безумовному середньому.

4.6 Прогнозування з мінімальною дисперсією

Функцію прогнозування можна будувати різними способами, які відрізняються припущеннями щодо характеристик процесу (який описується математично і прогнозується), критерію оптимальності оцінок (якщо він використовується) та вигляду самої функції. Побудова функції прогнозування, яка забезпечує мінімум дисперсії, ґрунтується на таких припущеннях:

1. Процес, що прогнозується, є стаціонарним, гаусовим і має дробово раціональну спектральну густину.
2. Оптимальною вважається функція прогнозування, яка генерує оцінку прогнозу шляхом мінімізації дисперсії похибки прогнозу, тобто:

$$E[y(k+s) - \hat{y}(k+s)]^2 \Rightarrow \min.$$

3. Значення $y(k+s)$, для якого оцінюється прогноз, є функцією всіх попередніх спостережень до моменту k включно.

Результат буде таким же, якщо замінити припущення про нормальність процесу припущенням про те, що він є двічі диференційованим, а функцією прогнозування є лінійна функція спостережень [9].

Розглянемо стохастичне різницеве рівняння:

$$y(k+1) = -a y(k) + \varepsilon(k+1) + \beta \varepsilon(k), \quad (4.6.1)$$

де $\{\varepsilon(k)\} \sim N(0,1)$, тобто нормально розподілений процес із нульовим середнім та одиничною дисперсією. Розглянемо спочатку функцію прогнозування на один крок за допомогою рівняння (4.6.1). Для того щоб знайти значення $y(k+1)$, необхідно знати $y(k)$, $\varepsilon(k)$. Значення $y(k)$ – це значення останнього виміру, а $\varepsilon(k)$ необхідно обчислити.

Нехай відомі початкові значення $\varepsilon(0) = \varepsilon_0$, $y(0) = y_0$. Запишемо рівняння (4.6.1) у вигляді:

$$\varepsilon(\tau+1) - y(\tau+1) + \beta \varepsilon(\tau) = a y(\tau), \quad 0 \leq \tau \leq k$$

і знайдемо розв'язок цього рівняння. Додамо до лівої і правої частин член $-\beta y(\tau)$ і отримаємо:

$$[\varepsilon(\tau+1) - y(\tau+1)] + \beta[\varepsilon(\tau) - y(\tau)] = (a - \beta) y(\tau). \quad (4.6.2)$$

Тепер введемо нову змінну $x(\tau) = \varepsilon(\tau) - y(\tau)$ і запишемо різницеве рівняння для нової змінної:

$$x(\tau + 1) = -\beta x(\tau) + (a - b) y(\tau). \quad (4.6.3)$$

Знайдемо розв'язок цього рівняння за методом прямої ітерації:

$$\begin{aligned} x(1) &= -\beta x(0) + (a - b) y(0), \\ x(2) &= -\beta x(1) + (a - b) y(1) = -\beta[-\beta x(0) + (a - b) y(0)] + (a - b) y(1) = \\ &= -\beta^2 x(0) - \beta(a - b) y(0) + (a - b) y(1), \\ x(3) &= -\beta x(2) + (a - b) y(2) = \\ &= -\beta^3 x(0) + \beta^2 (a - b) y(0) - \beta(a - b) y(1) + (a - b) y(2). \end{aligned}$$

Для довільного k розв'язок має такий вигляд:

$$x(k) = (-\beta)^k x(0) + (a - b) \sum_{i=0}^{k-1} (-\beta)^{k-1-i} y(i). \quad (4.6.4)$$

Зробимо перевірку індексів:

- при $k = 1$ останній член в правій частині має вигляд: $(a - b) y(0)$;
- при $k = 2$ маємо наступне рівняння:

$$(a - b) \sum_{i=0}^1 (-\beta)^{1-i} y(i) = -\beta(a - b) y(0) + (a - b) y(1),$$

тобто часові індекси в рівнянні (4.6.4) змінюються в належних інтервалах.

Використовуючи рівняння (4.6.4), можемо записати вираз для $\varepsilon(k)$:

$$\varepsilon(k) = [\varepsilon(0) - y(0)](-\beta)^k + y(k) + (a - b) \sum_{i=0}^{k-1} (-\beta)^{k-1-i} y(i), \quad (4.6.5)$$

а для довільного початкового моменту часу k_0 отримаємо:

$$\varepsilon(k) = [\varepsilon(0) - y(0)](-\beta)^{k-k_0} + y(k) + (a - b) \sum_{i=0}^{k-1} (-\beta)^{k-1-i} y(i). \quad (4.6.6)$$

Оскільки для стаціонарного процесу $|\beta| < 1$, то перший член (4.6.5) і (4.6.6) наближається до нуля при $k_0 \rightarrow \infty$ для довільних початкових умов. Таким чином, $\varepsilon(k)$ може бути обчислено безпосередньо за допомогою результатів вимірів із рівняння:

$$\varepsilon(k) = y(k) + (a - \beta) \sum_{i=-\infty}^{k-1} (-\beta)^{k-1-i} y(i).$$

Таким чином, можемо записати рівняння для прогнозування значення $y(k+1)$ на основі спостережень $y(k), y(k-1), y(k-2), \dots$ у вигляді:

$$\begin{aligned} \hat{y}(k+1|k) &= \hat{y}(k+1, k) = \hat{y}(k+1) = -a y(k) + \beta \varepsilon(k) = \\ &= -a y(k) + \beta y(k) + \beta (a - \beta) \sum_{i=-\infty}^{k-1} (-\beta)^{k-i-1} y(i) = \\ &= -(a - \beta) y(k) + \beta (a - \beta) \sum_{i=-\infty}^{k-1} (-\beta)^{k-i-1} y(i) = (\beta - a) \sum_{i=-\infty}^k (-\beta)^{k-i} y(i). \end{aligned} \quad (4.6.7)$$

Рівняння (4.6.7) можна переписати у зручнішій рекурсивній формі:

$$\begin{aligned} \hat{y}(k+1, k) &= (\beta - a) \left[y(k) + \sum_{i=-\infty}^{k-1} (-\beta)^{k-i} y(i) \right] = \\ &= (\beta - a) \left[y(k) + (-\beta) \sum_{i=-\infty}^{k-1} (-\beta)^{k-1-i} y(i) \right] = -\beta \hat{y}(k, k-1) + (\beta - a) y(k). \end{aligned} \quad (4.6.8)$$

Тобто функція прогнозування на один крок описується різницевою рівнянням першого порядку. Динаміка прогнозу визначається коефіцієнтом β .

За означенням, похибка прогнозу дорівнює:

$$f(k+1, k) = y(k+1) - \hat{y}(k+1, k) = \varepsilon(k+1),$$

а її математичне сподівання $E[f(k+1, k)] = 0$.

Альтернативна форма функції прогнозування на один крок

Введемо оператор зсуву $z x(k) = x(k+1)$ або $z^{-1} x(k) = x(k-1)$. Тепер рівняння,

$$y(k+1) = -a y(k) + \varepsilon(k+1) + \beta \varepsilon(k), \quad (4.6.9)$$

можна переписати у вигляді:

$$y(k+1) = \frac{1 + \beta z^{-1}}{1 + a z^{-1}} \varepsilon(k+1) = \varepsilon(k+1) - \varepsilon(k) + \frac{1 + \beta z^{-1}}{1 + a z^{-1}} \varepsilon(k+1) =$$

$$\begin{aligned}
&= \varepsilon(k+1) + \frac{-\varepsilon(k+1) - a\varepsilon(k) + \varepsilon(k+1) + \beta\varepsilon(k)}{1 + az^{-1}} = \\
&= \varepsilon(k+1) + \frac{\beta - a}{1 + az^{-1}} \varepsilon(k). \tag{4.6.10}
\end{aligned}$$

Із рівняння (4.6.9) знайдемо вираз для випадкової величини:

$$\varepsilon(k) = \frac{1 + az^{-1}}{1 + \beta z^{-1}} y(k). \tag{4.6.11}$$

Підставимо цей вираз в (4.6.10) щоб уникнути змінної $\varepsilon(k)$ з метою зменшення невизначеності прогнозу:

$$y(k+1) = \varepsilon(k+1) + \frac{\beta - a}{1 + \beta z^{-1}} y(k). \tag{4.6.12}$$

Відніmemo від обох частин цього рівняння прогнозоване значення $\hat{y}(k+1, k)$ і знайдемо математичне сподівання квадрата похибки прогнозу:

$$E_k [y(k+1) - \hat{y}(k+1, k)]^2 = E[\varepsilon^2(k+1)] + E \left[\frac{\beta - a}{1 + \beta z^{-1}} y(k) - \hat{y}(k+1, k) \right]^2$$

за умови, що $E[y(k)\varepsilon(k+1)] = 0$. Якщо дисперсію збурення нормувати до одиниці, тобто $E[\varepsilon^2(k)] = 1$, то можна записати таке співвідношення:

$$E[y(k+1) - \hat{y}(k+1, k)] \geq E[\varepsilon^2(k+1)] = 1,$$

де рівність виконується тільки за умови, що:

$$\hat{y}(k+1, k) = \frac{\beta - a}{1 + \beta z^{-1}} y(k).$$

Таким чином, оптимальне з точки зору мінімуму дисперсії похибки оцінки прогнозу значення при прогнозуванні на один крок можна знайти за допомогою рівняння:

$$\hat{y}(k+1, k) + \beta \hat{y}(k, k-1) = (\beta - a) y(k),$$

або

$$\hat{y}(k+1, k) = -\beta \hat{y}(k, k-1) + (\beta - a) y(k), \tag{4.6.13}$$

що співпадає з отриманим вище рівнянням (4.6.8).

Функція прогнозування на два кроки

Побудуємо функцію прогнозування на два кроки $y(k+2, k)$ для випадкового процесу АРКС(1,1):

$$y(k) = -a y(k-1) + \varepsilon(k) + \beta \varepsilon(k-1),$$

де $\{\varepsilon(k): N(0,1)\}$ – випадковий процес з одиничною дисперсією та нульовим середнім значенням. Для моменту $k+2$ можна записати попереднє рівняння за допомогою оператора зсуву у формі:

$$y(k+2) = \frac{1 + \beta z^{-1}}{1 + a z^{-1}} \varepsilon(k+2). \quad (4.6.14)$$

Права частина цього рівняння – лінійна функція від $\varepsilon(k+2), \varepsilon(k+1), \varepsilon(k), \varepsilon(k-1), \dots$. При цьому випадкові змінні $\varepsilon(k), \varepsilon(k-1), \dots$ можна визначити за допомогою спостережень $y(k), y(k-1), \dots$, а випадкові змінні $\varepsilon(k+1), \varepsilon(k+2)$ не залежать від отриманих спостережень на момент k . Виділимо у правій частині члени з $\varepsilon(k+1), \varepsilon(k+2)$; з цією метою додамо та віднімемо $\varepsilon(k+2)$ до правої частини:

$$\begin{aligned} y(k+2) &= \varepsilon(k+2) - \varepsilon(k+2) + \frac{1 + \beta z^{-1}}{1 + a z^{-1}} \varepsilon(k+2) = \\ &= \varepsilon(k+2) + \frac{-\varepsilon(k+2) - a\varepsilon(k+1) + \varepsilon(k+2) + \beta\varepsilon(k+1)}{1 + a z^{-1}} = \\ &= \varepsilon(k+2) + \frac{\beta - a}{1 + a z^{-1}} \varepsilon(k+1). \end{aligned}$$

Тепер домножимо чисельник на многочлен $(1 + a z^{-1} - a z^{-1})$:

$$\begin{aligned} y(k+2) &= \varepsilon(k+2) + \frac{(\beta - a)(1 + a z^{-1} - a z^{-1})\varepsilon(k+1)}{1 + a z^{-1}} = \\ &= \varepsilon(k+2) + \frac{(\beta - a)(1 + a z^{-1})}{1 + a z^{-1}} \varepsilon(k+1) - \frac{a(\beta - a)\varepsilon(k)}{1 + a z^{-1}} = \end{aligned}$$

$$= \varepsilon(k+2) + (\beta - a)\varepsilon(k+1) - \frac{a(\beta - a)}{1 + az^{-1}}\varepsilon(k), \quad (4.6.15)$$

де $\varepsilon(k)$ можна обчислити за спостереженнями:

$$\varepsilon(k) = \frac{1 + az^{-1}}{1 + \beta z^{-1}} y(k).$$

Якщо підставити цей вираз в (4.6.15), то отримаємо рівняння:

$$\begin{aligned} y(k+2) &= \varepsilon(k+2) + (\beta - a)\varepsilon(k+1) - \frac{a(\beta - a)}{1 + az^{-1}}\varepsilon(k) = \\ &= \varepsilon(k+2) + (\beta - a)\varepsilon(k+1) - \frac{a(\beta - a)}{1 + \beta z^{-1}} y(k). \end{aligned} \quad (4.6.16)$$

Позначимо через $\hat{y}(k+2, k)$ – прогноз на два кроки на основі інформації, що є в наявності на момент k , включно. Запишемо дисперсію похибки прогнозу за допомогою рівняння (4.6.16):

$$\begin{aligned} E[y(k+2) - \hat{y}(k+2, k)]^2 &= E[\varepsilon^2(k+2)] + (\beta - a)^2 E[\varepsilon^2(k+1)] + \\ &+ E[\hat{y}(k+2, k) + \frac{a(\beta - a)}{1 + \beta z^{-1}} y(k)]^2. \end{aligned} \quad (4.6.17)$$

Таким чином, можна записати, що

$$E[y(k+2) - \hat{y}(k+2, k)]^2 \geq 1 + (\beta - a)^2.$$

Із (4.6.17) випливає, що рівність можлива за умови:

$$\hat{y}(k+2, k) = -\frac{a(\beta - a)}{1 + \beta z^{-1}} y(k). \quad (4.6.18)$$

Тобто функція двокрокового прогнозу задовольняє різницевому рівнянню:

$$\hat{y}(k+2, k) = -\beta \hat{y}(k+1, k) - a(\beta - a) y(k). \quad (4.6.19)$$

Знайдемо похибку двокрокового прогнозу:

$$\begin{aligned}
f(k+2, k) &= y(k+2) - \hat{y}(k+2, k) = \\
&= \varepsilon(k+2) + (\beta - a)\varepsilon(k+1) - \frac{a(\beta - a)}{1 + az^{-1}}y(k) + \frac{a(\beta - a)}{1 + \beta z^{-1}}y(k) = \\
&= \varepsilon(k+2) + (\beta - a)\varepsilon(k+1),
\end{aligned} \tag{4.6.20}$$

тобто, це процес КС(2).

Загальна постановка задачі прогнозування з мінімумом дисперсії

Модель процесу можна представити у вигляді:

$$A(z^{-1})y(k) = B(z^{-1})\varepsilon(k), \tag{4.6.21}$$

де $A(z^{-1}), B(z^{-1})$ – поліноми від z такого вигляду:

$$A(z^{-1}) = 1 + a_1z^{-1} + \dots + a_nz^{-n}, \tag{4.6.22}$$

$$B(z^{-1}) = 1 + b_1z^{-1} + \dots + b_nz^{-n}, \tag{4.6.23}$$

$$\{\varepsilon(k)\} \sim N(0, 1),$$

тобто, $\varepsilon(k)$ – нормально розподілена випадкова величина з нульовим середнім та одиничною дисперсією.

Необхідно побудувати найкращу, в смислі мінімуму дисперсії похибки, функцію прогнозування на s кроків за умови, що відомі значення $y(k), y(k-1), y(k-2), \dots$, тобто:

$$E_k [y(k+s) - \hat{y}(k+s)]^2 \rightarrow \min_{\hat{y}},$$

де \hat{y} – оцінка прогнозу.

На основі рівняння (4.6.21) можна записати:

$$y(k+s) = \frac{B(z^{-1})}{A(z^{-1})}\varepsilon(k+s). \tag{4.6.24}$$

Права частина цього рівняння є лінійною комбінацією змінних $\varepsilon(k+s), \varepsilon(k+s-1), \varepsilon(k+s-2), \dots, \varepsilon(k+1), \varepsilon(k), \varepsilon(k-1), \varepsilon(k-2), \dots$.

Оцінки випадкових змінних $\varepsilon(k), \varepsilon(k-1), \dots$ можна обчислити за результатами вимірів $y(k), y(k-1), \dots$, а значення $\varepsilon(k+1), \dots, \varepsilon(k+s)$ не залежать від вимірів, тому що ці виміри ще невідомі. Для того щоб виділити вказані групи змінних, перепишемо праву частину (4.6.24) у вигляді (що допустимо для лінійних систем):

$$\frac{B(z^{-1})}{A(z^{-1})} \varepsilon(k+s) = F(z^{-1}) \varepsilon(k+s) + z^{-s} \frac{G(z^{-1})}{A(z^{-1})} \varepsilon(k+s), \quad (4.6.25)$$

де $F(z^{-1}), G(z^{-1})$ – поліноми від z^{-1} степені $s-1$ та $n-1$, відповідно, тобто:

$$\begin{aligned} F(z^{-1}) &= 1 + f_1 z^{-1} + \dots + f_{s-1} z^{-(s-1)}, \\ G(z^{-1}) &= g_0 + g_1 z^{-1} + \dots + g_{n-1} z^{-(n-1)}. \end{aligned} \quad (4.6.26)$$

Таке розщеплення є допустимим для лінійних систем, якщо правильно вибрати степені z та коефіцієнти поліномів F, G . Для першої складової у правій частині рівняння (4.6.25) можна записати:

$$F(z^{-1}) \varepsilon(k+s) = \varepsilon(k+s) + f_1 \varepsilon(k+s-1) + \dots + f_{s-1} \varepsilon(k+s-s+1),$$

або

$$F(z^{-1}) \varepsilon(k+s) = \varepsilon(k+s) + f_1 \varepsilon(k+s-1) + \dots + f_{s-1} \varepsilon(k+1).$$

Враховуючи, що $z^{-s} \varepsilon(k+s) = \varepsilon(k)$, для чисельника другого члена у правій частині (4.6.25) отримуємо ковзне середнє порядку $n-1$:

$$g_0 \varepsilon(k) + g_1 \varepsilon(k-1) + \dots + g_{n-1} \varepsilon(k-n+1).$$

Тепер рівняння (4.6.24) можна записати у вигляді:

$$y(k+s) = F(z^{-1}) \varepsilon(k+s) + \frac{G(z^{-1})}{A(z^{-1})} \varepsilon(k). \quad (4.6.27)$$

Другий член у правій частині цього рівняння є функцією випадкових змінних $\varepsilon(k), \varepsilon(k-1), \dots$, а тому його можна обчислити за результатами спостережень $y(k), y(k-1), \dots$, тобто скориставшись рівнянням (4.6.21), отримаємо:

$$\varepsilon(k) = \frac{A(z^{-1})}{B(z^{-1})} y(k).$$

Тепер рівняння (4.6.27) можна звести до вигляду:

$$y(k+s) = F(z^{-1}) \varepsilon(k+s) + \frac{G(z^{-1})}{B(z^{-1})} y(k). \quad (4.6.28)$$

Перший член у правій частині цього рівняння є лінійною комбінацією змінних $\varepsilon(k+1), \varepsilon(k+2), \dots, \varepsilon(k+s)$, які не залежать від спостережень, а другий представляє собою лінійну функцію результатів спостережень.

Нехай $\hat{y}(k+s, k)$ – оцінка прогнозу на s кроків. Запишемо похибку цього прогнозу:

$$f(s) = y(k+s) - \hat{y}(k+s, k),$$

де $y(k+s)$ – спостереження, яке точно описується рівнянням (4.6.28). Відніmemo $\hat{y}(k+s, k)$ від обох частин рівняння (4.6.28), що приводить до виразу:

$$y(k+s) - \hat{y}(k+s, k) = F(z^{-1})\varepsilon(k+s) + \frac{G(z^{-1})}{B(z^{-1})}y(k) - \hat{y}(k+s, k)$$

і знайдемо квадрат похибки прогнозу:

$$\begin{aligned} [y(k+s) - \hat{y}(k+s, k)]^2 &= \left[F(z^{-1})\varepsilon(k+s) + \frac{G(z^{-1})}{B(z^{-1})}y(k) - \hat{y}(k+s, k) \right]^2 = \\ &= \left[F(z^{-1})\varepsilon(k+s) \right]^2 + 2 \left[F(z^{-1})\varepsilon(k+s) \right] \left[\frac{G(z^{-1})}{B(z^{-1})}y(k) - \hat{y}(k+s, k) \right] + \\ &\quad + \left[\frac{G(z^{-1})}{B(z^{-1})}y(k) - \hat{y}(k+s, k) \right]^2. \end{aligned}$$

Тепер запишемо умовне математичне сподівання для правої частини:

$$\begin{aligned} E_k [F(z^{-1})\varepsilon(k+s)]^2 + 2 E_k \left\{ \left[F(z^{-1})\varepsilon(k+s) \right] \left[\frac{G(z^{-1})}{B(z^{-1})}y(k) - \hat{y}(k+s, k) \right] \right\} + \\ + E_k \left[\frac{G(z^{-1})}{B(z^{-1})}y(k) - \hat{y}(k+s, k) \right]^2. \end{aligned}$$

Оскільки випадкові величини $\varepsilon(k+s), \varepsilon(k+2), \dots$ не залежать від $y(k), y(k-1), \dots$, то середній член у правій частині дорівнює нулю, а тому отримуємо:

$$E[y(k+s) - \hat{y}(k+s, k)]^2 = E_k [F(z^{-1})\varepsilon(k+s)]^2 +$$

$$+ E_k \left[\frac{G(z^{-1})}{B(z^{-1})} y(k) - \hat{y}(k+s, k) \right]^2. \quad (4.6.29)$$

Якщо знехтувати другим членом у правій частині цієї рівності, то можна записати нерівність:

$$E[y(k+s) - \hat{y}(k+s, k)]^2 \geq [\varepsilon(k+s) + f_1 \varepsilon(k+s-1) + f_2 \varepsilon(k+s-2) + \dots + f_{s-1} \varepsilon(k+1)]^2.$$

В цій нерівності рівність досягається за умови, що:

$$E_k \left[\frac{G(z^{-1})}{B(z^{-1})} y(k) - \hat{y}(k+s, k) \right]^2 = 0,$$

або

$$\hat{y}(k+s, k) = \frac{G(z^{-1})}{B(z^{-1})} y(k). \quad (4.6.30)$$

Звідси випливає, що оптимальну (за мінімумом дисперсії) оцінку прогнозу на s кроків можна обчислити за виразом:

$$\begin{aligned} \hat{y}(k+s, k) + b_1 \hat{y}(k+s, k-1) + \dots + b_n \hat{y}(k+s-n, k-n) = \\ = g_0 y(k) + g_1 y(k-1) + \dots + g_{n-1} y(k-n-1). \end{aligned} \quad (4.6.31)$$

Похибка прогнозування:

$$\begin{aligned} f(k+s, k) = y(k+s) - \hat{y}(k+s, k) = F(z^{-1}) \varepsilon(k+s) = \\ = \varepsilon(k+s) + f_1 \varepsilon(k+s-1) + \dots + f_{s-1} \varepsilon(k+1). \end{aligned} \quad (4.6.32)$$

Мінімально можливе значення дисперсії похибки оцінки прогнозу при даному підході таке:

$$\text{Var}[f(k+s, k)] = 1 + f_1^2 + f_2^2 + \dots + f_{s-1}^2.$$

Для того щоб скористатись формулами (4.6.31) і (4.6.32), необхідно знати коефіцієнти поліномів $F(z^{-1})$ і $G(z^{-1})$. Знайти невідомі коефіцієнти можна за допомогою тотожності:

$$B(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-s} G(z^{-1}). \quad (4.6.33)$$

Фактично, $F(z^{-1})$ – це частка від відношення $B(z^{-1})/A(z^{-1})$, а поліном $G(z^{-1})$ – залишок від ділення. Коефіцієнти поліномів $F(z^{-1})$ і $G(z^{-1})$ можна визначити шляхом прирівнювання коефіцієнтів при однакових степенях z^{-1} .

Для заданої структури поліномів $A(z^{-1})$, $B(z^{-1})$, $F(z^{-1})$ і $G(z^{-1})$:

$$\begin{aligned}
 B(z^{-1}) &= b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_n z^{-n}, \\
 A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_n z^{-n}, \\
 F(z^{-1}) &= 1 + f_1 z^{-1} + f_2 z^{-2} + \dots + f_{s-1} z^{-(s-1)}, \\
 G(z^{-1}) &= g_0 + g_1 z^{-1} + \dots + g_{n-1} z^{-(n-1)}, \\
 z^{-s} G(z^{-1}) &= g_0 z^{-s} + g_1 z^{-(s+1)} + g_2 z^{-(s+2)} + \dots + g_{n-1} z^{-(n-1+s)}
 \end{aligned} \tag{4.6.34}$$

отримаємо такі рівності для того щоб визначити коефіцієнти:

$$\begin{aligned}
 \text{при } z^{-1}: \quad b_1 &= a_1 + f_1, \\
 \text{при } z^{-2}: \quad b_2 &= a_2 + a_1 f_1 + f_2, \\
 &\quad \vdots \\
 \text{при } z^{-(s-1)}: \quad b_{s-1} &= a_{s-1} + a_{s-2} f_1 + a_{s-3} f_2 + \dots + a_1 f_{s-2} + f_{s-1}, \\
 \text{при } z^{-s}: \quad b_s &= a_k + a_{k-1} f_1 + a_{k-2} f_2 + \dots + a_1 f_{s-1} + g_0, \\
 \text{при } z^{-(s+1)}: \quad b_{s+1} &= a_{k+1} + a_k f_1 + a_{k-1} f_2 + \dots + a_2 f_{s-1} + g_1, \\
 &\quad \vdots \\
 \text{при } z^{-n}: \quad b_n &= a_n + a_{n-1} f_1 + a_{n-2} f_2 + \dots + a_{n-s-1} f_{s-1} + g_{n-s}, \\
 \text{при } z^{-(n+1)}: \quad 0 &= a_n f_1 + a_{n-1} f_2 + \dots + a_{n-s+2} f_{s-1} + g_{n-s+1}, \\
 &\quad \vdots \\
 \text{при } z^{-(n+s-1)}: \quad 0 &= a_n f_{s-1} + g_{n-1}.
 \end{aligned} \tag{4.6.35}$$

Таким чином, коефіцієнти f_i, g_i можна обчислити за допомогою простих рекурсивних співвідношень при відомих $A(z^{-1})$ і $B(z^{-1})$, а також вибраній кількості кроків s для прогнозування. Очевидно, що це число не може бути великим.

Примітка 1: Для даної постановки задачі кращою є лінійна оцінка прогнозу і результат прогнозування не залежить критично від вибору критерію мінімізації похибки прогнозу. Якщо послідовність $\{y(k)\}$ має нормальний розподіл, то функція прогнозування не зміниться за умови вибору критерій мінімізації у вигляді:

$$E w[y(k+s) - \hat{y}(k+s, k)]^2,$$

де w – довільна симетрична функція.

Примітка 2. Вирішальну роль для коректності наведеного вище виводу функції прогнозування відіграє той факт, що $\varepsilon(k)$ і $\varepsilon(l)$ є взаємно незалежними при $k \neq l$. Якщо випадкові змінні $\varepsilon(k)$ і $\varepsilon(l)$ взаємозалежні одна від одної, то математичне сподівання добутку $\varepsilon(k+l)$ на довільну функцію від $y(k), y(k-1), \dots$ не обов'язково буде дорівнювати нулю при $l > 0$.

Приклад 4.2: Для стохастичного процесу $\{y(k)\}$, який визначається рівнянням

$$y(k) = 1,5y(k-1) + 0,5y(k-2) = 2[\varepsilon(k) - 1,2\varepsilon(k-1) + 0,6\varepsilon(k-2)],$$

де $\{\varepsilon(k)\} \sim N(0,1)$, записати функцію прогнозування на три кроки, $s=3$, яка мінімізує середньоквадратичну похибку прогнозу.

Для даного прикладу поліноми моделі:

$$A(z^{-1}) = 1 - 1,5z^{-1} + 0,5z^{-2},$$

$$B(z^{-1}) = 1 - 1,2z^{-1} + 0,6z^{-2}.$$

Використовуючи співвідношення (4.6.35), можна визначити коефіцієнти функції прогнозування:

$$b_1 = a_1 + f_1 \quad \text{або} \quad f_1 = b_1 - a_1 = -1,2 + 1,5 = 0,3;$$

$$b_2 = a_2 + a_1 f_1 + f_2 \quad \text{або} \quad f_2 = b_2 - a_2 - a_1 f_1 = 0,6 - 0,5 + 1,5 \cdot 0,3 = 0,55;$$

Таким чином, $F(z^{-1}) = 1 + 0,3z^{-1} + 0,55z^{-2}$.

Використовуючи тотожність (4.6.33), запишемо:

$$1 - 1,2z^{-1} + 0,6z^{-2} = (1 - 1,5z^{-1} + 0,5z^{-2}) \cdot (1 + 0,3z^{-1} + 0,55z^{-2}) + z^{-3}(g_0 + g_1z^{-1}).$$

Звідси маємо:

$$G(z^{-1}) = 0,675 + 0,275z^{-1},$$

тобто всі коефіцієнти, необхідні для функції прогнозування, знайдено.

4.7 Функції прогнозування, отримані у даному розділі

Запишемо функції прогнозування динаміки процесу на довільне число s кроків, отримані в даному розділі:

- для рівняння першого порядку без розв'язку:

$$E_k[y(k+s)] = \hat{y}(k+s, k) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k);$$

- для рівняння АР(2) без розв'язку:

$$E_k[y(k+s)] = \hat{y}(k+s, k) = a_0 + a_1 E_k[y(k+s-1)] + a_2 E_k[y(k+s-2)];$$

- на основі повного розв'язку рівняння АРКС(1,1):

$$E_k[y(k+s)] = \frac{a_0}{1-a_1}(1-a_1^s) + \beta_1 a_1^{s-1} \varepsilon(k) + a_1^s y(k);$$

- на основі повного розв'язку рівняння АР(1):

$$E_k[y(k+s)] = \frac{a_0}{1-a_1}(1-a_1^s) + a_1^s y(k);$$

- функція прогнозування з мінімальною дисперсією:

$$\hat{y}(k+s, k) = \frac{G(z^{-1})}{B(z^{-1})} y(k).$$

Інші види прогнозуючих моделей будуть розглянуті нижче в розділах, присвячених моделюванню нестационарних гетероскедастичних та коінтегрованих процесів, а також оптимальному оцінюванню станів за допомогою фільтра Калмана (ФК). ФК представляє собою рекурсивну процедуру для обчислення оптимальних оцінок стану процесу в умовах впливу випадкових збурень стану і шумів вимірів. В процесі визначення поточних оцінок стану фільтр також дає можливість оцінити однокроковий прогноз оцінок стану на наступний крок на основі попередньої оцінки.

Існують також інші методи прогнозування, які суттєво відрізняються від наведених в даному розділі. Зокрема, це підхід на основі теорії формування Байєсового висновку, який забезпечує отримання високоякісних оцінок прогнозів в умовах наявності суттєвих невизначеностей, коли задача погано формалізується. Для визначення оцінки прогнозу необхідно розглядати декілька взаємозв'язаних процесів, коли статистичні та інші типи невизначеностей не можуть бути описані з прийнятною точністю за допомогою рівнянь авторегресії, АРКС та у просторі станів.

Відомий Байєсів метод використовують у різних формах постановки задачі та застосовують різні підходи до їх розв'язання. Однією із популярних модифікацій даного методу є представлення задачі прогнозування і прийняття рішення у вигляді динамічної байєсівської мережі, яка формально зображається спрямованим ациклічним графом. Вершинами графа виступають змінні стану процесу. Для отримання остаточного результату можна скористатись різними методами формування висновку.

4.8 Оцінювання якості прогнозу

Важливим моментом процесу прогнозування є об'єктивне визначення якості отриманого прогнозу. Оскільки оцінки прогнозів – випадкові величини, то для встановлення їх якості необхідно використовувати декілька статистичних критеріїв. Рис. 6.1 ілюструє часову вісь та відрізки часу, на яких виконується оцінювання моделі і перевірка якості прогнозу.

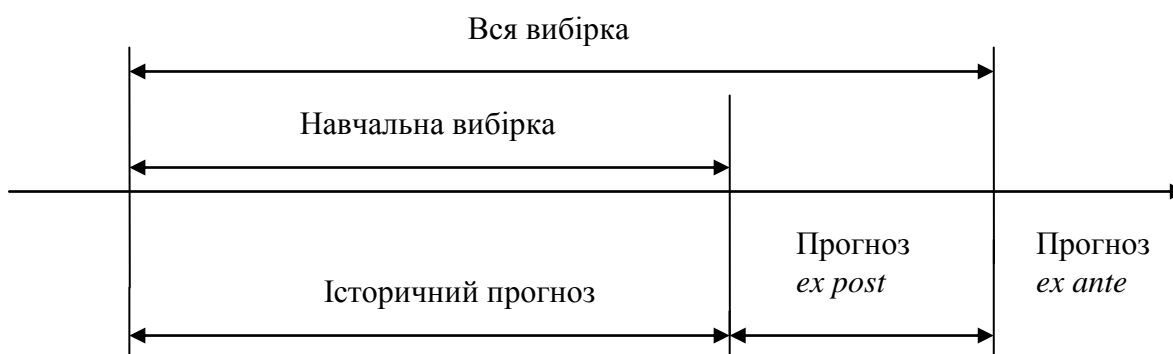


Рис. 4.1. Види прогнозування за часовим рядом

На наявну вибірку даних доцільно розділити на навчальну та перевірочну. На навчальній вибірці виконується оцінювання параметрів моделі процесу і реалізується так званий „історичний” однокроковий прогноз, який дає змогу встановити якість оцінок прогнозування на цьому участку ряду. Прогноз на перевірочній частині вибірки даних в науковій літературі називають ще прогнозом *ex post*. В різних емпіричних дослідженнях рекомендують залишати для перевірки (5 – 40) % значень ряду даних. Хоча при аналізі коротких рядів для оцінювання параметрів моделі доцільно використовувати значно більшу частину ряду. Прогнозування значень поза вибіркою даних називають прогнозом *ex ante* (рис. 4.1).

Як правило, для оцінювання якості прогнозів використовують множину взаємно доповнюючих статистичних критеріїв. Наприклад, значення середньоквадратичної похибки залежить від масштабу даних, а тому недостатньо використовувати тільки цей статистичний параметр для аналізу якості прогнозу. Поглиблене оцінювання якості прогнозів досягається за рахунок використання критеріїв, які дають відносні оцінки в процентах. Розглянемо деякі статистичні критерії якості прогнозу та їх призначення.

Дисперсія і стандартне відхилення прогнозу

Досить часто проста сума похибок прогнозів дорівнює нулю, оскільки похибки приймають різні знаки, а тому необхідно використовувати інші міри похибки. Ступінь розсіювання значень змінної навколо її середнього можна виміряти за допомогою стандартного відхилення $\sigma(k)$, яке дорівнює квадратному кореню з дисперсії, тобто:

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}, \quad (4.8.1)$$

де $y(k)$ – фактичне значення змінної; $\hat{y}(k)$ – прогноз. (Ділення на $N-1$, а не на N має смисл при невеликій потужності (довжині) ряду.)

Стандартне відхилення залишків – один із основних показників якості прогнозу. Це досить широко застосовувана статистична характеристика, яка виявляється корисною при аналізі поведінки процесів різної природи. Вона використовується, наприклад, як міра ризику (волатильність) при аналізі фінансових процесів, як характеристика надійності в управлінні запасами та інших прикладних задачах управління, а також як міра розсіювання значень змінних стану при аналізі систем різної природи.

Оскільки дисперсія є квадратичною характеристикою процесу, то в процесі її обчислення не відбувається взаємна компенсація значень відхилень від середнього. Крім того, дисперсію $\sigma^2(k)$ і стандартне відхилення можна перевіряти на статистичну значущість, що в деякій мірі сприяє поглибленому аналізу похибки прогнозу. Вище було показано, що при застосуванні для прогнозування рівнянь АРКС прогноз незміщений (математичне сподівання похибок прогнозів дорівнює нулю), але дисперсія прогнозованих значень прямо пропорційна числу кроків прогнозування. Однак вона збігається до скінченної величини для стаціонарних процесів.

При відносно невеликому горизонті прогнозування можна стверджувати, що майбутнє значення прогнозованого показника попадає в інтервал, який визначається як \pm два стандартних відхилення від обчисленого значення прогнозу.

Очевидно, що дисперсії та стандартного відхилення недостатньо для аналізу якості прогнозів. Наприклад, якщо прогноз попиту дорівнює 2500 одиниць товару із стандартним відхиленням похибки 100, то інтервал $2300 \div 2700$ буде достатньо інформативним. Однак, якщо при такому ж значенні прогнозу стандартне відхилення складе 400 одиниць, то відповідний інтервал $1700 \div 3300$ навряд чи буде корисним для практичного використання.

Середнє абсолютне значення похибки

Рекурсивне обчислення середнього абсолютного значення похибки (САП) ґрунтується на виразі для експоненціального середнього і має вигляд:

$$\begin{aligned} \text{САП}(k) &= \alpha \cdot |y(k) - \hat{y}(k)| + (1 - \alpha) \text{САП}(k-1) = \\ &= \alpha |e(k)| + (1 - \alpha) \text{САП}(k-1), \end{aligned} \quad (4.8.2)$$

де $0 < \alpha < 1$; $e(k)$ – похибка прогнозу. При такому обчисленні середнє абсолютне значення завжди невід’ємне.

САП можна обчислити також за іншим виразом:

$$САП = \frac{1}{N} \sum_{k=1}^N |y(k) - \hat{y}(k, k)|.$$

Для досить широкого класу статистичних розподілів випадкових величин значення стандартного відхилення є дещо більшим САП і строго пропорціональним йому. Коефіцієнт пропорціональності коливається для різних розподілів між 1,2 і 1,3 (для нормального розподілу це значення дорівнює $\sqrt{\pi/2} = 1,2533$) [45]. Таким чином, можна записати, що

$$\sigma(k) \approx 1,25САП. \quad (4.8.3)$$

Виходячи з цього, процедуру оцінювання *стандартної похибки прогнозу* можна сформулювати так.

1. Обчислити поточне значення похибки прогнозу:
 $e(k) = y(k) - \hat{y}(k)$.
2. Обчислити нове значення САП за допомогою рівняння (4.8.2).
3. Знайти поточне стандартне відхилення за виразом (4.8.3).

Середню абсолютну похибку і її стандартне відхилення використовують, як правило, для визначення якості оцінок прогнозу одночасно.

Середній квадрат похибки і сума квадратів похибок

Якщо середній квадрат похибок (СКП) визначається для розповсюдженого випадку однокрокового прогнозування при довжині часового ряду N , то СКП визначається за формулою:

$$СКП = \frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2. \quad (4.8.4)$$

У випадках, коли прогнозування виконується на s кроків відносно моменту часу k , то СКП обчислюється за виразом:

$$СКП = \frac{1}{s} \sum_{i=1}^s [y(k+i) - \hat{y}(k+i, k)]^2. \quad (4.8.5)$$

Очевидно, що формули (4.8.4) і (4.8.5) однакові, але кожна форма чітко відображає на якому відрізку ряду даних відбувається прогнозування.

Більшість відомих пакетів для математичного моделювання і виконання статистичних розрахунків розраховують суму квадратів похибок за виразом:

$$C_{MKП} = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2. \quad (4.8.6)$$

Це корисний інтегральний показник якості, який використовують для порівняльного аналізу різних методів прогнозування.

Саме СКП і $C_{MKП}$ – найбільш поширені критерії, які використовуються при порівняльному аналізі і виборі кращої моделі для обчислення оцінок прогнозів. Однак їх також недостатньо для поглибленого аналізу результатів прогнозування.

Середня абсолютна похибка у процентах

Середня абсолютна похибка в процентах (САПП) – це середнє абсолютних значень похибок оцінок прогнозу в процентах відносно фактичного значення показника:

$$САПП = \frac{1}{N} \sum_{k=1}^N \frac{|y(k) - \hat{y}(k)|}{|y(k)|} \times 100\% = \frac{1}{N} \sum_{k=1}^N \frac{|e(k)|}{|y(k)|} \times 100\%, \quad (4.8.7)$$

або у випадку прогнозування на s кроків відносно k –го моменту:

$$САПП = \frac{1}{s} \sum_{i=1}^s \frac{|y(k+i) - \hat{y}(k+i, k)|}{|y(k+i)|} \times 100\% = \frac{1}{s} \sum_{i=1}^s \frac{|e(k+i)|}{|y(k+i)|} \times 100\%. \quad (4.8.8)$$

Оскільки ця міра характеризує відносну якість прогнозу, то її використовують, в основному, для порівняння точності прогнозів різнорідних об'єктів (процесів) прогнозування. Однак вона завжди корисна при виконанні порівняльного аналізу якості прогнозування одного й того ж процесу різними методами, оскільки відносна міра є чіткою і зрозумілою для дослідника і практичного користувача. Типові значення САПП та їх запропонована інтерпретація стосовно якості прогнозу наведені в таблиці 4.1.

Якщо у формулах (4.8.7) і (4.8.8) $y(k)$ або $y(k+i)$ прямує до нуля, то значення САПП прямуватиме до нескінченності. Про це необхідно пам'ятати при застосуванні даного критерію якості прогнозу. Для того щоб виконати обчислення даного критерію в таких випадках, нульові значення $y(k)$ або $y(k+i)$ необхідно пропускати з відповідним корегуванням значення N або s . Можливо, що такий підхід не відповідає деяким вимогам статистичного аналізу даних, але він дає можливість з практично достатньою повнотою виконати аналіз якості прогнозування.

Таблиця 4.1

Інтерпретація типових значень критерію САПП	
САПП, %	Інтерпретація
< 10	Висока точність
10 – 15	Хороша точність
15 – 20	Задовільна точність
> 20	Незадовільна (неприйнятна) точність

Середня похибка (СП) і середня похибка в процентах (СПП)

Середня похибка – це не відносний показник, вона характеризує *ступінь зміщення* прогнозованих значень від фактичних і для навчальної вибірки розраховується за формулою:

$$СП = \frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)] = \frac{1}{N} \sum_{k=1}^N e(k), \quad (4.8.9)$$

або при прогнозуванні відносно k -го моменту часу:

$$СП = \frac{1}{s} \sum_{i=1}^s [y(k+i) - \hat{y}(k+i, k)]. \quad (4.8.10)$$

Очевидно, що СП буде зменшуватись у випадках, коли похибки мають різні знаки.

Середню похибку в процентах (СПП) обчислюють за виразом:

$$СПП = \frac{1}{N} \sum_{k=1}^N \frac{[y(k) - \hat{y}(k)]}{y(k)} \times 100\%, \quad (4.8.11)$$

або

$$СПП = \frac{1}{s} \sum_{i=1}^s \frac{[y(k+i) - \hat{y}(k+i, k)]}{y(k+i)} \times 100\%. \quad (4.8.12)$$

СПП також характеризує *зміщеність* прогнозу. Якщо втрати при прогнозуванні, які зв'язані із завищенням фактичного майбутнього значення, врівноважуються заниженням, то ідеальний прогноз повинен бути незміщеним. У такому випадку СП і СПП повинні прямувати до нуля. Очевидно, що нуль – це ідеальне значення і забезпечити його на практиці неможливо. Емпірично встановлено, що прийнятними значеннями для СПП (так само як і для САПП) є $\leq 10\%$.

Максимальна і мінімальна абсолютна похибка

Очевидно, що максимальна абсолютна похибка (МАП) може бути визначена так:

$$МАП = \max_k \{ |y(k) - \hat{y}(k)| \}, \quad 1 \leq k \leq N, \quad (4.8.13)$$

або

$$МАП = \max_i \{ |y(k+1) - \hat{y}(k+1, k)|, \dots, |y(k+s) - \hat{y}(k+s, k)| \}; \quad (4.8.14)$$

$$1 \leq i \leq s.$$

Мінімальна абсолютна похибка (МіАП) визначається як

$$МіАП = \min_k \{ |y(k) - \hat{y}(k)| \}, \quad 1 \leq k \leq N, \quad (4.8.15)$$

або

$$МіАП = \min_i \{ |y(k+1) - \hat{y}(k+1, k)|, \dots, |y(k+s) - \hat{y}(k+s, k)| \}, \quad (4.8.16)$$

$$1 \leq i \leq s.$$

Критерії МАП і МіАП також можуть бути корисними при виконанні порівняльного аналізу декількох методів прогнозування, особливо якщо нас цікавлять максимально або мінімально можливі відхилення прогнозів від фактичних значень на заданому інтервалі.

Коефіцієнт нерівності Тейла

Коефіцієнт нерівності Тейла U – це важливий індикатор якості моделі і прогнозу; за означенням: $0 \leq U \leq 1$. Якщо $U = 1$, то модель має практично нульові (неприйнятні) прогнозуючі властивості, що впливає із формули для обчислення U :

$$U = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}}{\sqrt{\frac{1}{N} \sum_{k=1}^N y^2(k)} + \sqrt{\frac{1}{N} \sum_{k=1}^N \hat{y}^2(k)}}. \quad (4.8.17)$$

При $U = 0$ прогнозовані значення співпадають з фактичними значеннями ряду – модель ідеальна. Тобто U дає можливість встановити придатність моделі (методу) в принципі для оцінювання прогнозу. Емпіричні дослідження свідчать, що високоякісним оцінкам прогнозів наближено відповідають значення $U \leq 0,005$.

Коефіцієнт Тейла можна розкласти на три складові; складова, пропорційна зміщенню:

$$U^M = \frac{(\bar{y} - \hat{\bar{y}})^2}{\frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2}; \quad (4.8.18)$$

складова, пропорційна дисперсії:

$$U^S = \frac{(\sigma_{\text{факт}} - \sigma_{\text{моделі}})^2}{\frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2}; \quad (4.8.19)$$

і складова, пропорційна коваріації:

$$U^C = \frac{2(1-\rho)(\sigma_{\text{факт}} \cdot \sigma_{\text{моделі}})}{\frac{1}{N} \sum_{i=1}^N [y(i) - \hat{y}(i)]^2}, \quad (4.8.20)$$

де ρ – коефіцієнт кореляції між залишками моделі.

Величина U^M використовується для перевірки факту, чи існує систематичне відхилення середніх фактичного і прогнозованого ряду. Іншими словами, чи існує систематичне зміщення на виході моделі у той чи інший бік. Чим меншим значення приймає U^M , тим кращою є модель. Якщо $U^M = 0$, то прогнозовані значення не містять зміщення і модель адекватна процесу за цим показником.

Величина U^S використовується для тестування динамічних властивостей моделі, тобто чи відповідає її дисперсія дисперсії фактичного ряду. Наприклад, модель може відтворювати систематично менші за амплітудою коливання ніж ті, які має фактичний ряд. Чим менше значення U^S , тим меншим є відхилення дисперсії виходу моделі від дисперсії ряду.

Накінець, складова, пропорційна коваріації, є мірою корельованості фактичного і прогнозованого за моделлю ряду. Зазначимо, що за побудовою останнього критерію для нього виконується рівність:

$$U^M + U^S + U^C = 1.$$

Якість моделі визначається також тим, наскільки точно вона може прогнозувати зміни напряму розвитку процесу, тобто нахил або знак тренду. Моделі можуть мати високу точність відтворення ряду, але погано прогнозувати тренди або цикли. Інші моделі, навпаки, можуть мати меншу точність (адекватність опису), але кращі динамічні властивості. Таким чином, завжди необхідно шукати компроміс між точністю моделі, тобто якістю апроксимації ряду, і її динамічними властивостями. Однак, для аналізу цієї

властивості формальних тестів немає. Можна наближено встановити якість моделі стосовно відтворення динаміки ряду шляхом візуального аналізу фактичного та спрогнозованого ряду.

Ще однією характеристикою якості моделі є її чутливість до початкового (стартового) періоду імітаційного моделювання. Якщо модель генерує результати, які можна наближено класифікувати як інваріантні до початкових умов, то вона вважається хорошою. Інакше, якщо результати імітаційного моделювання залежать від початкових умов, можна припустити, що модель неякісна. Наприклад, вона може мати нестационарність деякого типу.

4.9 Довірчий інтервал для оцінок прогнозу

При використанні для обчислення прогнозу регресійних та рівнянь інших типів ми отримуємо точкову оцінку. Однак, така оцінка є далеко не завжди значущою. При прогнозуванні необхідно визначити інтервал, всередині якого з достатньою ступінню впевненості можна очікувати появу фактичного значення показника. В регресійному аналізі границі цього інтервалу задають за допомогою довірчого інтервалу.

Довірчий інтервал – це інтервал, який містить сам прогноз і в якому з визначеною ступінню впевненості можна очікувати появу фактичного значення прогнозованої змінної.

Так, значення прогнозу $\hat{y}(k+1) = 1500$ з довірчим інтервалом ± 150 і ступенем впевненості 95% означає, що з ймовірністю $\approx 0,95$ очікується, що наступне значення прогнозованої змінної буде знаходитись в межах $1350 \div 1650$. Якби довірчий інтервал складав ± 500 , то з ймовірністю $\approx 0,95$ очікується, що майбутнє значення буде знаходитись в інтервалі $1000 \div 2000$.

В регресійному аналізі *мінімальна ширина довірчого інтервалу* відповідає точці (\bar{y}, \bar{k}) – “середині спостережень”[45]. По обидва боки від цієї середини довжина інтервалу збільшується. Для того щоб визначити довірчий інтервал, необхідно знайти стандартну похибку рівняння регресії S_r , яка обчислюється всіма пакетами статистичної обробки даних. Величину S_r ще називають середньоквадратичним відхиленням і обчислюють за формулою:

$$S_r = \sqrt{\frac{1}{N-p} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}, \quad (4.9.1)$$

де p – кількість параметрів моделі, які оцінюються в процесі її побудови. Суму $\sum [y(k) - \hat{y}(k)]^2$ можна обчислити також як різницю між загальною сумою квадратів і сумою квадратів значень, знайденою за регресією.

Тепер можна визначити стандартну похибку прогнозу так:

$$S_{\hat{y}(k)} = S_r \sqrt{1 + \frac{1}{N} + \frac{(k - \bar{k})^2}{\sum_{k=1}^N (k - \bar{k})^2}}, \quad (4.9.2)$$

тобто стандартна похибка прогнозу залежить від довжини ряду N і віддалі (в часі) від середини часового періоду, що розглядається, до моменту прогнозування [45]. Видно, що збільшення довжини ряду приводить до зменшення стандартної похибки прогнозу. Якщо кількість кроків прогнозування відносно N -го значення ряду дорівнює τ , то член $(k - \bar{k})$ в виразі (4.9.2) необхідно замінити на $(\tau + (N - 1)/2)$. Таку процедуру введення нового позначення називають ще *стандартним прогностичним перепозначенням*.

Із врахуванням введеного позначення вираз для стандартної похибки приймає вигляд:

$$S_{\hat{y}(k+\tau)} = S_r \sqrt{1 + \frac{1}{N} + \frac{(\tau + (N - 1)/2)^2}{\sum_{k=1}^N k^2 - (\sum k)^2 / N}}, \quad (4.9.3)$$

де S_r для рівняння регресії $y(k) = a_0 + bk$ можна обчислити за виразом:

$$S_r = \sqrt{\frac{\sum y^2(k) - a_0 \sum y(k) - b \sum k y(k)}{N - 2}}. \quad (4.9.4)$$

Скориставшись стандартною похибкою прогнозу, можна визначити довірчі інтервали. Так, наближеними 99%, 95% і 68%-ми довірчими інтервалами будуть значення $\pm 3S_{\hat{y}(k+\tau)}$, $\pm 2S_{\hat{y}(k+\tau)}$ і $\pm S_{\hat{y}(k+\tau)}$, відповідно.

4.10 Комбінування оцінок прогнозів, отриманих за різними методами

4.10.1 Усереднення прогнозів (вагові коефіцієнти однакові)

Для двох методів прогнозування середнє визначається досить просто:

$$\hat{y}_c(k) = \frac{\hat{y}_1(k) + \hat{y}_2(k)}{2}, \quad (4.10.1)$$

де $\hat{y}_c(k)$ – комбінований прогноз; $\hat{y}_1(k), \hat{y}_2(k)$ – прогнози, отримані за різними методами. Якщо окремі прогнози незміщені (це повинен забезпечувати метод прогнозування), то комбінований прогноз також буде незміщеним. Похибка комбінованого прогнозу визначається як середнє похибок окремих оцінок:

$$e_c(k) = y(k) - \hat{y}_c(k) = y(k) - \frac{\hat{y}_1(k) + \hat{y}_2(k)}{2} = \frac{e_1(k) + e_2(k)}{2}, \quad (4.10.2)$$

де $y(k)$ – фактичне значення прогнозованої змінної.

Дисперсія похибки комбінованого прогнозу:

$$\begin{aligned} \text{Var} \left[\frac{e_1(k) + e_2(k)}{2} \right] &= E \left[\frac{e_1(k) + e_2(k)}{2} \right]^2 = \frac{1}{4} E [e_1^2(k) + 2e_1(k)e_2(k) + e_2^2(k)] = \\ &= \frac{1}{4} \{ E [e_1^2(k)] + 2E [e_1(k)e_2(k)] + E [e_2^2(k)] \} = \\ &= \frac{1}{4} \left[\sigma_1^2 + 2 \frac{E [e_1(k)e_2(k)]}{\sigma_1 \sigma_2} \sigma_1 \sigma_2 + \sigma_2^2 \right] = \frac{\sigma_1^2 + 2\rho \sigma_1 \sigma_2 + \sigma_2^2}{4}. \end{aligned}$$

Таким чином, дисперсія комбінованого прогнозу обчислюється за виразом:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho \sigma_1 \sigma_2}{4}, \quad (4.10.3)$$

де ρ – коефіцієнт кореляції між похибками прогнозу. Якщо похибки оцінок прогнозів, отриманих за двома моделями, незалежні, тобто, $\rho = 0$, то формула (4.10.3) спрощується:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2}{4}. \quad (4.10.4)$$

Таким чином, якщо дисперсії близькі за значеннями і похибки оцінок прогнозів незалежні, то дисперсія комбінованої похибки буде значно меншою будь-якої з двох дисперсій. Наприклад, нехай $\sigma_1^2 = \sigma_2^2 = 100$:

$$\sigma_c^2 = \frac{100+100}{4} = 50.$$

Але навіть при існуванні досить високої кореляції між похибками прогнозування дисперсія похибки комбінованого прогнозу буде меншою ніж дисперсія кожного методу окремо. Наприклад, нехай $\sigma_1^2 = \sigma_2^2 = 100$ і $\rho = 0,8$:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{100 + 100 + 2 \cdot 0,8 \cdot 10 \cdot 10}{4} = 90.$$

Навіть в цій ситуації спостерігається зменшення дисперсії похибки прогнозу після усереднення оцінок, отриманих за двома методами.

Однак ситуація змінюється у випадку, коли дисперсії індивідуальних похибок сильно відрізняються. Наприклад, нехай $\sigma_1^2 = 100$, $\sigma_2^2 = 16$ і $\rho = 0,8$:

$$\sigma_c^2 = \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{4} = \frac{100 + 16 + 2 \cdot 0,8 \cdot 10 \cdot 4}{4} = 45.$$

Таким чином, якщо дисперсії похибок сильно відрізняються одна від одної, то просте усереднення результатів не потрібно робити. Висновок такий: просте усереднення можна застосовувати у випадках, коли дисперсії індивідуальних похибок прогнозування приблизно рівні або не дуже сильно відрізняються за своїми значеннями.

4.10.2 Зважене комбінування прогнозів

Якщо інформація стосовно характеристик індивідуальних оцінок прогнозів відсутня, то можна присвоїти різні вагові коефіцієнти окремим прогнозам на основі суб'єктивних або експертних суджень:

$$\hat{y}_c(k) = w_1 \hat{y}_1(k) + w_2 \hat{y}_2(k),$$

де w_1, w_2 – вагові коефіцієнти. Очевидно, що більші значення вагових коефіцієнтів необхідно присвоювати тим індивідуальним прогнозам, які мають меншу дисперсію похибок. При цьому для коректності обчислень необхідно, щоб виконувалась умова: $w_1 + w_2 = 1$.

4.10.3 Вибір вагових коефіцієнтів за допомогою похибок оцінок прогнозів

Як правило, похибки прогнозів для конкретних моделей і процесів відомі, або їх можна визначити на навчальній вибірці. Це дає можливість об'єктивно підійти до розв'язання задачі вибору вагових коефіцієнтів. Оскільки моделі, які дають менші суми квадратів похибок прогнозів, генерують якісніші прогнози, то логічно прийняти цю міру за основу для визначення вагових коефіцієнтів. Позначимо суму квадратів похибок прогнозування (для історичного прогнозу) через

$$sse = \sum_{k=1}^N e^2(k).$$

Тепер можна записати вирази для вагових коефіцієнтів окремих прогнозів:

$$w_1 = \frac{1/sse_1}{1/sse_1 + 1/sse_2}, \quad (4.10.5)$$

$$w_2 = \frac{1/sse_2}{1/sse_1 + 1/sse_2}, \quad (4.10.6)$$

де sse_1, sse_2 – суми квадратів похибок для кожного з методів, що використовуються.

Наприклад, нехай $sse_1 = 100, sse_2 = 16$:

$$w_1 = \frac{1/100}{1/100 + 1/16} = \frac{0.01}{0.01 + 0.0625} = 0,1379,$$

$$w_2 = \frac{1/16}{1/100 + 1/16} = \frac{0.0625}{0.01 + 0.0625} = 0,8621.$$

Тобто ми об'єктивно присвоїли більший ваговий коефіцієнт точнішій оцінці прогнозу. При цьому виконується умова, $\sum w_i = 1$, що необхідно для досягнення коректності застосування методу.

Таким чином, можна зробити загальний висновок, що при прогнозуванні процесів довільної природи необхідно застосовувати як окремі методи, так і комбінування оцінок прогнозів, обчислених за допомогою різних методів. При цьому вагові коефіцієнти для окремих оцінок можна отримати різними способами, що також сприяє пошуку кращого варіанту формування прогнозів. Очевидно, що такі підходи до прогнозування краще реалізовувати у відповідних системах підтримки прийняття рішень з автоматизацією функцій обробки даних, оцінювання структур і параметрів моделей та прогнозів на їх основі.

Запитання і вправи до четвертого розділу

1. Наведіть приклади практичних ділових задач, при розв'язанні яких необхідно користуватись прогнозами? Яка різниця між прогнозом і передбаченням?
2. Наведіть приклади задач прогнозування, які виникають при прийнятті особистих рішень?
3. На які частини і в якій пропорції доцільно ділити вибірку даних при побудові моделі і оцінюванні прогнозів?
4. Що означає термін „історичний прогноз”? Чи існує необхідність у його виконанні, обґрунтуйте відповідь. Які існують інші види прогнозів?
5. Чи існує різниця між якістю прогнозу, оціненого без знаходження розв'язку і на основі розв'язку різницевого рівняння? У чому полягають переваги функції прогнозування, знайденої на основі розв'язку?
6. Побудуйте функції прогнозування на основі розв'язків таких рівнянь:
 - $y(k) = 0,35 - 0,75 y(k - 1) + 0,25 y(k - 2) + \varepsilon(k)$;
 - $y(k) = a_0 + 1,5y(k - 1) - 0,5y(k - 2) + \varepsilon(k)$;
 - $y(k) = a_0 + y(k - 2) + \varepsilon(k)$;
 - $y(k) = a_0 + 2y(k - 1) - y(k - 2) + \varepsilon(k)$;
 - $y(k) = a_0 + y(k - 1) + 0,25y(k - 2) - 0,25y(k - 3) + \varepsilon(k)$.
7. Запишіть модель (поліном) кубічного тренду + AP(2), поясніть які прогнози можна оцінити за допомогою цієї моделі?
8. Які асимптотичні статистичні характеристики має похибка прогнозу, оціненого за допомогою рівнянь АРКС? Поясніть на прикладі.
9. Поясніть суть методу прогнозування з мінімальною дисперсією? Наведіть приклад його доцільного застосування.
10. Порівняйте дисперсію похибки прогнозу, оціненого на основі розв'язку рівняння, з дисперсією прогнозу, отриманого за методом мінімальної дисперсії.

11. Чому для оцінювання якості прогнозу необхідно використовувати множину критеріїв? Поясніть суть середньої і середньої абсолютної похибки в процентах?
12. Який недолік має використання середньоквадратичної похибки для оцінювання якості оцінок прогнозів?
13. Поясніть суть коефіцієнта Тейла? Яке значення приймає цей коефіцієнт у випадку, коли модель має високі характеристики якості прогнозу?
14. Які складові має коефіцієнт Тейла? Як їх можна інтерпретувати?
15. Чому модель, яка має високі апроксимаційні характеристики (тобто з високою точністю описує дані), може виявитись непридатною для знаходження високоякісних оцінок прогнозу?
16. Які основні статистичні характеристики повинна зберігати модель процесу?
17. У яких випадках доцільно комбінувати (усереднювати) оцінки прогнозів, отримані за допомогою різних методів?
18. Чи можливе коректне (з точки зору теорії оцінювання) багатокрокове прогнозування?
19. Яким чином можна врахувати якісні змінні при прогнозуванні за допомогою регресійних моделей?
20. Назвіть відомі вам із спеціальної літератури методи, які забезпечують врахування якісних змінних у прогнозуючій моделі?
21. Які функції стосовно обробки статистичних (експериментальних) даних можна автоматизувати за допомогою відповідної системи підтримки прийняття рішень? Якими можуть бути потенційні вигоди від автоматизації процесів обробки даних, моделювання і прогнозування?

Розділ 5

МЕТОДИКА ПОБУДОВИ МАТЕМАТИЧНИХ МОДЕЛЕЙ ЧАСОВИХ РЯДІВ

Розглянемо методику побудови математичних моделей, яка більше орієнтована на побудову моделей фінансово-економічних, соціальних та екологічних процесів, для яких, як правило, набагато складніше поставити експеримент (або неможливо взагалі) та отримати інформативні експериментальні дані у достатньому об'ємі. У наведеному вигляді ця методика також може бути успішно застосована до побудови моделей динаміки технічних систем і технологічних процесів при належному плануванні та реалізації експерименту з метою збору даних, необхідних для побудови високоякісних математичних моделей.

Основи методики побудови моделей і аналізу часових рядів запропоновані Боксом і Дженкінсом у роботі [7]. Модифікована авторами методика побудови математичної моделі процесу з використанням даних у вигляді часового ряду і часового перерізу, складається з таких кроків:

– Виконання аналізу процесу, для якого будується модель, на основі спеціальних літературних джерел, експертних оцінок протікання процесу, візуального дослідження графіків вимірів вхідних і вихідних змінних, представлених часовими рядами або часовими перерізами, та іншої доступної інформації.

– Попередня обробка експериментальних даних з метою їх приведення до форми, найбільш придатної для оцінювання параметрів (коефіцієнтів) моделі.

– Аналіз часових рядів на можливу наявність нестационарності і нелінійності за допомогою множини відповідних статистичних критеріїв (статистик).

– Вибір (оцінювання) структури моделей-кандидатів, для чого необхідно виконати такі дії: (1) обчислити та виконати аналіз кореляційної матриці для часових рядів залежної та незалежних змінних з метою визначення тих екзогенних змінних, які необхідно включити в модель; (2) обчислити автокореляційну (АКФ) та часткову автокореляційну функції (ЧАКФ) залежної змінної з метою визначення оцінки порядку авторегресійної частини моделі та ковзного середнього; (3) оцінити характеристики інших елементів структури математичної моделі, що буде розглянуто нижче.

– Вибрати метод (методи) оцінювання параметрів математичних моделей вибраних структур і обчислити оцінки векторів їх параметрів. Найчастіше це метод найменших квадратів (МНК), метод максимальної правдоподібності (ММП) та їх модифікації (рекурсивні та нелінійні). В окремих випадках застосовують метод Монте Карло для марковських ланцюгів (МКМЛ), який придатний для оцінювання нелінійних моделей.

– Вибрати кращу з оцінених моделей-кандидатів за допомогою множини статистичних критеріїв адекватності (якості) моделі. Застосувати модель до розв'язання основної задачі – прогнозування, синтезу системи керування або поглибленого дослідження процесу і остаточно встановити її придатність.

Тепер розглянемо докладніше кожний з етапів побудови моделі.

5.1 Аналіз процесу

Аналіз процесу – це надзвичайно важливий етап побудови моделі, коректне виконання якого потребує досвіду дослідження реальних процесів різної природи. Ігнорування цього етапу призводить до неможливості побудови моделі високого ступеня адекватності процесу та її придатності для розв'язання задач, згаданих вище. Аналіз процесу спрямовується на розв'язання таких задач:

- визначення кількості входів і виходів, тобто визначення розмірності процесу; як правило, вимірність визначається кількістю виходів процесу, кожний із яких описують окремим рівнянням;
- встановлення логічних зв'язків між змінними та аналіз можливостей їх спільного математичного опису (коректного об'єднання в одному математичному виразі); для цього необхідно використати всю наявну інформацію про процес із спеціальної літератури, наукових звітів, від

експертів, а також у результаті спостережень за функціонуванням процесу (у тих випадках, коли це можливо);

- визначення кількості зовнішніх збурень та їх типу (детерміноване чи стохастичне) та попереднє встановлення можливості їх статистичного опису за допомогою конкретних типів розподілів випадкових величин;
- встановлення можливості декомпозиції процесу на окремі підпроцеси, які є простішими як з точки зору їх функціонування, так і з точки зору математичного опису; декомпозиція – це досить складний процес, який ґрунтується на спеціальних математичних методах;
- якщо процес має ієрархічну структуру (верхній та нижній рівень функціонування, або більшу кількість рівнів), то необхідно чітко розмежувати ці рівні, визначити функції кожного з них і встановити які типи зв'язків існують між ними; наприклад, технологічні процеси часто можна розмежувати на два і більше рівнів, які зв'язані між собою інформаційними потоками, за логічними ознаками, методами досягненням кінцевої мети і т. ін.;
- використання знань із спеціальної літератури стосовно особливостей функціонування процесу, відомих законів та закономірностей його протікання, виявлення існуючих моделей процесу та досвіду його теоретичного чи експериментального дослідження;
- за наявності раніше розроблених моделей досліджуваного процесу необхідно встановити їх недоліки та переваги, а також визначити можливість подальшого використання (модифікації); аналіз і використання існуючих моделей надає можливість дослідникам суттєво скоротити час та матеріальні витрати на побудову і використання моделі.

Отриману інформацію максимально використовують для попереднього оцінювання структури моделі або декількох моделей-кандидатів, параметри яких оцінюють за допомогою експериментальних даних. При виконанні аналізу функціонування досліджуваного процесу доцільно використовувати та порівнювати інформацію з різних джерел. Це особливо стосується фінансово-економічних процесів, щодо функціонування яких може надходити інформація з суттєвими протиріччями, пропусками, похибками і затримкою.

5.2 Попередня обробка даних

Процес попередньої обробки експериментальних (статистичних) даних складається, як правило, з таких операцій:

- нормування та візуальна перевірка даних i , 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 рівнів викидів (екстремальних імпульсних значень), що виходять за основний (усталений) діапазон значень змінних;
- бутстреп аналіз з метою збільшення об'ємів вибірок (розмноження вибірки);
- заміна некоректних вимірів інтерпольованими, усередненими або коректно згенерованими значеннями;
- формування перших та різниць вищих порядків, які необхідні для аналізу відповідних складових процесу, представленого часовим рядом;
- ортогональні перетворення і цифрова фільтрація даних з метою вилучення шумових складових.

Поширеним методом нормування даних є їх логарифмування з наступним формуванням додаткових часових рядів з перших чи других різниць. Нагадаємо, що перші різниці представляють собою наближений дискретний аналог першої похідної, а другі різниці – другої похідної. Використання різниць дає можливість будувати моделі для швидкості та прискорення основної змінної. Часто із значень ряду віднімають його середнє для того щоб отримати можливість працювати з відхиленнями, а не повними значеннями змінних. Такий підхід застосовують, наприклад, при побудові моделей у просторі станів з їх подальшим використанням для оптимальної фільтрації або оптимального керування процесом.

– Досить хороші результати нормування при оцінюванні множинної регресії

$$y(k) = \beta_0 + \beta_1 x_1(k) + \beta_2 x_2(k) + \dots + \beta_p x_p(k) + \varepsilon(k) \quad (5.2.1)$$

можна досягти завдяки одночасному нормуванню і центруванню даних таким чином:

$$x_{iH}(k) = \frac{x_i(k) - \bar{x}_i}{\sqrt{S_i}} = \frac{x_i(k) - \bar{x}_i}{\left(\sum_{k=1}^N (x_i(k) - \bar{x}_i)^2\right)^{1/2}}, \quad k = 1, \dots, N, \quad i = 1, \dots, p;$$

$$y_H(k) = \frac{y(k) - \bar{y}}{\sqrt{S_y}} = \frac{y(k) - \bar{y}}{\left(\sum_{k=1}^N (y(k) - \bar{y})^2\right)^{1/2}}, \quad (5.2.2)$$

де $x_i(k)$ – значення i -го стовпчика матриці вимірів (виміри незалежних змінних); $y(k)$ – виміри залежної змінної; $x_{iH}(k)$, $y_H(k)$ – нормовані значення

змінних; \bar{x}_i, \bar{y} – вибіркові середні значення незалежних і залежної змінних, відповідно; N – кількість вимірів; p – кількість незалежних змінних (регресорів) x_i . Якщо ввести позначення для центрованих змінних

$$\tilde{x}_i(k) = x_i(k) - \bar{x}_i; \quad \tilde{y}(k) = y(k) - \bar{y}, \quad (5.2.3)$$

то регресія для центрованих змінних матиме вигляд:

$$\tilde{y}(k) = \beta_1 \tilde{x}_1(k) + \beta_2 \tilde{x}_2(k) + \dots + \beta_p \tilde{x}_p(k) + \varepsilon(k). \quad (5.2.4)$$

Якщо підставити (5.2.2) в (5.2.4), то рівняння множинної регресії прийме вигляд:

$$y_H(k) S_y^{1/2} = \beta_1 S_1^{1/2} x_{1H}(k) + \beta_2 S_2^{1/2} x_{2H}(k) + \dots + \beta_p S_p^{1/2} x_{pH}(k) + \varepsilon'(k). \quad (5.2.5)$$

Тепер поділимо ліву і праву частини на $S_y^{1/2}$:

$$y_H(k) = \alpha_1 x_{1H}(k) + \alpha_2 x_{2H}(k) + \dots + \alpha_p x_{pH}(k) + \varepsilon''(k), \quad (5.2.6)$$

де $\alpha_1 = \beta_1 (S_1 / S_y)^{1/2}, \dots, \alpha_p = \beta_p (S_p / S_y)^{1/2}$. Отримане рівняння (5.2.6) – це рівняння для нормованих вимірів.

В результаті центрування і нормування покращується ступінь обумовленості матриці вимірів, яка вимірюється відношенням:

$$\eta = \left| \frac{\lambda_{\max}}{\lambda_{\min}} \right|,$$

де $\lambda_{\max}, \lambda_{\min}$ – максимальне і мінімальне власні числа матриці вимірів. Для забезпечення належних умов оцінювання параметрів необхідно задовольнити умову: $\eta < 10$. Іноді використовують таку умову: $\eta < m$, де m – кількість регресорів у правій частині моделі.

Застосування того чи іншого методу підготовки даних для моделювання визначається у кожному випадку по-своєму.

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 або замінюють згенерованими з розподілу, до якого відносяться дані.

Екстремальні значення – це правильно виміряні (зібрані) дані, які характеризують фактичні раптові (стрибкоподібні) зміни процесу. Підхід до розв’язання задачі дослідження 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 оцінок статистичних параметрів і моделей), то виникає задача – як правильно виконати обробку даних. Так, спираючись на критерій для визначення екстремальних значень, можна визначити величину зміщення оцінок параметрів.

Критерії аналізу екстремальних значень застосовують з метою:

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

Класичний підхід до виявлення аномальних спостережень полягає в тому, що вибіркові спостереження розглядають як випадкові, нормально розподілені величини. При цьому для аналізу (виявлення екстремальних значень) формується статистика (статистичний тест, який ґрунтується на статистичних даних), яка є чутливою до різких відхилень такого типу. Необхідно встановити розподіл цієї статистики при нульовій гіпотезі, що всі спостереження належать нормальній сукупності, а потім відхилити цю гіпотезу, якщо виявиться, що обчислена статистика їй протирічить.

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

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

Якщо виділити одне значення із спостережень, то вибіркове середнє для спостережень, що залишились, визначається як

$$\sum_{\substack{k=1 \\ k \neq i}}^N \frac{x_i}{N-1} = \bar{x} - \frac{\tilde{x}}{N-1}. \quad (5.2.7)$$

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

$$\bar{x} - \frac{\tilde{x}_1 + \tilde{x}_2 + \dots + \tilde{x}_r}{N-r}. \quad (5.2.8)$$

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 – вибіркоче 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 таким виразом:

$$\sqrt{\frac{N c^2 (v + v_0 - 1)}{v(v + v_0 - \frac{N c^2}{v})}} \approx t_{1-\alpha/2}^{v_0+v-1}, \quad (5.2.9)$$

де $v = N - 1$; α – рівень значущості; v_0 – будь-яке інше число додаткових ступенів свободи, яке зв'язане з оцінюванням σ_x^2 за вибіркою, об'єм якої не дорівнює N ($v_0 = 0$, якщо такої інформації немає). Також існує наближений вираз для c через розподіл F :

$$c \approx \left(\frac{v}{N}\right)^{1/2} \left(\frac{3F_{1-q}}{1 + [3F_{1-q} - 1]/(v + v_0)}\right)^{1/2}, \quad (5.2.10)$$

де $q = \Delta \hat{\sigma}_x^2 \frac{v}{N}$; $\Delta \hat{\sigma}_x^2$ – очікуваний приріст дисперсії внаслідок появи екстремальних значень. При використанні (5.2.10) значення c визначається додатнім значенням квадратного кореня.

Виразом (5.2.10) можна скористатись наступним чином: якщо з ряду значення не видалялись, то допустимий (очікуваний) відносний приріст дисперсії (“премію”) $\Delta \hat{\sigma}_x^2$ необхідно помножити на v/N і, таким чином, отримаємо q . За його допомогою знайдемо відповідну верхню процентну точку для відношення дисперсій F_{1-q} при трьох і $v + v_0 - 1$ ступенях свободи. За виразом (5.2.10) обчислимо значення c і застосуємо критерій до x_m .

Очікуваний відносний приріст дисперсії (“премію”) 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 ступенях свободи $F_{1-q} = F_{1-0,05} = 9,28$. Тепер знайдемо значення c :

$$c = (0,75)^2 \left(\frac{3F_{0,95}}{1 + (3F_{0,95} - 1)/3} \right)^{1/2} = 1,449.$$

Спостереження x_m необхідно видалити, якщо $|\tilde{x}_m| > 1,449S_x$. Можливі інші підходи до аналізу екстремальних значень.

Приклад 5.1. Обчислення критерію для виявлення екстремального значення. 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 (4.2.9) маємо:

$$\left(\frac{15c^2}{16 - 5c^2} \right)^{1/2} = 2,776^3.$$

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

$$|1,6| > 1,49 \cdot 0,77 = 1,05,$$

тобто спостереження x_5 видаляється.

Видалення (фільтрація) шумів з вимірів

При необхідності застосовують *фільтрацію даних від шумів*. Для розв’язання цієї важливої задачі застосовують *цифрові* або *оптимальні* фільтри. Обидва методи фільтрації широко застосовуються у технічних системах, а також знаходять застосування в системах обробки економічних, фінансових та інших типів даних. Розглянемо схематично ці можливості попередньої обробки даних.

– **Цифровий фільтр** (ЦФ) можна представити, наприклад, рівнянням типу АР(p):

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + \dots + a_p y(k-p).$$

Такий фільтр має амплітудно-частотну характеристику (АЧХ), яка визначається значеннями коефіцієнтів рівняння. Мета застосування ЦФ: пропустити корисну частину спектру і затримати шумову або просто непотрібну для аналізу складову. Сьогодні існують високорозвинені методи оптимізаційного проектування ЦФ, які дозволяють спроектувати ефективні структури фільтрів з частотними характеристиками заданої форми. Зокрема, корисний інструментарій для проектування ЦФ містить система Matlab.

– **Оптимальний фільтр** потребує, як правило, модель процесу, представлену у просторі станів. Так, фільтр Калмана ґрунтується саме на моделі такого типу. Розглянемо коротко її структуру.

Нехай нестационарна лінійна система описується у дискретному часі рівняннями з непостійними в часі коефіцієнтами (непостійність тобто не стаціонарність означає залежність від дискретного часу $k=0, 1, 2, \dots$):

$$\mathbf{x}(k) = \mathbf{A}(k, k-1)\mathbf{x}(k-1) + \mathbf{B}(k, k-1)\mathbf{u}(k-1) + \mathbf{w}(k),$$

де $\mathbf{x}(k)$ – n -вимірний вектор станів системи; $\mathbf{u}(k-1)$ – m -вимірний вектор детермінованих вхідних величин (сигнали керування); $\mathbf{w}(k-1)$ – n -вимірний вектор випадкових зовнішніх збурень; $\mathbf{A}(k, k-1)$ – $(n \times n)$ матриця динаміки системи (вона містить коефіцієнти, що характеризують динаміку, тобто швидкість зміни станів у часі); $\mathbf{B}(k, k-1)$ – $(n \times m)$ матриця коефіцієнтів керування. Подвійний часовий аргумент у вигляді $(k, k-1)$ означає, що величина з цим аргументом використовується в момент k , але її значення ґрунтується на попередніх даних, які відомі на момент $k-1$, включно. Далі будемо записувати для простоти представлення матриці \mathbf{A} і \mathbf{B} з одним аргументом, тобто $\mathbf{A}(k)$ та $\mathbf{B}(k)$. Очевидно, що стаціонарна система описується матрицями з постійними коефіцієнтами, які записують просто \mathbf{A} і \mathbf{B} . Оскільки матриця \mathbf{A} зв'язує поточний стан із попереднім, то її називають ще перехідною матрицею станів. Нагадаємо, що дискретний час k зв'язаний з неперервним часом t періодом дискретизації вимірів T_s : $t = kT_s$.

У класичній постановці задачі оптимальної фільтрації послідовність зовнішніх збурень $\mathbf{w}(k)$ задовольняє властивостям білого гаусового шуму з нульовим середнім значенням і коваріаційною матрицею \mathbf{Q} , тобто статистики шуму мають вигляд:

$$\begin{aligned} E[\mathbf{w}(k)] &= \mathbf{0}, \quad \forall k, \\ E[\mathbf{w}(k)\mathbf{w}^T(j)] &= \mathbf{Q}(k)\delta_{kj}, \end{aligned}$$

де δ_{kj} – дельта-функція Кронекера, що визначається так: $\delta_{kj} = \begin{cases} 0 & \text{для } k \neq j \\ 1 & \text{для } k = j \end{cases}$

$\mathbf{Q}(k)$ – додатно визначена коваріаційна матриця зовнішніх збурень стану розмірності $(n \times n)$. Діагональні елементи матриці представляють собою дисперсії компонент вектора збурень $\mathbf{w}(k)$.

Початковим станом системи \mathbf{x}_0 будемо вважати випадкові змінні з відомими статистиками:

$$E[\mathbf{x}_0] = \bar{\mathbf{x}}_0; \quad E[\mathbf{x}_0 \mathbf{x}_0^T] = \mathbf{M}; \quad E[\mathbf{w}(k) \mathbf{x}_0^T] = 0, \quad \forall k.$$

Нехай вектор вимірів $\mathbf{z}(k)$ вихідних змінних доступний в будь-який момент часу t_k , а його компоненти лінійно зв'язані з вектором стану і на них впливає шум вимірів, тобто

$$\mathbf{z}(k) = \mathbf{H}(k) \mathbf{x}(k) + \mathbf{v}(k),$$

де $\mathbf{H}(k)$ – матриця спостережень вимірності $(r \times n)$, $\mathbf{v}(k)$ – r -вимірний вектор випадкових величин шуму вимірів з відомими статистиками:

$$E[\mathbf{v}(k)] = 0, \quad E[\mathbf{v}(k) \mathbf{v}^T(j)] = \mathbf{R}(k) \delta_{kj},$$

де $\mathbf{R}(k)$ – додатно визначена коваріаційна матриця шумів вимірів вимірності $(r \times r)$, діагональні елементи якої є дисперсіями адитивного шуму у кожному каналі вимірів. Шум вимірів також задовольняє властивостям білого гаусового шуму. Він вважається некорельованим із зовнішнім збуренням $\mathbf{w}(k)$ і початковим станом системи тобто:

$$E[\mathbf{v}(k) \mathbf{w}^T(j)] = 0 \quad \forall k, j;$$

$$E[\mathbf{v}(k) \mathbf{x}_0^T] = 0 \quad \forall k.$$

Для визначеної вище системи з вектором стану $\mathbf{x}(k)$ необхідно знайти оцінку стану $\hat{\mathbf{x}}(k)$ в момент t_k як лінійну комбінацію оцінки $\hat{\mathbf{x}}(k-1)$ в момент t_{k-1} і самого останнього виміру (статистичних даних) $\mathbf{z}(k)$.

Оцінка $\hat{\mathbf{x}}(k)$ повинна обчислюватися як найкраща за мінімумом середнього значення суми квадратів похибок оцінок. Іншими словами, оцінка повинна бути такою, щоб

$$E[(\hat{\mathbf{x}}(k) - \mathbf{x}(k))^T (\hat{\mathbf{x}}(k) - \mathbf{x}(k))] = \min_K,$$

де $\mathbf{x}(k)$ – точне значення вектора стану, яке може бути обчислене за допомогою детермінованої складової математичної моделі процесу (тобто без врахування випадкових складових – збурень стану і шумів вимірів); \mathbf{K} –

оптимальний матричний коефіцієнт фільтра, який необхідно обчислити в результаті розв'язання оптимізаційної задачі. Основне рівняння оптимального фільтра має вигляд:

$$\hat{\mathbf{x}}(k) = \mathbf{A}(k)\hat{\mathbf{x}}(k-1) + \mathbf{B}(k)\mathbf{u}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}\mathbf{A}(k)\hat{\mathbf{x}}(k-1)],$$

тобто до детермінованої частини оцінки стану додається корегуючи величина, яка визначається добутком коефіцієнта фільтра на інновацію (у квадратних дужках): різницею між поточним виміром і прогнозом значення оцінки на один крок.

Таким чином, фільтр необхідно будувати та використовувати для уточнення оцінок стану процесу в умовах впливу випадкових зовнішніх збурень та наявності шумів (похибок) вимірів. На сьогодні оптимальні фільтри – це практично обов'язкова складова комп'ютерних систем обробки експериментальних і статистичних даних. Докладніше оптимальний фільтр буде розглянуто нижче.

5.3 Аналіз наявності нелінійностей

Однією з проблем при визначенні структури моделі є встановлення факту наявності нелінійностей в досліджуваному процесі та їх типу. Для розв'язання цієї проблеми обов'язково використовують *візуальний аналіз* даних та формальні тести на наявність нелінійностей. Досвідченому фахівцю з моделювання візуальний аналіз дозволяє оперативно виявити наявність участків з лінійним або нелінійним трендом, в якійсь мірі наявність гетероскедастичності та значних імпульсних викидів (екстремальних значень), які можуть суттєво впливати на якість моделі. Необхідно зазначити, що існують окремі навчальні курси з візуального аналізу даних. Це свідчить про те, що не варто нехтувати такою доступною, але прискореною та ефективною можливістю дослідження даних. Психологи встановили, що один інформативний рисунок може замінити до двох тисяч слів.

Існує також ряд формальних тестів на наявність нелінійності. Розглянемо простий тест для визначення наявності нелінійності. Цей тест можна застосувати у випадку, коли можна набрати декілька груп (реалізацій) спостережень для одного і того ж процесу:

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

де \bar{y}_i – середнє значення для i -ї реалізації (вибірки або групи) даних; \hat{y}_i – середнє для лінійної апроксимації даних; m – кількість груп даних; n_i – кількість вимірів в i -й групі; n – загальна кількість вимірів. Фактично, наведена статистика представляє собою таке відношення:

$$\hat{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{Відхилення значень } y(k) \text{ від групових середніх}}.$$

Якщо статистика \hat{F} з $\nu_1 = m - 2$ та $\nu_2 = n - m$ степенями свободи досягає або перевищує рівень значущості, то гіпотезу стосовно лінійності процесу необхідно відхилити. Недоліком даного підходу є те, що для його застосування необхідно мати декілька (не менше трьох) груп (реалізацій) даних для одного і того ж процесу, які можна отримати в результаті виконання повторних експериментів. Очевидно, що це не завжди можливо.

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

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

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

Наявність нелінійного детермінованого тренду у процесі можна визначити шляхом оцінювання рівняння:

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

яке представляє собою поліном порядку m відносно часу; $\varepsilon(k)$ – випадковий процес, причини появи якого будуть розглянуті нижче у даному розділі. Якщо хоча б один із коефіцієнтів моделі, c_i , $i = 1, \dots, m$, є статистично значущим, то гіпотеза щодо відсутності тренду відхиляється. Якщо тренд відносно швидко змінює свій напрямок руху і для нього важко знайти адекватний функціональний опис, то застосовують моделі випадкових трендів, які ґрунтуються на комбінаціях випадкових величин. Наявність нелінійності стосовно регресора $x(k)$ можна встановити за допомогою відповідного полінома:

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

Автоматично оцінює структуру математичної моделі *метод групового врахування аргументів* (МГВА), запропонований академіком Івахненком О.Г. (Київський інститут кібернетики) [11]. Цей метод вже багаторазово застосовували до опису широкого класу процесів з метою оцінювання прогнозів та створення систем керування. Його успішно застосовують і сьогодні до моделювання процесів різної природи з нелінійностями та нестационарностями. Подальшим розвитком даного методу є нечіткий МГВА, який ґрунтується на нечіткому представленні параметрів оцінюваної моделі. Загалом задача встановлення наявності та визначення типу нелінійності у більшості практичних випадків залишається предметом дослідження.

1.4 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; імпульсні та неперервні); *обмеження* на параметри і змінні моделі.

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, дає можливість 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. У 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 вищою ступінню 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 :

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

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

$$y = f(x, z),$$

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

$$y(k) = a_0 + a_1x(k) + a_2z(k) + \varepsilon(k),$$

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 формально корельовані із залежною змінною, але фактично вони не впливають на неї;

– для будь-якого методу оцінювання параметрів рівнянь властиві методичні обчислювальні похибки, які повинні бути по можливості враховані у 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

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

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

Для обчислення елементів матриці \mathbf{R} необхідно мати синхронізовані у часі вибірки значень всіх трьох змінних y, x, z . Формула для розрахунку вибірових коефіцієнтів кореляції має вигляд:

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

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,

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

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

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

Очевидно, що перед формальним обчисленням коефіцієнтів кореляції необхідно зробити аналіз процесу і визначити наявність (або відсутність) логічних зв'язків між змінними. Це дозволяє обмежитись розглядом тільки тих змінних, які дійсно впливають на залежну змінну, наприклад, на показник якості продукції. Очевидно, що для правильного вибору незалежних (екзогенних або пояснюючих) змінних необхідно глибоко знати технологічний або інший процес, що моделюється.

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

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

яке може бути представлено в загальному вигляді так:

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

Можна показати, що між коефіцієнтами регресії b_1, b_2 і коефіцієнтами кореляції r_{yx}, r_{yz} існує однозначний аналітичний взаємозв'язок. Емпіричні дослідження свідчать, що незалежні змінні необхідно включати в регресію, якщо $|r_{yx}| \geq 0,4$. Однак, необхідно пам'ятати, що ця рекомендація має досить наближений характер. Існують випадки, коли у модель необхідно включати регресори, які мають менші значення коефіцієнта кореляції із залежною змінною.

Останнє рівняння представляє собою *лінійну регресію*, яка містить p параметрів (коефіцієнтів), але досить часто необхідно застосовувати складніші нелінійні моделі. Характерним представником нелінійної стосовно змінних регресії є поліноміальна регресія порядку $p - 1$:

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

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

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

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

тобто, в рівняння регресії введено авторегресійну (АР) складову другого порядку. Порядок авторегресії визначається за допомогою автокореляційної функції (АКФ) і часткової АКФ (ЧАКФ). Кількість значень автокореляційної функції, які відмінні від нуля в статистичному сенсі і визначаються параметром зсуву s (що використовується при обчисленні кореляційних функцій), і буде складати оцінку порядку авторегресії.

Природа авторегресії пояснюється існуванням так званої “пам’яті” процесу, яка проявляється у тому, що його поточний стан може в значній мірі визначатись попередніми станами. Наприклад, стан людини вранці залежить від того, яким було самопочуття ввечері та у попередні дні. На формування ринкових цін суттєво впливають значення цін у попередні періоди часу. Очевидно, що поточний рівень валового внутрішнього продукту (ВВП) залежить від його попередніх значень, а поточний стан технічної системи або технологічного процесу також залежить від їх стану у попередні моменти часу.

АКФ та ЧАКФ використовують для визначення попередньої оцінки порядку авторегресійної частини моделі, тобто скільки затриманих в часі значень необхідно брати для опису процесу. При цьому необхідно врахувати, що ЧАКФ дає «чіткішу» оцінку порядку процесу ніж АКФ. Наприклад, для процесу АР(1) значення основної змінної та будуть корельованими, незважаючи на те, що не наявне у моделі. Кореляція між i , тобто $y(i)$, дорівнює коефіцієнту кореляції між значеннями i помноженому на коефіцієнт кореляції між i або Подібні «непрямі» кореляції наявні в АКФ будь-якого

процесу авторегресії.

Вибіркова АКФ обчислюється за виразом:

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

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

Кількість значень АКФ, відмінних від нуля в статистичному смислі, вказує на порядок авторегресійної частини моделі. Для стаціонарного процесу (це процес із постійними середнім, дисперсією та коваріацією) коефіцієнти $r_y(s)$ мають нормальний розподіл та нульове середнє.

На відміну від АКФ, часткова АКФ між значеннями та виключає вплив величин $y(k-1) \dots y(k-s+1)$, а це означає, що коефіцієнти ЧАКФ чіткіше відображають зв'язок між окремими значеннями основної змінної. Так, для процесу АР(1) ЧАКФ між та дорівнює нулю за визначенням, що підтверджується обчисленими значеннями ЧАКФ. Для того щоб знайти попередню оцінку порядку моделі, вибіркові коефіцієнти ЧАКФ (тобто, коефіцієнти, знайдені за вибіркою даних) можна обчислити також за допомогою простого методу, який полягає в наступному.

а) Обчислюють додатковий часовий ряд з відхилень основної змінної:

$$\{y'(k)\} = \{y(k)\} - \mu,$$

де μ – середнє значення ряду.

б) Формують рівняння першого порядку:

$$y'(k) = \Phi_{11} y'(k-1) + e(k),$$

де $e(k)$ – похибка моделі. У такому рівнянні Φ_{11} відіграє роль коефіцієнта АКФ та ЧАКФ між $y(k)$ та $y(k-1)$. Для оцінювання двох коефіцієнтів можна сформулювати рівняння другого порядку:

$$y'(k) = \Phi_{11} y'(k-1) + \Phi_{22} y'(k-2) + e(k),$$

де Φ_{22} – коефіцієнт ЧАКФ між $y(k)$ та $y(k-2)$.

Таким чином, дискретні значення ЧАКФ можна обчислити за допомогою значень АКФ, використовуючи такі вирази [41]:

$$\Phi_{11} = r(1), \quad \Phi_{22} = \frac{r_2 - r_1^2}{1 - r_1^2};$$

$$\Phi_{ss} = \frac{r_s - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_{s-j}}{1 - \sum_{j=1}^{s-1} \Phi_{s-1,j} r_j}.$$

У загальному випадку коефіцієнти ЧАКФ стаціонарного процесу АРКС(p, q) повинні збігатися до нуля, починаючи із p -го значення. АКФ процесу АРКС(p, q) починається збігатися до нуля при значеннях часового зміщення

Коли ми говоримо, що значення 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 значення АКФ суттєво відмінні від нуля в статистичному сенсі, є обчислення та аналіз значущості статистичного параметра (або просто статистики) Льюнга-Бокса $Q(r_k)$ за формулою [41]:

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

де N – довжина вибірки даних змінної, для якої обчислено значення автокореляційної функції r_k ; s – кількість відліків 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 значущості 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 ступеня близькості залишків моделі до білого шуму. Однак, необхідно пам'ятати, що при обчисленні s значень кореляційної функції кількість ступенів свободи зменшується на кількість 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 – число кандидатів. Кандидатів може бути декілька, оскільки встановити структуру точно за один раз, як правило, неможливо. Загалом оцінювання структури моделі високого ступеня адекватності – це трудомісткий ітераційний процес, який вимагає поглиблених знань, досвіду моделювання та значних зусиль. На наступному етапі оцінюють числові значення параметрів знайдених моделей-кандидатів.

5.5 Оцінювання параметрів моделей-кандидатів

На *четвертому етапі* оцінюють параметри (коефіцієнти) рівняння, використовуючи принцип економії або збереження. Цей принцип означає, що *кількість коефіцієнтів, що оцінюються, не повинна перевищувати їх необхідну кількість* (“необхідність” можна встановити, 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. Наприклад, рівняння авторегресії першого порядку:

$$y(k) = 0,5y(k-1) + \varepsilon(k),$$

еквівалентне ковзному середньому у вигляді

$$y(k) = \varepsilon(k) + 0,5\varepsilon(k-1) + 0,25\varepsilon(k-2) + 0,125\varepsilon(k-3) + \dots$$

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 співпадати 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

$$y(k) = Y(k) - \bar{y},$$

де значення виміру, \bar{y} – середнє значення ряду. Якщо для оцінювання параметрів використовується рекурсивна процедура, то поточне середнє

можна обчислювати за формулою:

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

Найбільш поширеними методами оцінювання параметрів моделі є такі:

1. *Метод найменших квадратів* (МНК);
2. *Метод максимальної правдоподібності* (ММП),
3. *Метод допоміжної (інструментальної) змінної* (МДП),
4. *Нелінійний метод найменших квадратів* (НМНК),
5. *Метод Монте Карло для марковських ланцюгів* (МКМЛ);

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

$$\hat{\theta} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y},$$

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

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

матриця вимірів має вигляд:

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

Одиниці у 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:

$$\mathbf{X}' = \begin{bmatrix} N & \sum_{k=1}^N x(k) & \sum_{k=1}^N x^2(k) & \dots & \sum_{k=1}^N x^p(k) \\ \sum_{k=1}^N x(k) & \sum_{k=1}^N x^2(k) & \sum_{k=1}^N x^3(k) & \dots & \sum_{k=1}^N x^{p+1}(k) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \sum_{k=1}^N x^p(k) & \sum_{k=1}^N x^{p+1}(k) & \sum_{k=1}^N x^{p+2}(k) & \dots & \sum_{k=1}^N x^{2p}(k) \end{bmatrix}.$$

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

$$\hat{\theta} = [\mathbf{X}']^{-1} \mathbf{y}', \text{ де } \mathbf{y}' = \left[\sum_1^N y(k) \quad \sum_1^N x(k)y(k) \quad \dots \quad \sum_1^N x^p(k)y(k) \right]^T; \hat{\theta} - \text{вектор}$$

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 θ лінійної регресійної математичної моделі, наприклад, моделі змішаної регресії:

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

за допомогою методу найменших квадратів необхідно задовольнити наступні умови:

а) $\varepsilon(k)$ – некорельована послідовність однаково (нормально) розподілених 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 (апостеріорно), а до оцінювання (апріорно) можна тільки постулювати їх виконання. Тобто після оцінювання параметрів моделі оцінка значень випадкового процесу визначається похибками моделі:

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

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

5.6 Діагностика моделей – вибір кращої з множини оцінених кандидатів

На *n'*ятому етапі аналізується якість моделі, тобто виконується перевірка оцінених кандидатів на адекватність процесу. Діагностика складається із кроків, наведених нижче.

а) *Візуальне дослідження графіка похибок моделі* $e(k) = y(k) - \hat{y}(k)$, де оцінка змінної, отримана 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). У випадку застосування рекурсивних методів оцінювання найбільші похибки будуть у перехідному процесі, коли інформаційна матриця ще не містить достатньо інформації про процес. Однак для моделей низьких порядків (1-2) перехідний процес повинен закінчуватись через 20-30 кроків (не більше), а оцінки параметрів повинні прямувати до точних значень.

Для аналізу характеристик похибок необхідно застосовувати статистичні характеристики (статистики), які свідчать про близькість цього випадкового процесу до білого шуму. Це коефіцієнт асиметрії, ексцес, статистика Жак-Бера.

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

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

$$DW = 2 - 2\rho,$$

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 і Watson) запропонували скористатись для перевірки корельованості похибок моделі таким виразом:

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

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

Розкриваючи квадрат різниці у чисельнику, отримаємо:

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

де

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

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 (не більше) ітерацій алгоритму оцінювання. Якщо кількість ітерацій набагато перевищує вказані числа, то це свідчить про те, що процес може бути нестационарним.

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

$$t = \frac{\hat{a} - a^0}{SE_{\hat{a}}},$$

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 значущості оцінки можна висувати будь-яку: що коефіцієнт значущий, тобто ($H_0: a^0 \neq 0$), або незначущий ($H_0: a^0 = 0$). Статистична теорія перевірки гіпотез пропонує висувати нуль-гіпотезу, яка є протилежною бажаному результату. У даному випадку бажаним результатом є статистична значущість коефіцієнтів математичної моделі. Таким чином, необхідно висувати нульову гіпотезу, що коефіцієнт незначущий. Це дає можливість коректно підійти до визначення значущості оцінок коефіцієнтів та дещо спростити розрахунки.

Для того щоб встановити, чи оцінка коефіцієнта значуща, необхідно знати довжину вибірки даних N (потужність вибірки); кількість ступенів свободи $f = N - n$, де n – число коефіцієнтів моделі, які оцінюються на основі ряду даних, і вибрати рівень значущості $\alpha = 1\%$, або $\alpha = 5\%$, або $\alpha = 10\%$ (для цих значень існують розраховані таблиці критичних значень t -статистики). Фактично рівень значущості означає ймовірність припуститись помилки першого роду при перевірці гіпотези. Згадаємо, що

$$\alpha = p\{X \in G/\omega | H_0\} = \int_{n-m(G/\omega)} L_{H_0}(X) dx,$$

де $X = [x_1, \dots, x_n] \in R^n$ – вся вибірка, яка розбивається на дві множини, що перетинаються: ω і G/ω (ω – область прийняття нуль-гіпотези); G/ω – критична область: якщо $X \in G/\omega$, то H_0 відхиляється; $L_{H_0}(X)$ – закон розподілу X . Помилка першого роду означає відхилення коректно сформульованої 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

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

то нуль-гіпотеза стосовно незначущості 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 статистично значущим.

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

- сформулювати нуль-гіпотезу стосовно значущості 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:

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

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

Очевидно, що найкращим значенням є $R^2 = 1$, тобто коли дисперсії вимірів змінної та оцінок цієї ж змінної, отриманих за моделлю, збігаються. Цей параметр можна трактувати, також, як міру інформативності моделі, якщо за міру інформативності вибрати 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

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

у порівнянні з усіма іншими моделями.

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

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

та критерій Байєса-Шварца:

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

де $n = p + q + 1$ – кількість параметрів моделі, які оцінюються за допомогою статистичних даних (p – число параметрів авторегресійної частини моделі; q – кількість параметрів ковзного середнього; „1” з’являється тоді, коли оцінюється зміщення (або *перетин*, тобто a_0)).

Критерії Акайке і Байєса-Шварца містять у правій частині суму квадратів похибок, а тому за цими критеріями вибирають ту модель, для якої ці критерії приймають найменші значення. Введення нового регресора приводить до збільшення критерію (при цьому збільшується n), але одночасно 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 пропорційна відношенню:

$$F \sim \frac{R^2}{1 - R^2},$$

а для множинної (багатофакторної) регресії вона визначається за виразом

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

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 в даному випадку є припущення про те, що модель неадекватна в цілому, тобто:

$$H_0 : a_1 = a_2 = \dots = a_p = 0,$$

проти альтернативної гіпотези:

H_1 : хоча б одне значення a_i відмінне від нуля в статистичному смислі.

Значення $F_{крит}$ знаходять із таблиць для F -розподілу. Послідовність застосування цієї статистики можна представити так:

1. Сформулювати нуль-гіпотезу стосовно адекватності моделі в цілому. Наприклад, H_0 : 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. Вибрати рівень значущості $\alpha = 1\%$ або $\alpha = 5\%$ або $\alpha = 10\%$.
4. Користуючись значеннями N , f і α , знайти критичне значення $F_{крит}$ знаходять із таблиць для F -розподілу при $(p, N - p - 1)$ ступенях свободи.
5. Перевірити нуль-гіпотезу:

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

Коректне застосування запронованої методики забезпечує побудову 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 поведінку того режиму функціонування 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. Іноді інформативність формально оцінюють за допомогою величини дисперсії процесу, а також за кількістю гармонічних складових, які містяться у процесі. Чим більше гармонік містить вибірка, тим вищою є її інформативність.

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

5.7 Приклади побудови математичних моделей за допомогою статистичних даних

Приклад 5.2. Розглянемо можливості побудови математичних моделей динаміки для таких 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:

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 М3 складає приблизно 0,932 – це велике значення, яке може негативно вплинути на якість оцінок моделі при використанні цих змінних у правій частині рівняння.

5.7.1 Модель індексу споживчих цін

Авторегресійна модель індексу споживчих цін

Спочатку розглянемо можливість опису індексу споживчих цін моделлю авторегресії з ковзним середнім. Моделі парної регресії та авторегресійні моделі самі прості за своєю структурою, але досить часто вони дають можливість досягти високого ступеня адекватності досліджуваному процесу, прийнятної для подальшого застосування моделі.

Автокореляційна функція процесу наведена в таблиці 5.1. При побудові моделі індекс споживчих цін позначимо через isc . В таблиці 4.1 прийнято такі скорочення: АКФ – автокореляційна функція, а ЧАКФ – часткова АКФ.

Як видно із графіка 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 наведені в таблиці 5.2.

Таблиця 5.1. Автокореляційна функція процесу формування індексу споживчих цін (ICL)

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

Таблиця 5.2. Результати оцінювання моделі AP(1) для ICL

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):

$$\begin{aligned} isc(k) &= a_0 + a_1 isc(k-1) + \varepsilon(k) = \\ &= 43,294 + 0,57 isc(k-1) + e(k), \end{aligned}$$

де $e(k)$ – залишки (похибки) моделі, значення яких можна знайти у відповідному файлі пакету програм, що застосовується для побудови моделі. Основні статистичні характеристики якості моделі:

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

Коефіцієнт детермінації має низьке значення (0,415), сума квадратів похибок 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:

$$CeКП = 1,36; \quad САП = 1,02; \quad САПП = 1,008; \quad U = 0,0067,$$

тобто середньоквадратична похибка ($CeКП$), середня абсолютна похибка ($САП$), середня абсолютна похибка у процентах ($САПП$) і коефіцієнт Тейла U , який свідчить про загальну придатність моделі для прогнозування (його ідеальне значення – нуль).

В таблиці 5.3 наведено результати оцінювання авторегресійної моделі AP(3). Всі коефіцієнти моделі статистично значущі.

Таблиця 5.3. 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
Модель: $I=C(1)+C(2)*I(-1)+C(3)*I(-2)+C(4)*I(-3)$

	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 Шварта		3.256785
		Стат. Дарбіна-Уотсона		1.992389

Таким чином, можна записати таку модель:

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

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 значення статистики Дарбіна-Уотсона: від 1,931 до 1,992 (похибки моделі можна вважати практично некорельованими). Тобто отримані такі значення характеристик моделі:

$$R^2 = 0,317; \quad J = CKП = 135,148; \quad DW = 1,992.$$

Характеристики однокрокового прогнозу для даної моделі:

$$CeКП = 1,36; \quad САП = 1,02; \quad САПП = 1,01; \quad U = 0,0068,$$

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

Розглянемо характеристики моделі вищого порядку. В таблиці 5.4 наведені результати оцінювання моделі AP(7).

Таблиця 5.4. 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 коеф-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 AP(7):

$$CeKП = 1,36; \quad CAП = 1,012; \quad CAПП = 1,002; \quad U = 0,0067,$$

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

Результати оцінювання моделі 12-го порядку наведені в таблиці 5.5. Модель AP(12) має кращі характеристики ніж попередні моделі:

$$R^2 = 0,435; \quad J = CKП = 97,80; \quad DW = 1,94.$$

Значно зменшилась сума квадратів похибок, підвищилось значення R^2 , а значення статистики Дарбіна-Уотсона майже таке ж, як для моделі AP(3).

Таблиця 5.5. 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: $I=C(1)+C(2)*I(-1)+C(3)*I(-2)+C(4)*I(-3)+C(5)*I(-6)+C(6)*I(-7)+C(7)*I(-8)+C(8)*I(-9)+C(9)*I(-11)+C(10)*I(-12)$

	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	0.435002	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

Характеристики однокрокового прогнозу для цієї моделі такі:

$$CeKП = 1,337; \quad CAП = 1,02; \quad CAПП = 1,013; \quad U = 0,0066.$$

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 високою ступінню 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) наведені в таблиці 5.6.

Таблиця 5.6. 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

Три вибрані статистичні характеристики адекватності цієї моделі:

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

Характеристики якості однокрокового прогнозу:

$$CeKП = 1,36; \quad CAП = 1,02; \quad CAПП = 100,40; \quad U = 0,662.$$

У порівнянні з моделлю AP(1) для повних значень (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 непридатна для прогнозування. Це можна пояснити тим, що відхилення мають різні знаки, що утруднює побудову моделі.

Авторегресія 13-го порядку для відхилень ІСЦ від середнього наведена в табл. 5.7.

Таблиця 5.7. 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 коеф-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:

$$R^2 = 0,440; \quad J = CKII = 95,69; \quad DW = 1,965.$$

Характеристики якості однокрокового прогнозу:

$$CeKII = 1,346; \quad CAII = 1,02; \quad CAIII = 109,09; \quad U = 0,848.$$

Спостерігається погіршення характеристик прогнозу, особливо погіршилось значення CAIII 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 статистично незначущими, а тому їх можна вилучити з моделі.

Після вилучення цих коефіцієнтів із моделі та відповідних їм складових процесу отримаємо модель, параметри якої наведені в таблиці 5.8.

Таблиця 5.8. Результати оцінювання моделі AP(13) для відхилень

Метод оцінювання: метод найменших квадратів
 Скорегована часова вибірка даних: 1997:02 2005:01
 Всього використано спостережень після корегування крайніх значень: 96
 Модель: $IV=C(1)+C(2)*IV(-1)+C(3)*IV(-2)+C(4)*IV(-3)+C(5)*IV(-5)+C(6)*IV(-12)+C(7)*IV(-13)$

	Оцінки коеф-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:

$$R^2 = 0,439; \quad J = CKП = 96,05; \quad DW = 1,96.$$

Характеристики якості однокрокового прогнозу:

$$CeКП = 1,346; \quad CAП = 1,017; \quad CAПП = 108,2; \quad U = 0,854.$$

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 відбувається в двох квадрантах), складніше ніж для процесу, який змінюється в межах одного квадранту.

Авторегресія з ковзним середнім для ІСЦ

Розглянемо можливість описання ІСЦ за допомогою моделі АРКС. Характеристики моделі АРКС(1,1) наведені в таблиці 5.9.

Таблиця 5.9. 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.
С	100.6529	0.263645	381.7742	0.0000
AR(1)	0.546826	0.093191	5.867793	0.0000
КС(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 АР корені	.55			
Інвертовані КС корені	-.06			

Вибрані статистичні характеристики адекватності цієї моделі:

$$R^2 = 0,416; \quad J = СКП = 141,78; \quad DW = 1,996.$$

Характеристики якості однокрокового прогнозу:

$$СекП = 1,362; \quad САП = 1,016; \quad САПП = 1,005; \quad U = 0,0067.$$

Моделі АР(1) і АРКС(1,1) мають практично однакові характеристики адекватності та якості однокрокового прогнозу, а тому перевагу (при виборі з цих двох моделей) можна надати моделі АР(1), яка є простішою. Нижче наведена 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 причинний зв'язок між вибраними змінними. Характеристики змішаної моделі: авторегресія АР(1) + парна регресія наведені в таблиці 5.10.

Таблиця 5.10. Результати оцінювання моделі ІСЦ: АР(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 коеф-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 у порівнянні з АР(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.

Характеристики однокрокового прогнозу:

$$SeKП = 1,340; \quad CAП = 1,004; \quad CAПП = 0,994; \quad U = 0,0066.$$

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 стосовно об'ємів реалізації торговельних операцій у валюті підприємствами усіх форм власності. Очевидно, що отримати такі дані для тіньового обороту досить складно.

Визначення впливу ВВП на ІСЦ

Як було показано на початку цього параграфа, коефіцієнт кореляції між ІСЦ та ВВП складає - 0,304, тобто формально це невелике значення. Також логічно припустити, що зростання ВВП має приводити до зменшення ІСЦ (про це свідчить також знак коефіцієнта кореляції між цими змінними).

Характеристики змішаної моделі: авторегресія АР(1) + парна регресія наведені в таблиці 5.11.

Таблиця 5.11. Результати оцінювання змішаної регресії для ІСЦ і ВВП

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 коеф-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

Отримано таке рівняння:

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

з характеристиками:

$$R^2 = 0,419; \quad J = CKII = 141,05; \quad DW = 1,92.$$

Характеристики якості однокрокового прогнозу:

$$CeKII = 1,335; \quad CAI = 1,004; \quad CAIII = 0,993; \quad U = 0,0066.$$

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 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, у 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.

У таблиці 5.12 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 методики Бокса-Дженкінса, а також встановити можливість практичного використання цих результатів.

Таблиця 5.12. Результати моделювання і однокрокового прогнозування індексу оптових цін

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

Результати моделювання, наведені в таблиці 5.12, 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 AR(1)+ВВП. Для моделі AR(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 у 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 (за необхідності) необхідно знайти іншу структуру математичної моделі, придатну для адекватного опису різнознакових величин часового ряду.

Запитання і вправи до п'ятого розділу

1. Назвіть етапи побудови математичних моделей за методикою Бокса-Дженкінса. Що забезпечує коректне використання цієї методики на практиці?
2. Яка мета аналізу функціонування процесу? Які елементи структури моделі можна встановити за допомогою попереднього аналізу процесу на основі відомої інформації?
3. Яка мета попередньої обробки даних? Назвіть основні операції, які виконують в процесі попередньої обробки даних? До чого призводить визначення значень змінних у великому числовому діапазоні?
4. Які два основних типи нелінійностей зустрічаються в аналізі часових рядів? Який з них ускладнює процедуру оцінювання параметрів моделі? Поясніть на прикладах.
5. Яким чином можна встановити наявність нелінійностей у процесі? У яких випадках нелінійний процес можна описати лінійною моделлю?
6. Який метод дає можливість автоматизувати процес визначення та врахування нелінійностей процесу?
7. Яку інформацію можна отримати на основі візуального аналізу даних? Як можна нею скористатись?
8. Яким чином можна знайти оцінку порядку авторегресійної частини моделі?
9. У чому полягає відмінність між автокореляційною та частковою автокореляційною функціями процесу? Чи існує необхідність розрахунку обох функцій в процесі аналізу даних?
10. На чому ґрунтується відбір незалежних змінних (регресорів, екзогенних змінних) для включення у праву частину математичної моделі?
11. Назвіть три умови коректного застосування методу найменших квадратів до оцінювання параметрів математичної моделі? Як можна перевірити виконання цих умов?

12. Для чого призначена статистика Льюнга-Бокса і як вона обчислюється? Скористайтесь наявним пакетом програм для статистичного аналізу даних для обчислення цього статистичного параметру і поясніть його значущість (чи незначущість) на прикладі.
13. В чому полягає принципова різниця між методом найменших квадратів (МНК), призначеним для оцінювання лінійних моделей, та МНК для оцінювання моделей, нелінійних відносно параметрів? Які критерії якості мінімізують ці методи?
14. Виведіть формулу МНК для поліноміальної моделі?
15. Сформууйте матрицю вимірів для математичної моделі такого вигляду:

$$y(k) = a_0 + a_1 y(k-1) + b_1 x_1(k) + b_2 x_2(k) + \varepsilon(k).$$
16. Про що свідчить корельованість похибок моделі між собою? За допомогою якого статистичного параметра якості моделі можна визначити ступінь корельованості похибок? Яке значення приймає ця статистика в ідеальному випадку?
17. Що означає значущість оцінки параметра (коефіцієнта) моделі в статистичному сенсі? За допомогою якої статистики можна встановити значущість оцінки параметра моделі?
18. Що означають помилки першого і другого роду при перевірці статистичних гіпотез?
19. Поясніть фізичну суть коефіцієнта (множинної) детермінації, яке його ідеальне значення? Чому R^2 корегують із врахуванням кількості ступенів свободи?
20. У чому полягає різниця між критеріями Акайке та Байєса-Шварца? Чи можна обмежитись використанням тільки одного з цих критеріїв?
21. Яка мета розрахунку статистики Фішера? Наведіть послідовність розрахунку та аналізу цієї статистики.
22. Сформулюйте правила перевірки гіпотез для t – статистики Стьюдента і F – статистики Фішера. Яким чином можна визначити критичні (порогові) значення цих статистик?
23. Поясніть фізичну суть умови достатнього збудження процесу? До чого призводить невиконання цієї умови?
24. Яким чином досягають достатнього збудження процесів (об'єктів) на практиці? Які сигнали використовують для досягнення цієї мети? Які властивості цих сигналів використовують у даному випадку?
25. Нарисуйте наближений графік перетворення Фур'є для білого шуму і поясніть його.

26. Які інші ортогональні перетворення даних ви знаєте крім перетворення Фур'є? Чому на практиці використовують декілька різних ортогональних перетворень, а не одне?
27. Нарисуйте наблизений графік перетворення Фур'є для гармонічної функції: $y(k) = a_1 \sin(\omega_1 k + \varphi_1) + a_2 \cos(\omega_2 k + \varphi_2)$.

Р о з д і л 6

ОПТИМАЛЬНЕ ОЦІНЮВАННЯ І ПРОГНОЗУВАННЯ СТАНУ ЗА ДОПОМОГОЮ ФІЛЬТРА КАЛМАНА

6.1 Принцип рекурсивного оцінювання

Алгоритми оцінювання стану динамічних систем ґрунтуються на рекурсивних обчислювальних схемах, які зручні для практичного застосування та комп'ютерної реалізації. Без них неможливо обійтись при створенні систем керування та прогнозування, які працюють у реальному часі. Стан динамічної системи характеризують множиною змінних стану, смисл яких залежить від типу конкретного процесу, що аналізується. Наприклад, поточний стан макроекономіки можна характеризувати значеннями валового внутрішнього продукту, індексу інфляції, вартості споживчого кошика, рівнями середньої зарплати у галузях промисловості та у бюджетній сфері і т. ін. Поточний стан людини (у першому наближенні) можна описати такими змінними стану: температура тіла, артеріальний тиск, частота серцевих скорочень (пульс) і поверхневий опір шкіри.

Нижче розглянемо суть та алгоритм рекурсивного оцінювання, переваги і недоліки, які мають рекурсивні алгоритми порівняно із звичними нерекурсивними схемами обчислень. Загалом мета цього розділу полягає у тому, щоб надати необхідну початкову інформацію про рекурсивне оцінювання вимірів на фоні шумових складових (збурення стану і шумів або похибок вимірів), зокрема стосовно можливостей застосування популярного фільтра Калмана та його алгоритмізацію.

Для ілюстрації спочатку розглянемо загальний принцип рекурсивного оцінювання даних, який у подальшому буде використано при побудові алгоритмів оптимального оцінювання (фільтрації). Принцип рекурсивного оцінювання можна легко пояснити на прикладі обчислення поточного вибіркового середнього значення часового ряду статистичних або експериментальних даних:

$$\bar{x}(k) = \frac{1}{k} \sum_{i=1}^k x(i), \quad (6.1.1)$$

де $\bar{x}(k)$ – оцінка середнього значення послідовності $x(k)$. Формулу для середнього значення представимо у вигляді:

$$\begin{aligned} \bar{x}(k) &= \frac{k-1}{k-1} \cdot \frac{1}{k} \left[\sum_{i=1}^{k-1} \bar{x}(i) + x(k) \right] = \frac{k-1}{k} \cdot \frac{1}{k-1} \left[\sum_{i=1}^{k-1} x(i) + x(k) \right] = \\ &= \left(1 - \frac{1}{k} \right) \cdot \left[\frac{1}{k-1} \sum_{i=1}^{k-1} x(i) + \frac{1}{k-1} x(k) \right] = \left(1 - \frac{1}{k} \right) \left[\bar{x}(k-1) + \frac{1}{k-1} x(k) \right] = \\ &= \bar{x}(k-1) - \frac{1}{k} \bar{x}(k-1) + \frac{1}{k} x(k). \end{aligned}$$

Тобто тепер поточне середнє можна обчислити за формулою:

$$\bar{x}(k) = \bar{x}(k-1) + \frac{1}{k} \cdot [x(k) - \bar{x}(k-1)]. \quad (6.1.2)$$

Це рівняння має рекурсивну форму, тому що поточна оцінка $\bar{x}(k)$ ґрунтується на значенні попередньої оцінки $\bar{x}(k-1)$, до якої додається корекція (зважена нев'язка). Другий член у правій частині останньої формули призначений для корекції оцінки середнього в момент k . Таким чином, ми отримали рівняння у формі, яка подібна до основного рівняння оцінювання у фільтрі Калмана. Таке рівняння можна вважати рівнянням оцінювання поточного середнього у формі фільтра Калмана з нестационарним коефіцієнтом $1/k$. Такою ж формою ми скористаємось у наступному параграфі з метою отримання рівнянь оптимального фільтра.

Перевага рекурсивного рівняння у порівнянні з нерекурсивною формою (6.1.1) полягає в тому, що рекурсивне рівняння не потребує запам'ятовування

всієї вибірки значень $x(k)$, а середнє значення оцінки може знаходитись у будь-якому інтервалі. Воно ґрунтується на оцінці, отриманій у попередній момент часу, та на поточних вимірах. Очевидно, що із збільшенням значення k поточне середнє прямує до константи, а тому вираз (6.1.2) можна використовувати для оцінювання середнього тільки на коротких часових інтервалах. Оцінювання і прогнозування середнього на довгих часових інтервалах потребує застосування моделей поточного середнього, наприклад, поліномів від часу або інших детермінованих функцій.

6.2 Дискретний фільтр Калмана для вільної динамічної системи

Дискретний фільтр Калмана описується множиною рекурсивних матричних рівнянь, які у лінійному випадку відносно легко розв'язуються на комп'ютері. В основі фільтра лежить розв'язок рівняння Ріккаті, який дає можливість визначити оптимальне значення матричного коефіцієнта фільтра. У цьому розділі ми отримаємо рівняння алгоритму оптимальної фільтрації для лінійної динамічної системи, які можна використовувати для обчислювання оптимальних оцінок стану процесу, а також для короткострокового прогнозування.

Постановка задачі оптимальної фільтрації. Нестационарна лінійна система описується у дискретному часі рівняннями із змінними в часі коефіцієнтами (непостійність означає залежність значень коефіцієнтів від часу k):

$$\mathbf{x}(k) = \mathbf{A}(k, k-1)\mathbf{x}(k-1) + \mathbf{B}(k, k-1)\mathbf{u}(k-1) + \mathbf{w}(k-1), \quad (6.2.1)$$

де $\mathbf{x}(k)$ – n -вимірний вектор станів системи; $\mathbf{u}(k-1)$ – m -вимірний вектор детермінованих вхідних величин (сигнали керування); $\mathbf{w}(k-1)$ – n -вимірний вектор випадкових зовнішніх збурень; $\mathbf{A}(k, k-1)$ – $(n \times n)$ матриця динаміки системи (вона містить коефіцієнти, що характеризують динаміку, тобто швидкість зміни станів у часі); $\mathbf{B}(k, k-1)$ – $(n \times m)$ матриця коефіцієнтів керування. Подвійний часовий аргумент у вигляді $(k, k-1)$ означає, що величина з цим аргументом використовується в момент k , але її значення ґрунтується на попередніх даних, які відомі на момент $k-1$, включно. Далі будемо записувати для простоти матриці, вектори і змінні, наприклад \mathbf{A} і \mathbf{B} , з одним аргументом, тобто $\mathbf{A}(k)$ та $\mathbf{B}(k)$. Очевидно, що стаціонарна система описується матрицями з постійними коефіцієнтами, які записують просто \mathbf{A} і

В. Оскільки матриця \mathbf{A} зв'язує поточний стан із попереднім, то її називають ще перехідною матрицею станів. Нагадаємо, що дискретний час k зв'язаний з неперервним часом t періодом дискретизації вимірів T_s : $t = kT_s$.

У класичній постановці задачі оптимальної фільтрації послідовність зовнішніх збурень $\mathbf{w}(k)$ задовольняє властивостям білого гаусового шуму з нульовим середнім значенням і коваріаційною матрицею \mathbf{Q} , тобто статистики шуму мають вигляд:

$$\begin{aligned} E[\mathbf{w}(k)] &= \mathbf{0}, \quad \forall k, \\ E[\mathbf{w}(k)\mathbf{w}^T(j)] &= \mathbf{Q}(k)\delta_{kj}, \end{aligned} \quad (6.2.2)$$

де δ_{kj} – дельта-функція Кронекера, що визначається так: $\delta_{kj} = \begin{cases} 0 & \text{для } k \neq j; \\ 1 & \text{для } k = j; \end{cases}$

$\mathbf{Q}(k)$ – додатно визначена коваріаційна матриця зовнішніх збурень стану розмірності $(n \times n)$. Діагональні елементи матриці – це дисперсії компонент вектора збурень $\mathbf{w}(k)$.

Початковим станом системи \mathbf{x}_0 будемо вважати випадкові змінні з відомими статистиками:

$$E[\mathbf{x}_0] = \bar{\mathbf{x}}_0; \quad E[\mathbf{x}_0\mathbf{x}_0^T] = \mathbf{M}; \quad E[\mathbf{w}(k)\mathbf{x}_0^T] = \mathbf{0}, \quad \forall k. \quad (6.2.3)$$

Нехай вектор вимірів $\mathbf{z}(k)$ вихідних змінних доступний в будь-який момент часу k (або t_k), а його компоненти лінійно зв'язані з вектором стану і на них впливає шум вимірів, тобто

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k), \quad (6.2.4)$$

де $\mathbf{H}(k)$ – матриця спостережень вимірності $(r \times n)$, $\mathbf{v}(k)$ – r -вимірний вектор випадкових величин шумів вимірів з відомими статистиками:

$$E[\mathbf{v}(k)] = \mathbf{0}, \quad E[\mathbf{v}(k)\mathbf{v}^T(j)] = \mathbf{R}(k)\delta_{kj},$$

де $\mathbf{R}(k)$ – додатно визначена коваріаційна матриця шумів вимірів вимірності $(r \times r)$, діагональні елементи якої є дисперсіями адитивного шуму в кожному каналі вимірів. Шум вимірів також задовольняє властивостям білого гаусового шуму. Він вважається некорельованим із зовнішнім збуренням $\mathbf{w}(k)$ і початковим станом системи, тобто

$$E[\mathbf{v}(k)\mathbf{w}^T(j)] = 0 \quad \forall k, j;$$

$$E[\mathbf{v}(k)\mathbf{x}_0^T] = 0 \quad \forall k.$$

Для зазначеної вище системи з вектором стану $\mathbf{x}(k)$ необхідно знайти оцінку стану $\hat{\mathbf{x}}(k)$ в момент t_k як лінійну комбінацію оцінки $\hat{\mathbf{x}}(k-1)$ в момент t_{k-1} і самого останнього виміру (статистичних даних) $\mathbf{z}(k)$.

Оцінка $\hat{\mathbf{x}}(k)$ обчислюється як найкраща за мінімумом середнього значення суми квадратів оцінок похибок. Іншими словами, оцінка повинна бути такою, щоб

$$E[(\hat{\mathbf{x}}(k) - \mathbf{x}(k))^T (\hat{\mathbf{x}}(k) - \mathbf{x}(k))] = \min_{\mathbf{K}}, \quad (6.2.5)$$

де E – символ математичного сподівання; $\mathbf{x}(k)$ – точне значення вектора стану, яке може бути обчислене за допомогою детермінованої складової математичної моделі процесу; \mathbf{K} – оптимальний матричний коефіцієнт фільтра, який необхідно обчислити в результаті розв'язання оптимізаційної задачі.

Концепція вільної динамічної системи

Вільна динамічна система – це система, простіша ніж повна система (6.2.1); в ній відсутні зовнішні вхідні впливи. Така система описується однорідним лінійним різницеvim рівнянням:

$$\mathbf{x}(k) = \mathbf{A}(k)\mathbf{x}(k-1); \quad (6.2.6)$$

фактично це однорідне рівняння для матричного рівняння стану.

Воно має простішу структуру ніж рівняння (6.2.1) і за його допомогою можна значно легше отримати рівняння фільтра Калмана. Тобто спочатку отримаємо рівняння фільтра для вільної динамічної системи (6.2.6), а потім узагальнимо їх на випадок системи, збудованої детермінованими та випадковими вхідними впливами.

Формування рівняння фільтрації

Виберемо структуру рівняння для оцінювання лінійної системи. Прогноз оцінки стану в момент часу k на основі попередньої оцінки можна знайти за допомогою матриці динаміки:

$$\hat{\mathbf{x}}(k) = \mathbf{A}(k)\hat{\mathbf{x}}(k-1). \quad (6.2.7)$$

Рівняння (6.2.7) показує, що прогноз оцінки в момент k визначається як результат множення оцінки вектора стану в момент $(k - 1)$ на матрицю $\mathbf{A}(k)$, тобто прогноз оцінки стану на один крок визначається тільки динамікою системи.

Для корегування оцінки прогнозу необхідно використати вимір $\mathbf{z}(k)$. З рівняння (6.2.4) видно, що очікуване значення $\hat{\mathbf{z}}(k)$ вихідного вектора вимірів $\mathbf{z}(k)$ в момент t_k буде $\mathbf{H}(k)\hat{\mathbf{x}}(k)$ або

$$\hat{\mathbf{z}}(k) = \mathbf{H}(k)\mathbf{A}(k)\hat{\mathbf{x}}(k-1).$$

Матриця вимірів $\mathbf{H}(k)$ дає можливість поставити у відповідність розмірність вектора стану $\mathbf{x}(k)$ вектору вимірів $\mathbf{z}(k)$, оскільки у більшості практичних випадків виконується нерівність: $\dim(\mathbf{x}) \geq \dim(\mathbf{z})$. Похибка прогнозованого значення виміру може бути знайдена як різниця між виміром та його очікуваною величиною, тобто

$$\mathbf{e}_p(k) = \mathbf{v}(k) = \mathbf{z}(k) - \mathbf{H}(k)\mathbf{A}(k)\hat{\mathbf{x}}(k-1).$$

Цю величину називають ще *нев'язкою* або *інновацією*, оскільки вона несе нову інформацію про процес.

Таким чином, оцінка $\hat{\mathbf{x}}(k)$ є лінійною комбінацією оцінки $\hat{\mathbf{x}}(k-1)$ в попередній момент і значення вихідного вектора $\mathbf{z}(k)$, що вимірюється, в момент t_k :

$$\hat{\mathbf{x}}(k) = \mathbf{A}(k)\hat{\mathbf{x}}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\mathbf{A}(k)\hat{\mathbf{x}}(k-1)]. \quad (6.2.8)$$

В цьому рівнянні на вибір матриці $\mathbf{K}(k)$ (матричний коефіцієнт фільтра) впливають декілька факторів, які вибирають таким чином, щоб мінімізувати математичне сподівання суми квадратів похибок оцінок вектора стану, тобто задовольняють критерію (6.2.5). Матрицю $\mathbf{K}(k)$ називають матричним оптимальним фільтром Калмана, оскільки він обчислюється за умови мінімізації функціоналу, який наведено нижче.

Похибка оцінки вектора стану за означенням обчислюється так:

$$\mathbf{e}(k) = \hat{\mathbf{x}}(k) - \mathbf{x}(k),$$

$$E[(\hat{\mathbf{x}}(k) - \mathbf{x}(k))^T (\hat{\mathbf{x}}(k) - \mathbf{x}(k))] = E[\mathbf{e}^T(k)\mathbf{e}(k)],$$

$$E[\mathbf{e}^T(k)\mathbf{e}(k)] = \text{trace } E[\mathbf{e}(k)\mathbf{e}^T(k)] = \text{trace } [\mathbf{P}(k)],$$

де $\mathbf{P}(k)$ – коваріаційна матриця похибок оцінок вектора стану, яка асоціюється з оцінкою стану, визначеною рівнянням (6.2.8). Ця матриця

необхідна для того щоб знайти у подальшому вираз для оптимального матричного коефіцієнта фільтра.

Для того щоб отримати вираз для матриці $\mathbf{P}(k)$, сформуємо спочатку вираз для вектора похибок оцінок $\mathbf{e}(k)$:

$$\begin{aligned}
 \mathbf{e}(k) &= [\mathbf{A} \hat{\mathbf{x}}(k-1) + \mathbf{K}(k)(\mathbf{z}(k) - \mathbf{H}\mathbf{A} \hat{\mathbf{x}}(k-1))] - \mathbf{A}(k-1)\mathbf{x}(k-1) = \\
 &= \mathbf{A} \mathbf{e}(k-1) - \mathbf{K}(k)\mathbf{H}\mathbf{A} \hat{\mathbf{x}}(k-1) + \mathbf{K}(k)\mathbf{z}(k) = \\
 &= \mathbf{A} \mathbf{e}(k-1) - \mathbf{K}(k)\mathbf{H}\mathbf{A} [\mathbf{e}(k-1) + \mathbf{x}(k-1)] + \mathbf{K}(k)[\mathbf{H}\mathbf{x}(k) + \mathbf{v}(k)] = \\
 &= [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{A} \mathbf{e}(k-1) - \mathbf{K}(k)\mathbf{H}\mathbf{A} \mathbf{x}(k-1) + \\
 &+ \mathbf{K}(k)\mathbf{H}\mathbf{A} \mathbf{x}(k-1) + \mathbf{K}(k)\mathbf{v}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{A} \mathbf{e}(k-1) + \mathbf{K}(k)\mathbf{v}(k).
 \end{aligned}
 \tag{6.2.9}$$

Тепер підставимо (6.2.9) у вираз для коваріаційної матриці $\mathbf{P}(k) = E[\mathbf{e}(k) \mathbf{e}^T(k)]$ і отримаємо:

$$\begin{aligned}
 \mathbf{P}(k) &= [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{A} E[\mathbf{e}(k-1) \mathbf{e}^T(k-1)] \mathbf{A}^T [\mathbf{I} - \mathbf{H}^T \mathbf{K}^T(k)] + \\
 &+ \mathbf{K}(k) E[\mathbf{v}(k) \mathbf{e}^T(k-1)] \mathbf{A}^T [\mathbf{I} - \mathbf{H}^T(k) \mathbf{K}^T(k)] + \\
 &+ [\mathbf{I} - \mathbf{K}(k)\mathbf{H}(k)]\mathbf{A} E[\mathbf{e}(k-1) \mathbf{v}^T(k)] \mathbf{K}^T(k) + \mathbf{K}(k) E[\mathbf{v}(k) \mathbf{v}^T(k)] \mathbf{K}^T(k)
 \end{aligned}$$

За означенням,

$$\begin{aligned}
 E[\mathbf{e}(k-1) \mathbf{e}^T(k-1)] &= \mathbf{P}(k-1), \\
 E[\mathbf{v}(k) \mathbf{v}^T(k)] &= \mathbf{R}(k).
 \end{aligned}
 \tag{6.2.10}$$

Беручи до уваги, що $E[\mathbf{v}(k) \mathbf{v}^T(k-1)] = 0$ та $E[\mathbf{v}(k) \hat{\mathbf{x}}_0^T] = 0$, отримаємо

$$\begin{aligned}
 E[\mathbf{v}(k) \mathbf{e}^T(k-1)] &= 0, \\
 E[\mathbf{e}(k-1) \mathbf{x}^T(k)] &= 0.
 \end{aligned}$$

Тобто значення шуму $\mathbf{v}(k)$ в момент t_k некорельоване з вектором стану $\mathbf{x}(k-1)$ в момент t_{k-1} . Вектор $\hat{\mathbf{x}}(k-1)$, який залежить від $\mathbf{v}(k-1)$, не корельований з $\mathbf{v}(k)$, тому що немає кореляції між $\mathbf{v}(k-1)$ та $\mathbf{v}(k)$. Звідси отримуємо, що вектор $\mathbf{e}(k-1) = \hat{\mathbf{x}}(k-1) - \mathbf{x}(k-1)$ некорельований з $\mathbf{v}(k)$.

Апріорна коваріаційна матриця похибок оцінок вектора стану

Беручи до уваги (6.2.10) і останній вираз, запишемо вираз для $\mathbf{P}(k)$ у формі:

$$\mathbf{P}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{P}'(k) [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]^T + \mathbf{K}(k)\mathbf{R}(k)\mathbf{K}^T(k), \quad (6.2.11)$$

де

$$\mathbf{P}'(k) = \mathbf{A}\mathbf{P}(k-1)\mathbf{A}^T.$$

Матрицю $\mathbf{P}'(k)$ називають *априорною* коваріаційною матрицею похибок оцінок вектора стану, тому що вона характеризує оцінку вектора стану до надходження вимірів $\mathbf{z}(k)$. Тому матрицю $\mathbf{P}(k)$ називають *апостеріорною* коваріаційною матрицею похибок оцінок.

Після деяких алгебраїчних маніпуляцій з рівнянням (6.2.11) його можна представити у вигляді:

$$\begin{aligned} \mathbf{P}(k) &= [\mathbf{P}'(k) - \mathbf{K}(k)\mathbf{H}\mathbf{P}'(k)][\mathbf{I} - \mathbf{H}^T\mathbf{K}^T(k)] + \mathbf{K}(k)\mathbf{R}(k)\mathbf{K}^T(k) = \\ &= \mathbf{P}'(k) - \mathbf{K}(k)\mathbf{H}\mathbf{P}'(k) - \mathbf{P}'(k)\mathbf{H}^T\mathbf{K}^T(k) + \\ &+ \mathbf{K}(k)[\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}(k)]\mathbf{K}^T(k). \end{aligned} \quad (6.2.12)$$

Рівняння такого типу називають рівнянням Ріккати.

Матриця $[\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}]$ – симетрична, оскільки

$$[\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}] = [\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}]^T,$$

та невід'ємно визначена. Таку матрицю можна представити у вигляді:

$$\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R} = \mathbf{S}(k)\mathbf{S}^T(k), \quad (6.2.13)$$

де матриця $\mathbf{S}(k)$ – на даний момент невідома, але буде визначена нижче.

Таким чином, вираз для $\mathbf{P}(k)$ приймає вигляд:

$$\mathbf{P}(k) = \mathbf{P}'(k) - \mathbf{K}(k)\mathbf{H}\mathbf{P}'(k) - \mathbf{P}'(k)\mathbf{H}^T\mathbf{K}^T(k) + \mathbf{K}(k)\mathbf{S}(k)\mathbf{S}^T(k)\mathbf{K}^T(k),$$

або у спрощеному вигляді без індексів:

$$\mathbf{P} = \mathbf{P}' - \mathbf{K}\mathbf{H}\mathbf{P}' - \mathbf{P}'\mathbf{H}^T\mathbf{K}^T + \mathbf{K}\mathbf{S}\mathbf{S}^T\mathbf{K}^T. \quad (6.2.14)$$

Праву частину цього рівняння необхідно привести до такого вигляду, щоб можна було:

1. визначити коефіцієнт фільтра $\mathbf{K}(k)$ за умови мінімуму сліду матриці $\mathbf{P}(k)$;
2. забезпечити додатну напіввизначеність діагональних елементів матриці в правій частині рівняння.

З цією метою введемо до розгляду такий добуток:

$$(\mathbf{KS} - \mathbf{G})(\mathbf{KS} - \mathbf{G})^T = \mathbf{KSS}^T \mathbf{K}^T - \mathbf{KSG}^T - \mathbf{GS}^T \mathbf{K}^T + \mathbf{GG}^T,$$

де \mathbf{G} – невідома матриця, що існує за припущенням. Діагональні елементи матриці $(\mathbf{KS} - \mathbf{G})(\mathbf{KS} - \mathbf{G})^T$ повинні бути додатно напіввизначеними, тобто $\text{diag}\{(\mathbf{KS} - \mathbf{G})(\mathbf{KS} - \mathbf{G})^T\} \geq 0$.

Необхідно знайти матрицю \mathbf{G} . Можна прирівняти члени:

$$\mathbf{KHP}' - \mathbf{P}'\mathbf{H}^T \mathbf{K}^T = \mathbf{KSG}^T + \mathbf{GS}^T \mathbf{K}^T, \quad (6.2.15)$$

і записати тепер рівняння (6.2.14) у вигляді:

$$\mathbf{P} = \mathbf{P}' + (\mathbf{KS} - \mathbf{G})(\mathbf{KS} - \mathbf{G})^T - \mathbf{GG}^T. \quad (6.2.16)$$

Перехід від (6.2.14) до (6.2.16) дозволяє записати такі рівності (з використанням (6.2.15)):

$$\begin{aligned} \mathbf{KSG}^T &= \mathbf{KHP}', \\ \mathbf{GS}^T \mathbf{K}^T &= \mathbf{P}'\mathbf{H}^T \mathbf{K}^T, \end{aligned}$$

або $\mathbf{GS}^T = \mathbf{P}'\mathbf{H}^T$.

Тепер запишемо вираз для \mathbf{G} :

$$\mathbf{G} = \mathbf{P}'\mathbf{H}^T (\mathbf{S}^T)^{-1} = \mathbf{P}'\mathbf{H}^T (\mathbf{S}^{-1})^T. \quad (6.2.17)$$

Обчислення оптимального матричного коефіцієнта фільтра

В рівнянні (6.2.16) від $\mathbf{K}(k)$ залежить тільки середній член, який є добутком матриці $(\mathbf{KS} - \mathbf{G})$ на її транспоновану. Це забезпечує невід'ємність елементів головної діагоналі середнього члена в (6.2.16).

Таким чином, слід матриці $\mathbf{P}(k)$ буде мінімальним, якщо покласти середній член (6.2.16) рівним нулю, тобто

$$\mathbf{KS} - \mathbf{G} = 0,$$

або $\mathbf{KS} = \mathbf{G}$, а звідси $\mathbf{K}(k)$ можна визначити так:

$$\mathbf{K}(k) = \mathbf{GS}^{-1}. \quad (6.2.18)$$

Підставимо в (6.2.18) значення, що визначається рівнянням (6.2.17) і отримаємо:

$$\mathbf{K}(k) = \mathbf{P}'\mathbf{H}^T (\mathbf{S}^{-1})^T \mathbf{S}^{-1} = \mathbf{P}'\mathbf{H}^T (\mathbf{S}\mathbf{S}^T)^{-1},$$

а із врахуванням (6.2.13) остаточно можна записати:

$$\mathbf{K}(k) = \mathbf{P}'\mathbf{H}^T [\mathbf{H}\mathbf{P}'\mathbf{H}^T + \mathbf{R}]^{-1}.$$

Враховуючи часовий аргумент, отримаємо таке рівняння:

$$\mathbf{K}(k) = \mathbf{P}'(k)\mathbf{H}^T(k) [\mathbf{H}(k)\mathbf{P}'(k)\mathbf{H}^T(k) + \mathbf{R}(k)]^{-1}. \quad (6.2.19)$$

Таким чином, коефіцієнт фільтра Калмана (6.2.19) знайдено за умови мінімуму математичного сподівання суми квадратів похибок оцінок вектора стану об'єкта.

Апостеріорна коваріаційна матриця похибок оцінок вектора станів

Знайдемо вираз для коваріаційної матриці, яка зв'язана з оптимальною оцінкою вектора стану. Розглянемо знову коваріаційну матрицю, що визначається рівнянням Ріккати (6.2.12), тобто:

$$\mathbf{P}(k) = \mathbf{P}'(k) - \mathbf{K}\mathbf{H}\mathbf{P}'(k) - \mathbf{P}'(k)\mathbf{H}^T \mathbf{K}^T(k) + \mathbf{K}(k) [\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}(k)] \mathbf{K}^T(k).$$

Підставимо в це рівняння значення $\mathbf{K}(k)$, що визначається (6.2.19), і отримаємо:

$$\begin{aligned} \mathbf{P}(k) &= \mathbf{P}'(k) - \mathbf{P}'(k)\mathbf{H}^T [\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}]^{-1} \mathbf{H}\mathbf{P}'(k) = \\ &= [\mathbf{I} - \mathbf{K}(k)\mathbf{H}] \mathbf{P}'(k). \end{aligned} \quad (6.2.20)$$

Тобто апостеріорна коваріаційна матриця похибок оцінок є меншою порівняно з апріорною на величину $\mathbf{K}(k)\mathbf{H}\mathbf{P}'(k)$. Це пояснюється тим, що отриманий вимір $\mathbf{z}(k)$ в момент часу t_k сприяє зменшенню невизначеності оцінки вектора станів.

Результати

Матричне рівняння (6.2.8) використовується для обчислення оптимальних оцінок вектора стану в момент t_k за допомогою оптимального коефіцієнта фільтра $\mathbf{K}(k)$, який можна знайти із системи матричних рівнянь. Матричні рівняння (6.2.8), (6.2.11) і (6.2.20) утворюють рекурсивний алгоритм фільтра Калмана. Зрозуміло, що процес оцінювання може продовжуватись до нескінченності. Послідовність обчислень показано нижче у вигляді

алгоритму. Нагадаємо, що рівняння фільтра Калмана, наведені вище, побудовані для вільної динамічної системи.

Алгоритм оптимального оцінювання стану (фільтрації) вільної динамічної системи):

1. Задати початкові умови \mathbf{x}_0 для вектора стану і коваріаційної матриці похибок оцінок \mathbf{P}_0 . Присвоїти значення коваріаційним матрицям збурень стану \mathbf{Q} та похибок вимірів \mathbf{R} .
2. Обчислити матричний оптимальний коефіцієнт фільтра:

$$\mathbf{K}(k) = \mathbf{P}'(k-1)\mathbf{H}^T [\mathbf{H}\mathbf{P}(k-1)\mathbf{H}^T + \mathbf{R}]^{-1}.$$

3. Skorистатись новими вимірами для обчислення поточної оцінки вектора стану:

$$\hat{\mathbf{x}}(k) = \mathbf{A}\hat{\mathbf{x}}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}\mathbf{A}\hat{\mathbf{x}}(k-1)].$$

4. Обчислити апостеріорну коваріаційну матрицю похибок для оновлених оцінок:

$$\mathbf{P}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{P}'(k).$$

5. Обчислити апіорну (для наступної оцінки вектора стану) коваріаційну матрицю похибок оцінок:

$$\mathbf{P}'(k+1) = \mathbf{A}\mathbf{P}(k)\mathbf{A}^T.$$

і перейти на крок 2 (наступний крок обчислення рівнянь фільтра).

Розглянутий фільтр може функціонувати нескінченно довго при надходженні нових вимірів. Після закінчення перехідного процесу коефіцієнти фільтра встановлюються, як правило, на деяких постійних значеннях, які залишаються незмінними на довгих проміжках часу якщо процес стаціонарний. Якщо фільтр працює нормально (у відповідності до його теоретичних характеристик), то діагональні елементи коваріаційної матриці похибок оцінок вектора стану залишаються додатно визначеними (оскільки це дисперсії похибок оцінок) і збігаються до невеликих усталених значень.

Якщо дисперсії похибок оцінок стану не зменшуються, а зростають з часом, то такий режим називають розбіжним. Із плином часу розрядна сітка

комп'ютера може переповнюватись і функціонування алгоритму повністю припиняється. Очевидно, що виникнення такого режиму функціонування фільтра необхідно попереджувати. Методи підвищення обчислювальної стійкості алгоритму фільтрації будуть розглянуті нижче в даному розділі, а в наступному параграфі отримаємо рівняння фільтра для динамічної системи, яка функціонує в умовах впливу зовнішніх детермінованих та випадкових впливів (збурень).

6.3 Дискретний фільтр Калмана для лінійної системи з детермінованими і стохастичними впливами

Будемо вважати розширеною математичною моделлю лінійну систему з детермінованими і стохастичними входами (збуреннями):

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{A}(k)\mathbf{x}(k-1) + \mathbf{B}(k)\mathbf{u}(k-1) + \mathbf{w}(k-1), \\ \mathbf{z}(k) &= \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k). \end{aligned} \quad (6.3.1)$$

Середній член першого рівняння системи (6.3.1) характеризує вплив детермінованого вхідного сигналу $\mathbf{u}(k)$; $\mathbf{w}(k)$ – зовнішні стохастичні збурення, які впливають на функціонування лінійної системи.

Для початку знайдемо прогнозовану оцінку вектора стану, використовуючи попередню оцінку $\hat{\mathbf{x}}(k-1)$ для моменту t_{k-1} , тобто не будемо враховувати виміри, отримані в момент t_k . Стохастичне збурення стану (або шум стану) $\mathbf{w}(k-1)$ за припущенням не залежить від вектора стану в момент t_{k-1} і має нульове середнє значення. Тому вважаємо, що цей шум не буде впливати на оцінку вектора стану в момент t_k . Але $\mathbf{B}(k)\mathbf{u}(k-1) \equiv \mathbf{f}(k-1)$ – це відома векторна функція на інтервалі $[t_{k-1}, t_k]$ і, згідно з рівнянням (6.3.1), прогноз оцінки вектора стану можна представити у вигляді:

$$\hat{\mathbf{x}}'(k) = \mathbf{A}(k)\hat{\mathbf{x}}(k-1) + \mathbf{f}(k-1). \quad (6.3.2)$$

Після появи виміру вихідного вектора $\mathbf{z}(k)$ в момент t_k можна обчислити нову оцінку вектора стану за виразом:

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}'(k) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}(k)\hat{\mathbf{x}}'(k)], \quad (6.3.3)$$

де $\mathbf{K}(k)$ – невідома матриця коефіцієнтів фільтра (у подальшому будемо записувати матриці \mathbf{A}, \mathbf{H} без аргументів для спрощення запису):

$$\begin{aligned}
\mathbf{e}(k) &= \hat{\mathbf{x}}(k) - \mathbf{x}(k) = \{ \mathbf{A}\hat{\mathbf{x}}(k-1) + \mathbf{f}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}\hat{\mathbf{x}}'(k)] \} - \\
&\quad - [\mathbf{A}\mathbf{x}(k-1) + \mathbf{f}(k-1) + \mathbf{w}(k-1)] = \\
&= [\mathbf{A}\mathbf{e}(k-1) + \mathbf{K}(k)[\mathbf{H}\mathbf{x}(k) + \mathbf{v}(k)] - \mathbf{K}(k)\mathbf{H}\hat{\mathbf{x}}'(k) - \mathbf{w}(k-1)] = \\
&= \mathbf{A}\mathbf{e}(k-1) + \mathbf{K}(k)\mathbf{H}[\mathbf{A}\mathbf{x}(k-1) + \mathbf{f}(k-1) + \mathbf{w}(k-1)] - \\
&\quad - \mathbf{K}(k)\mathbf{H}[\mathbf{A}\hat{\mathbf{x}}(k-1) + \mathbf{f}(k-1)] - \mathbf{w}(k-1) + \mathbf{K}(k)\mathbf{v}(k) = \\
&= \mathbf{A}\mathbf{e}(k-1) - \mathbf{K}(k)\mathbf{H}\mathbf{A}\mathbf{e}(k-1) + \mathbf{K}(k)\mathbf{H}\mathbf{w}(k-1) - \mathbf{w}(k-1) + \mathbf{K}(k)\mathbf{v}(k) = \\
&= [\mathbf{I} - \mathbf{K}(k)\mathbf{H}][\mathbf{A}\mathbf{e}(k-1) - \mathbf{w}(k-1)] + \mathbf{K}(k)\mathbf{v}(k).
\end{aligned} \tag{6.3.4}$$

Тепер знайдемо апостеріорну коваріаційну матрицю похибок оцінок вектора стану лінійного процесу:

$$\begin{aligned}
\mathbf{P}(k) &= E[\mathbf{e}(k)\mathbf{e}^T(k)] = \\
&= E \left\{ \begin{aligned} & [[\mathbf{I} - \mathbf{K}(k)\mathbf{H}][\mathbf{A}\mathbf{e}(k-1) - \mathbf{w}(k-1)] + \mathbf{K}(k)\mathbf{v}(k)] \times \\ & [[\mathbf{I} - \mathbf{K}(k)\mathbf{H}][\mathbf{A}\mathbf{e}(k-1) - \mathbf{w}(k-1) + \mathbf{K}(k)\mathbf{v}(k)]^T \end{aligned} \right\}.
\end{aligned} \tag{6.3.5}$$

Введемо апіорну коваріаційну матрицю похибок оцінок вектора стану $\mathbf{P}'(k)$:

$$\begin{aligned}
\mathbf{P}'(k) &= E \left\{ [\mathbf{A}\mathbf{e}(k-1) - \mathbf{w}(k-1)][\mathbf{A}\mathbf{e}(k-1) - \mathbf{w}(k-1)]^T \right\} = \\
&= \mathbf{A}\mathbf{P}(k-1)\mathbf{A}^T + \mathbf{Q}(k-1).
\end{aligned} \tag{6.3.6}$$

У відповідності до (6.3.5) апостеріорна коваріаційна матриця $\mathbf{P}(k)$ визначається за рівнянням:

$$\mathbf{P}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{P}'(k)[\mathbf{I} - \mathbf{K}(k)\mathbf{H}]^T + \mathbf{K}(k)\mathbf{R}(k)\mathbf{K}^T(k). \tag{6.3.7}$$

Структура цього рівняння ідентична структурі рівняння для апостеріорної матриці, отриманої у попередньому параграфі для вільної динамічної системи. Це означає, що оптимальний матричний коефіцієнт фільтра $\mathbf{K}(k)$ для системи (6.3.1) визначається таким же рівнянням, що і для вільної динамічної системи. Звідси випливає, що оптимальний дискретний фільтр для лінійної системи (6.3.1) визначається рівняннями: (6.3.2), (6.3.3), (6.2.19), (6.2.20) і (6.3.6).

Очевидно, що поява випадкового збурення $\mathbf{w}(k)$ призводить до погіршення якості оцінок, про що свідчить додатковий член $\mathbf{Q}(k-1)$ у правій частині рівняння (6.3.6) для обчислення $\mathbf{P}'(k)$. Детермінований сигнал $\mathbf{f}(k-1)$ впливає на прогноз оцінки вектора стану у відповідності до рівняння (6.3.2).

Якщо середнє значення $E[\mathbf{w}(k-1)]$ випадкового збурення $\mathbf{w}(k-1)$ не

дорівнює нулю і може бути виміряне (оцінене) в будь-який момент часу, то його можна досить просто врахувати в рівняннях фільтрації. В такому випадку значення $\mu_w = E[\mathbf{w}(k-1)]$ додається до детермінованого впливу $\mathbf{f}(k-1)$.

Крім сигналу керування $\mathbf{B}\mathbf{u}(k-1)$, детерміноване збурення $\mathbf{f}(k-1)$ може включати в себе вплив інших вхідних детермінованих сигналів, що діють на динамічну систему. Аналогічно можна врахувати ненульове середнє шуму вимірів $\mathbf{v}(k)$.

Система рівнянь оптимального фільтра Калмана для лінійної системи, яка функціонує в умовах впливу детермінованих та випадкових збурень, наведена нижче у вигляді алгоритму фільтрації.

Алгоритм оптимального оцінювання стану для лінійної системи з детермінованими та випадковими входами:

3. Математична модель процесу (лінійної системи):

$$\mathbf{x}(k) = \mathbf{A}(k)\mathbf{x}(k-1) + \mathbf{B}(k)\mathbf{u}(k-1) + \mathbf{w}(k-1),$$

$$\mathbf{z}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{v}(k);$$

$$E[\mathbf{w}(k)] = 0, \quad E[\mathbf{v}(k)] = 0, \quad E[\mathbf{x}_0] = \bar{\mathbf{x}}_0, \quad E[\bar{\mathbf{x}}_0 \bar{\mathbf{x}}_0^T] = \mathbf{P}_0 = \mathbf{P}'_0;$$

$$E[\mathbf{w}(k) \mathbf{w}^T(l)] = \mathbf{Q}(k) \delta(k-l), \quad E[\mathbf{v}(k) \mathbf{v}^T(l)] = \mathbf{R}(k) \delta(k-l);$$

$$E[\mathbf{w}(k) \mathbf{v}^T(l)] = E[\mathbf{w}(k) \mathbf{x}_0^T] = E[\mathbf{v}(k) \mathbf{x}_0^T] = 0.$$

4. Матричний коефіцієнт фільтра:

$$\mathbf{K}(k) = \mathbf{P}'(k) \mathbf{H}^T [\mathbf{H} \mathbf{P}'(k) \mathbf{H}^T + \mathbf{R}]^{-1}.$$

5. Рівняння фільтрації:

$$\hat{\mathbf{x}}(k) = \mathbf{A} \hat{\mathbf{x}}(k-1) + \mathbf{B} \mathbf{u}(k-1) + \mathbf{K}(k) [\mathbf{z}(k) - \mathbf{H} \mathbf{A} \hat{\mathbf{x}}(k-1)].$$

6. Апостеріорна коваріаційна матриця похибок оцінок:

$$\mathbf{P}(k) = [\mathbf{I} - \mathbf{K}(k) \mathbf{H}] \mathbf{P}'(k-1).$$

7. Апріорна коваріаційна матриця похибок оцінок для наступного циклу обчислень:

$$\mathbf{P}'(k+1) = \mathbf{A} \mathbf{P}(k) \mathbf{A}^T + \mathbf{Q}(k).$$

8. Перейти на крок 2.

Якщо значення коваріаційних матриць \mathbf{Q} і \mathbf{R} неможливо оцінити на

основі наявної апріорної інформації про процес, то необхідно застосувати алгоритм адаптивної фільтрації, який дозволяє одночасно із станом оцінювати невідомі статистичні параметри зовнішнього збурення і шумів вимірів процесу. Однак необхідно пам'ятати, що ускладнення алгоритму фільтрації потребує особливої уваги до забезпечення збіжності, оскільки він чутливий до похибок моделі процесу.

Статистичне моделювання алгоритму фільтрації

Перед практичним застосуванням оптимального фільтра для обробки фактичних даних необхідно виконати його статистичне моделювання, яке полягає у багатократному прогоні алгоритму з використанням множини реалізацій випадкових процесів, що описують вплив збурень стану та шуму вимірів. Суть статистичного моделювання фільтра полягає у виконанні таких кроків:

1. Задати кількість реалізацій алгоритму фільтрації, якому відповідає кількість реалізацій випадкових процесів. Як правило вибирають число реалізацій $M \geq 100$ (це досить умовна рекомендація). Покласти кількість реалізацій (циклів) алгоритму фільтрації рівною $M : C = M$.

2. Згенерувати вибірки випадкових процесів $\mathbf{w}(k)$ і $\mathbf{v}(k)$, потужність яких дорівнює довжині вибірки експериментальних (статистичних) даних, і врахувати їх у вимірах (згенерованих штучно чи фактичних).

3. Застосувати алгоритм фільтрації до вибірки вимірів потужністю N і зберегти на кожному кроці ($k = 1, \dots, N$) діагональні елементи коваріаційної матриці похибок фільтрації $\mathbf{P}(k)$ (тобто дисперсії похибок оцінок вектора стану). При кожному наступному прогоні алгоритму фільтрації до цих дисперсій необхідно додати нові (накопичувати значення), отримані для наступного прогону. Тобто, якщо вимірність вектора стану дорівнює n , то в результаті моделювання отримаємо n векторів значень дисперсій похибок оцінок вимірністю $[N \times 1]$.

4. Зменшити кількість реалізацій на 1: $C \leftarrow C - 1$. Якщо $C \neq 0$, то перейти на крок 2. Інакше, перейти на крок 5. Знайти середнє значення по ансамблю реалізацій для кожного діагонального елемента похибок фільтрації $\mathbf{P}(k)$, тобто,

$$\bar{p}_{ii}(k) = \frac{1}{M} \sum_{l=1}^M p_{ii}(l), \quad k = 1, \dots, N,$$

де N – довжина вибірки даних.

5. Побудувати графіки залежностей усереднених дисперсій похибок оцінок від часу k , тобто, $p_{ii}(k), k = 1, \dots, N$. Якщо алгоритм реалізовано коректно, то усереднені дисперсії будуть експоненціально спадати до невеликих усталених значень.

Статистичне моделювання алгоритму фільтрації дає можливість дослідити характеристики його функціонування на множині реалізацій випадкових процесів і тим самим отримати гарантію його нормального функціонування у системі обробки фактичних експериментальних даних.

6.4 Причини розбіжності та заходи щодо підвищення обчислювальної стійкості оптимального фільтра

Показники нормального функціонування оптимального фільтра

Нормальне (коректне) функціонування фільтра характеризується такими показниками:

1. Фільтр Калмана відноситься до рекурсивних алгоритмів, характерною ознакою яких є наявність перехідного процесу оцінювання. На протязі перехідного процесу значення оцінок стану можуть досить сильно коливатись, а похибки оцінок можуть приймати великі значення. Для стаціонарних процесів дисперсії похибок оцінок вектора стану повинні бути спадаючими функціями часу і збігатись до деяких невеликих постійних усталених значень після закінчення перехідного процесу. Таким чином, контролювати функціонування фільтра можна за допомогою діагональних елементів коваріаційної матриці похибок оцінок вектора стану $\mathbf{P}(k)$.
2. Для стаціонарного процесу коефіцієнти підсилення фільтра $\mathbf{K}(k)$ повинні приймати після закінчення перехідного процесу постійні усталені значення.
3. Графіки змінних стану $\mathbf{x}(k)$ не повинні мати значних імпульсів чи викидів після закінчення перехідного процесу.
4. Послідовність інновацій $\{v(k) = \mathbf{z}(k) - \mathbf{A}\mathbf{N}\hat{\mathbf{x}}(k-1)\}$ повинна задовольняти властивостям білого шуму: $E[v(k)] = 0$ і $E[v(k)v(k-l)] = 0, l \geq 1$.

Остання властивість дуже важлива і досить часто використовується у системах автоматичної діагностики для виявлення аномальних режимів функціонування процесів і систем в реальному часі. Як правило, вона використовується у комбінації з іншими процедурами аналізу, наприклад, процедурами формування ймовірнісного висновку. Системи такого типу надзвичайно популярні, оскільки сприяють значному підвищенню якості діагностики та скороченню часу, необхідного для аналізу функціонування системи.

Причини розбіжності алгоритму оптимальної фільтрації

У практиці впровадження алгоритмів оптимальної фільтрації досить часто зустрічаються випадки, коли фільтр характеризується “розбіжністю”. Тобто елементи коваріаційної матриці похибок оцінок вектора стану не збігаються до невеликих усталених значень, а з часом набувають таких великих значень, які можуть перевищувати максимально допустимі для

вибраного способу зберігання чисел у комп'ютері. Розбіжність алгоритму фільтрації може бути зумовлена такими причинами:

- *Математична модель процесу* (матриці **A**, **B**, **H**) має недостатню ступінь адекватності динамічній системі, а тому у процесі її побудови необхідно застосувати кілька статистичних критеріїв для аналізу якості моделі. Особливу увагу необхідно звернути на якість даних, на основі яких оцінено коефіцієнти (параметри) моделі. Якщо дані недостатньо інформативні, то побудувати модель, адекватну досліджуваному режиму функціонування процесу, практично неможливо. Докладна процедура побудови моделі на основі експериментальних (статистичних) даних представлена у розділі 5.
- Некоректне визначення початкових умов – наприклад, коваріаційної матриці похибок оцінок початкового стану системи $\mathbf{P}(0)$, тобто використовується наближене значення $\mathbf{P}_c(0)$, де індекс “*c*” означає субоптимальність. Початкові умови найбільш помітно впливають на якість функціонування фільтра в перехідному режимі.
- *Значення коваріаційних матриць Q і R* невідомі або відомі із значними похибками. У такому випадку необхідно збирати додаткову інформацію про процес та/або застосовувати алгоритм адаптивної фільтрації. У окремих випадках можна спробувати “підібрати” невідомі статистичні параметри, виходячи з конкретних значень вимірюваних сигналів (статистичних даних). При цьому якість фільтрації можна оцінювати за допомогою діагональних елементів коваріаційної матриці похибок фільтрації $\mathbf{P}(k)$.
- *Вимірювальні (статистичні) дані* не відповідають вимогам якості (наявні значні пропуски, великі викиди, існують значні періоди з постійними значеннями, наявні невраховані у моделі нелінійності).
- *Припущення стосовно розподілу та/або статистичних характеристик збурень стану та похибок (шуму) вимірів* не відповідають дійсності.
- Іноді для спрощення алгоритму при його застосуванні у реальному часі використовують *постійні коефіцієнти фільтра*, що також може призводити до великих похибок оцінок стану досліджуваного процесу.
- Комбінації вказаних причин та *помилки програмування* алгоритму фільтрації.

Можливі заходи для підвищення обчислювальної стійкості алгоритму

1. *Підвищення адекватності* математичної моделі процесу шляхом уточнення та ускладнення її структури. Уточнення структури і параметрів моделі з надходженням нових вимірів.
2. Застосування спеціальних методів *попередньої обробки даних* з метою їх нормування, заповнення пропусків та зменшення великих викидів.

3. Застосування алгоритмів адаптивної фільтрації у випадку, коли невідомі параметри коваріаційних матриць \mathbf{Q} і \mathbf{R} .
4. Якщо шум вимірів відсутній, то його необхідно ввести штучно (на деякому невеликому рівні по відношенню до амплітуди вимірів), що сприяє покращенню обчислювальної обумовленості матриці $[\mathbf{H}\mathbf{P}(k)\mathbf{H}^T + \mathbf{R}]$, для якої обчислюється обернена при визначенні коефіцієнта фільтра.
5. Застосування спеціальних алгоритмів обчислення обернених матриць при визначенні матричного коефіцієнта фільтра. Наприклад, алгоритму квадратного кореня, LU – факторизації та інших. За можливості можна використовувати алгоритм послідовної фільтрації, який не потребує операції обернення матриці.
6. Використання статистичного (Монте-Карло) моделювання з метою підвищення достовірності визначення характеристик алгоритму фільтрації перед його використанням за призначенням.
7. Необхідно постійно контролювати процес оцінювання шляхом аналізу значень діагональних елементів матриці $\mathbf{P}(k)$. Допускається примусове зменшення значень цих елементів, якщо це не пов'язано з похибками самого алгоритму оцінювання. Крім того, інноваційна послідовність $\mathbf{v}(k) = \mathbf{z}(k) - \mathbf{H}\mathbf{A}\hat{\mathbf{x}}(k-1)$ повинна задовольняти вимогам: $E[\mathbf{v}(k)] = 0$ і $E[\mathbf{v}(k)\mathbf{v}^T(l)] = \mathbf{G}\delta(k-l)$, тобто повинна бути некорельованою послідовністю з нульовим середнім.
8. Якщо конкретне застосування фільтра припускає примусове вилучення перехідного процесу оцінювання (в якому спостерігаються максимальні похибки оцінок вектора стану процесу), то необхідно використовувати фільтри з постійними коефіцієнтами, які визначають наперед під час статистичного моделювання.
9. Необхідно уникати випадків одночасного оцінювання невідомих параметрів моделі та стану процесу, тому що це призводить до появи нелінійностей у моделі і може призвести до розбіжності алгоритму фільтрації.
10. У випадку мікропроцесорної реалізації необхідно уникати використання цілочислової арифметики, яка суттєво обмежує діапазон значень змінних і констант.

6.5 Приклади побудови оптимального фільтра

Приклад 6.5.1. Розглянемо просту скалярну модель випадкового кроку

$$\begin{aligned}x(k) &= x(k-1) + w(k), & x(0) &= x_0, \\z(k) &= x(k) + v(k),\end{aligned}$$

де $w(k)$ і $v(k)$ – процеси білого шуму з нульовим середнім та дисперсіями Q і R , відповідно. Оскільки процес скалярний, то модель і рівняння фільтрації суттєво спрощуються. Так, рівняння фільтрації має вигляд:

$$\hat{x}(k) = \hat{x}(k-1) + K(k)[z(k) - \hat{x}(k-1)].$$

Апріорна коваріація оцінки вектора стану і коефіцієнт фільтра

$$P'(k) = P(k-1) + Q, \quad K(k) = \frac{P(k-1) + Q}{P(k-1) + Q + R}.$$

З останнього рівняння видно, що значення коефіцієнта $K(k)$ є обернено пропорціональним дисперсії шуму вимірів. Це означає, що чим вищий шум вимірів, тим нижчим буде значення коефіцієнта фільтра і оцінка стану $\hat{x}(k)$ в значній мірі визначається попередньою оцінкою $\hat{x}(k-1)$. Внаслідок малого значення коефіцієнта $K(k)$, інновація $v(k) = z(k) - \hat{x}(k-1)$ не буде суттєво впливати на оцінки стану. У випадку малого рівня шуму вимірів (або похибок статистичних даних) значення $K(k)$ буде зростати і на поточне значення оцінки стану $x(k)$ будуть суттєво впливати інновації (тобто поточні виміри).

Ця властивість фільтра Калмана узгоджується з поведінкою людини, яка рухається вночі по неосвітленій або погано освітленій дорозі. У такій ситуації “виміри” стану дороги дуже ненадійні внаслідок поганого освітлення (впадини та випадкові предмети можна не помітити). Як правило, людина рухається в таких випадках обережно, більше покладаючись на свою власну “внутрішню прогнозувальну машину”, ніж на вимірювальну інформацію від візуальної системи. Якщо надворі зовсім темно, то можна сказати, що $R \rightarrow \infty$ і коефіцієнт “фільтра” візуальної системи людини $K(k) \rightarrow 0$. У такому випадку людина може рухатись вперед тільки завдяки своїй “прогнозувальній машині”. Фільтр Калмана, реалізований програмно чи апаратно, фактично функціонує за подібною логікою.

Вплив збурення стану $w(k)$ на коефіцієнт фільтра також можна простежити за допомогою отриманих рівнянь. Однак його вплив не такий явний як шуму вимірів внаслідок того, що дисперсія збурення стану є в чисельнику і знаменнику рівняння для $K(k)$. Очевидно, що $w(k)$ безпосередньо впливає на апріорну коваріацію похибок оцінок $P'(k)$. Чим більше значення має дисперсія Q , тим значнішими будуть коваріації $P'(k)$ і $P(k)$. Це означає, що Q безпосередньо впливає на похибки оцінок і, таким чином, зменшує надійність оцінок $\hat{x}(k)$ в цілому.

Приклад 6.5.2. Класичний технічний приклад застосування фільтра у навігаційних системах [19].

Розглянемо прямолінійний рух літаючого апарата (літака) в повітряному

просторі. На рух апарата впливає випадкове прискорення, зумовлене неоднорідністю густини атмосфери. Положення літаючого апарата (ЛА) вимірюється в дискретні моменти часу з періодом дискретизації T_s . При цьому кожний вимір містить шумову складову, поява якої зумовлена методичними похибками вимірювального пристрою (радару) та впливом атмосфери на корисний радіосигнал.

Будемо вважати, що положення ЛА вимірюється у прямокутних координатах, а точність вимірів не залежить від часу та координат; похибки вимірів мають нульове середнє і некорельовані між собою. Таким чином, положення $z(k)$ можна описати рівнянням:

$$z(k) = x(k) + v(k), \quad (6.5.1)$$

де $z(k) = z(kT_s)$ – вимір положення ЛА; $x(k) = x(kT_s)$ – дійсне положення ЛА; $v(k)$ – випадкова завада вимірів, яка має такі статистичні характеристики:

$$E[v(k)] = 0, \quad \forall k; \quad E[v^2(k)] = \sigma_x^2 = \text{const}, \quad \forall k; \quad E[v(k)v(l)] = 0, \quad k \neq l. \quad (6.5.2)$$

Припустимо, що координати ЛА вимірюються незалежно, що дає можливість описати динаміку руху апарата по кожній координаті окремо за допомогою таких рівнянь:

$$x(k) = x(k-1) + v(k-1)T_s + \frac{1}{2}a(k-1)T_s^2, \quad (6.5.3)$$

$$v(k) = v(k-1) + a(k-1)T_s, \quad (6.5.4)$$

де $x(k)$ – положення ЛА в поточний момент kT_s ; $v(k) = \dot{x}(k)$ – швидкість руху ЛА; $a(k)$ – випадкове прискорення ЛА, величина якого покладається незмінною на протязі одного періоду дискретизації вимірів. Будемо також вважати, що середнє значення випадкового прискорення дорівнює нулю, а кореляція між значеннями $a(k)$ в різних періодах дискретизації відсутня, тобто

$$E[a(k)] = 0; \quad E[a^2(k)] = \sigma_a^2 = \text{const}, \quad \forall k; \quad E[a(k)a(l)] = 0, \quad k \neq l. \quad (6.5.5)$$

Для того щоб скористатись рівняннями оптимальної фільтрації, запишемо рівняння руху ЛА у формі простору станів:

$$\mathbf{x}(k) = \mathbf{F}\mathbf{x}(k-1) + \mathbf{G}a(k-1), \quad (6.5.6)$$

$$\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k) + v(k). \quad (6.5.7)$$

де

$$\mathbf{F} = \begin{bmatrix} 1 & T_s \\ 0 & 1 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \end{bmatrix}, \quad \mathbf{H} = [1 \ 0], \quad (6.5.8)$$

$\mathbf{z}(k) = [x(k) \ v(k)]^T$ – вектор стану ЛА.

Рівняння дискретного фільтра Калмана приймає вигляд:

$$\hat{\mathbf{x}}(k) = \mathbf{F}\hat{\mathbf{x}}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}\mathbf{F}\hat{\mathbf{x}}(k-1)]. \quad (6.5.9)$$

Величина $\hat{\mathbf{x}}(k, k-1) = \mathbf{F}\hat{\mathbf{x}}(k-1)$ – представляє прогноз значення вектора стану на основі попередньої оптимальної оцінки. Із врахуванням цього позначення рівняння фільтрації (6.5.9) можна переписати у вигляді:

$$\hat{\mathbf{x}}(k, k) = \hat{\mathbf{x}}(k, k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \hat{\mathbf{x}}(k, k-1)], \quad (6.5.10)$$

де $\hat{\mathbf{x}}(k, k)$ – оцінка вектора стану в момент k на основі інформації, наявної на цей момент, включно.

Матричний коефіцієнт підсилення фільтра $\mathbf{K}(k)$ визначається за формулою:

$$\mathbf{K}(k) = \mathbf{P}'(k)\mathbf{H}^T [\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}]^{-1}, \quad (6.5.11)$$

де $\mathbf{R} = \sigma_v^2$ – дисперсія шуму вимірів; $\mathbf{P}'(k)$ – апіорна (до отримання останнього виміру $\mathbf{z}(k)$) коваріаційна матриця похибок оцінювання, що оцінюється за формулою:

$$\mathbf{P}'(k) = \mathbf{F}\mathbf{P}(k-1, k-1)\mathbf{F}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T, \quad (6.5.12)$$

де

$$\mathbf{P}(k-1, k-1) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{P}'(k-1) \quad (6.5.13)$$

– апостеріорна коваріаційна матриця похибок оцінювання, отримана на попередньому кроці виконання алгоритму; $\mathbf{Q} = \sigma_a^2$ – дисперсія випадкового процесу (прискорення руху ЛА) $a(k)$.

Із врахуванням виразу (6.5.11) для обчислення коефіцієнта фільтра, (6.5.12) приймає вигляд:

$$\mathbf{P}(k-1, k-1) = \{\mathbf{I} - \mathbf{P}'\mathbf{H}^T [\mathbf{H}\mathbf{P}'\mathbf{H}^T + \mathbf{R}]^{-1}\mathbf{H}\}\mathbf{P}'. \quad (6.5.14)$$

Матриці \mathbf{P}' і \mathbf{P} мають різні значення, оскільки після врахування останнього виміру похибки оцінок зменшуються. Тобто між цими матрицями існує таке співвідношення:

$$\|\mathbf{P}\| < \|\mathbf{P}'\|.$$

Однак, у відповідності до рівняння (6.5.12), випадкові прискорення ЛА між моментами дискретизації вимірів призводять до збільшення похибок оцінювання. Фільтр переходить в усталений стан функціонування тоді, коли збільшення похибок на протязі періоду дискретизації внаслідок випадкових прискорень буде компенсуватися їх зменшенням завдяки приходу наступного виміру.

Аналіз фільтра в усталеному стані

В усталеному стані справедливі співвідношення:

$$\mathbf{P}'(k, k) = \mathbf{P}'(k-1, k-1) = \mathbf{P}', \quad \mathbf{P}(k, k) = \mathbf{P}(k-1, k-1) = \mathbf{P}, \quad (6.5.15)$$

а тому (6.5.12), (6.5.14) приймають вигляд:

$$\mathbf{P}' = \mathbf{F}\mathbf{P}\mathbf{F}^T + \mathbf{G}\mathbf{Q}\mathbf{G}^T, \quad (6.5.16)$$

$$\mathbf{P} = \{\mathbf{I} - \mathbf{P}'\mathbf{H}^T [\mathbf{H}\mathbf{P}'\mathbf{H}^T + \mathbf{R}]^{-1} \mathbf{H}\} \mathbf{P}', \quad (6.5.17)$$

Перепишемо (6.5.16) як

$$\mathbf{P}' - \mathbf{G}\mathbf{Q}\mathbf{G}^T = \mathbf{F}\mathbf{P}\mathbf{F}^T,$$

і підставимо у праву частину останньої рівності замість \mathbf{P} (6.5.17):

$$\mathbf{P}' - \mathbf{G}\mathbf{Q}\mathbf{G}^T = \mathbf{F}\{\mathbf{I} - \mathbf{P}'\mathbf{H}^T [\mathbf{H}\mathbf{P}'\mathbf{H}^T + \mathbf{R}]^{-1} \mathbf{H}\} \mathbf{P}' \mathbf{F}^T. \quad (6.5.18)$$

Якщо ввести позначення:

$$\mathbf{P}' = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \quad \text{де} \quad p_{12} = p_{21}, \quad (6.5.19)$$

і врахувати значення матриць, задані (6.5.8), то (6.5.18) можна переписати у вигляді:

$$\begin{bmatrix} p_{11} - \frac{\sigma_a^2 T_s^4}{4} & p_{12} - \frac{\sigma_a^2 T_s^3}{2} \\ p_{12} - \frac{\sigma_a^2 T_s^3}{2} & p_{22} - \sigma_a^2 T_s^2 \end{bmatrix} = \frac{1}{1 + \frac{p_{11}}{\sigma_x^2}} \times \quad (6.5.20)$$

$$\times \begin{bmatrix} p_{11} + 2p_{12}T_s + p_{22}T_s^2 + \frac{(p_{11}p_{22} - p_{12}^2)T_s^2}{\sigma_x^2} & p_{12} + p_{22}T_s + \frac{(p_{11}p_{22} - p_{12}^2)T_s}{\sigma_x^2} \\ p_{12} + p_{22}T_s + \frac{(p_{11}p_{22} - p_{12}^2)T_s}{\sigma_x^2} & p_{22} + \frac{p_{11}p_{22} - p_{12}^2}{\sigma_x^2} \end{bmatrix}.$$

Якщо порівняти елементи матриці зліва (6.5.20) елементам справа, то можна записати систему алгебраїчних рівнянь стосовно усталених коваріацій похибок оцінок стану ЛА:

$$\frac{p_{11}}{\sigma_x^2} = \frac{\sqrt{1+2r}(\sqrt{1+r}+1)^2}{r^2}, \quad (6.5.21)$$

$$\frac{p_{12}}{\sigma_x \sigma_a T_s} = \frac{1(\sqrt{1+2r}+1)^2}{2r}, \quad (6.5.22)$$

$$\frac{p_{22}}{\sigma_a^2 T_s^2} = \frac{1}{2}(\sqrt{1+2r}+1), \quad (6.5.23)$$

де $r = 4\sigma_x / (\sigma_a T_s^2)$ – параметр, який можна розглядати як відношення “шум/сигнал”, оскільки σ_x – середня квадратична похибка датчика, що вимірює положення ЛА; $\sigma_a T_s^2 / 2$ – середнє квадратичне відхилення ЛА щодо прискорення, яке зумовлене його випадковим прискоренням.

За допомогою рівнянь (6.5.21) – (6.5.24) можна знайти безрозмірні співвідношення, які характеризують елементи апостеріорної матриці дисперсій похибок оцінювання:

$$\frac{\hat{p}_{11}}{\sigma_x^2} = \frac{\sqrt{1+2r}}{r^2}(\sqrt{1+2r}-1)^2, \quad (6.5.25)$$

$$\frac{\hat{p}_{12}}{\sigma_x \sigma_a T_s} = \frac{1(\sqrt{1+2r}-1)^2}{2r}, \quad (6.5.26)$$

$$\frac{\hat{p}_{22}}{\sigma_a^2 T_s^2} = \frac{1}{2}(\sqrt{1+2r}-1). \quad (6.5.27)$$

Величини p_{11}/σ_x^2 і \hat{p}_{11}/σ_x^2 є відношеннями середнього значення квадрата похибки оцінки положення ЛА до дисперсії похибки σ_x^2 датчика положення до і після моменту вимірювання положення ЛА, відповідно.

Приклад 6.5.3. Припустимо, що лінійна система описується дискретною функцією передачі:

$$G(z) = \frac{Y(z)}{U(z)} = \frac{b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} = \frac{0,6 z^{-1}}{1 - 1,6 z^{-1} + z^{-2}}, \quad (6.5.28)$$

де u і y – вхід і вихід системи, відповідно; z^{-1} – оператор затримки у часі на один крок (ми використовували його раніше при формуванні функцій прогнозування методом мінімізації дисперсії оцінки прогнозу).

Модель (6.5.28) необхідно представити у просторі станів. Методика представлення моделей такого типу у просторі станів буде розглянута в наступному розділі, а зараз наведено тільки кінцевий результат – матриці \mathbf{A} і \mathbf{B} , побудовані на основі коефіцієнтів моделі:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ a_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix};$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 1,6 \end{bmatrix}; \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ -1,6 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0,6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0,6 \\ 0,96 \end{bmatrix}.$$

Модель (6.5.28) у просторі станів має вигляд:

$$\mathbf{x}(k) = \begin{bmatrix} 0 & 1 \\ -1 & 1,6 \end{bmatrix} \begin{bmatrix} x_1(k-1) \\ x_2(k-1) \end{bmatrix} + \begin{bmatrix} 0,6 \\ 0,96 \end{bmatrix} u(k-1), \quad (6.5.29)$$

де $x_1(k) = y(k)$; $x_2(k) = y(k-1)$ – компоненти вектора стану. Рівняння вимірів:

$$\mathbf{z}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{v}(k) = [1 \ 0] \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + v(k) = x_1(k) + v(k). \quad (6.5.30)$$

Для того щоб врахувати випадкові збурення, що діють на процес, у праву частину рівняння (6.5.29) необхідно ввести змінну $\mathbf{w}(k)$. Такі збурення, як правило, впливають на протікання реальних процесів, наприклад, у вигляді похибок неідеальної моделі, похибок виконання арифметичних операцій, неврахованих ефектів при моделюванні та інших факторів. Статистичні параметри випадкового процесу $\mathbf{w}(k)$ необхідно визначати апріорно, до впровадження фільтра в функціонуючу систему обробки даних. Якщо неможливо визначити точні значення статистичних параметрів, то використовують наближені, визначені у процесі статистичного моделювання фільтра.

Нехай коваріації похибок оцінок початкового стану процесу, а також коваріації збурень стану і похибок вимірів мають такі значення:

$$\mathbf{P}_0 = \mathbf{P}(0,0) = \begin{bmatrix} 0,5 & 0,2 \\ 0,2 & 0,5 \end{bmatrix}; \quad \mathbf{Q} = \begin{bmatrix} 0,3 & 0 \\ 0 & 0,1 \end{bmatrix}; \quad \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Використовуючи ці дані, обчислимо коефіцієнт фільтра для одного кроку:

$$\begin{aligned} \mathbf{P}(1,0) &= \mathbf{A} \mathbf{P}(0,0) \mathbf{A}^T + \mathbf{Q} = \begin{bmatrix} 0,8 & 0,6 \\ 0,6 & 1,24 \end{bmatrix}; \\ \mathbf{K}(1) &= \mathbf{P}(1,0) \mathbf{H}^T [\mathbf{H} \mathbf{P}(1,0) \mathbf{H}^T + \mathbf{R}]^{-1} = \begin{bmatrix} 0,55 & 0 \\ 0 & 0,33 \end{bmatrix}; \\ \mathbf{P}(1,1) &= [\mathbf{I} - \mathbf{K}(1) \mathbf{H}] \mathbf{P}(1,0) = \begin{bmatrix} 0,36 & 0,27 \\ 0,336 & 1,042 \end{bmatrix}. \end{aligned}$$

Далі можна обчислювати оцінку вектора стану з використанням виміру, який поступає в момент часу 1, і переходити до наступного циклу реалізації рівнянь фільтра.

Приклад 6.5.4. Розглянемо безперервний стаціонарний процес Гауса-Маркова з автокореляційною функцією

$$r_x(\tau) = \exp(-|\tau|).$$

Для одиничної дисперсії і нульового середнього спектральною функцією цього процесу є

$$S_x(s) = \frac{2}{1-s^2} = \frac{\sqrt{2}}{1+s} \cdot \frac{\sqrt{2}}{1-s},$$

а формуючий фільтр, який перетворює білий шум в процес Гауса-Маркова, має функцію передачі:

$$G(s) = \frac{x(s)}{w(s)} = \frac{\sqrt{2}}{1+s},$$

де $w(t)$ – білий гаусівський шум з одиничною дисперсією; $x(t)$ – процес Гауса-Маркова; s – змінна Лапласа. Рівняння стану для цього процесу має вигляд:

$$\dot{x}(t) = -x(t) + \sqrt{2} w(t).$$

Для того щоб застосувати фільтр Калмана необхідно згенерувати послідовність вимірів даного процесу. Виберемо період дискретизації вимірів $T_s = 0,045$ секунди з нульовим початковим значенням при $t = 0$.

Нехай дисперсія похибок вимірів складає $R = 0,65$. Перехідна матриця станів в даному випадку є скаляром і може бути визначена за допомогою експоненти:

$$A(k) = \exp(-0,045) \approx 0,956.$$

Оскільки між виміром вихідної величини процесу і станом $x(t)$ немає різниці, то матриця (скаляр) вимірів $H(k) = 1$. Дисперсія збурення (шуму) стану визначається за означенням так:

$$Q(k) = E[w^2(k)] = E \left[\int_0^{0,045} \sqrt{2} \exp(-u) w(u) du \times \int_0^{0,045} \sqrt{2} \exp(-u) w(u) du \right] =$$

$$= \int_0^{0,045} (\sqrt{2} \exp(-v))^2 dv \approx 0,165.$$

Для того щоб розпочати процес оцінювання необхідно задати початкові умови x_0 і P_0 . Оскільки процес починається в момент $t=0$, має нульове середнє та одиничну дисперсію, то покладемо початкові умови рівними $x_0 = 0$, $P_0 = 1$ при $t=0$. Тепер всі параметри фільтра відомі і можна розпочинати рекурсивну процедуру оцінювання.

В спеціальній літературі з проблем оцінювання параметрів і станів динамічних систем наведено багато прикладів успішного застосування оптимального фільтра. Спектр реальних задач, де він може бути використаний, є надзвичайно широким. Зокрема, відомі такі приклади його успішного застосування: навігаційні системи, аерокосмічна техніка, системи автоматичного керування технічними об'єктами і технологічними процесами, обробка часових рядів в експериментальній фізиці, сейсмологічних дослідженнях, економетриці, біології та інших галузях.

6.6 Оцінювання неспостережуваних компонент вектора стану за допомогою оптимального фільтра

При моделюванні динамічних систем виникають проблеми оцінювання невимірюваних компонент вектора стану. Наприклад, якщо до вектора стану входить третя похідна або вища, то вимірювати такі величини дуже складно або неможливо. Іншими прикладами можуть бути необхідність вимірювання високих температур при плавленні тугоплавких матеріалів, вимірювання відстаней до космічних об'єктів, вимірювання змінних, що характеризують функціонування центральної нервової системи.

Температури сплавів можуть бути настільки високими, що неможливо створити датчик, який може витримати таку температуру. Проблема оцінювання невимірюваних компонент вектора стану досить часто виникає при моделюванні економічних систем. Така задача виникає у випадках, коли відповідні статистичні дані не були зібрані або ці дані мають нерегулярний характер. Задачу оцінювання невимірюваних компонент можна розв'язати за допомогою оптимального фільтра.

Оцінювання невимірюваних компонент вектора стану можливе у випадку, коли коваріаційна матриця похибок оцінок має ненульові відповідні елементи, які дають можливість обчислити коефіцієнти фільтра, зв'язані з невимірюваними компонентами. Розглянемо цей механізм оцінювання. У даному випадку в рівнянні фільтрації

$$\hat{\mathbf{x}}(k) = \mathbf{A}(k)\hat{\mathbf{x}}(k-1) + \mathbf{K}(k)[\mathbf{z}(k) - \mathbf{H}\mathbf{A}(k)\hat{\mathbf{x}}(k-1)]$$

розмірність вектора вимірів $\mathbf{z}(k)$ є меншою розмірності вектора стану $\mathbf{x}(k)$, тобто, $\dim[\mathbf{z}] < \dim[\mathbf{x}]$.

Оптимальний коефіцієнт фільтра обчислюється за рівнянням:

$$\mathbf{K}(k) = \mathbf{P}'(k)\mathbf{H}^T [\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}]^{-1},$$

де $\dim[\mathbf{P}'(k)] = [n \times n]$; $\dim[\mathbf{H}^T] = [n \times r]$ за означенням; n – вимірність вектора стану і $\dim[\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}]^{-1} = [r \times r]$. Таким чином, $\dim[\mathbf{K}(k)] = [n \times r]$.

Вектор нев'язок (інновацій) $\mathbf{v}(k)$ в рівнянні оцінювання

$$\mathbf{v}(k) = \mathbf{z}(k) - \mathbf{H}\mathbf{A}\hat{\mathbf{x}}(k)$$

має розмірність $[r \times 1]$, а розмірність добутку $\dim[\mathbf{K}(k)\mathbf{v}(k)] = [n \times 1]$. Наприклад, якщо $\dim[\mathbf{x}] = [3 \times 1]$, а $\dim[\mathbf{z}] = [2 \times 1]$, то добуток $\mathbf{K}(k)\mathbf{v}(k)$ має вигляд:

$$\mathbf{K}(k)\mathbf{v}(k) = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \\ K_{31} & K_{32} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix},$$

а оцінка вектора стану визначається за рівняннями:

$$\begin{aligned} \hat{x}_1(k) &= \hat{x}_1(k, k-1) + K_{11}v_1 + K_{12}v_2, \\ \hat{x}_2(k) &= \hat{x}_2(k, k-1) + K_{21}v_1 + K_{22}v_2, \\ \hat{x}_3(k) &= \hat{x}_3(k, k-1) + K_{31}v_1 + K_{32}v_2. \end{aligned}$$

Таким чином, невимірювана третя компонента вектора стану може бути оцінена, якщо відповідні елементи матриць $\mathbf{P}'(k)$, $\mathbf{P}(k)$ і $\mathbf{K}(k)$ мають ненульові значення.

6.7 Функція прогнозування на основі оптимального фільтра

Успішне застосування фільтра Калмана розпочалось з навігаційних систем, у яких його використовують для фільтрації вимірів радіолокаційних станцій, оцінювання невимірюваних компонент сигналів, екстраполяції, розв'язання задачі комплексування інформації (об'єднання потоків даних, що поступають з різних джерел) з метою підвищення якості оцінок стану об'єкта та деяких інших задач. Успіх такого застосування пояснюється високою ступінню адекватності математичної моделі рухомого об'єкта, стан якого оцінюється, в тому числі досить високою точністю статистичного опису випадкових збурень і шумів вимірів. Область застосування алгоритмів оптимальної фільтрації постійно розширюється. На сьогодні фільтр застосовують, у тій чи іншій формі, практично в усіх областях техніки, експериментальній фізиці, біофізиці, економіці, фінансах та інших галузях досліджень.

Одним із напрямів сучасного застосування фільтра є оптимальне оцінювання та короткострокове прогнозування динаміки розвитку фінансово-економічних та процесів іншої природи. Така нагальна потреба застосування методів оптимальної обробки вибірок даних зумовлена тим, що у природі немає процесів (об'єктів), які функціонували б в ідеальних умовах. Фактично всі вони функціонують в умовах впливу різнорідних випадкових збурень та наявності шумів вимірів.

Структура рівняння у просторі станів є дуже зручною для використання з метою обчислення короткострокового прогнозу. Розглянемо дискретне рівняння динаміки стохастичної системи у просторі станів (детермінований керуючий вплив для простоти викладок не враховуємо):

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{w}(k), \quad (6.7.1)$$

де $\mathbf{w}(k)$ – збурюючий процес білого шуму з нульовим середнім та скінченною постійною коваріацією $\mathbf{Q}(k)$. Функцію прогнозування на один крок можна знайти як умовне математичне сподівання (6.7.1):

$$\hat{\mathbf{x}}(k+1|k) = E_k [\mathbf{x}(k+1)] = \mathbf{F}\mathbf{x}(k), \quad (6.7.2)$$

де $\hat{\mathbf{x}}(k+1)$ – прогноз на один крок на основі інформації на момент k , включно. Функцією (6.7.2) можна скористатись для обчислення оцінок прогнозів на довільну кількість кроків. Так, прогноз на два кроки має вигляд:

$$\hat{\mathbf{x}}(k+2) = \mathbf{F}\hat{\mathbf{x}}(k+1) = \mathbf{F} \cdot \mathbf{F}\mathbf{x}(k) = \mathbf{F}^2 \mathbf{x}(k),$$

і на довільну кількість кроків s :

$$\hat{\mathbf{x}}(k+s) = \mathbf{F}^s \mathbf{x}(k). \quad (6.7.3)$$

Очевидно, що дисперсія похибки прогнозу буде зростати пропорційно кількості кроків s . Похибка прогнозу на один і два кроки складає:

$$\begin{aligned} \mathbf{e}_f(1) &= \mathbf{x}(k+1) - \hat{\mathbf{x}}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{w}(k) - \mathbf{F}\mathbf{x}(k) = \mathbf{w}(k), \\ \mathbf{e}_f(2) &= \mathbf{x}(k+2) - \hat{\mathbf{x}}(k+2) = \mathbf{F}\mathbf{x}(k+1) + \mathbf{w}(k+1) - \mathbf{F}[\mathbf{F}\mathbf{x}(k)] = \\ &= \mathbf{F}\mathbf{x}(k+1) + \mathbf{w}(k+1) - \mathbf{F}[\mathbf{x}(k+1) - \mathbf{w}(k)] = \mathbf{w}(k+1) + \mathbf{w}(k). \end{aligned}$$

Тобто дисперсія прогнозу на s кроків визначається так:

$$\begin{aligned} \text{Var}[\mathbf{e}_f(s)] &= E\{[\mathbf{w}(k+s-1) + \mathbf{w}(k+s-2) + \dots + \mathbf{w}(k)] \times \\ &\quad \times [\mathbf{w}^T(k+s-1) + \mathbf{w}^T(k+s-2) + \dots + \mathbf{w}^T(k)]\} = \\ &= s \mathbf{Q}_w. \end{aligned}$$

Отриманий результат аналогічний тому, що був отриманий для прогнозування на основі різницевих рівнянь, оскільки в обох випадках ми користуємось лінійними моделями одного класу. Тобто (6.7.1) можна інтерпретувати як стохастичну авторегресію першого порядку. Незважаючи на те, що модель досить проста, вона дає можливість визначити високоякісну оцінку однокрокового прогнозу завдяки тому, що оцінки стану формуються з використанням оптимального коефіцієнта фільтра.

Прогнозування навантаження на телефонну лінію (неперервний час)

Приклад 6.7.1. Задача побудови моделі та прогнозування навантаження на телефонний канал розглядається у багатьох роботах. Навантаження на канал має коливальний характер з незначним додатним трендом. Максимальне навантаження припадає на зимові місяці, а мінімальне на середину літа, що пояснюється піком відпускового сезону. На лінійний тренд процесу накладається шумова компонента, зумовлена тим, що коливання мають не чисто гармонічний характер. В цілому процес описується сумою коливальної та лінійної компонент з адитивним білим шумом. Лінійна частина неперервної моделі має вигляд:

$$\ddot{x}(t) = f_1(t), \quad (6.7.4)$$

де $f_1(t)$ – білий шум з нульовим середнім. Коливання описані диференціальним рівнянням другого порядку:

$$\ddot{y}(t) + \omega^2 y(t) = f_2(t), \quad (6.7.5)$$

де $f_2(t)$ – білий шум, незалежний від $f_1(t)$. Введемо елементи вектора станів і запишемо рівняння (6.7.4) і (6.7.5) у формі простору станів:

$$x_1(t) = x(t), \quad (6.7.6)$$

$$x_2(t) = \dot{x}(t), \quad (6.7.7)$$

$$x_3(t) = y(t), \quad (6.7.8)$$

$$x_4(t) = \dot{y}(t). \quad (6.7.9)$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\omega^2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} 0 \\ f_1(t) \\ 0 \\ f_2(t) \end{bmatrix}. \quad (6.7.10)$$

Оскільки виміри мають дискретний характер з періодом дискретизації T_s , то модель перетворимо у дискретну форму:

$$\mathbf{x}(k+1) = \begin{bmatrix} 1 & T_s & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos(\omega T_s) & \sin(\omega T_s) \\ 0 & 0 & -\sin(\omega T_s) & \cos(\omega T_s) \end{bmatrix} \mathbf{x}(k) + \mathbf{w}(k). \quad (6.7.11)$$

Вигляд елементів правої нижньої субматриці розмірності $[2 \times 2]$ зумовлений виглядом розв'язку диференціального рівняння (6.7.5). Значення елементів коваріаційної матриці шуму стану об'єкта $\mathbf{Q}(k)$ залежать від амплітуд випадкових величин $f_1(t)$ і $f_2(t)$. З визначенням значень елементів матриці $\mathbf{Q}(k)$, як правило, існують проблеми, а тому частіше всього значення її елементів підбирають у процесі обчислювальних експериментів.

Рівняння вимірів в даному випадку має досить просту форму завдяки тому, що загальний вимір представляє собою суму лінійної та коливальної складових. Таким чином, вимір $z(k)$ – це скаляр, який визначається рівнянням:

$$z(k) = [1 \ 0 \ 1 \ 0] \mathbf{x}(k) + \mathbf{v}(k). \quad (6.7.12)$$

Оскільки виміри навантаження на телефонний канал є досить точними, то елементи коваріаційної матриці $\mathbf{R}(k)$ шумів вимірів $\mathbf{v}(k)$ в даному випадку мають відносно малі значення. Після визначення елементів матриць $\mathbf{F}(k)$, $\mathbf{Q}(k)$, $\mathbf{H}(k)$ і $\mathbf{R}(k)$ розв'язувалась задача прогнозування навантаження на телефонний канал на один крок. Як зазначалось вище, прогноз визначається добутком $\hat{\mathbf{x}}(k+1|k) = \mathbf{F}(k)\hat{\mathbf{x}}(k-1)$, тобто для прогнозування можна скористатись звичайним алгоритмом фільтрації, який складається з таких кроків:

1. Початкові умови (ініціалізація): задати $\hat{\mathbf{x}}(0)$, $\mathbf{P}(0)$, \mathbf{Q} і \mathbf{R} .
2. Обчислити коефіцієнт фільтра:

$$\mathbf{K}(k) = \mathbf{P}'(k)\mathbf{H}^T [\mathbf{H}\mathbf{P}'(k)\mathbf{H}^T + \mathbf{R}(k)]^{-1}.$$

3. Обчислити нову оцінку вектора стану:

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{K}(k)[\mathbf{z}(k) + \mathbf{H}\hat{\mathbf{x}}(k|k-1)].$$

4. Визначити елементи апостеріорної коваріаційної матриці:

$$\mathbf{P}(k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}]\mathbf{P}'(k).$$

5. Обчислити апіорну коваріаційну матрицю для наступного кроку і прогнозовану оцінку вектора стану:

$$\begin{aligned} \mathbf{P}'(k+1) &= \mathbf{F}(k)\mathbf{P}(k)\mathbf{F}^T(k) + \mathbf{Q}, \\ \hat{\mathbf{x}}(k+1|k) &= \mathbf{F}(k)\hat{\mathbf{x}}(k). \end{aligned}$$

Перейти на крок 2.

Для того щоб стабілізувати функціонування фільтра у перехідному процесі, використано перших 16 вимірів навантаження на канал для оцінювання елементів коваріаційної матриці похибок оцінок вектора стану $\mathbf{P}(0)$ (фактично це означає адаптацію моделі за допомогою експериментальних даних). Для отримання значень похибок оцінок використано зважений метод найменших квадратів з одночасним використанням всіх 16-и вимірів.

Використання наведеного вище алгоритму фільтрації розпочато з 17-го кроку і продовжувалось до того моменту, на який були відомі виміри. Іншим варіантом ініціалізації фільтра є використання довільних великих значень діагональних елементів матриці $\mathbf{P}(0)$. В результаті встановлено, що фільтр Калмана дозволяє обчислити досить точний (прийнятний) однокроковий

прогноз для вибраної змінної – навантаження на телефонну лінію. Загалом в спеціальній літературі можна знайти численні приклади застосування фільтра в різних галузях техніки (особливо в аерокосмічній), керуванні процесами різної природи і т. ін.

Фільтр Калмана також успішно застосовують для прогнозування динаміки змінних в багатьох сферах діяльності, в тому числі для прогнозування фінансових та економічних змінних і їх дисперсії, що підтверджено численними виконаними дослідженнями. Існуючі різновидності фільтра дають можливість охопити широкий клас стаціонарних та нестаціонарних процесів різної природи, а також нелінійні процеси, які можна апроксимувати лінійними моделями з прийнятною точністю.

Прогнозування фінансової змінної

Приклад 6.7.2. Розглянемо ряд даних, що описують дисперсію вартості акцій однієї з компаній, що входить до числа провідних на Нью-Йоркській фондовій біржі. На основі часового ряду, що характеризує динаміку ціни акцій, побудовано таку модель дисперсії ціни:

$$\varepsilon_1^2(k) = 0,0714 + 0,1187 \varepsilon_1^2(k-1) + 0,1123 \varepsilon_1^2(k-2) - 0,096 \varepsilon_1^2(k-4) + 0,1257 \varepsilon_1^2(k-5) + \varepsilon_2(k),$$

яка також представлена у просторі станів.

Прогноз дисперсії вартості акцій на 5 кроків виконано за допомогою трьох методів: безпосередньо за отриманою моделлю, за допомогою фільтра Калмана (який використовує ту ж модель) та за відомим методом подібних траєкторій.

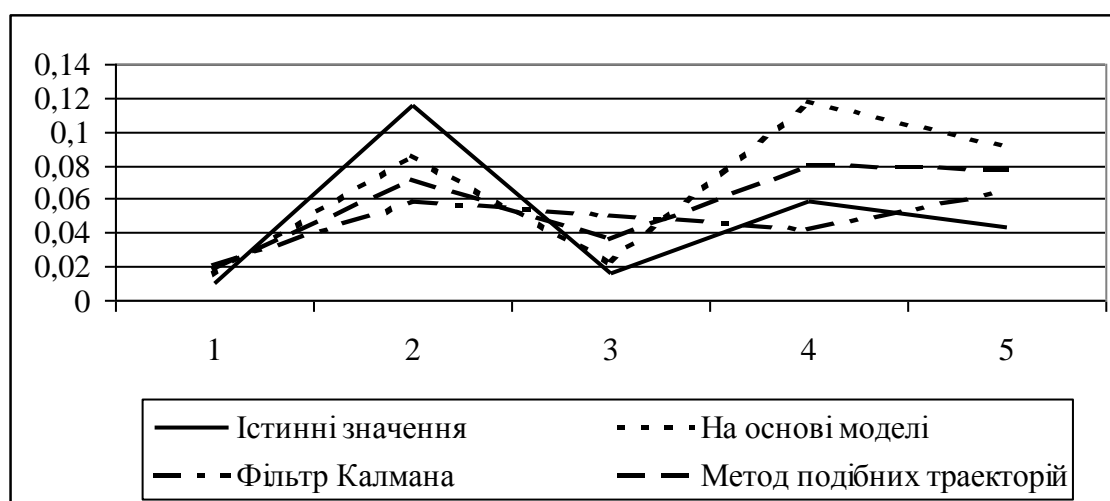


Рис. 6.1 Порівняльний графік прогнозу дисперсії вартості акцій компанії на Нью-Йоркській фондовій біржі

Таблиця 6.1

Метод прогнозування	Максимальне відхилення		Мінімальне відхилення		Сума квадратів похибок
	Абсолютне	%	Абсолютне	%	
За моделлю	0,055	73	0,003	8	0,0338
Фільтр Калмана	0,059	78	0,007	12	0,0337
Метод подібних траєкторій	0,058	77	0,007	10	0,0335

Графік, який ілюструє якість прогнозу, наведено на рис. 6.1 (по осі абсцис – кількість кроків прогнозування, а по осі ординат – значення похибки прогнозу).

Прогноз на один крок отримано за допомогою фільтра. З таблиці 6.1 видно (вона містить лише результати оцінювання якості прогнозу), що фільтр дозволяє також обчислити прийнятні значення прогнозу на п'ять кроків для досить складного процесу. Він є кращим за сумою квадратів похибок прогнозу. Численні обчислювальні експерименти з фільтром свідчать про те, що він, як правило, дає можливість знайти високоякісну оцінку однокрокового прогнозу. Однак, прогнозування на більше число кроків може супроводжуватись високими похибками оцінок. Це можна пояснити використанням відносно простої моделі у просторі станів, яка враховує тільки попередній стан системи. Основною проблемою застосування оптимального фільтра залишається побудова математичної моделі процесу високого ступеня адекватності. Досить непростою з обчислювальної точки зору є також задача оцінювання і прогнозування станів нелінійних систем, яка потребує коректної лінеаризації нелінійної моделі. Однак всі ці проблеми, як правило, вирішуються шляхом коректного застосування методів статистичної обробки даних і теорії оцінювання.

Запитання і вправи до шостого розділу

1. Запишіть вираз для поточного середнього і поясніть принцип рекурсивного оцінювання. Чи можна застосовувати рекурсивне оцінювання на довгих часових інтервалах, дайте докладне пояснення цій задачі?
2. Яка форма математичної моделі використовується у методі оптимальної фільтрації Калмана, які її переваги? Наведіть приклад моделі процесу, яка може бути використана для побудови алгоритму оптимальної фільтрації?
3. Дайте означення вільної динамічної системи? Що є причиною руху динамічної системи в даному випадку?
4. Запишіть модель у просторі станів загального вигляду для нестационарного динамічного процесу? Поясніть суть можливих нестационарностей.
5. Сформулюйте мету застосування оптимального фільтра. У чому полягає різниця між оптимальним і цифровим фільтром? Що необхідно знати про процес (дані вимірів) для того щоб застосувати цифровий фільтр?
6. Яким чином враховують статистичні параметри зовнішнього випадкового збурення $\mathbf{w}(k)$ і шумів вимірів $\mathbf{v}(k)$ в алгоритмі оптимальної фільтрації?
7. За яким методом можна отримати оцінки коваріаційних матриць \mathbf{Q} і \mathbf{R} ? Що представляють собою діагональні елементи цих матриць?
8. Поясніть суть квадратичного функціоналу:

$$J = E\{[\mathbf{x}(k) - \hat{\mathbf{x}}(k)]^T [\mathbf{x}(k) - \hat{\mathbf{x}}(k)]\},$$

на мінімізації якого критерію ґрунтується обчислення оптимального матричного коефіцієнта фільтра Калмана? Яким чином можна отримати точні значення вектора стану $\mathbf{x}(k)$?

9. Які початкові умови необхідно задати оптимальному фільтру? Як їх можна отримати? До чого призводять неправильні значення початкових умов?
10. Сформулюйте своїми словами суть процесу статистичного моделювання оптимального фільтра. З якою метою виконується статистичне моделювання?
11. Запишіть повністю алгоритм статистичного моделювання процесу оптимальної фільтрації для моделі AP(1).
12. Назвіть фактори, які можуть спричинити розбіжність процедури оптимальної фільтрації? Поясніть, що означає розбіжність на практиці?
13. Які заходи необхідно застосовувати для підвищення обчислювальної стійкості алгоритму фільтрації? Чи можна побудувати алгоритм оптимальної фільтрації без операції обернення матриці? У яких випадках?
14. Чому окремі компоненти вектора стану динамічної системи можуть бути не спостережуваними (не вимірюватись)?
15. Поясніть, яким чином можна оцінити невимірювані компоненти вектора стану системи за допомогою оптимального фільтра? Наведіть приклади, коли може виникнути така необхідність.
16. Рівняння для обчислення коефіцієнта фільтра можна отримати за допомогою диференціального числення. Для цього необхідно мінімізувати елементи головної діагоналі коваріаційної матриці похибок оцінок вектора стану, яка визначається рівнянням Ріккати:

$$\mathbf{P} = \mathbf{P}' - \mathbf{K}\mathbf{H}\mathbf{P}' - \mathbf{P}'\mathbf{H}^T\mathbf{K}^T + \mathbf{K}(\mathbf{H}\mathbf{P}'\mathbf{H}^T + \mathbf{R})\mathbf{K}^T.$$

Елементи головної діагоналі матриці \mathbf{P} необхідно мінімізувати за допомогою відповідного вибору значень коефіцієнта фільтра \mathbf{K} . Тобто можна мінімізувати слід цієї матриці $\text{trace } \mathbf{P} = p_{11} + p_{22} + \dots + p_{nn}$. При знаходженні похідних будуть корисними формули матричного числення:

$$(1) \frac{d[\text{trace}(\mathbf{A}\mathbf{B})]}{d\mathbf{A}} = \mathbf{B}^T, \quad (\text{матриця } \mathbf{A}\mathbf{B} \text{ повинна бути квадратною});$$

$$(2) \frac{d[\text{trace}(\mathbf{A}\mathbf{C}\mathbf{A}^T)]}{d\mathbf{A}} = 2\mathbf{A}\mathbf{C}, \quad (\text{матриця } \mathbf{C} \text{ повинна бути симетричною}).$$

Знайдіть вираз для оптимального коефіцієнта фільтра за цим методом.

Розділ 7

ПРЕДСТАВЛЕННЯ МАТЕМАТИЧНИХ МОДЕЛЕЙ У ПРОСТОРІ СТАНІВ

7.1 Вступ

У попередньому розділі ми використали представлення моделей у просторі станів (ПС) при формуванні рівнянь оптимального фільтра. Таке представлення є зручним та уніфікованим, а тому часто використовується в теорії моделювання, оцінювання, прогнозування і автоматичного керування. У цьому розділі розглянемо представлення у просторі станів дискретних моделей різних типів, зокрема, авторегресії та авторегресії з ковзним середнім, які вже були використані вище для опису динаміки процесів і побудови функцій прогнозування. Таке уніфіковане представлення математичних моделей у ПС спрощує їх подальше використання.

Від моделей у вигляді різницевих чи диференціальних рівнянь можна перейти до простору станів, який представляє собою уніфіковану форму представлення математичних моделей, зручною для застосування у багатьох випадках, наприклад, при прогнозуванні значень змінних та синтезі систем керування. Для неперервного часу лінійна модель у просторі станів має такий вигляд:

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{w}(t), \quad (7.1.1)$$

$$\mathbf{z}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{v}(t), \quad (7.1.2)$$

де $\mathbf{x}(t)$ – вектор змінних стану розмірності n ; $\dot{\mathbf{x}}(t)$ – перша похідна вектора стану відносно часу (швидкість зміни значень вектора стану); $\mathbf{A}(t)$ – матриця динаміки об'єкта (суб'єкта) чи процесу вимірності $[n \times n]$, яка у загальному випадку залежить від часу; $\mathbf{B}(t)$ – матриця коефіцієнтів керування вимірності $[n \times m]$, яка також може залежати від часу; $\mathbf{u}(t)$ – вектор сигналів керування вимірності $[m \times 1]$; $\mathbf{w}(t)$ – вектор збурень стану вимірності $[n \times 1]$; $\mathbf{z}(t)$ – вектор вимірів (експериментальних даних) вимірності $[r \times 1]$; $\mathbf{H}(t)$ – матриця коефіцієнтів вимірів розмірності $[r \times n]$; $\mathbf{v}(t)$ – вектор похибок (шумів) вимірів розмірності r . Наведена вище модель може бути, за необхідності, частково модифікована з метою її адаптації до конкретного типу процесів. Досить часто роблять припущення (яке необхідно апріорно чи апостеріорно довести для конкретного випадку) про те, що збурення стану $\mathbf{w}(t)$ та шуми вимірів $\mathbf{v}(t)$ – незалежні гаусові процеси з постійними дисперсіями та нульовими середніми значеннями, тобто:

$$E[\mathbf{w}(t)] = 0, \quad E[\mathbf{w}(t)\mathbf{w}^T(\tau)] = \mathbf{Q}(t)\delta(t - \tau); \quad (7.1.3)$$

$$E[\mathbf{v}(t)] = 0, \quad E[\mathbf{v}(t)\mathbf{v}^T(\tau)] = \mathbf{R}(t)\delta(t - \tau), \quad (7.1.4)$$

де $E[\cdot]$ – символ математичного сподівання; $\mathbf{w}^T(t)$ – означає операцію транспонування вектора або матриці; $\mathbf{Q}(t)$, $\mathbf{R}(t)$ – коваріаційні матриці збурень стану та шумів вимірів, відповідно (діагональні елементи цих матриць – це дисперсії складових векторів $\mathbf{w}(t)$ і $\mathbf{v}(t)$). Модель у просторі станів для дискретного часу має такий же вигляд як і (7.1.1), (7.1.2), але є різниця у методах отримання матриць \mathbf{A} , \mathbf{B} , \mathbf{H} . Нижче розглянемо приклади переходу від диференціальних та різницевих рівнянь до простору станів.

Приклад 7.1.1. Розглянемо модель моторної функції людини, яку часто описують диференціальним рівнянням другого порядку:

$$a_2 \frac{d^2 y(t)}{dt^2} + a_1 \frac{dy(t)}{dt} + y(t) = bu(t), \quad (7.1.5)$$

де $y(t)$ – сигнал (реакція) на виході суб'єкта; $u(t)$ – вхідний сигнал (через систему візуального або слухового сприйняття). Для того щоб представити рівняння (7.1.5) у просторі станів, приведемо його спочатку до вигляду:

$$\frac{d^2 y(t)}{dt^2} + \frac{a_1}{a_2} \frac{dy(t)}{dt} + \frac{1}{a_2} y(t) = \frac{b}{a_2} u(t), \quad (7.1.6)$$

і введемо позначення $x_1 = y(t)$, $x_2 = \dot{y}(t)$. Тепер,

$$\begin{aligned}\dot{x}_1(t) &= x_2(t), \\ \dot{x}_2(t) = \ddot{y}(t) &= -\frac{a_1}{a_2} \dot{y}(t) - \frac{1}{a_2} y(t) + \frac{b}{a_2} u(t).\end{aligned}$$

При таких позначеннях можна записати рівняння (7.1.5) у матричній формі так:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1/a_2 & -a_1/a_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ b/a_2 \end{bmatrix} u(t).$$

Таким чином, звичайне диференціальне рівняння другого порядку представлено у вигляді системи двох рівнянь першого порядку, тобто у формі Коші або у просторі станів. Рівняння в частинних похідних також можна представити в такій формі.

Розглянемо лінійне диференціальне рівняння n -го порядку, що описує систему з одним входом і одним виходом у неперервному часі:

$$(p^n + \alpha_{n-1}p^{n-1} + \dots + \alpha_1p + \alpha_0)y(t) = (\beta_n p^n + \beta_{n-1}p^{n-1} + \dots + \beta_1p + \beta_0)u(t), \quad (7.1.7)$$

де $u(t)$, $y(t)$ – вхід і вихід системи, відповідно; $p = d/dt$ – оператор диференціювання; $\alpha_i, \beta_i, i = 1, \dots, n$ – параметри моделі. На рис. 7.1 наведена структурна схема обчислень, яка відповідає цьому рівнянню.

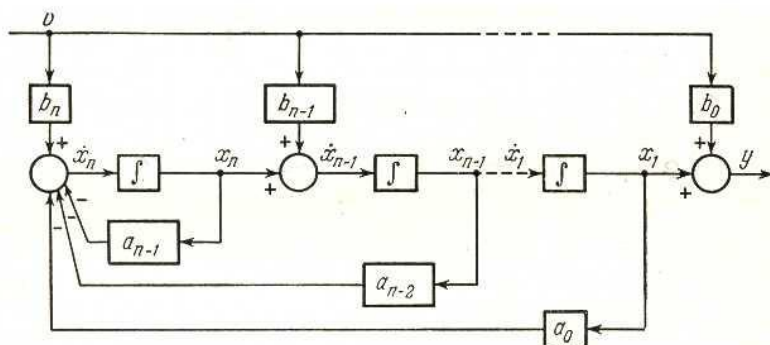


Рис. 7.1 Структурна схема рівняння (7.1.7)

Безпосередньо із цієї схеми можна записати, що:

$$\begin{aligned}y(t) &= x_1(t) + b_0 u(t), \\ \dot{x}_k(t) &= x_{k+1}(t) + b_k u(t), \quad \forall k < n, \\ \dot{x}_n(t) &= -[a_0 x_1(t) + a_1 x_2(t) + \dots + a_{n-1} x_n(t)] + b_n u(t).\end{aligned} \quad (7.1.8)$$

Запишемо похідну для $y(t)$:

$$p[y(t)] = \dot{y}(t) = \dot{x}_1(t) + b_0 \dot{u}(t),$$

і підставимо замість $\dot{x}_1(t)$ відповідний вираз із (7.1.8):

$$p[y(t)] = \dot{y}(t) = x_2(t) + b_1 u(t) + b_0 \dot{u}(t). \quad (7.1.9)$$

Тепер запишемо по аналогії старші похідні від $y(t)$:

$$\begin{aligned} p^2[y(t)] &= \dot{x}_2(t) + b_1 \dot{u}(t) + b_0 \ddot{u}(t) = x_3(t) + b_2 u(t) + b_1 \dot{u}(t) + b_0 \ddot{u}(t), \\ &\vdots \\ p^{n-1}[y(t)] &= x_n + b_{n-1} u(t) + b_{n-2} p[u(t)] + \dots + b_0 p^{n-1}[u(t)], \\ p^n[y(t)] &= -[a_0 x_1(t) + a_1 x_2(t) + \dots + a_{n-1} x_n(t)] + \\ &\quad + b_n u(t) + b_{n-1} p[u(t)] + \dots + b_0 p^n[u(t)]. \end{aligned} \quad (7.1.10)$$

Якщо підставити $y(t)$, $p[y(t)]$, ..., $p^{n-1}[y(t)]$, які визначаються рівняннями (7.1.8)–(7.1.10), в рівняння (7.1.7) і порівняти результат з виразом для $p^n[y(k)]$, що задається рівнянням (7.1.10), то отримаємо вирази для коефіцієнтів a_i , b_i у вигляді:

$$a_i = \alpha_i, \quad (7.1.11)$$

$$b_0 = \beta_n,$$

$$b_1 = \beta_{n-1} - \alpha_{n-1} b_0,$$

$$b_2 = \beta_{n-2} - \alpha_{n-1} b_1 - \alpha_{n-2} b_0,$$

$$\vdots$$

$$b_n = \beta_0 - \alpha_{n-1} b_{n-1} - \alpha_{n-2} b_{n-2} - \dots - \alpha_0 b_0. \quad (7.1.12)$$

Рівняння (7.1.12) – це зручна форма запису виразу для коефіцієнтів b_i , які можна знайти шляхом послідовних підстановок. Так, з рівняння (7.1.12) випливає, що β_i можна записати у вигляді:

$$\begin{bmatrix} \beta_n \\ \beta_{n-1} \\ \beta_{n-2} \\ \vdots \\ \beta_0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \alpha_{n-1} & 1 & 0 & \dots & 0 \\ \alpha_{n-2} & \alpha_{n-1} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \alpha_0 & \alpha_1 & \dots & \alpha_{n-1} & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

а звідси можна визначити коефіцієнти b_i за допомогою оберненої матриці. Таким чином, згідно із отриманими виразами, однією з можливих форм представлення матриць **A**, **B** і **H** моделі у просторі станів є така:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ -\alpha_0 & -\alpha_1 & -\alpha_2 & \cdots & -\alpha_{n-1} \end{bmatrix}, \quad (7.1.13)$$

$$\mathbf{B} = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \alpha_{n-1} & 1 & 0 & \cdots & 0 \\ \alpha_{n-2} & \alpha_{n-1} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ \alpha_0 & \alpha_1 & \cdots & \alpha_{n-1} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_n \\ \beta_{n-1} \\ \beta_{n-2} \\ \vdots \\ \beta_0 \end{bmatrix}, \quad (7.1.14)$$

$$\mathbf{H} = [1 \ 0 \ 0 \ \cdots \ 0]. \quad (7.1.15)$$

Приклад 7.1.2. Привести до форми простору станів диференціальне рівняння третього порядку:

$$\ddot{y}(t) + 3\dot{y}(t) + 4y(t) = 2\ddot{u}(t) + 3\dot{u}(t) + \dot{u}(t) + 2u(t).$$

За допомогою рівняння (7.1.12) отримаємо:

$$\begin{aligned} b_0 &= \beta_n = 2, \\ b_1 &= \beta_{n-1} - \alpha_{n-1}b_0 = 3 - 3 \cdot 2 = -3, \\ b_2 &= \beta_{n-2} - \alpha_{n-1}b_1 - \alpha_{n-2}b_0 = 1 - 3 \cdot (-3) - 4 \cdot 2 = 2, \\ b_3 &= \beta_{n-3} - \alpha_{n-1}b_2 - \alpha_{n-2}b_1 - \alpha_{n-3}b_0 = 2 - 3 \cdot 2 - 4 \cdot (-3) - 1 \cdot 2 = 6. \end{aligned}$$

Згідно з представленням (7.1.13), (7.1.14) запишемо модель в просторі станів:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -4 & -3 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} + \begin{bmatrix} -3 \\ 2 \\ 6 \end{bmatrix} u(t).$$

З попередніх підрозділів видно, що методи, які ми розглядаємо, спрямовані на дослідження дискретних систем у просторі станів. Перед тим як перейти до представлення дискретних систем у просторі станів, розглянемо перехід від лінійних неперервних до лінійних дискретних систем.

7.2 Перехід від лінійних неперервних до лінійних дискретних систем

Багатовимірну лінійну стаціонарну систему можна описати системою лінійних диференціальних рівнянь у векторній формі

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t), \quad (7.2.1)$$

де \mathbf{A} – матриця з постійними коефіцієнтами вимірності $[n \times n]$; $\mathbf{x}(t)$ – вектор-стовпчик змінних стану вимірності $[n \times 1]$. Подібно розв'язку скалярних диференціальних рівнянь розв'язок рівняння (7.2.1) можна записати у вигляді:

$$\mathbf{x}(t) = \mathbf{e}^{\mathbf{A}(t-\tau)} \mathbf{x}(\tau), \quad (7.2.2)$$

де матриця $\mathbf{e}^{\mathbf{A}t}$ може бути представлена у вигляді розкладу в нескінченний ряд:

$$\mathbf{e}^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \dots \quad (7.2.3)$$

При $t = \tau$ матриця $\mathbf{e}^{\mathbf{A}(t-\tau)}$ стає одиничною, тобто граничні умови виконуються:

$$\mathbf{x}(\tau) = \mathbf{e}^{\mathbf{A}0} \mathbf{x}(\tau) = \mathbf{I}\mathbf{x}(\tau) = \mathbf{x}(\tau).$$

Матрицю $\mathbf{F}(t) = \mathbf{e}^{\mathbf{A}t}$ називають *перехідною матрицею стану* або *фундаментальною матрицею* системи (7.2.1).

Перехідна матриця стану “описує” рух кінця вектора стану у просторі станів з початкового значення до наступного (заданого кінцевим моментом часу), а тому вона характеризує (описує) перехід системи із поточного стану у наступний. Очевидно, що вектор стану $\mathbf{x}(t) = [x_1(t) \dots x_n(t)]^T$ містить у собі значну інформацію про лінійну систему. Перехідну матрицю стану $\mathbf{F}(t)$ можна визначати різними методами. Відомі підходи ґрунтуються на теоремі Сільвестра та Келі-Гамільтона, методі розкладання в нескінченний ряд, перетворенні Лапласа та методі передавальної функції.

Метод розкладання в нескінченний ряд. За означенням, $\mathbf{e}^{\mathbf{A}t}$, перехідна матриця обчислюється за нескінченним рядом:

$$\mathbf{e}^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \dots$$

Цей метод потребує найбільше обчислювальних витрат, якщо значення \mathbf{A}^k не збігаються до нуля при збільшенні k . Після обчислення суми

необхідно знайти ще прийнятні („замкнені”) вирази для кожного елемента матриці $\mathbf{F}(t)$.

Приклад 7.2.1. Знайти перехідну матрицю стану для матриці:

$$\mathbf{A} = \begin{bmatrix} 0 & -2 \\ 1 & -3 \end{bmatrix}.$$

Знайдемо \mathbf{A}^2 і \mathbf{A}^3 :

$$\mathbf{A}^2 = \mathbf{A} \cdot \mathbf{A} = \begin{bmatrix} -2 & 6 \\ -3 & 7 \end{bmatrix}, \quad \mathbf{A}^3 = \mathbf{A}^2 \cdot \mathbf{A} = \begin{bmatrix} 6 & 14 \\ 7 & -15 \end{bmatrix}.$$

Таким чином, перехідна матриця визначається як

$$\mathbf{F}(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -2 \\ 1 & -3 \end{bmatrix} t + \begin{bmatrix} -2 & 6 \\ -3 & 7 \end{bmatrix} \frac{t^2}{2!} + \begin{bmatrix} 6 & 14 \\ 7 & -15 \end{bmatrix} \frac{t^3}{3!} + \dots$$

або

$$\mathbf{F}(t) = \begin{bmatrix} 1 - \frac{2t^2}{2!} + \frac{6t^3}{3!} + \dots & -2t + \frac{6t^2}{2!} + \frac{14t^3}{3!} + \dots \\ t - \frac{3t^2}{2!} + \frac{7t^3}{3!} + \dots & 1 - 3t + \frac{7t^2}{2!} - \frac{15t^3}{3!} + \dots \end{bmatrix}.$$

Обчислюючи нескінченні ряди для кожного з елементів матриці (це основний недолік даного методу), отримаємо:

$$\mathbf{F}(t) = \begin{bmatrix} 2e^{-t} - e^{2t} & 2(e^{-2t} - e^{-t}) \\ e^{-t} - e^{-2t} & 2e^{-2t} - e^{-t} \end{bmatrix}.$$

Якщо позначити період дискретизації вимірів через T_s , то перехід лінійної системи з одного стану в інший на протязі періоду дискретизації легко визначити за допомогою підстановки $t = T_s$.

Метод на основі перетворення Лапласа (метод комплексної частоти). Перетворення Лапласа для рівняння динаміки (7.2.1) має вигляд: $s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s)$, тобто,

$$\mathbf{X}(s) = [s\mathbf{I} - \mathbf{A}]^{-1} \mathbf{x}(0). \quad (7.2.4)$$

Застосуємо до (7.2.4) зворотне перетворення Лапласа:

$$\mathbf{x}(t) = L^{-1} \{ [s\mathbf{I} - \mathbf{A}]^{-1} \} \mathbf{x}(0). \quad (7.2.5)$$

Порівнюючи (7.2.5) з (7.2.2), отримаємо:

$$\mathbf{F}(t) = e^{\mathbf{A}t} = L^{-1} \{ [s\mathbf{I} - \mathbf{a}]^{-1} \}. \quad (7.2.6)$$

Даний метод є одним із самих зручних при розв'язуванні багатьох задач. Очевидно, що труднощі обчислювального характеру можуть виникати при знаходженні оберненої матриці.

Приклад 7.2.2. Обчислити $\mathbf{F}(t)$ для матриці \mathbf{A} з попереднього прикладу за методом перетворення Лапласа. Для матриці

$$[s\mathbf{I} - \mathbf{A}] = \begin{bmatrix} s & 2 \\ -1 & s+3 \end{bmatrix}$$

обернена має вигляд:

$$\Phi(s) = [s\mathbf{I} - \mathbf{A}]^{-1} = \frac{1}{\Delta} \begin{bmatrix} s+3 & -2 \\ 1 & s \end{bmatrix},$$

де $\Delta = s^2 + 3s + 2$, а матриця $\mathbf{F}(t) = L^{-1}[\Phi(s)]$.

Крім розглянутих методів, перехідну матрицю станів можна знайти за допомогою деяких інших методів, наприклад, передавальної функції.

7.3 Дискретні рівняння типу АРКС у просторі станів

У дискретному часі динаміка систем описується різницевиими рівняннями, для яких також розроблено процедури переходу до простору станів. Наприклад, можна побудувати авторегресійні моделі для змінних, що характеризують поточний стан людини і представити отриману модель у просторі станів.

Приклад 7.3.1. Нехай для характеристики поточного стану людини вибрано такі змінні: $x_1(k)$ – температура тіла ($k=0,1,2,\dots$ – дискретний час); $x_2(k)$ – артеріальний тиск крові; $x_3(k)$ – частота пульсу; $x_4(k)$ – електричний опір шкіри; $x_5(k)$ – вміст цукру в крові. Якщо вимірювати ці змінні два рази на добу, то за місяць отримаємо для кожної із них вибірку потужністю 60 вимірів, яку можна використати для побудови моделей у вигляді авторегресійних рівнянь, що описуватимуть динаміку отриманих часових рядів. Очевидно, що для поглибленого аналізу процесів в організмі людини

необхідно вибрати складніші структури моделей, які будуть містити у собі незалежні змінні у правій частині рівнянь та можливі нелінійності. Але якщо зупинитися на авторегресії першого порядку, то отримаємо таку систему рівнянь:

$$\begin{aligned}x_1(k) &= a_{10} + a_{11}x_1(k-1) + \varepsilon_1(k), \\x_2(k) &= a_{20} + a_{21}x_2(k-1) + \varepsilon_2(k), \\x_3(k) &= a_{30} + a_{31}x_3(k-1) + \varepsilon_3(k), \\x_4(k) &= a_{40} + a_{41}x_4(k-1) + \varepsilon_4(k), \\x_5(k) &= a_{50} + a_{51}x_5(k-1) + \varepsilon_5(k),\end{aligned}$$

яку можна представити у просторі станів у вигляді

$$\mathbf{x}(k) = \mathbf{F}(k)\mathbf{x}(k-1) + \mathbf{a} + \mathbf{w}(k),$$

де $\mathbf{x}^T(k) = [x_1(k) \ x_2(k) \ x_3(k) \ x_4(k) \ x_5(k)]$ – вектор стану; $\mathbf{a}^T = [a_{10} \ a_{20} \ a_{30} \ a_{40} \ a_{50}]$ – вектор постійних значень; $\mathbf{F}(k)$ – матриця динаміки дискретної системи; $\mathbf{w}^T(k) = [\varepsilon_1(k) \ \varepsilon_2(k) \ \varepsilon_3(k) \ \varepsilon_4(k) \ \varepsilon_5(k)]$. Матриця динаміки і вектор $\mathbf{x}(k-1)$ для даної системи мають вигляд:

$$\begin{aligned}\mathbf{F}(k) &= \text{diag}[a_{11} \ a_{21} \ a_{31} \ a_{41} \ a_{51}]; \\ \mathbf{x}^T(k) &= [x_1(k) \ x_2(k) \ x_3(k) \ x_4(k) \ x_5(k)].\end{aligned}$$

Очевидно, що розглянута модель, але побудована для відхилень від середнього, не буде містити вектора констант \mathbf{a} .

Цей приклад можна розглядати як спрощене представлення у просторі станів, оскільки керуючого впливу у правій частині нема. У цілому отриману систему у просторі станів не можна вважати вільною, тому що в правій частині наявні вектори зміщення \mathbf{a} та збурень $\mathbf{w}(k)$. Нагадаємо, що вільною називають таку динамічну систему, рух якої відбувається під впливом тільки початкових умов.

Авторегресія з ковзним середнім у просторі станів

Розглянемо лінійну систему, яка описується авторегресією з ковзним середнім довільного порядку. Нехай зв'язок між входом і виходом процесу визначається за допомогою різницевого рівняння:

$$y(k) + a_1 y(k-1) + a_2 y(k-2) + \dots + a_n y(k-n) =$$

$$= b_0 u(k) + b_1 u(k-1) + b_2 u(k-2) + \dots + b_m u(k-m),$$

або

$$\begin{aligned} y(k+n) + a_1 y(k+n-1) + \dots + a_n y(k) &= \\ &= b_0 u(k+n) + b_1 u(k+n-1) + b_2 u(k+n-2) + \dots + b_m u(k), \end{aligned} \quad (7.3.1)$$

де $u(k)$, $y(k)$ – вхід та вихід процесу, відповідно. Якщо ввести позначення $y(k-1) = z^{-1}y(k)$ (тобто скористаємось оператором зсуву), то наведене рівняння можна переписати у вигляді:

$$(1 + a_1 z^{-1} + \dots + a_n z^{-n})y(k) = (b_0 + b_1 z^{-1} + \dots + b_m z^{-m})u(k).$$

Оператор z^{-1} називають оператором зсуву (в часі). Рівняння (7.3.1) можна представити у вигляді дискретної функції передачі між виходом і входом:

$$G(z) = \frac{y(z)}{u(z)} = \frac{b_1 z^{-1} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + \dots + a_n z^{-n}},$$

або при $m = n$:

$$G(z) = \frac{y(z)}{u(z)} = \frac{b_0 z^n + b_1 z^{n-1} + \dots + b_n}{z^n + a_1 z^{n-1} + \dots + a_n}. \quad (7.3.2)$$

Зазначимо, що для процесів, які можуть бути фізично реалізованими, необхідно виконувати умову $m \leq n$, що забезпечує стійкість їх функціонування. Інакше буде спостерігатись розбіжність, яка характеризує аномальний розвиток динаміки процесу.

Введемо такі змінні стану:

$$\left. \begin{aligned} x_1(k) &= y(k) \\ x_2(k) &= x_1(k+1) = y(k+1) \\ x_3(k) &= x_2(k+1) = y(k+2) \\ &\vdots \\ x_n(k) &= x_{n-1}(k+1) = y(k+n-1) \\ x_n(k+1) &= y(k+n) \end{aligned} \right\} \quad (7.3.3)$$

Підставимо вирази (7.3.3) в рівняння (7.3.1) при $b_n = 1$ і $b_0, b_1, \dots, b_{n-1} = 0$. В результаті отримаємо:

$$y(k+n) = x_n(k+1) = -a_1 x_n(k) - a_2 x_{n-1}(k) - \dots - a_n x_1(k) + 1u(k). \quad (7.3.4)$$

Це співвідношення можна представити у формі простору станів:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \vdots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} u(k), \quad (7.2.5)$$

а рівняння вимірів має вигляд:

$$y(k) = [1 \ 0 \ \cdots \ 0] \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix}. \quad (7.3.6)$$

Тепер рівняння (7.3.1) можна представити у просторі станів таким чином:

$$\mathbf{x}(k+1) = \mathbf{F}\mathbf{x}(k) + \mathbf{B}u(k), \quad (7.3.7)$$

$$\mathbf{y}(k) = \mathbf{H}^T \mathbf{x}(k), \quad (7.3.8)$$

де матриці \mathbf{F} , \mathbf{B} , \mathbf{H} визначають у відповідності до (7.3.5) та (7.3.6).

Якщо $b_n = 1$, а $b_0, b_1, \dots, b_{n-1} = 0$, то рівняння (7.3.2), (7.3.3) можна представити у вигляді:

$$y(z) = \frac{1}{z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_n} u(z) = x_1(z). \quad (7.3.9)$$

Якщо ж $b_n \neq 1$, а $b_0, b_1, \dots, b_{n-1} \neq 0$, то із умови стійкості можна записати, що

$$b_n + b_{n-1}z + b_{n-2}z^2 + \dots + b_0z^n = 1. \quad (7.3.10)$$

Якщо помножити ліву і праву частини (7.3.10) на $y(z)$, то отримаємо:

$$y(z) = b_n y(z) + b_{n-1}z y(z) + \dots + b_0 z^n y(z),$$

а із врахуванням того, що $y(z) = x_1(z)$ маємо:

$$y(z) = b_n x_1(z) + b_{n-1} z x_1(z) + \dots + b_0 z^n x_1(z)$$

або

$$y(k) = b_n x_1(k) + b_{n-1} x_1(k+1) + \dots + b_0 x_1(k+n). \quad (7.3.11)$$

Використовуючи позначення (7.3.3), перейдемо до такого запису (7.3.11):

$$y(k) = b_n x_1(k) + b_{n-1} x_2(k) + b_{n-2} x_3(k) + \dots + b_0 x_n(k+1). \quad (7.3.12)$$

З рівняння (7.3.4) отримуємо:

$$x_n(k+1) = -a_1 x_n(k) - a_2 x_{n-1}(k) - \dots - a_n x_1(k) + 1u(k), \quad (7.3.13)$$

а в результаті підстановки правої частини (7.3.13) в (7.3.12) замість $x_n(k+1)$ маємо:

$$y(k) = (b_n - b_0 a_n) x_1(k) + (b_{n-1} - b_0 a_{n-1}) x_2(k) + \dots + (b_1 - b_0 a_1) x_n(k) + b_0 u(k). \quad (7.3.14)$$

Останнє рівняння представляє собою рівняння вимірів, в якому враховано вплив чисельника передаточної функції, тобто воно приймає вигляд:

$$y(k) = [(b_n - b_0 a_n) \dots (b_1 - b_0 a_1)] \begin{bmatrix} x_1(k) \\ \vdots \\ x_n(k) \end{bmatrix} + b_0 u(k),$$

або

$$y(k) = \mathbf{H}^T \mathbf{x}(k) + d u(k). \quad (7.3.15)$$

При $b_0 = 0$, тобто для систем без прямої передачі керуючого сигналу, рівняння (7.3.15) приймає спрощений вигляд:

$$y(k) = [b_n \dots b_1] \begin{bmatrix} x_1(k) \\ \vdots \\ x_n(k) \end{bmatrix}. \quad (7.3.16)$$

Другий підхід до представлення моделі (7.3.1) у просторі станів

У даному випадку матриця \mathbf{F} в (7.3.7) визначається так само як і в попередньому випадку, а матриця \mathbf{B} визначається за виразом:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdot & \cdot & 0 \\ a_1 & 1 & 0 & 0 & \cdot & \cdot & 0 \\ a_2 & a_1 & 1 & 0 & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & 0 \\ a_{n-1} & \cdot & \cdot & \cdot & \cdot & a_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix} = (\mathbf{A}')^{-1} \mathbf{b}. \quad (7.3.17)$$

Матрицю \mathbf{F} називають коагульованою матричною формою; \mathbf{A}' – нижня трикутна матриця [9]. Покажемо, що представлення у просторі станів,

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \vdots \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \vdots \\ x_n(k) \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} u(k), \quad (7.3.18)$$

відповідає початковій моделі (7.3.1). В (7.3.18)

$$\mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_n], \quad \mathbf{B}^T = [\beta_1 \ \beta_2 \ \dots \ \beta_n]. \quad (7.3.19)$$

Вище було показано, що

$$x_1(k) = y(k). \quad (7.3.20)$$

У відповідності до (7.3.18) та (7.3.19) можна записати:

$$\begin{aligned} x_1(k) &= x_2(k-1) + \beta_1 u(k-1) \\ \text{або} \quad x_2(k) &= x_1(k+1) - \beta_1 u(k) = y(k+1) - \beta_1 u(k). \end{aligned}$$

За аналогією можна отримати:

$$x_i(k) = y(k+i-1) - \sum_{j=2}^i \beta_{j-1} u(k+i-j), \quad i \leq n. \quad (7.3.21)$$

Відповідно, $x_n(k)$ визначається за рівнянням:

$$x_n(k) = y(k+n-1) - \beta_1 u(k+n-2) - \dots - \beta_{n-1} u(k-1). \quad (7.3.22)$$

Прирівнюємо нижній рядок матричного рівняння (7.3.18) до виразу (7.3.22) і скористаємося матрицею \mathbf{F} , що визначається (7.3.18). В результаті отримаємо:

$$\begin{aligned} -a_n x_1(k-1) - \dots - a_1 x_n(k-1) + \beta_n u(k-1) = \\ = y(k+n-1) - \beta_1 u(k+n-2) - \dots - \beta_{n-1} u(k). \end{aligned} \quad (7.3.23)$$

Тепер підставимо вираз для $x_i(k)$, $i=1, \dots, n-1$ із рівняння (7.3.21) в останнє рівняння:

$$\begin{aligned} y(k+n-1) - \beta_1 u(k+n-2) - \dots - \beta_{n-1} u(k) = -a_n y(k-1) - \dots - a_1 y(k+n-2) + \\ + a_{n-1} \beta_1 u(k-1) + a_{n-2} [\beta_1 u(k) + \beta_2 u(k-1)] + \\ + \dots + a_1 [\beta_1 u(k+n-3) + \dots + \beta_{n-1} u(k-1)] + \beta_n u(k-1). \end{aligned} \quad (7.3.24)$$

Зробимо перегрупування членів в останньому рівнянні:

$$\begin{aligned} y(k+n-1) + a_1 y(k+n-2) + a_2 y(k+n-3) + \dots + a_n y(k-1) = \\ = a_{n-1} \beta_1 u(k-1) + a_{n-2} [\beta_1 u(k) + \beta_2 u(k-1)] + \dots + \\ + a_1 [\beta_1 u(k+n-3) + \dots + \beta_{n-1} u(k-1)] + \\ + \beta_1 u(k+n-2) + \dots + \beta_n u(k-1) = \\ = b_1 u(k+n-2) + b_2 u(k+n-3) + \dots + b_m u(k+n-m-1). \end{aligned} \quad (7.3.25)$$

Таким чином,

$$\begin{aligned} b_1 &= \beta_1, \\ b_2 &= \beta_2 + a_1 \beta_1, \\ &\vdots \\ b_m &= \beta_m + a_1 \beta_{m-1} + \dots + a_{m-1} \beta_1, \\ 0 &= \beta_j + a_1 \beta_{j-1} + \dots + a_{j-1} \beta_1, \quad n \geq j > m. \end{aligned} \quad (7.3.26)$$

Рівняння (7.3.26) можна записати таким чином:

$$\begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ b_m \\ 0 \\ \cdot \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdot & \cdot & 0 \\ a_1 & 1 & 0 & 0 & \cdot & \cdot & 0 \\ a_2 & a_1 & 1 & 0 & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & 0 \\ a_{n-1} & \cdot & \cdot & \cdot & \cdot & a_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix} = \mathbf{A}' \mathbf{B} \quad (7.3.27)$$

або

$$\mathbf{B} = (\mathbf{A}')^{-1} \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ b_m \\ 0 \\ \cdot \\ 0 \end{bmatrix} = (\mathbf{A}')^{-1} \mathbf{b}, \quad (7.3.28)$$

де \mathbf{b} – n – вимірний вектор; $m \leq n$. Оскільки рівняння (7.3.25) задовольняє рівнянням (7.3.20) – (7.3.21), то перетворення (7.3.1) до простору станів закінчено.

Приклад 7.3.2. Розглянемо функцію передачі другого порядку в дискретному часі

$$\begin{aligned} G(z) &= \frac{0,5z}{z^2 - 1,3z + 0,9} = \frac{0,5z^{-1}}{1 - 1,3z^{-1} + 0,9z^{-2}} = \\ &= \frac{b_1z^{-1}}{1 + a_1z^{-1} + a_2z^{-2}} = \frac{y(z)}{u(z)}. \end{aligned}$$

Для того щоб отримати дискретну модель цієї системи у просторі станів, обчислимо матриці \mathbf{F} , \mathbf{B} у відповідності до (7.3.5) та (7.3.6):

$$\mathbf{F} = \begin{bmatrix} 0 & 1 \\ -0,9 & +1,3 \end{bmatrix};$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 \\ -1,3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0,5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1,3 & 1 \end{bmatrix} \begin{bmatrix} 0,5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0,5 \\ 0,65 \end{bmatrix}.$$

В результаті отримаємо:

$$\begin{aligned} x_1(k) &= x_2(k-1) + 0,5u(k-1), \\ x_2(k) &= -0,9x_1(k-1) + 1,3x_2(k-2) + 0,65u(k-1), \\ y(k) &= x_1(k). \end{aligned}$$

Приклад 7.3.3. Розглянемо різницеве рівняння третього порядку

$$y(k) = 0,5y(k-1) - 0,25y(k-2) + 0,4y(k-3) + 0,8u(k-1) - 0,6u(k-2),$$

якому відповідає функція передачі

$$G(z) = \frac{y(z)}{u(z)} = \frac{0,8z^{-1} - 0,6z^{-2}}{1 - 0,5z^{-1} + 0,25z^{-2} - 0,4z^{-3}}.$$

Отримаємо модель цієї системи в просторі станів у відповідності до (7.3.5) та (7.3.6):

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0,4 & -0,25 & +0,5 \end{bmatrix}; \quad \mathbf{A}' = \begin{bmatrix} 1 & 0 & 0 \\ -0,5 & 1 & 0 \\ 0,25 & -0,5 & 1 \end{bmatrix};$$

$$(\mathbf{A}')^{-1} = \frac{1}{\det(\mathbf{A}')} \begin{bmatrix} \text{транспонована матриця} \\ \text{алгебраїчних доповнень} \\ \text{для } \mathbf{A}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0,5 & 1 & 0 \\ 0 & 0,5 & 1 \end{bmatrix};$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0,5 & 1 & 0 \\ 0 & 0,5 & 1 \end{bmatrix} \begin{bmatrix} 0,8 \\ -0,6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0,8 \\ -0,2 \\ -0,3 \end{bmatrix}.$$

В результаті отримаємо:

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{F}\mathbf{x}(k-1) + \mathbf{B}\mathbf{u}(k-1) = \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0,4 & -0,25 & +0,5 \end{bmatrix} \begin{bmatrix} x_1(k-1) \\ x_2(k-1) \\ x_3(k-1) \end{bmatrix} + \begin{bmatrix} 0,8 \\ -0,2 \\ -0,3 \end{bmatrix} u(k-1), \end{aligned}$$

або

$$\begin{aligned} x_1(k) &= x_2(k-1) + 0,8u(k-1), \\ x_2(k) &= x_3(k-1) - 0,2u(k-1), \\ x_3(k) &= 0,4x_1(k-1) - 0,25x_2(k-1) + 0,5x_3(k-1) - 0,3u(k-1), \\ y(k) &= x_1(k). \end{aligned}$$

Розглянуті вище математичні моделі складають лише незначну частину тих форм, що використовуються і можуть використовуватись для математичного опису досліджуваних об'єктів в економетричному аналізі і технічних системах. Для поглибленого вивчення цієї теми необхідно звертатись до відповідної спеціальної літератури.

7.4 Приклади представлення моделей у просторі станів

В теорії аналізу часових рядів часто використовують таку форму представлення дискретних моделей у просторі станів (рівняння стану і рівняння вимірів):

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{F}\mathbf{x}(k) + \mathbf{w}(k), \\ \mathbf{z}(k) &= \mathbf{A}^T \mathbf{u}(k) + \mathbf{H}^T \mathbf{x}(k) + \mathbf{v}(k), \end{aligned}$$

де \mathbf{F} , \mathbf{A}^T , \mathbf{H}^T – матриці параметрів розмірності $[n \times n]$, $[r \times p]$ і $[r \times n]$, відповідно; $\mathbf{u}(k)$ – вектор регресорів (екзогенних змінних) вимірності $[p \times 1]$. Як і раніше, $\mathbf{w}(k)$, $\mathbf{v}(k)$ – вектори збурень стану і шумів вимірів, відповідно, з такими статистичними характеристиками:

$$E[\mathbf{w}(k) \mathbf{w}^T(l)] = \begin{cases} \mathbf{Q}, & k = l \\ 0, & k \neq l \end{cases}, \quad E[\mathbf{v}(k) \mathbf{v}^T(l)] = \begin{cases} \mathbf{R}, & k = l \\ 0, & k \neq l \end{cases}$$

де $\mathbf{Q}[n \times n]$, $\mathbf{R}[r \times r]$ – коваріаційні матриці збурень та шумів вимірів, відповідно. Послідовності $\{w(k)\}$, $\{v(k)\}$ некорельовані для будь-яких моментів часу:

$$E[\mathbf{w}(k) \mathbf{v}^T(l)] = 0, \quad \forall k, l.$$

Припускається, що вектор екзогенних змінних $\mathbf{u}(k)$ не містить інформації стосовно $\mathbf{x}(k+s)$ та $\mathbf{v}(k+s)$ для $s=0,1,2,\dots$ окрім тієї, що міститься в $y(k-1)$, $y(k-2)$, \dots , $y(1)$. Наприклад, $\mathbf{u}(k)$ може містити у собі затримані у часі значення y або змінні, що некорельовані з $\mathbf{x}(j)$ або $\mathbf{v}(j)$, $\forall j$.

Модель у просторі станів використовують, як правило, для опису динаміки деякої скінченної кількості спостережень, $\{y(1), y(2), \dots, y(T)\}$, з припущенням, що відоме початкове значення вектора стану $\mathbf{x}(k)$. Як правило, для зручності подальшого аналізу робиться припущення щодо некорельованості початкового значення $\mathbf{x}(1)$ вектора стану $\mathbf{x}(k)$ з будь-якою реалізацією векторів $\mathbf{w}(k)$, $\mathbf{v}(k)$, тобто:

$$E[\mathbf{x}(1) \mathbf{w}^T(k)] = 0, \quad \forall k \in [1, \dots, T], \quad E[\mathbf{x}(1) \mathbf{v}^T(k)] = 0, \quad \forall k \in [1, \dots, T].$$

З рівняння станів випливає, що $\mathbf{x}(k)$ можна записати як лінійну функцію від $[\mathbf{x}(1), \mathbf{w}(2), \mathbf{w}(3), \dots, \mathbf{w}(k)]$:

$$\mathbf{x}(k) = \mathbf{w}(k) + \mathbf{F}\mathbf{w}(k-1) + \mathbf{F}^2\mathbf{w}(k-2) + \dots + \mathbf{F}^{k-2}\mathbf{w}(2) + \mathbf{F}^{k-1}\mathbf{x}(1), \\ \forall k = 2, 3, \dots, T.$$

Таким чином, із $E[\mathbf{w}(k) \mathbf{w}^T(l)] = 0$, $k \neq l$ і $E[\mathbf{x}(1) \mathbf{w}^T(k)] = 0$, $k = 1, 2, \dots, T$ випливає, що $\mathbf{w}(k)$ некорельовано із затриманими значеннями \mathbf{x} :

$$E[\mathbf{x}(l) \mathbf{w}^T(k)] = 0 \quad \text{при} \quad l = k-1, k-2, \dots, 1.$$

Аналогічно:

$$E[\mathbf{x}(l)\mathbf{v}^T(k)] = 0 \quad \text{при } l=1,2,\dots,T,$$

$$E[\mathbf{y}(l)\mathbf{v}^T(k)] = E\{[\mathbf{A}^T\mathbf{u}(l) + \mathbf{H}^T\mathbf{x}(l) + \mathbf{v}(l)]\mathbf{v}^T(k)\} = 0 \quad \text{при } l=k-1, k-2, \dots, 1,$$

$$E[\mathbf{y}(l)\mathbf{w}^T(k)] = 0 \quad \text{при } l=k-1, k-2, \dots, 1.$$

Очевидно, що матриці параметрів \mathbf{F} , \mathbf{Q} , \mathbf{A} , \mathbf{H} , \mathbf{R} можуть бути функціями часу для нестационарних об'єктів та процесів.

Приклад 7.4.1. Розглянемо процес авторегресії p -го порядку:

$$y(k+1) = a_0 + a_1 y(k) + a_2 y(k-1) + \dots + a_p y(k-p+1) + \varepsilon(k+1), \quad (7.4.1)$$

де $\{\varepsilon(k)\}$ – нормально розподілений процес незалежних величин, скорочено можна записати так: $\{\varepsilon(k)\} \sim \text{ННР}(0, \sigma_\varepsilon^2)$, тобто $E[\varepsilon(k)] = 0$, а

$$E[\varepsilon(k)\varepsilon(l)] = \begin{cases} \sigma_\varepsilon^2, & k=l \\ 0, & k \neq l \end{cases}.$$

Процес (7.4.1) можна представити також у вигляді:

$$y(k+1) - \mu = a_1[y(k) - \mu] + a_2[y(k-1) - \mu] + \dots + a_p[y(k-p+1) - \mu] + \varepsilon(k+1), \quad (7.4.2)$$

де μ – відоме середнє значення процесу.

Цей процес можна записати у формі простору станів таким чином (рівняння стану):

$$\begin{bmatrix} y(k+1) - \mu \\ y(k) - \mu \\ y(k-1) - \mu \\ \vdots \\ y(k-p+2) - \mu \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \dots & a_{p-1} & a_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} y(k) - \mu \\ y(k-1) - \mu \\ y(k-2) - \mu \\ \vdots \\ y(k-p+1) - \mu \end{bmatrix} + \begin{bmatrix} \varepsilon(k+1) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (7.4.3)$$

Тобто вектор стану має вигляд:

$$x(k) = [y(k) - \mu, y(k-1) - \mu, y(k-2) - \mu, \dots, y(k-p+1) - \mu]^T.$$

Рівняння вимірів:

$$z(k) = y(k) = \mu + [1 \ 0 \ \dots \ 0] \begin{bmatrix} y(k) - \mu \\ y(k-1) - \mu \\ \vdots \\ y(k-p+1) - \mu \end{bmatrix}. \quad (7.4.4)$$

В рівнянні вимірів шуми відсутні $\mathbf{R} = 0$, а коваріаційна матриця збурень стану має вигляд:

$$\mathbf{Q} = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Приклад 7.4.2. Процес ковзного середнього першого порядку, КС(1):

$$y(k) = \mu + \varepsilon(k) + \beta_1 \varepsilon(k-1), \quad (7.4.5)$$

який може бути представлений у просторі станів наступним чином:

Рівняння стану:

$$\begin{bmatrix} x(k+1) \\ x(k) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon(k) \\ \beta_1 \varepsilon(k-1) \end{bmatrix} + \begin{bmatrix} \varepsilon(k+1) \\ 0 \end{bmatrix}. \quad (7.4.6)$$

Рівняння вимірів:

$$z(k) = y(k) = \mu + [1 \ \beta_1] \begin{bmatrix} \varepsilon(k) \\ \varepsilon(k-1) \end{bmatrix}, \quad (7.4.7)$$

тобто,

$$\mathbf{x}(k) = \begin{bmatrix} \varepsilon(k) \\ \varepsilon(k-1) \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{w}(k+1) = \begin{bmatrix} \varepsilon(k+1) \\ 0 \end{bmatrix},$$

$$\mathbf{Q} = \begin{bmatrix} \sigma_\varepsilon^2 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{z}(k) = y(k), \quad \mathbf{v}(k) = 0, \quad \mathbf{R} = 0.$$

Процес КС(1) можна представити в просторі станів і в такий спосіб:

Рівняння стану:

$$\begin{bmatrix} \varepsilon(k+1) + \beta_1 \varepsilon(k) \\ \beta_1 \varepsilon(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon(k) + \beta_1 \varepsilon(k-1) \\ \beta_1 \varepsilon(k) \end{bmatrix} + \begin{bmatrix} \varepsilon(k+1) \\ \beta_1 \varepsilon(k+1) \end{bmatrix}. \quad (7.4.8)$$

Рівняння вимірів:

$$z(k) = y(k) = \mu + [1 \ 0] \begin{bmatrix} \varepsilon(k) + \beta_1 \varepsilon(k-1) \\ \beta_1 \varepsilon(k) \end{bmatrix}. \quad (7.4.9)$$

Приклад 7.4.3. Процес авторегресії з ковзним середнім довільного порядку, АРКС(p, q). В даному випадку вимірність вектора стану буде $n \equiv \max\{p, q+1\}$. В рівнянні

$$\begin{aligned} y(k) - \mu &= a_1[y(k-1) - \mu] + a_2[y(k-2) - \mu] + \dots + a_n[y(k-n) - \mu] + \\ &+ \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2) + \dots + \beta_{n-1} \varepsilon(k-n+1) \end{aligned} \quad (7.4.10)$$

$a_j = 0, j > p$ і $\beta_j = 0, j > q$. Розглянемо варіант його представлення у просторі станів.

Рівняння стану ($n = \max\{p, q+1\}$):

$$\mathbf{x}(k+1) = \begin{bmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} \varepsilon(k+1) \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (7.4.11)$$

Рівняння вимірів:

$$z(k) = y(k) = \mu + [1 \ \beta_1 \ \beta_2 \ \dots \ \beta_{n-1}] \mathbf{x}(k). \quad (7.4.12)$$

Покажемо, що (7.4.10) і (7.4.11), (7.4.12) описують один і той же процес. Нехай $x_j(k)$ – елемент вектора стану $\mathbf{x}(k)$. Відповідно до другого рядка рівняння стану (7.4.11) можна записати

$$x_2(k+1) = x_1(k).$$

З третього рядка випливає, що $x_3(k+1) = x_2(k) = x_1(k-1)$, а з довільного j -го рядка випливає, що

$$x_j(k+1) = L^{j-1} x_1(k+1).$$

Таким чином, для першого рядка рівняння стану справедливе таке рівняння:

$$x_1(k+1) = (a_1 + a_2L + a_3L^2 + \dots + a_nL^{n-1}) x_1(k) + \varepsilon(k+1),$$

або

$$(1 - a_1L - a_2L^2 - a_3L^3 - \dots - a_nL^n) x_1(k+1) = \varepsilon(k+1). \quad (7.4.13)$$

З рівняння вимірів можна записати:

$$z(k) = y(k) = \mu + (1 + \beta_1L + \beta_2L^2 + \dots + \beta_{n-1}L^{n-1}) x_1(k). \quad (7.4.14)$$

Якщо останнє рівняння помножити на многочлен $(1 - a_1L - a_2L^2 - \dots - a_nL^n)$ і скористатись рівнянням (7.3.13), то отримаємо:

$$(1 - a_1L - a_2L^2 - \dots - a_nL^n)[y(k) - \mu] = (1 + \beta_1L + \beta_2L^2 + \dots + \beta_{n-1}L^{n-1})\varepsilon(k),$$

що збігається з рівнянням (7.4.10).

Приклад 7.4.4. *Динаміка реальної процентної ставки.* Реальну процентну ставку (РПС) можна розглядати як різницю між номінальною процентною ставкою і очікуваним індексом інфляції $\pi^e(k)$. РПС не спостерігається (тобто не вимірюється), оскільки неможливо отримати дані щодо індексу інфляції, який очікується на фондовому ринку.

Таким чином, змінна стану в даному випадку є скаляром:

$$x(k) = i(k) - \pi^e(k) - \mu,$$

де μ – середнє значення РПС. В літературі для опису динаміки очікуваної РПС запропонована модель АР(1):

$$x(k+1) = a x(k) + w(k+1), \quad (7.4.15)$$

де $w(k)$ – шум стану. Рівняння вимірів можна записати із врахуванням того, що за минулий час можна зібрати дані, які характеризують відомі значення РПС (номінальна процентна ставка $i(k)$ мінус дійсне значення індексу інфляції $\pi(k)$), тобто, $i(k) - \pi(k)$. Таким чином, рівняння вимірів має вигляд:

$$i(k) - \pi(k) = [i(k) - \pi^e(k)] + [\pi^e(k) - \pi(k)] = \mu + x(k) + v(k), \quad (7.4.16)$$

де $v(k) \equiv [\pi^e(k) - \pi(k)]$ – похибка прогнозу індекса інфляції. Якщо прогнозоване значення індекса інфляції обчислюється за оптимальним алгоритмом, то $v(k)$ не повинно корелюватись із своїми минулими значеннями і очікуваними значеннями РПС. Таким чином, для даного прикладу матриці моделі у просторі станів мають вигляд:

$$\mathbf{F} = \mathbf{a}, \quad z(k) = y(k) = i(k) - \pi(k), \quad \mathbf{H} = 1, \quad v(k) = [\pi^e(k) - \pi(k)].$$

Приклад 7.4.5. Для опису динаміки ділового циклу в модель вводять скаляр $c(k)$, значення якого не спостерігається (тобто не вимірюється). Припускається, що на множину n різних вимірюваних макроекономічних змінних $[y_1(k), y_2(k), \dots, y_n(k)]$ впливає діловий цикл, а також те, що кожна з цих компонент має свою специфічну структурну складову, яку ми позначимо через $\chi_i(k)$ і яка не зв'язана із значеннями $y_j(k)$ при $i \neq j$. Якщо діловий цикл і кожна із специфічних структурних складових можуть бути описані за допомогою рівняння АР(1), то отримуємо такий вектор стану:

$$\mathbf{x}(k) = [c(k) \ \chi_1(k) \ \chi_2(k) \ \dots \ \chi_n(k)]^T,$$

вимірністю $[(n+1) \times 1]$ і рівняння стану:

$$\begin{bmatrix} c(k+1) \\ \chi_1(k+1) \\ \chi_2(k+1) \\ \vdots \\ \chi_n(k+1) \end{bmatrix} = \begin{bmatrix} a_c & 0 & 0 & \dots & 0 \\ 0 & a_1 & 0 & \dots & 0 \\ 0 & 0 & a_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & a_n \end{bmatrix} \begin{bmatrix} c(k) \\ \chi_1(k) \\ \chi_2(k) \\ \vdots \\ \chi_n(k) \end{bmatrix} + \begin{bmatrix} w_c(k+1) \\ w_1(k+1) \\ w_2(k+1) \\ \vdots \\ w_n(k+1) \end{bmatrix}. \quad (7.4.17)$$

Рівняння вимірів має вигляд:

$$\begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_n(k) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \gamma_1 & 1 & 0 & \dots & 0 \\ \gamma_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \gamma_n & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} c(k) \\ \chi_1(k) \\ \chi_2(k) \\ \vdots \\ \chi_n(k) \end{bmatrix}. \quad (7.4.18)$$

Параметр γ_i описує чутливість i -го ряду до ділового циклу. Для того щоб описати динаміку p -го порядку, величини $c(k)$ і $\chi_i(k)$ у векторі стану

замінюють на $(p \times 1)$ -вимірні вектори $[c(k) \ c(k-1) \dots \ c(k-p+1)]^T$ та $[\chi_i(k) \ \chi_i(k-1) \ \dots \ \chi_i(k-p+1)]^T$. Таким чином вектор стану буде мати вимірність $[(n+1)p \times 1]$. Скаляри a_i в рівняння стану (7.4.17) замінюють матрицями \mathbf{F}_i вимірності $[p \times p]$, а між стовпчиками матриці \mathbf{H}^T в рівнянні вимірів (7.4.18) вставляють блоки нулів вимірністю $[n \times (p-1)]$.

Приклад 7.4.6. *Лінійна гаусівська модель у просторі станів:*

$$\mathbf{x}(k+1) = \mathbf{d}(k) + \mathbf{T}(k)\mathbf{x}(k) + \mathbf{B}(k)\varepsilon(k), \quad \{\mathbf{x}(k)\} \sim N\{\mu_x, \mathbf{P}\}; \quad (7.4.19)$$

$$\mathbf{x}_1 \sim N(a, \mathbf{P}_1); \quad \{\varepsilon(k)\} \sim NID(0, \mathbf{I});$$

$$\mathbf{y}(k) = \mathbf{c}(k) + \mathbf{H}(k)\mathbf{x}(k), \quad (7.4.20)$$

$$\mathbf{z}(k) = \mathbf{y}(k) + \mathbf{G}(k)\varepsilon(k), \quad (7.4.21)$$

де μ_x – вектор середніх значень нормально розподіленої послідовності векторів $\{\mathbf{x}(k)\}$; матриця спостережень вимірності $[N \times n]$ задається векторами $[\mathbf{z}_1, \dots, \mathbf{z}_n]$; вектор стану $\mathbf{x}(k)$ вимірності $[m \times 1]$ містить стохастичний процес, що не спостерігається, а також невідомі ефекти від постійних впливів; $\mathbf{x}(k)$ – послідовність нормально розподілених векторів з середнім значенням μ_x та коваріаційною матрицею \mathbf{P} ; початковий стан \mathbf{x}_1 – нормально розподілений вектор із середнім a та коваріаційною матрицею \mathbf{P}_1 . Рівняння (7.4.19) має так звану марковську структуру, яка є досить ефективною для описання структури послідовної кореляції часового ряду $\mathbf{z}(k)$.

Сигнал $\mathbf{y}(k)$ – деякий проміжний сигнал між вектором стану $\mathbf{x}(k)$ і вектором вимірів $\mathbf{z}(k)$; $\varepsilon(k)$ – нормально розподілений вектор збурень стану з нульовим середнім значенням та одиничною дисперсією.

Детерміновані матриці $\mathbf{T}(k)$, $\mathbf{B}(k)$, $\mathbf{H}(k)$ і $\mathbf{G}(k)$ називають матрицями системи; вони мають, як правило, розріджену структуру. Вектори $\mathbf{d}(k)$ і $\mathbf{c}(t)$ – частіше всього відомі константи або мають нульові значення.

Якщо система (об'єкт) інваріантний в часі (стаціонарний), то розглянуті вище матриці системи записують без часового аргументу: \mathbf{T} , \mathbf{B} , \mathbf{H} і \mathbf{G} . В такому випадку систему у просторі станів (7.4.19) – (7.4.21) називають інваріантною стосовно часу.

Запитання і вправи до сьомого розділу

1. Поясніть переваги представлення математичних моделей динамічних систем у просторі станів? Яким чином враховується шум вимірів у моделі АРКС та в моделі, представленій у просторі станів? Яка з цих моделей представляється вам більш адекватною модельованому процесу і чому?
2. Поясніть, яку роль відіграють матриці моделі **A**, **B**, **H**? Запишіть розмірності цих матриць, якщо розмірність вектора стану дорівнює n .
3. Яке співвідношення існує, як правило, між розмірністю вектора стану $\mathbf{x}(k)$ і вектора вимірів $\mathbf{z}(k)$? Поясніть це на прикладах.
4. Яка причина існування, в деяких випадках, невимірюваних компонент вектора стану?
5. Якщо матриця вимірів **H** діагональна, то що це означає по відношенню до вимірів? Поясніть характер взаємозв'язку між векторами стану та вимірів у випадку, коли вона недіагональна? Наведіть приклад такої моделі.
6. У якому випадку для аналізу процесу можна скористатись моделлю вільної динамічної системи? Наскільки вона зручна для практичного застосування? У чому полягає її недолік?
7. Наведіть приклади випадкових збурень стану динамічної системи, які можна описати нормальним розподілом?
8. Якщо на динамічну систему впливає випадкове збурення, яке має розподіл, відмінний від нормального, то яким чином це вплине на модель у просторі станів?
9. Запишіть у формі простору станів звичайне диференціальне рівняння третього порядку:

$$a_3 \frac{d^3 y(t)}{dt^3} + a_2 \frac{d^2 y(t)}{dt^2} + a_1 \frac{d y(t)}{dt} + y(t) = b_1 u_1(t) + b_2 u_2(t).$$

10. Запишіть у формі простору станів рівняння АРКС(3, 2):

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + a_3 y(k-3) + b_1 \varepsilon(k-1) + b_2 \varepsilon(k-2) + \varepsilon(k).$$

11. Як можна визначити стійкість динамічної системи, яка описується моделлю у просторі станів? Зверніть увагу на відмінності аналізу стійкості у неперервному та дискретному часі.
12. Які труднощі щодо використання моделі у ПС можуть виникнути у випадку, якщо розмірність вектора стану висока ($n = 20-30$)?
13. Запропонуйте структуру моделі у просторі станів для опису сукупності макроекономічних процесів. Яким чином можна знайти оптимальні оцінки стану такої системи? Поясніть природу шумів вимірів у випадку побудови такої математичної моделі. Опишіть можливі збурення стану, особливо зверніть увагу на збурення для макроекономічних процесів перехідного періоду.
14. Для розв'язання яких задач можна скористатись математичною моделлю у просторі станів, побудованою для макроекономіки країни?
15. Виберіть три макроекономічних процеси (наприклад, формування валового внутрішнього продукту, інфляцію та надходження від податку на додану вартість) і побудуйте для них векторно-матричну модель у просторі станів на основі статистичних даних. Визначте особливості цієї моделі у порівнянні із скалярним випадком.

Р о з д і л 8

МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ СЕЗОННИХ ЕФЕКТІВ

8.1 Особливості перебігу реальних фінансово-економічних процесів

Фактичні фінансово-економічні процеси мають ряд відмінних характеристик, які необхідно враховувати при побудові моделі та обчисленні оцінок прогнозів. Знання цих характеристик надає можливість виділити деякі специфічні ефекти, притаманні реальним процесам, та спрогнозувати їх окремо чи за допомогою повної (структурної) моделі процесу. Наприклад, для процесу з трендом і сезонним ефектом можна будувати повну математичну модель, яка враховує всі притаманні їй ефекти, а можна будувати окремо модель для тренду, модель для коливань, що накладаються на тренд, і модель для сезонного ефекту. Який з цих підходів кращий, можна виявити у процесі побудови та аналізу якості моделі конкретного процесу.

Але також потрібно пам'ятати, що будь-які перетворення даних супроводжуються деякими похибками обчислень, які можуть накопичуватись при збільшенні об'ємів обчислювальних операцій. А тому можна очікувати, що кращі результати моделювання і прогнозування можна отримати за допомогою структурних моделей досліджуваних процесів, які враховують наявні ефекти у одній моделі. Це відбувається незважаючи на те, що простіші моделі, які описують окремі ефекти, мали б практично завжди приводити до кращих результатів.

Розглянемо деякі особливості протікання реальних фінансово-економічних процесів та можливості їх математичного опису за допомогою різницевого рівнянь типу АР, АРКС та АРІКС.

Більшість процесів мають явно *виражений тренд*.

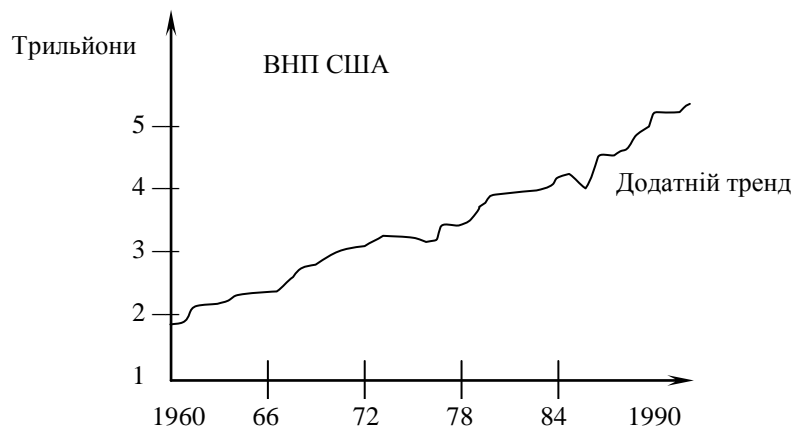


Рис. 8.1 Приклад зростаючого (позитивного) тренду

Тренд (або поточне середнє процесу) може бути зростаючим (позитивним) або спадаючим (негативним). Так, тренди процесів перехідної економіки мають, як правило, досить складну форму із періодами спадання та зростання. При цьому досить часто спостерігаються досить суттєві стрибкоподібні зміни.

У подальшому *трендом* будемо вважати поточне середнє значення вибірки даних, яке вказує на напрямок довгострокового руху процесу. Вважається, що процес не має тренду на деякій ділянці його розвитку, якщо його поточне середнє залишається константою.

У загальному випадку тренд може бути детермінованим або стохастичним (нерегулярним). Визначити характер тренду можна тільки під час побудови моделі процесу. Якщо після вилучення тренду за допомогою різниць модель для залишків має кращі статистичні характеристики ніж модель для залишків детермінованого тренду, то тренд вважають стохастичним.

На коротких часових інтервалах поточне середнє можна обчислити за допомогою виразу:

$$\bar{y}(k) = \bar{y}(k-1) + \frac{1}{k} [y(k) - \bar{y}(k-1)], \quad (8.1.1)$$

де $y(k)$ – фактичне значення процесу в момент k . Очевидно, що із збільшенням значення k поправка у правій частині буде прямувати до нуля, тобто

$$\lim_{k \rightarrow \infty} \frac{1}{k} [y(k) - \bar{y}(k-1)] = 0,$$

що обмежує застосування виразу (8.1.1) відносно короткими часовими інтервалами, на яких зазначена величина буде приймати скінченне значення. Ряд з трендом називають нестационарним стосовно тренду, процесом з одиничними коренями або інтегрованим процесом. Формально для ряду з трендом можна записати: $E[y(k)] \neq \text{const}$, $0 \leq k \leq N-1$ (всього N значень вибірки даних).

Процеси з трендами особливо характерні для нестійкої високодинамічної перехідної економіки, яка може тривати кілька десятків років (наприклад, в Україні). Більшість макроекономічних процесів в перехідних економіках є нестационарними, вони містять тренд та/або мають змінну у часі дисперсію (гетероскедастичні процеси). Це зумовлено надзвичайно високою чутливістю перехідної економіки до внутрішніх та зовнішніх збурень. Приклади спадаючого та зростаючого трендів для економіки України наведено на рис. 8.2.

Іноді тренд неможливо описати детермінованою функцією з необхідним ступенем адекватності. Наприклад, тренди фінансових процесів на біржі часто мають випадковий (стохастичний) характер або високий порядок. На коротких часових інтервалах випадкові тренди можна описати за допомогою комбінації кількох випадкових процесів в одній моделі або за допомогою інших підходів.

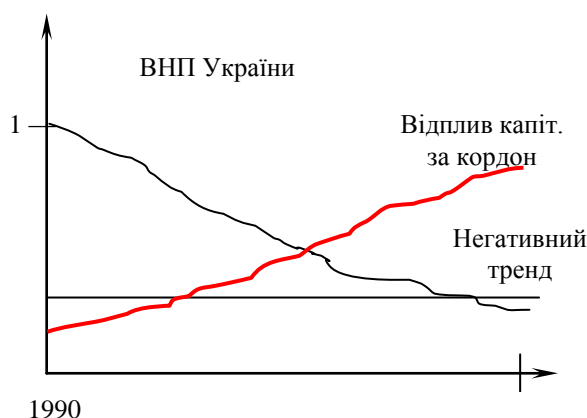


Рис. 8.2 Приклади трендів для економіки України

Зазначимо, що коректна модель тренду надає можливість обчислювати оцінки довгострокових прогнозів і використовувати їх при прийнятті стратегічних рішень, наприклад, у межах застосування систем підтримки прийняття рішень. Такими оцінками можна скористатись, також, у системах технологічного передбачення, динамічного планування, системах аналізу сталого розвитку і т. ін. Для підвищення достовірності оцінок довгострокових

прогнозів необхідно застосовувати кілька ідеологічно різних методів одночасно з метою обчислення комбінованих прогнозів. За виконання деяких нескладних умов оцінки комбінованих прогнозів, як правило, перевищують якість оцінок окремих прогнозів.

2. Деякі ряди мають *характер меандра*. В строгому смислі меандр – це послідовність прямокутних імпульсів з постійним періодом, кожний з яких складається з додатної та від'ємної напівхвиль однакової тривалості. Звичайно, що фінансово-економічні часові ряди можуть тільки наближено відповідати строгому означенню.

Прикладами рядів з меандром можуть бути сезонний продаж овочів (рис. 8.3), сезонний наплив туристів в теплу пору (на морі) або холодну пору року (в горах). Процеси такого типу досить складні з точки зору їх точного математичного опису, оскільки піки можуть мати суттєво різні значення, що призводить до нестационарності стосовно дисперсії процесу. Тобто процес може містити тренд і гетероскедастичність одночасно.

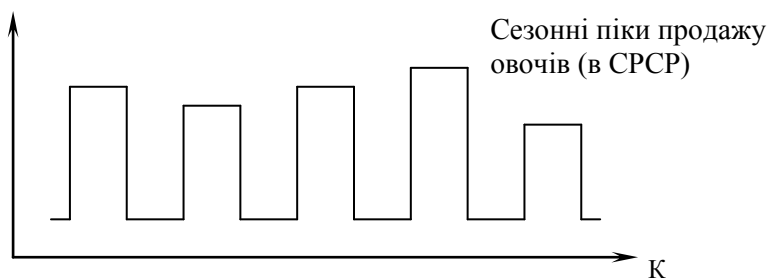


Рис. 8.3 Приклад ряду з меандром

Достатньо повний математичний опис процесів цього класу повинен містити модель ряду у цілому, а також, окремо тренду і умовної дисперсії як складових, що спричиняють нестационарність, яка відіграє значну роль у прийнятті відповідних рішень. Наприклад, дисперсія – один із визначальних параметрів, що впливають на прийняття рішень у процесі виконання торговельних операцій на біржі. Тренд, у свою чергу, свідчить про напрям руху ціни – на зростання, спадання або утримання постійної позиції.

3. *Макроекономічні явища*, які виникають внаслідок внутрішніх або зовнішніх факторів, *носять, як правило, затяжний характер*. Прикладами можуть бути такі явища:

- поточний перехідний період в економіці країн Східної Європи, Китаю та Латинської Америки;
- поточний перехідний період на пострадянському просторі (для цих країн економісти прогнозують тривалість перехідного періоду щонайменше 35-40 років – це наслідок неефективного, так званого соціалістичного підходу, керування економікою);

- криза у капіталістичному світі в 30-і роки ХХ століття (щонайменше 8 – 10 років);
- падіння промислового виробництва в Англії наприкінці 1970-х і до середини 1980-х.

Що стосується перехідного періоду на пострадянському просторі, то у даному випадку для багатьох країн не можна визначити (спрогнозувати) строки, коли цей період закінчиться. Це зумовлено особливостями їх попереднього розвитку, ефективністю державного управління в поточний перехідний період, впливом значного числа збурень на фінансово-економічні та соціально-політичні процеси, великою роллю особистостей (на керівних посадах) та іншими факторами.

Разом з тим у більшості країн Східної Європи (Польща, Чехія, Словаччина, Угорщина) спостерігається досить висока і відносно стабільна динаміка економічного розвитку, яка свідчить про те, що перехідний період може бути подолано в найближчі 10-12 років і ці країни вийдуть на рівень високорозвинених країн світу.

Стосовно України, Білорусі, Росії та інших держав, що утворились в результаті розпаду СРСР, нелегко визначити тривалість перехідного періоду в економіці. Він може складати, за оцінками провідних економістів, 35-40 років і більше. Це пояснюється тим, що занадто багато негативних факторів (збурень) впливає на їх поточний розвиток. Це, зокрема, такі:

- критична нестача керівних кадрів високої кваліфікації та відсутність належного рівня їх загальної підготовки і розвитку, які вкрай необхідні на сучасному етапі розвитку людства у цілому;
- відсутність наукових шкіл належного рівня розвитку (це стосується не тільки економіки, але й багатьох інших напрямів розвитку науки);
- використання застарілих технологій у сільському господарстві, промисловості, навчальному процесі;
- перетікання значних об'ємів інтелектуального та грошового капіталів за кордон;
- жорстка конкуренція з високорозвиненими країнами за ринки збуту та у всіх інших сферах економічної і науково-технічної діяльності;
- енергетичні та фінансові кризи локального та глобального масштабів;
- жорстка внутрішня боротьба за владу, яка спрямована на перерозподіл капіталу і власності, а не на конструктивне планування розвитку держави з метою досягнення високих результатів у економіці, науці, освіті;
- олігархічно-бюрократичний характер влади, який спрямований на задоволення потреб невеликої частки населення, що вкрай

негативно впливає на розвиток науки, техніки, технологій, освіти і культури та соціально-економічних процесів в цілому;

- глобалізація розвитку світової економіки, яка примушує підкорюватись загальним (часто несприятливим для слабких економік) зовнішнім тенденціям та впливам.

Теорія економіки перехідного періоду також стверджує, що перехідний період не обов'язково закінчується переходом на вищу фазу розвитку суспільства.

4. На різних ділянках часового ряду *може спостерігатись різна динаміка*, тобто різна швидкість падіння або зростання значень змінної. Це особливо стосується інфляції, об'єму промислового та сільськогосподарського виробництва, податкових надходжень та багатьох інших процесів.

Ця проблема є особливо гострою для процесів перехідної економіки та фінансових процесів, зокрема, процесів формування цін на біржах. Побудова математичних моделей процесів формування біржових цін залишається складною проблемою. Яскравим прикладом процесів з нестаціонарною динамікою є індекс інфляції в Україні, величина якого склала 10250% в 1993 році, а в 2003 році біля 8%. Іншим прикладом може бути падіння ВВП України в 2,5 рази на протязі 1991 – 1997 років. На 45% спостерігалось падіння промислового виробництва у США в період кризи 30-х років ХХ століття.

Очевидно, що побудова прогнозуючих моделей процесів з високою і різною динамікою (з відповідними переходами з одного режиму функціонування в інший) потребує виконання їх докладного аналізу, використання всіх доступних статистичних даних та альтернативних структур математичних моделей.

5. Існує певний *взаємозв'язок між макроекономічними явищами* в різних країнах.

Так, рівень промислового виробництва в США впливає на рівень промислового виробництва в Англії та Канаді та багатьох інших країн світу. Макроекономіка Росії суттєво впливає на розвиток макроекономіки України, Білорусі, Казахстану, а фінансові кризи в одній частині світу впливають, фактично, на всі інші країни світу (яскравим прикладом є фінансова криза 2008 року). При цьому, чим слабшою (нестійкою) є економіка, тим сильнішими будуть для неї впливи (збурення) з боку зовнішніх факторів.

Математичний опис таких взаємозв'язків потребує застосування системного підходу до аналізу макроекономічних явищ, високоякісного експертного оцінювання та великих об'ємів достовірних статистичних даних, які, як правило, не можна знайти. Створення математичних моделей взаємозв'язків між процесами на основі теоретичного структурного підходу також пов'язане із значними труднощами, оскільки воно потребує глибоких знань процесів, що відбуваються одночасно у багатьох країнах світу,

особливостей, логіки та ієрархії їх взаємодії, а також глибини і характеру змін, що відбуваються під впливом взаємодіючих факторів. Одним із центральних завдань є, у даному випадку, моделювання, прогнозування і висококваліфікований менеджмент відповідних економічних ризиків.

6. Макроекономічні процеси підпорядковані впливу випадкових факторів, особливо в умовах взаємозалежності між країнами. Знову ж таки, це в більшій мірі стосується країн з перехідною (слабкою) економікою. Сюди можна віднести такі фактори:

- неплатежі (несвоєчасність виконання фінансових зобов'язань);
- значний рівень тіньової економіки;
- несвоєчасні поставки енергетичних ресурсів, сировини, комплектуючих;
- значні коливання цін на енергоресурси, а також різке зростання цін;
- коливання рівня податків і відповідно коливання цін та рівня споживання;
- нестійкість та жорсткість митних правил і митного законодавства;
- нестійкість існуючого законодавства в цілому;
- застосування бартеру і взаємозаліків (це особливо характерно для перехідної економіки);
- великий вплив демпінгових цін (наприклад, ціни на китайські товари в Україні є нижчими ніж на аналогічні українські товари);
- інтенсивне використання декількох валют на території однієї країни (наприклад, доларизація української економіки не сприяє зміцненню національної валюти та економіки країни, а навпаки – підтримує валюту США);
- зовнішня та внутрішня політична нестабільність (рішення приймаються не з точки зору їх економічної доцільності для держави, а з політичних або особистих міркувань);
- вплив однієї особистості на протікання економічних та політичних процесів (це явище є особливо характерним для слабких країн, але може мати місце в будь-якій економіці – вплив правління Д. Буша на розвиток економіки США виявився настільки негативним, що держава з величезним профіцитом бюджету за два роки перейшла в стан значного дефіциту з відповідними наслідками для населення країни);
- негативний вплив локальних воєн (локальні війни на Близькому Сході та в інших частинах світу не сприяють зниженню цін на нафтопродукти, посиленню міжнародної економічної кооперації або піднесенню економіки країн, що приймають у них участь);
- відвертий економічний та політичний тиск сильних держав на слабкіших з метою прийняття рішень, в яких зацікавлений сильніший сусід.

У зв'язку з реальним існуванням наведених вище особливостей протікання фінансово-економічних процесів, їх можна розділити на такі види:

- нестационарні процеси із *сезонними ефектами*;
- *гетероскедастичні процеси* (процеси із змінною в часі дисперсією);
- процеси з *детермінованими та випадковими (стохастичними) трендами*;
- *коінтегровані процеси* (однією моделлю одночасно описуються два або більше процесів з трендами однакового порядку);
- процеси з *різкими стрибкоподібними переходами* з одного режиму функціонування в інший;
- *нелінійні процеси*, які можуть містити згадані ефекти;
- процеси з *комбінованими ефектами*.

Наявність сезонного ефекту в процесі не обов'язково приводить до нестационарності ряду, але цілком можливо, що суттєва різниця у пікових значеннях ряду може призвести до гетероскедастичності. Крім того, такі процеси містять, як правило, тренд. Очевидним фактом також є те, що сезонні ефекти суттєво підвищують дисперсію ряду, а тому при використанні моделей стаціонарних процесів для опису їх динаміки необхідно застосовувати так зване “сезонне” диференціювання з метою зменшення рівня дисперсії ряду даних.

Гетероскедастичні, процеси з трендами та коінтегровані процеси – все це нестационарні процеси, для яких розроблено спеціальні методи побудови моделей та прогнозуючих функцій. Методики побудови математичних моделей та прогнозування процесів цих класів будуть докладно розглянуті нижче.

8.2 Моделювання процесів із сезонними ефектами

Сезонні ефекти (СЕ), тобто вплив тієї чи іншої пори року на протікання економічних процесів, можна пояснити таким чином:

- так, погода впливає на темпи будівництва – продуктивність праці у цій галузі максимальна у теплу пору року;
- спостерігається сезонна активність у туризмі: максимальну кількість відпочиваючих (і туристів) на морях можна виявити у теплу року, а на гірськолижних курортах – взимку;
- існує святковий розпродаж товарів перед різдвяними святами і Новим Роком, на День Подяки в США та на багато інших свят;
- сезонний розпродаж овочів, пік якого приходить на кінець літа і початок осені.

У наведених прикладах сезонний ефект проявляється значним і досить різким підвищенням значень відповідного часового ряду, що може призвести до його нелінійності та нестационарності. Загалом поява сезонного ефекту може суттєво змінити статистичні характеристики даних та їх математичної чи статистичної моделі. Зокрема, може з'явитись сезонна гетероскедастичність. Прогнозування процесів без врахування сезонних ефектів призводить до великих значень дисперсії похибок оцінок прогнозів і, відповідно, до їх неприйнятної якості.

Врахування сезонного ефекту в моделі

Сезонний ефект враховується шляхом включення в модель змінної із запізненням, яке відповідає періодичності сезонного ефекту. Так, якщо СЕ проявляється один раз на рік протягом одного кварталу, то при обробці квартальних даних ($T_s = 3$ місяці) у модель необхідно ввести змінну із запізненням $d = 4$.

СЕ можна врахувати в *адитивній формі* за допомогою основної (залежної) змінної або випадкової змінної у ковзному середньому, тобто, можливі такі структури рівняння:

$$y(k) = a_4 y(k-4) + \varepsilon(k), \quad |a_4| < 1; \quad (8.2.1)$$

$$y(k) = \varepsilon(k) + \beta_4 \varepsilon(k-4). \quad (8.2.2)$$

Для рівняння (8.2.1) автокореляційна функція має вигляд:

$$\rho(s) = \begin{cases} (a_4)^{s/4}, & \text{якщо } s/4 \text{ ціле число,} \\ 0, & \text{якщо } s/4 \text{ дробове.} \end{cases}$$

Тобто, АКФ має ненульові значення при $s = 4, 8, 12, \dots$. Для рівняння (8.2.2) АКФ має одне ненульове значення при $s = 4$.

Очевидно, що для підвищення ступеня адекватності моделі процесу в цілому рівняння (8.2.1), (8.2.2) необхідно розширювати, наприклад, додатковими пояснюючими змінними, авторегресією та ковзним середнім першого порядку:

$$y(k) = a_1 y(k-1) + a_4 y(k-4) + \varepsilon(k) + \beta_1 \varepsilon(k-1),$$

$$y(k) = a_1 y(k-1) + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_4 \varepsilon(k-4).$$

Крім того, СЕ можна враховувати у *мультиплікативній формі*. На відміну від рівнянь (8.2.1) і (8.2.2), мультиплікативна форма має вигляд:

$$y(k)(1 - a_1L) = \varepsilon(k)(1 + \beta_1L)(1 + \beta_4L^4), \quad (8.2.3)$$

або

$$y(k)(1 - a_1L)(1 - a_4L^4) = \varepsilon(k)(1 + \beta_1L). \quad (8.2.4)$$

Для зменшення дисперсії процесу з сезонним ефектом необхідно застосовувати так звані “сезонні” різниці, тобто обчислювати різниці типу $\Delta y(k - s)$, де s – періодичність сезонного ефекту. В результаті отримуємо згладжений ряд даних з меншим ступенем нестационарності.

Якщо ряд має тренд, то після обчислення сезонних різниць обчислюють ще перші різниці або різниці вищого порядку залежно від кількості одиничних коренів характеристичного рівняння, записаного для моделі процесу. Застосування різниць приводить до видалення тренду, тобто такі дані вже не описують тренд. Послідовність побудови математичної моделі процесу з сезонним ефектом наведена нижче.

Послідовність побудови моделі часового ряду з сезонним ефектом

Обчислити “сезонні” різниці ряду відповідно до періодичності сезонного ефекту з метою зменшення („згладжування”) дисперсії процесу. Якщо отриманий в результаті сезонного диференціювання ряд нестационарний стосовно тренду, то необхідно обчислити ще звичайні перші різниці або різниці вищого порядку залежно від характеру тренду (лінійний, квадратичний і так далі). Можливий також варіант побудови моделі, яка одночасно включає опис тренду та коливань, що накладаються на нього (це модель структурного типу).

Обчислити автокореляційну та часткову автокореляційну функції для отриманого ряду і шляхом їх візуального або автоматизованого дослідження визначити можливі структури математичних моделей-кандидатів. При цьому перевагу віддають (за принципом економії) моделям АР та АРКС низького порядку.

Обчислити оцінки параметрів моделей-кандидатів з адитивною та мультиплікативною формами врахування сезонності.

Вибрати кращу модель за допомогою статистичних критеріїв, наведених в розглянутій вище методиці побудови моделей ФЕП.

Примітка: При побудові та використанні моделей процесів з сезонними ефектами застосовують такі позначення:

Δ_s^l – “сезонна різниця” порядку l з періодичністю ефекту s .

Наприклад,

$$\begin{aligned}\Delta_{12}^2 y(k) &= \Delta_{12}[y(k) - y(k-12)] = \Delta_{12}y(k) - \Delta_{12}y(k-12) = \\ &= y(k) - y(k-12) - [y(k-12) - y(k-24)] = \\ &= y(k) - 2y(k-12) + y(k-24).\end{aligned}$$

Мультиплікативні моделі записують у такій загальній формі:

$$APIKC(p, d, q)(P, D, Q)_s,$$

де p, q – порядок авторегресії і ковзного середнього, які не відносяться до описання сезонного ефекту; d – число звичайних (несезонних) різниць; D – число сезонних різниць; P – число коефіцієнтів авторегресії в мультиплікативній частині моделі; Q – число коефіцієнтів ковзного середнього в мультиплікативній частині; s – періодичність сезонного ефекту.

Наприклад, для моделі $(1 - L^{12})(1 - L)y(k) = (1 + \beta_1 L)(1 + \beta_{12} L^{12})\varepsilon(k)$ вказані параметри мають такі значення: $p = 0, d = 1, q = 1, P = 0, D = 1, Q = 1, s = 12$.

Задачі і вправи до восьмого розділу

1. Якими особливостями характеризується протікання реальних фінансово-економічних процесів? Чи стосуються ці особливості економічних процесів перехідного періоду?
2. Поясніть, що означає термін „процес із трендом”? Наведіть приклади з макроекономіки України. Що означають терміни «процес з одиничними коренями» та «інтегрований процес»?
3. У чому полягає різниця між детермінованим і стохастичним трендами? Які можуть бути причини виникнення стохастичного тренду?
4. За допомогою яких моделей можна описати детермінований тренд? Які види прогнозів можна розрахувати за допомогою таких моделей?
5. Якими засобами можна описати поведінку стохастичного тренду? На яких часових інтервалах це можливо?
6. Що таке ідеальний меандр? В чому полягають відмінності реального меандру від ідеального? Наведіть приклад процесу з меандром?
7. Економіка яких країн найбільше впливає на економічні процеси в Україні? В чому проявляється цей вплив?
8. Назвіть причини виникнення сезонних ефектів? Наведіть декілька прикладів сезонних ефектів?
9. Як змінюються статистичні характеристики процесу при появі сезонного ефекту? Як можуть впливати сезонні ефекти на стаціонарність процесу?
10. Запишіть можливі математичні моделі, які враховують сезонний ефект з періодичністю $s = 12$. Наведіть приклад процесу, який має таку періодичність сезонного ефекту.
11. Яка мета застосування „сезонних” різниць до статистичних даних, що містять сезонний ефект?
12. Назвіть недоліки та переваги описання сезонного ефекту в мультиплікативній формі?
13. Чи можна врахувати сезонний ефект в ковзному середньому?
14. Поясніть термін „висока динаміка процесу”? Наведіть приклади з макроекономіки.

15. Чому сучасні посткомуністичні держави називають державами з перехідною економікою? Що це означає на практиці? Яку тривалість сучасних перехідних процесів прогнозують економісти? Чим зумовлені такі значення тривалості перехідного періоду?
16. Наведіть приклади випадкових збурень, які діють на макроекономіку держав, що знаходяться в перехідному періоді розвитку? Який вплив вони мають на темпи економічного розвитку? Як можна описати математично випадкові збурення?
17. Знайдіть другі різниці для процесу, що описується рівнянням:

$$y(k) = a_0 + ck + ck^2 + a_1y(k-1) + \varepsilon(k).$$

Поясніть отриманий результат.

18. Поясніть процедуру побудови моделі процесу з сезонним ефектом. Виберіть макро- або мікроекономічний процес з сезонним ефектом і побудуйте для нього математичну модель. Зверніть увагу на адекватність моделі.
19. Обчисліть і поясніть автокореляційну функцію такого рівняння:

$$y(k) = a_0 + a_4y(k-4) + \varepsilon(k),$$

де $\{\varepsilon(k)\}$ – процес білого шуму з нульовим середнім. Чи зміниться вигляд автокореляційної функції процесу після додавання у модель складової AR(1)?

20. Які інші ефекти, крім сезонних, зустрічаються в реальних фінансово-економічних процесах? Яким чином їх можна спрогнозувати?
21. Чи впливає збільшення кількості параметрів моделі на якість оцінок її параметрів?
22. Чи може бути процес із сезонним ефектом гетероскедастичним? Чому?
23. Наведіть приклад скороченого опису моделі процесу з мультиплікативною формою сезонного ефекту?

Розділ 9

ПРОГНОЗУВАННЯ ГЕТЕРОСКЕДАСТИЧНИХ ПРОЦЕСІВ

9.1 Загальні відомості про гетероскедастичні процеси

Гетероскедастичними називають процеси із змінною в часі дисперсією, а *гомоскедастичними* – процеси із сталою дисперсією на відрізок часу, що розглядається при побудові моделі та оцінюванні прогнозу. Гетероскедастичні процеси відносяться до широкого класу нестационарних процесів, який включає в себе такі процеси:

- процеси із детермінованими і стохастичними трендами;
- процеси із змінною дисперсією;
- процеси, які характеризуються змінним в часі математичним сподіванням та змінною дисперсією одночасно;
- процеси із змінною коваріацією (дисперсійно-коваріаційна нестационарність).

Дисперсію та стандартне відхилення часто використовують як *міру ризику* при дослідженні фінансово-економічних процесів, а тому цій проблемі приділяється значна увага у спеціальній літературі. При дослідженні фінансових процесів дисперсію та стандартне відхилення використовують як *міру волатильності* (мінливості) процесу.

У технічних процесах дисперсія – це *міра розсіювання вимірів*. Вона може характеризувати поточний стан механізмів, технологічних процесів, інтенсивність випадкових збурень, що впливають на технічні системи; інтенсивність похибок (шумів) вимірів і т. ін., а тому це надзвичайно

важливий статистичний параметр з точки зору дослідження поточного та прогнозування майбутнього стану системи.

В соціальних дослідженнях дисперсію і стандартне відхилення використовують для порівняння між собою різних соціальних груп або різних множин індивідуумів, що відносяться до однієї й тієї ж соціальної групи з подальшим прийняттям відповідних рішень.

У дослідженнях фінансових ризиків дисперсія – один із основних параметрів, які використовуються для оцінювання величини можливих втрат. Наприклад, існує підхід до оцінювання фінансових ризиків, який спрямований на оцінювання величини Value-at-Risk (VaR), яка характеризує об'єм можливих втрат з деякою (вибраною ОПР) ймовірністю на заданому часовому інтервалі. Цей підхід використовує прогноз волатильності (стандартного відхилення відповідного фінансового процесу) для прогнозування втрат. Прогноз волатильності і напряму руху процесу ціноутворення активів також використовують у правилах виконання торгових операцій на біржі.

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 (періодів дискретизації вимірів) наперед. Це дасть можливість покращувати якість рішень стосовно управління процесами із змінною в часі дисперсією. Наприклад, коректно формулювати правила прийняття рішень щодо купівлі/продажу акцій на біржі, коректно оцінювати міру кредитного, валютного та інших ризиків у 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 діагностики, аналізу якості продукції при побудові контрольних карт, при визначенні точності розмірів і т. ін.

Формально процес вважається гетероскедастичним (ГСП), якщо

$$\text{Var}[\varepsilon(k)] = \sigma_{\varepsilon}^2 \neq \text{const}, \quad k=0, \dots, N-1. \quad (9.1.1)$$

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 із плином часу, або є більш складною функцією часу. Тобто вона може змінюватись за досить складним законом, який і потрібно знайти при створенні моделі процесу. Іноді висувають припущення, що гетероскедастичність має відносно просту форму:

$$\sigma_{\varepsilon(k)}^2 = c^2 x^2(k),$$

де c – константа, яку необхідно оцінити за допомогою експериментальних даних та вибраного методу оцінювання параметрів; $x(k)$ – випадкова змінна, яка визначає дисперсію процесу. Розглянемо простий випадок мультиплікативного збурення, тобто:

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

де $y(k+1)$ – основна залежна змінна; $\varepsilon(k+1)$ – збурення у вигляді (наприклад) білого шуму із скінченною дисперсією σ^2 ; $x(k)$ – незалежна змінна, виміри якої відомі на відріжку часу, що розглядається. У випадку, якщо $x(k) = x(k-1) = x(k-2) = \dots = \text{const}$, послідовність $\{y(k)\}$ представляє собою білий шум з постійною дисперсією. Однак у випадку, коли процес $\{x(k)\}$ – послідовність різних за значеннями чисел, то умовна дисперсія $y(k+1)$ при відомих значеннях $x(k)$ визначається так:

$$\text{Var}[y(k+1)|x(k)] = x^2(k) \sigma_\varepsilon^2.$$

Таким чином, умовна дисперсія величини $y(k+1)$ залежить від конкретної реалізації $x(k)$. Оскільки $x(k)$ можна виміряти в момент k , то дисперсію величини $y(k)$ можна визначити як умовну за умови, що відоме значення $x(k)$. Якщо значення $x^2(k)$ велике (мале), то дисперсія $y(k+1)$ також буде великою (малою). З цього випливає, що за наявності додатної кореляції між значеннями послідовності $\{x(k)\}$ (тобто є тенденція, що за великим значенням $x(k)$ знову іде велике значення), то умовна дисперсія послідовності $\{y(k)\}$ також матиме тенденцію до послідовної додатної кореляції. Іноді розглядають модифіковану модель типу:

$$\ln(y(k)) = a_0 + a_1 \ln(x(k-1)) + \ln e(k), \quad (9.1.2)$$

де $e(k)$ – похибка моделі. Недоліком такої моделі є те, що потрібно дійсно знайти таку змінну $x(k)$, яка суттєво впливає на дисперсію процесу $y(k)$ 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. У наведеному прикладі (9.1.2) покладається, що послідовність $e(k)$ має постійну дисперсію. Якщо ця умова не виконується, то необхідно зробити додаткову обробку даних.

Однак замість того щоб шукати одну змінну $x(k)$, яка визначає дисперсію процесу $\{y(k)\}$, і робити додаткові перетворення даних, можна одночасно моделювати (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

Нехай оцінюється модель стаціонарного процесу авторегресії з ковзним середнім (АРКС) такого вигляду (для спрощення аналізу в даному випадку це просто АР(1)):

$$\begin{aligned} y(k) &= a_0 + a_1 y(k-1) + \varepsilon(k), \quad |a_1| < 1, \\ E[\varepsilon(k)] &= 0, \quad E[\varepsilon^2(k)] = \sigma_\varepsilon^2, \quad E[\varepsilon(k) y(k-l)] = 0, \quad l > 0, \\ E[\varepsilon(k) y(k)] &\approx \sigma_\varepsilon^2. \end{aligned}$$

Необхідно оцінити 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:

$$E_k[y(k+1)] = a_0 + a_1 y(k).$$

Якщо використати це *умовне середнє* для прогнозування значення $y(k+1)$, то дисперсію похибки прогнозу можна визначити так:

$$E_k \{ [y(k+1) - a_0 - a_1 y(k)]^2 \} = E_k [\varepsilon^2(k+1)] = \sigma_\varepsilon^2.$$

Знайдемо *безумовне середнє*, тобто:

$$E[y(k)] = a_0 + a_1 E[y(k-1)] + E[\varepsilon(k)];$$

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):

$$y(k) = a_0 \sum_{i=0}^{\infty} a_1^i + y_0 a_1^k + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i),$$

а при $k \rightarrow \infty$ маємо:

$$y(k) = a_0 \sum_{i=0}^{\infty} a_1^i + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i) = \frac{a_0}{1-a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i), \quad |a_1| < 1.$$

Таким чином, дисперсія похибки безумовного прогнозу визначається за виразом:

$$\begin{aligned}
E\{[y(k+1) - a_0/(1-a_1)]^2\} &= \\
&= E\{[\varepsilon(k+1) + a_1\varepsilon(k) + a_1^2\varepsilon(k-1) + a_1^3\varepsilon(k-2) + \dots]^2\} = \\
&= \sigma^2/(1-a_1^2), \quad |a_1| < 1.
\end{aligned}$$

Послідовність у 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*. У 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 у часі, то тенденцію зміни дисперсії можна описати за допомогою моделі АРКС. Наприклад, позначимо через $\{\hat{\varepsilon}(k)\}$ оцінки залишків (похибок) моделі першого порядку $y(k) = a_0 + a_1y(k-1) + \varepsilon(k)$. У такому випадку умовна дисперсія основної змінної визначається так:

$$Var [y(k+1) | y(k)] = E_k \{ [y(k+1) - a_0 - a_1y(k)]^2 \} = E_k [\hat{\varepsilon}^2(k+1)].$$

До цих пір вважалось, що $E_k[\varepsilon^2(k+1)] = \sigma_\varepsilon^2$ є постійною величиною. Покладемо тепер, що умовна дисперсія – величина змінна. Одним із простих підходів до опису такої змінної величини є застосування моделі типу АР(q) до квадратів оцінок залишків, наприклад:

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1\hat{\varepsilon}^2(k-1) + \alpha_2\hat{\varepsilon}^2(k-2) + \dots + \alpha_q\hat{\varepsilon}^2(k-q) + v(k), \quad (9.1.3)$$

де $v(k)$ – процес білого шуму з нульовим середнім для адекватної моделі.

Якщо всі коефіцієнти $\alpha_1, \alpha_2, \dots, \alpha_q$ дорівнюють нулю, то оцінка дисперсії буде просто константою. Інакше, умовна дисперсія для $y(k)$ описується рівнянням (9.1.3). Це рівняння можна використати для прогнозування умовної дисперсії на один крок таким чином:

$$E_k[\hat{\varepsilon}^2(k+1)] = \alpha_0 + \alpha_1\hat{\varepsilon}^2(k-1) + \alpha_2\hat{\varepsilon}^2(k-2) + \dots + \alpha_q\hat{\varepsilon}^2(k-q).$$

Оскільки (9.1.3) може бути побудоване за умови, що $Var [y(k)] \neq const$, то (9.1.3) називають *авторегресією з умовною гетероскедастичністю* (АРУГ). Залишки (збурення), які використовуються у рівнянні (9.1.3), можна отримати 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 (9.1.3) можна вибрати і складніші форми опису поведінки дисперсії. Наприклад, майже ніколи наперед невідомо як впливає збурення на процес – адитивно чи мультиплікативно. Тому його можна ввести в модель у мультиплікативній формі:

$$\varepsilon^2(k) = v^2(k)[\alpha_0 + \alpha_1 \varepsilon^2(k-1)], \quad (9.1.4)$$

де $v(k)$ – мультиплікативне збурення у формі білого шуму, причому для спрощення аналізу можна задати умову $\{v(k)\} \sim (0,1)$, тобто воно має нульове середнє і одиничну дисперсію; змінні $\varepsilon(k-1)$ і $v(k)$ – статистично незалежні (некорельовані) величини. Можна показати, що елементи послідовності $\{\varepsilon(k)\}$ некорельовані між собою і мають нульове середнє. Знайдемо умовне математичне сподівання для $\varepsilon(k)$. Оскільки $E[v(k)] = 0$, то:

$$\begin{aligned} E[\varepsilon(k)] &= E[v(k)(\alpha_0 + \alpha_1 \varepsilon^2(k-1))^{1/2}] = \\ &= E[v(k)] E[\alpha_0 + \alpha_1 \varepsilon^2(k-1)]^{1/2} = 0. \end{aligned} \quad (9.1.5)$$

При $E[v(k)v(k-1)] = 0$ маємо, що

$$E[\varepsilon(k)\varepsilon(k-i)] = 0, \quad i \neq 0. \quad (9.1.6)$$

Безумовну дисперсію для $\varepsilon(k)$ можна визначити, також, за виразом:

$$E[\varepsilon^2(k)] = E[v^2(k)(\alpha_0 + \alpha_1 \varepsilon^2(k-1))] = E[v^2(k)] E[\alpha_0 + \alpha_1 \varepsilon^2(k-1)].$$

Враховуючи те, що $\sigma_v^2 = 1$, а безумовна дисперсія змінної $\varepsilon(k)$ така ж як і для $\varepsilon(k-1)$, тобто $E[\varepsilon^2(k)] = E[\varepsilon^2(k-1)]$, то безумовна дисперсія приймає значення:

$$E[\varepsilon^2(k)] = \frac{\alpha_0}{1 - \alpha_1}. \quad (9.1.7)$$

Таким чином, на безумовне середнє та дисперсію не впливає процес $\varepsilon(k)$, який визначений рівнянням (9.1.4). Аналогічно можна досить просто показати, що умовне середнє процесу $\varepsilon(k)$ дорівнює нулю. Якщо $v(k)$ і $\varepsilon(k-1)$ – незалежні величини і $E[v(k)] = 0$, то умовне середнє для $\varepsilon(k)$ можна знайти так:

$$E_{k-1}[\varepsilon(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = E_{k-1}[v(k)] E_{k-1}[\alpha_0 + \alpha_1 \varepsilon^2(k-1)]^{1/2} = 0.$$

Очевидно, що процес, який визначається виразом (9.1.4), впливає на умовну дисперсію. Оскільки дисперсія $\sigma_v^2 = 1$, то для умовної дисперсії змінної можна записати:

$$E_{k-1}[\varepsilon^2(k) | \varepsilon(k-1), \varepsilon(k-2), \dots] = \alpha_0 + \alpha_1 \varepsilon^2(k-1). \quad (9.1.8)$$

Звідси видно, що умовна дисперсія змінної $\varepsilon(k)$ залежить від значення $\varepsilon^2(k-1)$. Якщо остання змінна приймає велике значення, то і умовна дисперсія в момент k також буде мати велике значення. Рівняння (9.1.8) – це авторегресія першого порядку, але вона відрізняється від звичайної авторегресії тим, що коефіцієнти α_0 , α_1 повинні мати певні обмеження. Для того щоб забезпечити додатність дисперсії, обидва коефіцієнти α_0 , α_1 повинні бути додатними. Крім того, для забезпечення стійкості (збіжності) процесу авторегресії необхідно щоб виконувалась умова: $0 < \alpha_1 < 1$.

Рівняння (9.1.5) – (9.1.8) ілюструють загальні властивості будь-якого процесу АРУГ. Структура моделі АРУГ така, що умовне та безумовне середнє значення похибок цієї моделі дорівнюють нулю (фактично вони мають дорівнювати нулю для адекватної моделі). Крім того, для послідовності $\varepsilon(k)$ відсутня автокореляція, оскільки $E[\varepsilon(k)\varepsilon(k-s)] = 0$, $\forall s \neq 0$. Головною властивістю розглянутої моделі у даному випадку є те, що елементи послідовності $\{\varepsilon(k)\}$ зв'язані між собою через їх другий центральний момент, тобто умовну дисперсію. Якщо значення елемента $\varepsilon(k-1)$ суттєво відрізняється від нуля, що приводить до відносно великого значення $\alpha_1 \varepsilon^2(k-1)$, то дисперсія змінної $\varepsilon(k)$ має тенденцію до зростання. Таким чином, умовна гетероскедастичність процесу $\{\varepsilon(k)\}$ свідчить про те, що послідовність $\{y(k)\}$ також є процесом типу АРУГ.

Яким же чином структура похибки впливає на послідовність $\{y(k)\}$? Суттєвим моментом у цьому випадку є те, що структура похибки процесу АРУГ і параметри автокореляції послідовності $\{y(k)\}$ взаємодіють між собою. Про це свідчать численні результати відповідних обчислювальних експериментів. Цьому явищу можна дати пояснення на інтуїтивному рівні. Будь-яке різке зростання модуля змінної $v(k)$ спричиняє збільшення дисперсії послідовності $\{\varepsilon(k)\}$; при цьому чим більше значення має коефіцієнт α_1 , тим більшим буде вплив $v(k)$ на дисперсію. Більше того, при більших значеннях параметра авторегресії a_1 у моделі для основної змінної, збільшується вплив правої частини на $y(k)$. І чим більшою буде тенденція до зміни середнього значення послідовності $\{y(k)\}$, тим вищою буде її дисперсія.

Умовні значення середнього та дисперсії для послідовності $\{y(k)\}$ можна записати так:

$$E_{k-1}[y(k)] = a_0 + a_1 y(k-1) \quad ,$$

і

$$\begin{aligned} \text{Var}[y(k) | y(k-1), y(k-2), \dots] &= E_{k-1}[y(k) - a_0 - a_1 y(k-1)]^2 = \\ &= E_{k-1}[\varepsilon^2(k)] = \alpha_0 + \alpha_1 \varepsilon^2(k-1). \end{aligned}$$

Оскільки α_1 і $\varepsilon^2(k-1)$ не можуть приймати від'ємних значень, то мінімальним значенням умовної дисперсії буде α_0 . Будь-яке ненульове значення змінної $\varepsilon(k-1)$ додатно зв'язане з умовною дисперсією змінної $y(k)$ через коефіцієнт α_1 . Безумовні значення середнього та дисперсії для $y(k)$ можна отримати за допомогою розв'язку різницевого рівняння, яке описує процес $\{y(k)\}$ та математичного сподівання. Як було показано вище, для рівняння першого порядку (за умови, що процес починається досить далеко в минулому і можна знехтувати довільною константою A у однорідному розв'язку) розв'язок має вигляд:

$$y(k) = \frac{a_0}{1 - a_1} + \sum_{i=0}^{\infty} a_1^i \varepsilon(k-i). \quad (9.1.9)$$

Оскільки $E[\varepsilon(k)] = 0, \forall k$, то безумовне математичне сподівання для останнього рівняння матиме вигляд:

$$E[y(k)] = \frac{a_0}{1 - a_1}.$$

Безумовну дисперсію для $y(k)$ також можна отримати на основі (9.1.9):

$$\text{Var}[y(k)] = \sum_{i=0}^{\infty} a_1^{2i} \text{Var}[\varepsilon(k-i)].$$

Якщо врахувати те, що безумовна дисперсія для $\varepsilon(k)$ – величина постійна, тобто $\text{Var}[\varepsilon(k)] = \text{Var}[\varepsilon(k-1)] = \text{Var}[\varepsilon(k-2)] = \dots = \alpha_0 / (1 - \alpha_1)$, то

$$\text{Var}[y(k)] = \frac{\alpha_0}{(1 - \alpha_1)} \frac{1}{(1 - a_1^2)}.$$

Модель гетероскедастичного процесу (9.1.4) може бути розширена до довільного порядку і записана у вигляді:

$$\varepsilon(k) = v(k) \left(\alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) \right)^{1/2}. \quad (9.1.10)$$

У цьому рівнянні на змінну впливають всі значення від до а тому умовну дисперсію можна формально розглядати як процес авторегресії порядку

9.2 Узагальнений авторегресійний умовно гетероскедастичний (УАРУГ) процес

Подальшим розширенням моделі АРУГ є опис умовної дисперсії як процесу авторегресії з ковзним середнім, що сприяє підвищенню адекватності цієї моделі стосовно опису динаміки дисперсії. Нехай похибки описуються рівнянням:

$$\varepsilon(k) = v(k) [h(k)]^{1/2},$$

де $\sigma_v^2 = 1$, а умовна дисперсія $h(k)$ визначається за виразом:

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (9.2.1)$$

Оскільки процес $\{v(k)\}$ визначено як білий шум, некорельований із значеннями $\varepsilon(k-i)$, то умовне і безумовне середнє для $\varepsilon(k)$ дорівнюють нулю. Очевидно, що безумовне математичне сподівання визначається так:

$$E[\varepsilon(k)] = E[v(k) (h(k))^{1/2}] = 0.$$

Звідси умовна дисперсія змінної $\varepsilon(k)$ визначиться так: $E_{k-1}[\varepsilon^2(k)] = h(k)$. Вибіркову умовну дисперсію $h_g(k)$ можна досить наближено визначити за виразом:

$$h_g(k) = \frac{1}{k-1} \sum_{i=1}^k [y(i) - \bar{y}_g(k)]^2, \quad k = 2, 3, \dots, N,$$

де $\bar{y}_g(k)$ – вибіркоче умовне середнє.

Узагальнена модель АРУГ, яку скорочено називають УАРУГ(p, q), складається з двох компонент: авторегресії та ковзного середнього стосовно дисперсії гетероскедастичного процесу. Процес першого порядку отримуємо

при $p=0, q=1$; формально його можна визначити як процес УАРУГ(0,1). Якщо всі коефіцієнти $\beta_i=0$, то модель УАРУГ(p,q) еквівалентна моделі АРУГ(q). Для того щоб забезпечити скінченність умовної дисперсії, корені характеристичного рівняння, записаного для (9.2.1), повинні знаходитися всередині кола одиничного радіусу на комплексній площині.

Основним характерним моментом моделі УАРУГ є те, що збурення, яке діє на процес $\{y(k)\}$, – це процес АРКС. Тому можна очікувати, що залишки (похибки) моделі АРКС (попередньої моделі процесу) будуть відповідати за своїми характеристиками гетероскедастичному процесу.

Це твердження можна пояснити таким чином. Нехай $\{y(k)\}$ – процес АРКС. Якщо модель АРКС адекватна процесу $\{y(k)\}$, 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. Оскільки $E_{k-1}[\varepsilon(k)] = (h(k))^{1/2}$, то рівняння (9.1.11) можна записати у вигляді:

$$E_{k-1}[\varepsilon^2(k)] = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i). \quad (9.2.2)$$

Останнє рівняння формально схоже на модель АРКС(p,q) стосовно послідовності $\{\varepsilon^2(k)\}$. Якщо процес містить умовну гетероскедастичність, то корелограма повинна вказувати на цей факт. Алгоритм аналізу корелограми квадратів залишків та побудови моделі динаміки дисперсії можна сформулювати таким чином:

Крок 1:

Побудувати просту модель низького порядку АРКС (або АР) для основної досліджуваної послідовності $\{y(k)\}$ (як правило, це модель першого або другого порядку) і обчислити додатковий ряд із квадратів похибок $\hat{\varepsilon}^2(k)$. Обчислити вибірку (на основі значень елементів ряду) дисперсію залишків:

$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{k=1}^N \hat{\varepsilon}^2(k),$$

де N – кількість елементів ряду залишків.

Крок 2:

Обчислити вибірккову автокореляційну функцію квадратів залишків за виразом:

$$\rho(s) = \frac{\sum_{k=s+1}^N [\hat{\varepsilon}^2(k) - \hat{\sigma}^2][\hat{\varepsilon}^2(k-s) - \hat{\sigma}^2]}{\sum_{k=1}^N [\hat{\varepsilon}^2(k) - \hat{\sigma}^2]^2}, \quad s = 0, 1, 2, \dots$$

Крок 3:

Для вибірок досить великої довжини стандартне відхилення для змінної $\rho(s)$ приблизно прирівнюють значенню $N^{1/2}$. Індивідуальні значення змінної $\rho(s)$, які суттєво відрізняються від нуля, вказують на наявність процесу УАРУГ. Для того щоб встановити, наскільки значення вибіркової АКФ є відмінними від нуля (тобто чи вони значущі в статистичному сенсі), можна застосувати Q – статистику Льюнга-Бокса:

$$Q = N(N+2) \sum_{i=1}^s \rho(i)/(N-i).$$

Якщо значення $\hat{\varepsilon}^2(k)$ некорельовані між собою, то Q – статистика асимптотично має розподіл χ^2 з n ступенями свободи. За нульову гіпотезу можна прийняти те, що значення $\hat{\varepsilon}^2(k)$ не корельовані між собою. Відмова від цієї нульової гіпотези еквівалентна тому, що процес АРУГ чи УАРУГ у конкретному випадку наявний.

Простіше перевірку на наявність гетероскедастичності можна здійснити за допомогою такої двокрокової процедури:

Крок 1:

За допомогою МНК побудувати модель $AR(p)$ для основної змінної:

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + \dots + a_n y(k-p) + \varepsilon(k).$$

При цьому вибір кращої моделі виконується за наведеною вище 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. За допомогою МНК оцінити авторегресію для нового ряду:

$$\boxed{\hspace{15em}} \quad (9.2.3)$$

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 $\alpha_1, \dots, \alpha_q$ 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.

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 дохід є нижчим очікуваного; це призводить до зростання волатильності. Навпаки, додатні інновації, $\varepsilon(k) > 0$, тобто “хороші новини” – це ситуація, коли фактичний дохід є вищим за очікуваний, приводять 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.

По відношенню до узагальнених умовно гетероскедастичних процесів не узгоджено означення стаціонарності, тобто сильно стаціонарний процес УАРУГ не завжди буде слабо стаціонарним. Таким чином, виникає проблема коректного означення стаціонарності для 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 виконання таких умов: $E[y(k)] = const$, $E[y^2(k)] = const$ та $E[y(k)y(k-s)] = const$.

Обмеження області допустимих значень параметрів α і β додатними значеннями створюють додаткові труднощі з оцінюванням параметрів моделей АРУГ.

9.3 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 $g(\cdot)$:

$$\log[h(k)] = c_0 + \sum_{i=1}^{\infty} c_i g[y(k-i)], \quad (9.3.1)$$

$$g(y) = \alpha y(k) + \beta[|y(k)| - E|y(k)|], \quad (9.3.2)$$

де $E[g(y)] = c_i$, α, β – параметри моделі; $y(k)$ – основна змінна, що моделюється.

Оскільки $E[g(y)] = 0$, то $\log[h(k)]$ – це процес лінійної фільтрації, а рівняння (9.3.1) – ковзне середнє нескінченного порядку. На практиці для умовної дисперсії застосовують представлення типу АРКС(p, q):

$$\log[h(k)] = c_0 + \frac{1 + \alpha(L)}{1 - \beta(L)} g[y(k)], \quad (9.3.3)$$

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 (9.3.1) – (9.3.3) не накладаються. Обидві складові функції $g(y)$, яка визначається виразом (9.3.2), мають нульове середнє. Якщо процес $\{y(k)\}$ має симетричний розподіл, то вони ортогональні.

В інтервалі $0 < y(k) < +\infty$ функція $g(y)$ лінійна стосовно $y(k)$ з коефіцієнтом нахилу $\alpha + \beta$, а в інтервалі $-\infty < y(k) < 0$ вона лінійна стосовно $y(k)$ 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 $\alpha < 0, \beta = 0$. У такому випадку інновація умовної дисперсії додатна (від'ємна), якщо похибка прогнозу доходності від'ємна (додатна).

Член $\beta [|y(k)| - E|y(k)|]$ представляє в моделі “ефект абсолютного значення” за аналогією з моделями АРУГ. Припустимо, що $\alpha = 0, \beta > 0$; 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

$$\text{Cov} [\log(h(k)), y(k-i)] = c_k \alpha.$$

Ефект левериджу має місце при $\alpha < 0$.

Ще одна форма моделі ЕУАРУГ має вигляд:

$$\log[h(k)] = \alpha_0 + \sum_{i=1}^p \alpha_i \frac{|\varepsilon(k-i)|}{\sqrt{h(k-i)}} + \sum_{i=1}^p \beta_i \frac{\varepsilon(k-i)}{\sqrt{h(k-i)}} + \sum_{i=1}^q \gamma_i \log[h(k-i)] + v(k). \quad (9.3.4)$$

В даній моделі є “стандартна” частина, яка враховує модуль інновацій $\varepsilon(k)$, і частина, яка враховує знак цієї інновації. Значення $\varepsilon(k)$ нормовані стосовно волатильності, що дозволяє зменшити великі імпульсні значення цієї величини у випадку їх наявності. Значення умовної дисперсії згладжені шляхом логарифмування.

Модель стохастичної волатильності

В моделі такого типу мінлива структура дисперсії описується стохастичним, наприклад, нормально розподіленим процесом. Структуру стохастичної моделі волатильності можна представити таким чином:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k), \quad (9.3.5)$$

$$\varepsilon(k) = \sqrt{h(k)} v(k), \quad \{v(k)\} \sim N(0,1), \quad (9.3.6)$$

$$\ln[h(k+1)] = \alpha_0 + \alpha_1 \ln[h(k)] + \eta(k), \quad (9.3.7)$$

$$\{\eta(k)\} \sim N(0, \sigma_\eta^2),$$

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 (9.3.5) з інновацією $\varepsilon(k)$ пояснює можливе існування авторегресії у процесі $\{y(k)\}$. Для оцінювання таких моделей застосовують метод максимальної правдоподібності, метод моментів та байєсівський підхід у поєднанні з методом Монте Карло. Один із методів оцінювання цієї моделі буде розглянуто у наступному розділі.

9.4 Врахування великих імпульсних значень у моделях ГСП

Моделі АРУГ дозволяють 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 ексцес (куртозис). Стандартну 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 даних 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 у даному параграфі пропонується підхід до побудови моделей УАРУГ, які враховують вказаний ефект, що дозволяє підвищити адекватність моделі.

Розглянемо часовий ряд $\{y(k)\}$, $k = 1, 2, \dots, N$, який можна описати з деякою ступінню адекватності моделлю авторегресії з ковзним середнім АРКС(p, q), тобто:

$$y(k) = \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j), \quad (9.4.1)$$

де $\varepsilon(k)$ – випадкова послідовність типу білого шуму з нульовим середнім та дисперсією σ_ε^2 . В моделі (9.4.1) 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 (9.4.1) має збіжний розв'язок, тобто корені його характеристичного рівняння знаходяться в колі одиничного радіуса на комплексній площині. Рівняння

(9.4.1) можна записати також із застосування оператора зсуву L :

$$(1 - a_1L - a_2L^2 - \dots - a_pL^p)y(k) = (1 + b_1L + b_2L^2 + \dots + b_qL^q)\varepsilon(k),$$

де L – оператор затримки в часі, $Ly(k) = y(k-1)$.

Нехай модель, що враховує значні адитивні імпульсні складові часового ряду, має вигляд:

$$z(k) = y(k) + \beta u(k, l), \quad (9.4.2)$$

де $u(k, l) = \begin{cases} 1, & k = l, \\ 0, & k \neq l. \end{cases}$ Тобто фактичні спостереження описуються рівнянням

(9.4.2), а не рівнянням (9.4.1). 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:

$$\hat{\varepsilon}(k) = \frac{A(L)}{B(L)} y(k) = D(L) y(k), \quad (9.4.3)$$

де $D(L) = A(L)/B(L) = (1 - d_1L - d_2L^2 - \dots)$. Значення порядку авторегресії p та ковзного середнього q можна визначити в процесі побудови моделі (9.4.1). Якщо ряд містить значний імпульс, то рівняння (9.4.3) приймає вигляд:

$$\hat{\varepsilon}(k) = \varepsilon(k) + \beta D(L)u(k, l), \quad (9.4.4)$$

яке можна розглядати як модель регресії для $\hat{\varepsilon}(k)$:

$$\hat{\varepsilon}(k) = \beta \omega(k) + \varepsilon(k), \quad (9.4.5)$$

де $\omega(k) = \begin{cases} 0, & k < l, \\ 1, & k = l, \end{cases}$ і $\omega(k+j) = -D(j)$, $k > l$, $j = 1, 2, \dots$

Тепер вплив суттєвого імпульсного значення на процес в момент $k = l$ можна представити у моделі таким чином:

$$\hat{\beta}(l) = \frac{\sum_{k=l}^N \hat{\varepsilon}(k)\omega(k)}{\sum_{k=l}^N \omega^2(k)}. \quad (9.4.6)$$

Функцію $\hat{\beta}(l)$ можна представити у вигляді, зручному для тестування її значущості. Для цього необхідно визначити оцінку дисперсії залишків, отриманих при побудові моделі (9.4.1). При цьому бажано, щоб оцінка $\hat{\beta}(l)$ не була суттєво зміщеною внаслідок наявності імпульсів у початковому (первісному) ряді значень. Оцінку дисперсії залишків можна отримати різними способами. Один з них полягає у тому, що дисперсія обчислюється для залишків моделі, з яких вилючається спостереження (імпульс) при $k=l$. Якщо позначити цю дисперсію через $\hat{\sigma}_e^2$, то можна записати стандартний статистичний параметр у вигляді:

$$\hat{\tau} = \frac{\hat{\beta}(l) / \hat{\sigma}_e}{\left[\sum_{k=l}^N \omega^2(k) \right]^{1/2}}. \quad (9.4.7)$$

Вплив імпульсного значення ряду вважається значним, якщо $\hat{\tau} > c$, де c – деяке визначене порогове значення. Оскільки в асимптотиці $\hat{\tau}$ 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 (9.4.6), 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 незначущою. 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, $d(k) = \log[p(k)] - \log[p(k-1)]$, і запишемо можливу структуру узагальненої авторегресійної умовно гетероскедастичної моделі у вигляді:

$$d(k) = \eta(k) h^{1/2}(k), \quad (9.4.8)$$

$$h(k) = \alpha_0 + \alpha_1 d^2(k-1) + \gamma_1 h(k-1), \quad (9.4.9)$$

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 (9.4.8), (9.4.9)

представляють собою модель УАРУГ. Але якщо $\eta(k)$ має умовний t – розподіл, то модель (9.4.8), (9.4.9) називають УАРУГ – t . Для покращення опису ряду даних з імпульсними значеннями скористаємось стандартною моделлю УАРУГ, але будемо корегувати її із врахуванням наявності імпульсних значень в ряді $\eta(k)$.

Рівняння (9.4.9) описує умовну мінливість величин $d(k)$, а тому ним можна скористатись для прогнозування мінливості цін, доходів і т. ін. Представимо модель УАРУГ у вигляді:

$$d^2(k) = \alpha_0 + (\alpha_1 + \gamma_1)d^2(k-1) + x(k) - \gamma_1x(k-1), \quad (9.4.10)$$

де $x(k) \equiv d^2(k) - h(k)$. Рівняння (9.4.10) – 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 врахуванням наявності імпульсних значень можна представити у вигляді наступних кроків:

Крок 1. Обчислити оцінки параметрів рівнянь (9.4.8), (9.4.9) за методом максимальної правдоподібності. В результаті отримаємо оцінки $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\gamma}_1$, а також оцінки значень рядів $\hat{h}(k), \hat{\eta}(k)$. Це дозволить знайти ряд значень оцінок, $\hat{x}(k) = d^2(k) - \hat{h}(k)$, який використовується на наступному кроці при визначенні наявності імпульсних значень.

Крок 2. З рівняння (9.4.3) 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 (9.4.10) 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 в момент часу $k = l$. Для того 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 (9.4.7), обчислюють для $\hat{x}(k)$ і порівнюють 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 (9.4.2)) з використанням вагового коефіцієнта $\beta(l)$ (вираз (9.4.6)).

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 $k = l$. Скорегувати ряд даних стосовно імпульсних значень таким чином:

$$d^*(k) = \begin{cases} d(k), & k \neq l, \\ \text{sign}[d(k)] \cdot [d^{*2}(k)]^{1/2}, & k = l. \end{cases} \quad (9.4.11)$$

Наведений вираз показує, що $\text{sign}[d^2(k)] = \text{sign}[d(k)]$ при $k = l$. З цього випливає, що для членів ряду, які корегуються 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: $\hat{t} > 4$, 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 (9.4.9).

Якщо $d(k)$ в рівнянні (9.4.8) 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 обробляють за розглянутим алгоритмом. Однак, необхідно пам'ятати, що коефіцієнт корекції β в рівнянні (9.4.2) 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 скорегованого стосовно імпульсних значень ряду $d^*(k)$. Для прикладу візьмемо дані, які характеризують щотижневий доход російської 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 даного часового ряду оцінюються моделі $AR(p)$, потім $УАРУГ(1,1)$, а також модель для доходу, що має структуру:

$$\begin{aligned}d(k) &= \mu + a_1 d(k-1) + a_2 d(k-2) + \varepsilon(k), \\ \varepsilon(k) &= \eta(k) h^{1/2}(k), \quad \eta(k) \sim N(0,1), \\ h(k) &= \alpha_0 + \alpha_1 \varepsilon^2(k-1) + \gamma_1 h(k-1).\end{aligned}$$

Результати оцінювання наведено в таблиці 9.1. Параметри моделей оцінені за методом максимальної правдоподібності. Діагностичні статистичні параметри Q_1 , Q_2 , які наведені у двох останніх стовпчиках таблиці, свідчать про те, що моделі мають достатню ступінь адекватності. Параметри $Q_1(10)$ і $Q_2(10)$ – це статистики Льюнга-Бокса, 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 9.1

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 9.2 наведено результати оцінювання тих же моделей, але вже із врахуванням корекції внаслідок наявності суттєвих імпульсних значень, тобто модель побудована для $d^*(k)$.

Таблиця 9.2

Результати оцінювання моделей $AR(p)$ та $УАРУГ(1,1)$ після корекції

Вибірка даних	Параметри моделі					
	$\hat{\mu}$	\hat{a}_1	\hat{a}_2	α_0	α_1	γ_1
1996-	0,0043	0,0021	0,353	0,0002	0,118	0,694

1999	(2,45)	(2,39)	(3,19)	(2,18)	(4,02)	(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 9.1 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 стосовно імпульсних значень. На основі отриманих даних моделювання можна зробити висновок, що застосування корекції даних приводить до підвищення стійкості оцінок параметрів для різних часових вибірок даних.

9.5 Приклади побудови математичних моделей гетероскедастичних процесів

Приклад 1. Модель інфляції.

Моделі типу АРУГ та УАРУГ зручні тим, що дають можливість 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, при встановленні об'ємів інвестицій на майбутнє, об'ємів договорів на поставки продукції і т. ін.

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.

Взято відомі поквартальні дані стосовно інфляції в Англії за період з другого кварталу 1958 р. по другий квартал 1977р. (всього 80 значень). Введемо такі позначення:

- $p(k)$ – логарифм індексу споживчих цін;
- $W(k)$ – логарифм індексу номінальної (середньої) заробітної плати;
- рівень інфляції визначимо першою різницею: $\pi(k) = p(k) - p(k-1)$.
- реальний прибуток: $r(k) = w(k) - p(k)$.

Перший етап побудови моделі

За часовим рядом побудована така модель:

$$\pi(k) = 0,0257 + 0,394 \pi(k-1) + 0,408 \pi(k-4) - 0,404 \pi(k-5) + 0,0559 r(k-1) + \varepsilon_1(k), \quad (9.5.1)$$

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:

$$\varepsilon_1^2(k) = \alpha_0 + \alpha_1 [0,4 \varepsilon_1^2(k-1) + 0,3 \varepsilon_1^2(k-2) + 0,2 \varepsilon_1^2(k-3) + 0,1 \varepsilon_1^2(k-4)]. \quad (9.5.2)$$

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 складається з двох кроків (у відповідності з кроком 5 методики, наведеної вище):

1-й крок:

- спочатку оцінюють коефіцієнти рівняння (9.5.1) і залишки;
- на основі залишків оцінюють коефіцієнти α_0 і α_1 рівняння (9.5.2);

2-й крок:

- уточнення коефіцієнтів моделі (9.5.1) після знаходження коефіцієнтів моделі (9.5.2).

Остаточна отримана модель такої структури:

$$\pi(k) = 0,0328 + 0,162 \pi(k-1) + 0,264 \pi(k-4) - 0,325 \pi(k-5) + 0,0707 r(k-1) + \varepsilon_2(k), \quad (9.5.3)$$

$$h(k) = \varepsilon_2^2(k) + 1,4 E^{-5} + 0,955 [0,4 \varepsilon_2^2(k-1) + 0,3 \varepsilon_2^2(k-2) + 0,2 \varepsilon_2^2(k-3) + 0,1 \varepsilon_2^2(k-4)];$$

$h(k)$ – спрогнозована 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 за допомогою одного з відомих методів, то це часто призводить до порушення обмеження невід'ємності.

Приклад 2. Моделювання утворення цін на продукцію фірми.

Вхідні та вихідні змінні виробничої фірми схематично зображені на рис. 9.1. 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.

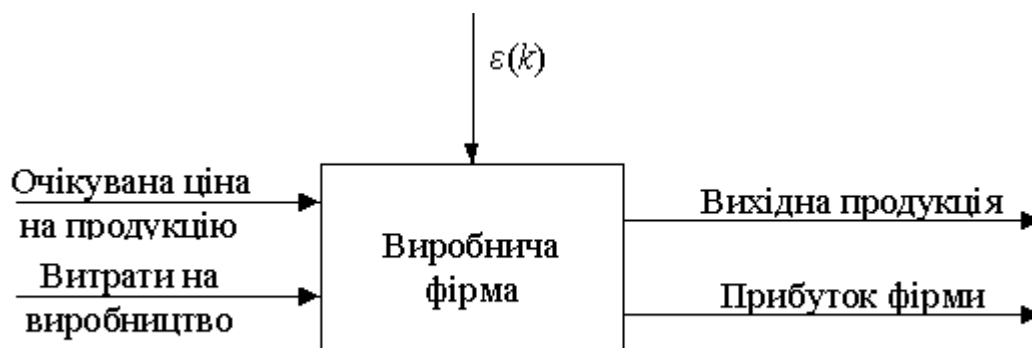


Рис. 9.1 Вхідні та вихідні змінні виробничої фірми

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 $(k-1)$;

$pp(k-1)$ – витрати 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 опису прибутку обрана така структура моделі:

$$q(k) = a_0 + a_1 p^e(k) + a_2 h(k) + a_3 pp(k-1) + a_4 p(k-1) + a_5 q(k-4) + \varepsilon_1(k). \quad (9.5.4)$$

Для обчислення $p^e(k)$ і $h(k)$ будуються окремі моделі на основі кварталних 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 $q(k-4)$ введено в рівняння із лагом 4 для того 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 AP(4):

$$(1 - \beta_1 L - \beta_2 L^2 - \beta_3 L^3 - \beta_4 L^4) p(k) = \beta_0 + \varepsilon_2(k). \quad (9.5.5)$$

Оцінена модель має вигляд:

$$(1 - 0,511L - 0,129 L^2 - 0,130 L^3 - 0,138 L^4) p(k) = 1,632 + \varepsilon_2(k). \quad (9.5.6)$$

Тепер підставимо $p^e(k) = p(k)$ в (9.5.5):

$$h(k) = 1,353 + 0,162 \varepsilon_2^2(k-1) + 0,591 h(k-1). \quad (9.5.7)$$

Значення, обчислені в (9.5.6) і (9.5.7) підставляються в (9.5.4) для того щоб оцінити рівень ціни на продукцію. Коефіцієнти в (9.5.6) і (9.5.7) відносяться до прибутку птахофабрики з вирощування бройлерів.

Для обчислення оцінок $\hat{a}_0, \hat{a}_1, \dots, \hat{a}_5$ необхідно мати 5 рядів змінних:

$$q(k) = \underbrace{2,767 p^e(k)}_{\text{Якщо, очікується ріст ціни, то прибуток зростає}} - \underbrace{0,521 h(k) + 4,325 pp(k-1)}_{\text{Коливання ціни і витрати зменшують прибуток}} + \underbrace{1,887 p(k-1) + 0,603 q(k-4)}_{\text{Позитивні фактори}} + \varepsilon_1(k).$$

Другий підхід до прогнозу ціни ґрунтується на відомому виразі для ковзного середнього:

$$p^e(k) = \alpha p(k-1) + (1-\alpha) p(k-1),$$

$$p^e(k) = \alpha \sum_{i=0}^{\infty} (1-\alpha)^i p(k-1-i).$$

За аналогією з цим рівнянням можемо записати рівняння для дисперсії похибок прогнозованої ціни:

$$h(k) = \beta \sum_{i=0}^{\infty} (1-\beta)^i [p(k-1-i) - p^e(k-1-i)]^2, \quad 0 < \beta < 1.$$

Очевидно, що другий підхід являється більш формальним (не пов'язаним з вхідними даними).

Приклад 3. Модель прибутку від продажу майна або цінних паперів (модель АРУГ-М). Даний підхід до моделювання може бути використаний у випадках, коли необхідно прогнозувати дохід 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 опису приросту прибутку від затримки продажу ризикованого майна або цінних паперів скористаємось рівнянням:

$$y(k) = \mu(k) + \varepsilon(k), \quad (9.5.8)$$

де $y(k)$ – приріст прибутку за рахунок затримки продажу ризикованого майна за умови, 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 короткий строк; $\varepsilon(k)$ – випадкове коливання прибутку на даному інтервалі часу.

Очікуваний прибуток визначається умовним математичним сподіванням:

$$E_{k-1}[y(k)] = \mu(k).$$

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:

$$\mu(k) = \beta + \gamma h(k), \quad \gamma > 0, \quad (9.5.9)$$

де $h(k)$ – умовна дисперсія процесу $\varepsilon(k)$; $\beta = const$. Для опису умовної дисперсії виберемо модель АРУГ:

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i). \quad (9.5.10)$$

Рівняння (9.5.8)-(9.5.10) 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 (9.5.8) є

$$\begin{aligned}y(k) &= 0,185 + \varepsilon_1(k), \\ \mu(k) &= -0,0368 + 0,745h(k) + \varepsilon_2(k), \\ h(k) &= 0,0031 + 1,73[0,4\varepsilon_1^2(k-1) + 0,3\varepsilon_1^2(k-2) + 0,2\varepsilon_1^2(k-3) + 0,1\varepsilon_1^2(k-4)],\end{aligned}$$

де коефіцієнт 0,185 – характеризує міру прибутку в процентах за квартал (при використанні кварталних 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 4. Дисперсія – міра валютного ризику банку. Оцінюванню валютного ризику приділяють у 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 на протязі деякого 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 означенням *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:

$$c = F(x) = \int_{-\infty}^x f(u) du.$$

Тобто ймовірність того, що $x_i < x$ дорівнює c : $p(x_i < x) = c$. Ймовірність того, що $x_i > x$ визначається так: $p(x_i \geq x) = 1 - c$. Квантиль позначають через $Q(X, c)$. 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 імовірністю, скажемо, $p = 95\%$. Якщо $f(u)$ – розподіл втрат (прибутків), то означення VaR зв'язане з виразом:

$$F(x) = \int_{-\infty}^x f(u) du = 1 - p,$$

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:

$$VaR(c) = E[X] - Q(X, c),$$

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 оцінювання валютних ризиків, тобто ризиків, пов'язаних із використанням іноземних валют. Для оцінювання валютних позицій використовують курс НБУ.

Вихідні (початкові) дані, необхідні для розрахунку VaR:

- значення курсів НБ для валют 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 стосовно курсів валют та відкритих валютних позицій.

Крок 2. На другому кроці розраховують темпи зміни курсів валют за виразом:

$$x^i(k) = \ln \left(\frac{\text{Курс}^i(k)}{\text{Курс}^i(k-1)} \right), k = 1, 2, 3, \dots, N, \quad i = 1, 2, \dots, n,$$

де $\text{Курс}^i(k)$ – значення курсу i -ї валюти в момент k ; N – загальне число значень статистичних даних; n – кількість 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} :

$$x_{ij}^H = \frac{x_{ij} - \mu_i}{\sqrt{\text{var}[x_i]}} = \frac{x_{ij} - \mu_i}{\sigma_{x_i}}, \quad i = 1, \dots, n; \quad j = 1, \dots, N,$$

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 отримуємо 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 $\mathbf{R}[n \times n]$:

$$\mathbf{R} = \mathbf{X}_H^T \mathbf{X}_H.$$

Крок 4. Прогнозування волатильності. Деякі методики обчислення прогнозу пропонують такий вираз: $\sigma_i^{\text{прогн}} = a_i \sigma_i$, де $a_i = F(\mathbf{R}, \sigma_i)$. 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:

$$VaR_i = \alpha \sigma_i^{прогн} V_i \sqrt{N}, \quad i = 1, \dots, n,$$

де α – квантиль для довірчого інтервалу (для інтервалу 95% він складає 1,65, а для 99% дорівнює 2,33); $\sigma_i^{прогн}$ – 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:

$$V_i = \text{Позиція}_i \cdot \text{Курс}_i.$$

Обчислені значення VaR_i утворюють вектор:

$$VaR = [VaR_1 \quad VaR_2 \quad \dots \quad VaR_n]^T,$$

який використовують для обчислення загального значення $VaR_{заг.}$:

$$VaR_{заг.} = \sqrt{VaR^T \cdot \mathbf{R} \cdot VaR}.$$

Значення $\sigma_i^{прогн}$ можуть корегуватись у бік зростання в залежності від того, яка є додаткова інформація стосовно можливих змін курсів.

Виконані розрахунки свідчать, що можливі втрати (наступного дня) за валютними позиціями не перевищать значення VaR_i . Існують модифіковані варіанти розглянутої методики, які ґрунтуються на інших розподілах випадкових величин.

9.6 Оцінювання моделі стохастичної волатильності

Одновимірна модель стохастичної волатильності. Розглянемо задачу оцінювання параметрів стохастичної моделі волатильності (МСВ) для моделі процесу $\{y(k)\}$ та його дисперсії загального вигляду:

$$y(k) = \beta_0 + \beta_1 x_1(k) + \beta_2 x_2(k) + \beta_p x_p(k) + u(k), \quad (9.6.1)$$

$$u(k) = \sqrt{h(k)} \varepsilon(k),$$

$$\ln[h(k)] = \alpha_0 + \alpha_1 \ln[h(k-1)] + v(k), \quad (9.6.2)$$

$$\{\varepsilon(k)\} \sim N(0, 1); \quad \{v(k)\} \sim N(0, \sigma_v^2);$$

$$E[\varepsilon(k)v(k-l)] = 0, \quad \forall k, l, \quad (9.6.3)$$

де $\{x_i(k), i=1, \dots, p\}$ – пояснючі змінні. Логарифмування змінної $h(k)$ застосовується для того, щоб забезпечити згладжування всіх її додатних значень, оскільки вона має смисл умовної дисперсії. Пояснючі змінні $x_i(k)$ можуть включати затримані у часі значення залежної змінної, тобто $x_i(k) = y(k-i)$. Для забезпечення стаціонарності процесу $\ln[h(k)]$ необхідно щоб виконувалась умова: $|\alpha_1| < 1$. Очевидно, що рівняння (9.6.2) може мати, при необхідності, вищий порядок, тобто це може бути авторегресія $AR(p)$.

Позначимо вектор параметрів моделі (9.6.1) через $\beta = [\beta_0, \beta_1, \dots, \beta_p]^T$, а вектор параметрів моделі (9.6.2) через $\theta = [\alpha_0, \alpha_1, \sigma_v^2]^T$. Нехай $\mathbf{y} = [y(1), y(2), \dots, y(N)]^T$ – вектор спостережень залежної змінної; \mathbf{X} – матриця спостережень пояснюючих змінних; $\mathbf{h} = [h(1), h(2), \dots, h(N)]^T$ – вектор невимірюваних значень волатильності процесу $\{y(k)\}$, тобто вектор умовних дисперсій цього процесу, який ми будемо розглядати як допоміжну змінну математичної моделі (9.6.1) – (9.6.3). Оцінювання такої моделі за допомогою методу максимальної правдоподібності пов'язане із значними труднощами, оскільки функція правдоподібності в цій задачі представляє собою N -вимірний розподіл вектора \mathbf{h} такого вигляду:

$$f(\mathbf{y} | \mathbf{X}, \beta, \theta) = \int f(\mathbf{y} | \mathbf{X}, \beta, \mathbf{h}) f(\mathbf{h} | \theta) d\mathbf{h}. \quad (9.6.4)$$

Однак у байєсовій постановці задачі вектор волатильностей \mathbf{h} включає в себе так звані розширені параметри, а процес в цілому характеризується розподілами умовних ймовірностей $f(\mathbf{y} | \mathbf{h}, \beta)$ і $f(\mathbf{h} | \theta)$ з апіорним розподілом параметрів моделі $p(\beta, \theta)$. Припустимо, що апіорний розподіл параметрів можна представити у вигляді добутку: $p(\beta, \theta) = p(\beta) p(\theta)$, тобто апіорні розподіли для параметрів основної регресії і параметрів для рівняння, що описує волатильність, – незалежні. Тепер за методом Гіббса необхідно згенерувати випадкові вибірки для оцінювання параметрів рівнянь (9.6.1) і (9.6.2) з таких умовних апостеріорних розподілів:

$$f(\beta | \mathbf{y}, \mathbf{X}, \mathbf{h}, \theta), \quad f(\mathbf{h} | \mathbf{y}, \mathbf{X}, \beta, \theta), \quad f(\theta | \mathbf{y}, \mathbf{X}, \beta, \mathbf{h}).$$

Розглянемо спочатку одновимірний випадок (одна залежна змінна). Якщо поділити обидві частини рівняння (9.6.1) на $\sqrt{h(k)}$, то отримаємо таку

МОДЕЛЬ:

$$y_0(k) = \mathbf{x}_0^T(k)\beta + \varepsilon(k), \quad k = 1, 2, \dots, N, \quad (9.6.5)$$

де $y_0(k) = y(k)/\sqrt{h(k)}$; $\mathbf{x}_0(k) = \mathbf{x}(k)/\sqrt{h(k)}$; $\mathbf{x}(k) = [1, x_1(k), \dots, x_p(k)]^T$. Припустимо, що вектор β має багатовимірний нормальний апіорний розподіл: $\beta \sim \mathbf{N}(\beta_0, \mathbf{B}_0)$, де \mathbf{B} – відповідна коваріаційна матриця. В такому випадку апостеріорний розподіл вектора β також буде нормальним із середнім β^* та коваріаційною матрицею \mathbf{B}^* . Ці дві величини можна обчислити за виразами:

$$\mathbf{B}^* = \sum_{k=1}^N \mathbf{x}_0(k) \mathbf{x}_0^T(k) + \mathbf{B}_0^{-1}, \quad \beta^* = \mathbf{B}^* \left(\sum_{k=1}^N \mathbf{x}_0(k) y_0(k) + \mathbf{B}_0^{-1} \beta_0 \right).$$

Обчислення починаються із $(p+1)$ -го значення для того щоб врахувати $y(k-p)$ попередніх значень.

Елементи вектора волатильностей обчислюються послідовно один за одним за допомогою умовного апостеріорного розподілу $f[h(k) | \mathbf{y}, \mathbf{X}, \mathbf{h}(-k), \beta, \theta]$, який формується з нормального розподілу для $u(k)$ та логнормального розподілу для волатильності:

$$\begin{aligned} & f[h(k) | \mathbf{y}, \mathbf{X}, \mathbf{h}(-k), \beta, \theta] \propto \\ & \propto f[u(k) | h(k), y(k), \mathbf{x}(k), \beta] f[h(k) | h(k-1), \theta] f[h(k+1) | h(k), \theta] \propto \\ & \propto \frac{1}{\sqrt{h(k)}} \exp\left[-\frac{[y(k) - \mathbf{x}^T(k)\beta]^2}{2h(k)}\right] h^{-1}(k) \exp\left[-\frac{[\ln h(k) - \mu(k)]^2}{2\sigma^2}\right] \propto \\ & \propto \frac{1}{\sqrt{[h(k)]^3}} \exp\left[-\frac{[y(k) - \mathbf{x}^T(k)\beta]^2}{2h(k)} - \frac{[\ln h(k) - \mu(k)]^2}{2\sigma^2}\right], \quad (9.6.6) \end{aligned}$$

$$\text{де } \mu(k) = \frac{\alpha_0(1 - \alpha_1) + \alpha_1[\ln h(k+1)] + \ln h(k-1)}{1 + \alpha_1^2}; \quad \sigma^2 = \frac{\sigma_v^2}{1 + \alpha_1^2}.$$

Для отримання цих характеристик використано такі властивості процесу, що моделюється:

$$\text{а) } \{u(k) | h(k)\} \sim N(0, h(k));$$

$$\text{б) } \{\ln h(k) | \ln h(k-1)\} \sim N(\alpha_0 + \alpha_1 \ln h(k-1), \sigma_v^2);$$

$$в) \{ \ln h(k+1) | \ln h(k) \} \sim N(\alpha_0 + \alpha_1 \ln h(k), \sigma_v^2);$$

$$г) d \ln h(k) = h^{-1}(k) d h(k);$$

$$д) (x-a)^2 A + (x-b)^2 C = (x-c)^2 (A+C) + (a-b)^2 AC / (A+C),$$

де d – оператор диференціювання; $c = (Aa - Cb) / (A + C)$ при умові, що $A + C \neq 0$. Ця рівність є скалярною версією відомої лемми [18]. В даному випадку елементи цієї рівності мають такі значення: $A = 1$, $a = \alpha_0 + \ln h(k-1)$, $C = \alpha_1^2$ і $b = [\ln h(k+1) - \alpha_0] / \alpha_1$. Член $(a-b)^2 AC / (A+C)$ не містить випадкової змінної $h(k)$ і не впливає на умовний апостеріорний розподіл. Для генерування випадкових значень застосуємо алгоритм Гіббса на решітці, при цьому діапазон зміни $h(k)$ вибирається кратним безумовній вибірковій дисперсії процесу $y(k)$.

Для того щоб згенерувати випадкову вибірку вектора параметрів θ , розділимо параметри таким чином: $\alpha = [\alpha_0, \alpha_1]^T$ і σ_v^2 . Відповідно до цього розділимо також функцію апіорного розподілу: $p(\theta) = p(\alpha) p(\sigma_v^2)$. Тепер умовні апостеріорні розподіли можна описати так:

– $f(\alpha | \mathbf{y}, \mathbf{X}, \mathbf{h}, \beta, \sigma_v^2) = f(\alpha | \mathbf{h}, \sigma_v^2)$: при наявності \mathbf{h} можна оцінити параметри моделі АР(1) для умовної дисперсії y відповідності до процедури, розробленої для авторегресійних моделей. За апіорний розподіл для α виберемо багатовимірний нормальний розподіл із середнім α_0 та коваріаційною матрицею \mathbf{C}_0 , що приводить до багатовимірного нормального розподілу $f(\alpha | \mathbf{h}, \sigma_v^2)$ із середнім α^* і коваріаційною матрицею \mathbf{C}^* :

$$(\mathbf{C}^*)^{-1} = \frac{\sum_{k=2}^N \mathbf{z}(k) \mathbf{z}^T(k)}{\sigma_v^2} + \mathbf{C}_0^{-1}, \quad \alpha^* = \mathbf{C}^* \left[\frac{\sum_{k=2}^N \mathbf{z}(k) \ln h(k)}{\sigma_v^2} + \mathbf{C}_0^{-1} \alpha_0 \right],$$

де $\mathbf{z}(k) = [1, \ln h(k-1)]^T$.

– $f(\sigma_v^2 | \mathbf{y}, \mathbf{X}, \mathbf{h}, \beta, \alpha) = f(\sigma_v^2 | \mathbf{h}, \alpha)$: При відомих значеннях \mathbf{h} і α можна обчислити $v(k) = \ln h(k) - \alpha_0 - \alpha_1 \ln h(k-1)$, $k = 2, 3, \dots, N$. Якщо апіорним розподілом для σ_v^2 є $\left(\frac{m\lambda}{\sigma_v^2} \right) \sim \chi_m^2$, то умовним апостеріорним розподілом для σ_v^2 буде інвертований розподіл хі-квадрат з $m+n-1$ ступенями свободи, тобто:

$$\frac{m\lambda + \sum_{k=2}^N v^2(k)}{\sigma_v^2} \sim \chi_{m+n-1}^2.$$

Необхідно зазначити, що вираз (9.6.6) справедливий для $1 < k < N$, де N – довжина вибірки. Крайні значення $h(1)$ і $h(N)$ необхідно модифікувати, наприклад, $h(1) = \text{const}$, а обчислення $h(k)$ починається з $k = 2$. При $k = N$ можна скористатись виразом:

$$\ln[h(N)] \sim [\alpha_0 + \alpha_1 \ln[h(N-1), \sigma_v^2].$$

Можна також скористатись прогнозом $h(N+1)$, а також зворотним прогнозом $h(0)$, а далі вже застосовувати вираз (9.6.6). Оскільки нас цікавить $h(N)$, то значення $h(N+1)$ обчислюється як двокроковий прогноз на основі наявного значення $h(N-1)$. За допомогою моделі (9.6.2), можна записати:

$$\hat{h}_{N-1}(N+1) = \alpha_0 + \alpha_1 [\alpha_0 + \alpha_1 \ln h(N-1)].$$

Зворотний прогноз з метою визначення $h(0)$ ґрунтується на моделі:

$$\ln h(k) - b = \alpha_1 [\ln h(k-1) - b] + v(k),$$

де $b = \frac{\alpha_0}{1 - \alpha_1}$ при $|\alpha_1| < 1$. Модель зворотного часового ряду має вигляд:

$$\ln h(k) - b = \alpha_1 [\ln h(k+1) - b] + v^*(k),$$

де $\{v^*(k)\} \sim N(0, \sigma_v^2)$. Таким чином, прогноз на два кроки назад визначається за виразом:

$$h_2(-2) - b = \alpha_1^2 [\ln h(2) - b].$$

Вираз (9.6.6) можна також отримати за допомогою результатів, які стосуються обробки пропусків даних при побудові моделі AP(1). Припустимо, що пропущеним значенням є $\ln h(k)$. Це пропущене значення є сусіднім для $\ln h(k-1)$ і $\ln h(k+1)$, $1 < k < N$. Скористаємось моделлю процесу:

$$\ln h(k) = \alpha_0 + \alpha_1 \ln h(k-1) + u(k).$$

Введемо позначення: $z(k) = \alpha_0 + \alpha_1 \ln h(k-1)$, $x(k) = 1$, $\varphi(k) = -u(k)$. Тепер можна записати:

$$z(k) = x(k) \ln h(k) + \varphi(k). \quad (9.6.7)$$

З рівняння $\ln h(k+1) = \alpha_0 + \alpha_1 \ln h(k) + u(k+1)$ можна визначити, що $z(k+1) = \ln h(k+1) - \alpha_0$, $x(k+1) = \alpha_1$ і $\varphi(k+1) = u(k+1)$, а тому

$$z(k+1) = x(k+1) \ln h(k) + u(k+1). \quad (9.6.8)$$

Таким чином, рівняння (9.6.7) і (9.6.8) утворюють просту лінійну регресію з двома спостереженнями і невідомим параметром $\ln h(k)$. Зазначимо, що $u(k)$ і $u(k+1)$ мають однаковий розподіл, оскільки $-u(k)$ також має нормальний розподіл $N(0, \sigma_v^2)$. Оцінку $\ln h(k)$ позначимо $\hat{lh}(k)$; користуючись методом найменших квадратів, її можна знайти за виразом:

$$\hat{lh}(k) = \frac{x(k)z(k) - x(k+1)z(k+1)}{x^2(k) + x^2(k+1)} = \frac{\alpha_0(1 - \alpha_1) + \alpha_1[\ln h(k+1) + \ln h(k-1)]}{1 + \alpha_1^2},$$

що представляє собою умовне середнє для $\ln h(k)$, яке визначається рівнянням (9.6.6). Крім того, ця оцінка є нормально розподіленою із середнім $\ln h(k)$ та дисперсією $\frac{\sigma_v^2}{1 + \alpha_1^2}$. Вираз (9.6.6) представляє собою добуток розподілів $\{u(k)\}$

$\sim N(0, h(k))$ і $\{\hat{lh}(k)\} \sim N\left(\ln h(k), \frac{\sigma_v^2}{1 + \alpha_1^2}\right)$ із застосуванням перетворення

$d \ln h(k) = h^{-1}(k) dh(k)$. Наведений метод обчислення оцінки для $\ln h(k)$ можна узагальнити на модель довільного порядку $AR(p)$. Початкове значення для $\ln h(k)$ можна отримати в результаті оцінювання моделі (9.6.1) за часовим рядом.

Приклад 9.6.1. Побудуємо математичну модель для 456 спостережень індексу місячного доходу для біржі S&P 500. Узагальнена авторегресійна умовно гетероскедастична модель процесу (УАРУГ(1, 1)) має вигляд:

$$y(k) = 0,583 + h(k)\varepsilon(k), \quad (9.6.9)$$

$$h(k) = 2,975 + 0,079 u^2(k-1) + 0,684 h(k-1), \quad (9.6.10)$$

де $u(k) = h^{1/2}(k)\varepsilon(k)$. Застосування статистики Льюнга-Бокса до нормованих

залишків моделі та їх квадратів показало, що модель (9.6.9)-(9.6.10) має високу ступінь адекватності.

Стохастична модель волатильності для цього ж процесу:

$$\begin{aligned} y(k) &= \mu + u(k), \\ \ln h(k) &= \alpha_0 + \alpha_1 \ln h(k-1) + v(k), \end{aligned} \quad (9.6.11)$$

де $\{v(k)\} \sim N(0, \sigma_v^2)$. Для того щоб застосувати дискретизацію Гіббса (метод Монте Карло для марковських ланцюгів (МКМЛ)), скористаємось апіорними розподілами:

$$\mu \sim N(0; 7), \quad \alpha \sim N[\alpha_0, \text{diag}(0,08; 0,04)];$$

де $\alpha_0 = [0,3; 0,7]^T$ (початкові значення оцінок взяті з моделі (9.6.10) для умовної дисперсії $h(k)$). На основі вибірки даних обчислено наступні значення дисперсії випадкового процесу та середнього: $\sigma_v^2 = 0,5$ і $\mu = 0,73$. Значення $h(k)$ отримано за допомогою методу Монте Карло (дискретизація Гіббса) з обмеженням $0 \leq h(k) \leq 1,5s^2$, де s^2 – вибіркова дисперсія ряду $\log[y(k)]$.

Для того щоб оцінити параметри моделі (9.6.11), виконано 7000 ітерацій алгоритму дискретизації Гіббса, при цьому перші 500 ітерацій не враховано з огляду на перехідний процес, характерний для даної процедури оцінювання. Апостеріорні середні і стандартні похибки оцінок чотирьох коефіцієнтів моделі (9.6.11) наведені в таблиці 9.6.1.

Таблиця 9.6.1. Результати оцінювання параметрів моделі

Параметри	μ	α_0	α_1	σ_v^2
Середнє	0,784	0,829	0,575	0,351
Станд. Похибка	0,196	0,207	0,077	0,064

Оскільки коефіцієнт $\hat{\alpha}_1 = 0,575$, це свідчить про існування високої кореляції між значеннями волатильності досліджуваного фінансового ряду. Виконано додаткові обчислювальні експерименти з оцінювання параметрів стохастичної моделі волатильності при різних значеннях кількості циклів методу МКМЛ, які свідчать про повторюваність та стійкість отриманих оцінок $h(k)$.

Багатовимірна МСВ. Розглянемо багатовимірну стохастичну модель на прикладі двовимірної моделі волатильності. Модель двовимірного фінансового ряду $\mathbf{y}(k) = [y_1(k) \ y_2(k)]^T$ можна записати у вигляді:

$$\mathbf{y}(k) = \beta_0 + \beta_1 \mathbf{x}(k) + \mathbf{u}(k), \quad (9.6.12)$$

$$\ln g_{ii}(k) = \alpha_{i0} + \alpha_{i1} \ln g_{ii}(k-1) + v_i(k), \quad i=1,2, \quad (9.6.13)$$

$$q_{21}(k) = \gamma_0 + \gamma_1 q_{21}(k-1) + \eta(k), \quad (9.6.14)$$

де $\{\mathbf{u}(k)\}$ – послідовність некорельованих випадкових гаусових векторів з нульовим середнім та матрицею умовних коваріацій $\Sigma(k)$; β_0 – двовимірний вектор констант; $\mathbf{x}(k)$ – вектор пояснюючих змінних; $\{v_1(k)\}, \{v_2(k)\}, \{\eta(k)\}$ – незалежні послідовності білого шуму такі, що $Var[v_i(k)] = \sigma_{iv}^2$ і $Var[\eta(k)] = \sigma_{\eta}^2$. Логарифмування у рівнянні (9.6.13) застосовується з метою забезпечення додатної визначеності значень $g_{ii}(k)$.

Матриця $\Sigma(k)$ визначається рівнянням:

$$\begin{bmatrix} \sigma_{11}(k) & \sigma_{12}(k) \\ \sigma_{21}(k) & \sigma_{22}(k) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ q_{21}(k) & 1 \end{bmatrix} \begin{bmatrix} g_{11}(k) & 0 \\ 0 & g_{22}(k) \end{bmatrix} \begin{bmatrix} 1 & q_{21}(k) \\ 0 & 1 \end{bmatrix}, \quad (9.6.16)$$

де $g_{ii}(k) = Var[\varphi_i(k) | F(k-1)]$ і $\varphi_1(k) \perp \varphi_2(k)$. В даному випадку при моделюванні використовуються $g_{11}(k), g_{22}(k)$ і $q_{21}(k)$.

Застосування декомпозиції Холецького дає можливість виконати перетворення:

$$\varphi_1(k) = u_1(k), \quad \varphi_2(k) = u_2(k) - q_{21}(k)\varphi_1(k),$$

де $\varphi_2(k), q_{21}(k)$ можна інтерпретувати як залишки та оцінку за МНК параметра регресії:

$$u_2(k) = \varphi_2(k) + q_{21}(k)u_1(k).$$

Для зручності введемо такі позначення: $\mathbf{G}_i = [g_{ii}(1), \dots, g_{ii}(N)]^T$, $\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2]$ і $\mathbf{Q} = [q_{21}(1), \dots, q_{21}(N)]^T$. Параметрами моделі (9.6.12)–(9.6.14), які необхідно оцінити, є такі: $\beta = [\beta_0, \beta_1]^T$, $\theta_i = [\alpha_{i0}, \alpha_{i1}, \sigma_{iv}^2]^T$, $i=1,2$ і $\gamma = [\gamma_0, \gamma_1, \sigma_{\eta}^2]^T$. Розширеними векторами параметрів є \mathbf{Q}, \mathbf{G}_1 і \mathbf{G}_2 . Для того щоб оцінити параметри двовимірної моделі стохастичної волатильності з використанням алгоритму дискретизації Гіббса, скористаємось результатами оцінювання одновимірної моделі, отриманими вище, а також двома додатковими умовними апостеріорними розподілами. Таким чином, випадкові

вибірки можна отримати так:

для параметрів β_0 і β_1 скористаємось результатом, отриманим для рівняння (9.6.5);

для $g_{11}(k)$ скористаємось виразом (9.6.6) із заміною $u(k)$ на $u_1(k)$;

для вектора параметрів θ_1 скористаємось методом, розглянутим в одновимірному випадку, за умови заміни $u(k)$ на $u_1(k)$.

Для того щоб згенерувати випадкові вибірки для ω_2 і $g_{22}(k)$, необхідно обчислити значення $\varphi_2(k)$. Це можна зробити за допомогою виразу: $\varphi_2(k) = u_2(k) - q_{21}(k)u_1(k)$ при відомому векторі \mathbf{Q} . Також відомо, що $\varphi_2(k)$ має нормальний розподіл з нульовим середнім та умовною дисперсією $g_{22}(k)$.

Тепер необхідно знайти апостеріорні розподіли для параметрів:

$$f(\bar{\theta} | \mathbf{Q}, \sigma_\eta^2), \quad f(\sigma_\eta^2 | \mathbf{Q}, \bar{\theta}), \quad f(q_{21}(k) | \mathbf{A}, \mathbf{G}, \mathbf{Q}(-k), \gamma),$$

де $\bar{\theta} = [\gamma_0, \gamma_1]^T$ – вектор коефіцієнтів рівняння (9.6.14); \mathbf{A} – містить множину значень $\mathbf{u}(k)$, яку можна знайти, якщо відомі значення \mathbf{y} , \mathbf{X} , β_0 і β_1 . Якщо відомі значення \mathbf{Q} і σ_η^2 , то модель (9.6.14) – це проста модель АР(1) з гаусовим шумом. Таким чином, якщо $\bar{\theta}$ має двовимірний нормальний апріорний розподіл із середнім $\bar{\theta}_0$ та коваріаційною матрицею \mathbf{D}_0 , то умовний апостеріорний розподіл для $\bar{\theta}$ також буде двовимірним нормальним розподілом із середнім $\bar{\theta}_*$ та коваріаційною матрицею \mathbf{D}_* , де

$$\mathbf{D}_*^{-1} = \frac{\sum_{k=2}^N \mathbf{z}(k) \mathbf{z}^T(k)}{\sigma_\eta^2} + \mathbf{D}_0^{-1}; \quad \bar{\theta}_* = \mathbf{D}_* \left(\frac{\sum_{k=2}^N \mathbf{z}(k) q_{21}(k)}{\sigma_\eta^2} + \mathbf{D}_0^{-1} \bar{\theta}_0 \right),$$

де $\mathbf{z}(k) = [1, q_{21}(k-1)]^T$. Аналогічно, якщо апріорним розподілом для σ_η^2 є $\frac{m\lambda}{\sigma_\eta^2} \sim \chi_m^2$, то умовним апостеріорним розподілом для σ_η^2 буде:

$$\frac{m\lambda + \sum_{k=2}^N \eta^2(k)}{\sigma_\eta^2} \sim \chi_{m+N-1}^2,$$

де $\eta(k) = q_{21}(k) - \gamma_0 - \gamma_1 q_{21}(k-1)$. Останній з розподілів параметрів можна визначити так:

$$\begin{aligned} f(q_{21}(k) | \mathbf{A}, \mathbf{G}, \mathbf{Q}(-k), \sigma_\eta^2, \bar{\theta}) &\propto \\ &\propto f[(\varphi_2(k) | g_{22}(k))] f[q_{21}(k) | q_{21}(k-1), \bar{\theta}, \sigma_\eta^2] f[q_{21}(k+1) | q_{21}(k), \bar{\theta}, \sigma_\eta^2] \propto \end{aligned}$$

$$\propto g_{22}^{-0.5}(k) \exp[-(u_2(k) - q_{21}(k)u_1(k))^2 / 2g_{22}(k)] \exp[-(q_{21}(k) - \mu(k))^2 / (2\sigma^2)], \quad (9.6.17)$$

де $\mu(k) = [\gamma_0(1 - \gamma_1) + \gamma_1(q_{21}(k-1) + q_{21}(k+1))] / (1 + \gamma_1^2)$ і $\sigma^2 = \sigma_\eta^2 / (1 + \gamma_1^2)$.

В загальному випадку значення $\mu(k)$ і σ^2 можна отримати за допомогою результату аналізу пропущених значень в процесі $AR(p)$.

Вираз (9.6.17) представляє собою замкнену форму розподілу для $q_{21}(k)$. Перший член цього виразу – це умовний розподіл для $q_{21}(k)$ при відомих $g_{22}(k)$ і $\mathbf{u}(k)$; він представляє собою нормальний розподіл із середнім $u_2(k)/u_1(k)$ і дисперсією $g_{22}(k)/(u_1(k))^2$. Другий член цього виразу – це також щільність нормального розподілу з середнім $\mu(k)$ і дисперсією σ^2 . Таким чином, умовний апостеріорний розподіл $q_{21}(k)$ є нормальним розподілом з середнім μ_* і дисперсією σ_*^2 , де

$$\frac{1}{\sigma_*^2} = \frac{u_1^2(k)}{g_{22}(k)} + \frac{1 + \gamma_1^2}{\sigma_\eta^2}; \quad \mu_* = \sigma_*^2 \left(\frac{1 + \gamma_1^2}{\sigma_\eta^2} \mu(k) + \frac{u_1^2(k)}{g_{22}(k)} \frac{u_2(k)}{u_1(k)} \right).$$

Приклад 9.6.2. В цьому прикладі розглядається двовимірна модель волатильності, побудована для двох змінних: логарифмів доходів по акціях фірми ІВМ за період часу з січня 1962 по грудень 1999 року і індексу S&P 500 за той же період часу. Таким чином, $\mathbf{y}(k) = [ibm(k), sp(k)]^T$. Виберемо модель УАРУГ зі змінними коефіцієнтами кореляції та застосуємо розклад Холецького:

$$\mathbf{y}(k) = \beta_0 + \mathbf{u}(k), \quad (9.6.18)$$

$$g_{11}(k) = \alpha_{10} + \alpha_{11}g_{11}(k-1) + \alpha_{12}u_1^2(k-1), \quad (9.6.19)$$

$$g_{22}(k) = \alpha_{20} + \alpha_{21}u_1^2(k-1), \quad (9.6.20)$$

$$q_{21}(k) = \gamma_0, \quad (9.6.21)$$

Оцінки параметрів та їх стандартні похибки наведені в таблиці 9.6.2. Для визначення оцінок виконано 1500 ітерацій алгоритму оцінювання, при цьому перші п'ятсот ітерацій пропущено.

Таблиця 9.6.2. Результати оцінювання двовимірної моделі

Двовимірна модель УАРУГ(1,1) із змінними коефіцієнтами кореляції									
Параметр	β_{01}	β_{02}	α_{01}	α_{11}	α_{12}	α_{20}	α_{21}	γ_0	
Оцінка	1,07	0,84	3,78	0,81	0,14	11,42	0,06	0,41	
Станд. пох.	0,337	0,25	1,89	0,08	0,05	0,99	0,03	0,03	

Стохастична модель волатильності									
Параметр	β_{01}	β_{02}	α_{01}	α_{11}	σ_{1v}^2	α_{20}	σ_{2v}^2	γ_0	σ_{η}^2
Апост. сер.	0,89	0,91	0,52	0,82	0,07	1,69	0,43	0,37	0,09
Станд. пох.	0,37	0,25	0,26	0,06	0,03	0,13	0,05	0,03	0,03

З метою порівняння використано таке ж рівняння (9.6.18) для основної змінної і стохастична модель волатильності такого вигляду:

$$\ln g_{11}(k) = \alpha_{10} + \alpha_{11} \ln g_{11}(k-1) + v_1(k), \quad \text{Var}[v_1(k)] = \sigma_{1v}^2, \quad (9.6.22)$$

$$\ln g_{22}(k) = \alpha_{20} + v_2(k), \quad \text{Var}[v_2(k)] = \sigma_{2v}^2, \quad (9.6.23)$$

$$q_{21}(k) = \gamma_0 + \eta(k), \quad \text{Var}[\eta(k)] = \sigma_{\eta}^2. \quad (9.6.24)$$

Апріорні розподіли:

$$\beta_{i0} \sim N(0,8; 4); \quad \alpha_1 \sim N[(0,4; 0,8)^T; \text{diag}(0,15; 0,05)]; \quad \alpha_{20} \sim N(5,5; 25);$$

$$\gamma_0 \sim N(0,5; 0,4); \quad \frac{10 \times 0,1}{\sigma_{1v}^2} \sim \chi_{10}^2; \quad \frac{5 \times 0,2}{\sigma_{2v}^2} \sim \chi_5^2; \quad \frac{5 \times 0,2}{\sigma_{\eta}^2} \sim \chi_5^2.$$

Дані апріорні розподіли є відносно неінформативними. Алгоритм дискретизації Гіббса було застосовано для реалізації 1500 ітерацій, при цьому перші 400 ітерацій в подальшому не використовувались для знаходження (усереднення) оцінок. Випадкові вибірки параметра $g_{ii}(k)$ отримано за допомогою решітчастої процедури Гіббса з використанням 500 точок в інтервалі $[0; 1,5s_i^2]$, де s_i^2 – вибіркова дисперсія логарифму $y_i(k)$, $i=1,2$. Апостеріорні середні та стандартні похибки оцінок параметрів двовимірної стохастичної моделі також наведені в таблиці 3.5.2. Для перевірки збіжності алгоритму оцінювання з використанням дискретизації Гіббса сеанси оцінювання реалізовано 10 разів по 1500 ітерацій. Встановлено, що процедура оцінювання генерує стійкі повторювані оцінки.

Запитання і вправи до дев'ятого 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? Порівняйте вирази для розрахунку безумовної та умовної дисперсій; в чому принципова різниця між отриманими в результаті розрахунку значеннями?
3. Чому для прогнозування значень змінної та її дисперсії використовують умовне, а не безумовне математичне сподівання?
4. Для чого необхідно прогнозувати дисперсію? 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 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 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 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 дисперсії ринкової ціни на прибуток виробничої фірми? Запишіть відповідну математичну модель.
20. Як можна математично змодельовати плату за ризик при виконанні фінансових операцій, інвестуванні і т. ін.?
21. Поясніть суть методики VaR, яка застосовується для оцінювання валютного або іншого фінансового ризику? Який недолік має стандартна методика VaR?
22. В чому полягає суть баєсівського методу оцінювання стохастичної моделі волатильності, наведеного в розділі 9.6? Сформулюйте процедуру реалізації цього методу.
23. Яким чином оцінюють апіорні та апостеріорні розподіли оцінок параметрів регресійної моделі?
24. Чи можна застосувати для оцінювання параметрів моделей гетероскедастичних процесів нелінійний МНК та метод максимальної правдоподібності? Які обмеження має нелінійний МНК?
25. Який статистичний метод оцінювання параметрів є незалежним від структури моделі і може бути використаний для оцінювання моделей гетероскедастичних процесів?

Р о з д і л 10

МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ ДЕТЕРМІНОВАНИХ І СТОХАСТИЧНИХ ТРЕНДІВ

10.1 Вступ

Процеси, нестационарні стосовно дисперсії, описують рівняннями АРУГ, АРУГ-М, УАРУГ, стохастичними моделями волатильності та іншими модифікаціями моделей умовної дисперсії. Вони дають можливість описувати волатильність ряду, тобто ступінь його мінливості за допомогою рівняння динаміки дисперсії. Загальні тенденції до зростання чи спадання того чи іншого процесу характеризують його *трендом*. Так, зростання валового внутрішнього продукту розвинених країн на довгих часових інтервалах характеризується детермінованим позитивним 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. Якщо тренд відносно швидко змінюється в часі у 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 війни, енергетичні кризи і т. ін.

Вище ми широко застосовували для опису динаміки процесів з трендами рівняння авторегресії з ковзним середнім. Воно має таку узагальнену структуру:

$$y(k) = \text{дрейф} + \text{детермінований тренд} + \text{авторегресія} + \\ + \text{сезонна компонента} + \text{ковзне середнє.}$$

Подібна структура надає можливість описувати різні ефекти, які наявні у динамічних процесах, в тому числі і тренд. Розв'язки рівнянь АРКС містять складові, які описують тренди, детерміновані і нерегулярні компоненти процесу. Прикладами рівнянь, що описують детермінований тренд, можуть бути рівняння прямої, поліноми, експоненти та комбінації гармонічних складових:

$$y(k) = a_0 + c_1 t,$$

$$y(k) = a_0 + c_1 t + c_2 t^2 + \dots + c_m t^m,$$

$$y(k) = \exp(ck),$$

$$y(k) = a_0 + \sum_{i=1}^p a_i \sin(\omega_0 i) + \sum_{i=1}^q b_i \cos(\omega_0 i).$$

Їх використовують для опису відносно тривалих періодів стійкого спаду/зростання процесів або детерміновані періодичні коливання (останнє рівняння). Очевидно, що для повного опису процесу наведені рівняння необхідно розширювати відповідними складовими. Крім ВВП, тренд є характерним також для таких процесів:

- зростання туристичної активності у багатьох країнах світу;
- підвищення добробуту у розвиненій економіці;
- неперервне зростання об'єму науково-технічної інформації у глобальному масштабі;
- розвиток демографічних процесів (при цьому тренди можуть мати різний характер в локальному та глобальному масштабах);
- фінансові процеси ціноутворення на біржах (складні стохастичні процеси);
- інфляційні процеси.

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.
- 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 (особливо у 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 у різних країнах світу;
 - техногенні катастрофи, локальні та глобальні військові дії, природні аномалії.

Наприклад, 6-7 років засухи підряд (у глобальному масштабі) може зменшити населення Землі в 3-4 рази. Чорнобильська катастрофа призвела до тяжких демографічних та соціально-економічних наслідків для України і Білорусі за відносно короткий період часу. Вона створила значний негативний вплив на розвиток соціально-економічних процесів, на здоров'я великих груп населення.

Випадкові тренди можна описати комбінаціями випадкових величин з параметрами, які можна оцінити за допомогою характеристик фактичних випадкових процесів. У загальному випадку процес може містити як детерміновану, так і випадкову складові. Розглянемо деякі можливості математичного опису випадкових трендів.

10.2 Модель випадкового кроку

Детермінований тренд називають також *глобальним трендом* процесу. Однак, на сьогодні існує тенденція до вироблення більш загального підходу до 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:

$$y(k) = a(k) + d_1(k)k, \quad (10.1.1)$$

де $a(k)$ – локальна константа; $d_1(k)$ – коефіцієнт, який визначає локальний нахил тренду. Отримано результати моделювання, які показують, що такі функції є більш адекватними і робастними ніж детерміновані функції поліноміального типу.

Альтернативним підходом до опису локального тренду є використання рекурсивних рівнянь типу $y(k) = a_0 + y(k-1)$ або в ускладненому варіанті це може бути процес випадкового кроку із змінним перетином:

$$y(k) = a(k) + y(k-1) + \varepsilon(k), \quad (10.1.2)$$

де $\varepsilon(k)$ – випадкова змінна, яку для спрощення аналізу можна прийняти послідовністю білого шуму з відомою дисперсією. Це рівняння можна назвати рівнянням випадкового кроку із змінним у часі перетином $a(k)$. Модель (10.1.2) необхідно доповнити рівнянням, яке описує зміну у часі перетину, тобто $a(k)$. Наприклад, $a(k) = a(k-1) + v(k)$, де $v(k)$ – випадковий збурюючий процес. Зазначимо, що точне значення рівня $y(k)$ та швидкість зростання $a(k)$ безпосередньо не спостерігаються. Наприклад, у моделі лінійного зростання спостереження часового ряду визначається сумою $(y(k) + e(k))$, де $e(k)$ – похибка моделі, зумовлена рядом факторів, що виникають при її побудові. Зокрема, це похибки вимірів та обчислювальні похибки.

Модель випадкового кроку є однією з найпростіших, яка дає можливість описати випадковий тренд у деяких випадках. Вона має такий вигляд:

$$y(k) = y(k-1) + \varepsilon(k), \quad (10.1.3)$$

або

$$\Delta y(k) = \varepsilon(k),$$

де $\varepsilon(k)$ – білий шум з нульовим середнім. Приростом значення основної змінної у даній моделі є випадкова величина. Запишемо розв'язок цього рівняння:

$$\begin{aligned} y(1) &= y_0 + \varepsilon(1), \\ y(2) &= y(1) + \varepsilon(2) = y_0 + \varepsilon(1) + \varepsilon(2), \\ &\vdots \\ y(k) &= y_0 + \sum_{i=1}^k \varepsilon(i). \end{aligned} \quad (10.1.4)$$

Розв'язок (10.1.4) 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:

$$y(k+1) = y(k) + \varepsilon(k+1),$$

$$E_k[y(k+1) | y(k), y(k-1), \dots, \varepsilon(k), \varepsilon(k-1), \dots] = E[y(k) + \varepsilon(k+1)] = y(k).$$

Функція прогнозування на довільне число кроків на основі розв'язку:

$$\begin{aligned}
 y(k) &= y_0 + \sum_{i=1}^k \varepsilon(i), \\
 y(k+s) &= y_0 + \sum_{i=1}^{k+s} \varepsilon(i) = y_0 + \sum_{i=1}^k \varepsilon(i) + \sum_{i=k+1}^{k+s} \varepsilon(i) = y(k) + \sum_{i=k+1}^{k+s} \varepsilon(i), \\
 E_k[y(k+s)] &= y(k) + E_k \left[\sum_{i=k+1}^{k+s} \varepsilon(i) \right] = y(k). \tag{10.1.5}
 \end{aligned}$$

Тобто величина прогнозу не залежить від числа кроків. Знайдемо значення дисперсії для розв'язку $y(k)$. Для моменту k :

$$\begin{aligned}
 \text{var}[y(k)] &= E\{[y(k) - \mu_y][y(k) - \mu_y]\} = \\
 &= E\left\{ \left[y(0) + \sum_{i=1}^k \varepsilon(i) - y(0) \right] \left[y(0) + \sum_{i=1}^k \varepsilon(i) - y(0) \right] \right\} = \\
 &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k)][\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k)]\} = k\sigma_\varepsilon^2.
 \end{aligned}$$

Для моменту часу $(k-s)$:

$$\begin{aligned}
 \text{var}[y(k-s)] &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k-s)][\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k-s)]\} = \\
 &= (k-s)\sigma_\varepsilon^2. \tag{10.1.6}
 \end{aligned}$$

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:

$$\rho(s) = \frac{\gamma(s)}{\sqrt{\text{var}[y(k)] \text{var}[y(k-s)]}} = \frac{(k-s)\sigma^2}{\sqrt{k\sigma^2(k-s)\sigma^2}} = \sqrt{\frac{k-s}{k}}. \tag{10.1.7}$$

Тобто, АКФ процесу випадкового кроку є повільно спадаючою функцією.

Модель випадкового кроку плюс дрейф (зміщення або перетин). У даному випадку до моделі випадкового кроку додається константа a_0 :

$$y(k) = a_0 + y(k-1) + \varepsilon(k). \tag{10.1.8}$$

При відомій початковій умові $y(0) = y_0$ розв'язок рівняння (10.1.8) має вигляд:

$$y(k) = y_0 + a_0 k + \sum_{i=1}^k \varepsilon(k). \quad (10.1.9)$$

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:

$$E[y(k)] = y_0 + a_0 k, \quad E_k[y(k+s)] = y_0 + a_0(k+s).$$

Для того щоб отримати функцію прогнозування запишемо рівняння (30) для моменту $(k+s)$:

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

і умовне математичне сподівання

$$E_k[y(k+s)|k] = y(k) + a_0 s. \quad (10.1.10)$$

Таким чином, отримана функція прогнозування відрізняється від функції прогнозування для випадкового кроку (10.1.5) 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,

$$y(k) = \mu(k) + \eta(k), \quad (10.1.11)$$

$$\mu(k) = \mu(k-1) + \varepsilon(k), \quad (10.1.12)$$

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 (10.1.12) має вигляд: $\mu(k) = \mu_0 + \sum_{i=1}^k \varepsilon(k)$, що представляє

собой випадковий тренд для $y(k)$. Тепер $y(k) = \mu_0 + \sum_{i=1}^k \varepsilon(k) + \eta(k)$.

Використовуючи початкову умову $y(0) = y_0 = \mu_0 + \eta_0$, запишемо розв'язок:

$$y(k) = y_0 - \eta_0 + \sum_{i=1}^k \varepsilon(k) + \eta(k). \quad (10.1.13)$$

Таким чином, безумовне середнє процесу $\{y(k)\}$ є константою $E[y(k)] = E[y(k+s)] = y_0 - \eta_0$. У кожний момент часу на цей процес впливає чисто 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. Так, коваріація між $y(k)$ і $y(k-s)$ визначається як

$$\begin{aligned} cov[y(k) y(k-s)] &= E\{[y(k) - y_0 - \eta_0][y(k-s) - y_0 - \eta_0]\} = \\ &= E\{[\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k) + \eta(k)][\varepsilon(1) + \varepsilon(2) + \dots + \varepsilon(k-s) + \eta(k-s)]\} = \\ &= (k-s)\sigma_\varepsilon^2, \end{aligned}$$

оскільки $\{\varepsilon(k)\}$ і $\{\eta(k)\}$ – незалежні послідовності. Таким чином, автокореляційна функція $\rho(s)$ визначається за виразом:

$$\rho(s) = \frac{(k-s)\sigma_\varepsilon^2}{\sqrt{(k\sigma_\varepsilon^2 + \sigma_\eta^2)[(k-s)\sigma_\varepsilon^2 + \sigma_\eta^2]}}. \quad (10.1.14)$$

Порівнюючи (10.1.14) з виразом (10.1.7) 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 на s кроків скористаємось рівнянням (10.1.13):

$$\begin{aligned} y(k+s) &= y_0 - \eta_0 + \sum_{i=1}^{k+s} \varepsilon(i) + \eta(k+s) = \\ &= y(k) - \eta(k) + \sum_{i=1}^s \varepsilon(k+i) + \eta(k+s). \end{aligned}$$

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 $\sum \varepsilon(k)$.

Випадковий крок + шум + дрейф

Модель випадкового кроку з шумом та дрейфом має вигляд:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= a_0 + \mu(k-1) + \varepsilon(k), \end{aligned} \quad (10.1.15)$$

де a_0 – константа; $\{\varepsilon(k)\}, \{\eta(k)\}$ – незалежні процеси білого шуму. В даному випадку тренд $\mu(k)$ містить випадкову складову $\varepsilon(k)$ та детерміновану a_0 .

Запишемо розв'язок для $\mu(k)$: $\mu(k) = \mu_0 + a_0k + \sum_{i=1}^k \varepsilon(i)$, що приводить до виразу для $y(k)$ у вигляді:

$$y(k) = \mu_0 + a_0k + \sum_{i=1}^k \varepsilon(i) + \eta(k). \quad (10.1.16)$$

Враховуючи початкову умову $y_0 = \mu_0 + \eta_0$, запишемо рівняння для $y(k)$:

$$y(k) = y_0 - \eta_0 + a_0k + \sum_{i=1}^k \varepsilon(i) + \eta(k). \quad (10.1.17)$$

Таким чином, $y(k)$ представляє собою суму детермінованого і стохастичного трендів, а також чисто випадкової складової $\eta(k)$. Шумова складова не обов'язково повинна бути білим шумом. Рівняння (10.1.17) можна переписати у вигляді:

$$y(k) = \mu_0 + a_0k + \sum_{i=1}^k \varepsilon(i) + A(L)\eta(k), \quad (10.1.18)$$

де $A(L)$ – поліном відносно оператора затримки L ; $A(L)\eta(k)$ – стаціонарний шумовий процес. Рівняння (10.1.18) називають *моделлю узагальненого тренду з нерегулярною складовою*.

Локальний лінійний тренд

Модель локального лінійного тренду (ЛЛТ) поєднує в собі кілька процесів випадкового кроку з шумом:

$$y(k) = \mu(k) + \eta(k),$$

$$\begin{aligned}\mu(k) &= \mu(k-1) + \lambda(k) + \varepsilon(k), \\ \lambda(k) &= \lambda(k-1) + v(k),\end{aligned}\quad (10.1.19)$$

де $\{\eta(k)\}$, $\{\varepsilon(k)\}$, $\{v(k)\}$ – незалежні процеси білого шуму. В даному 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 легко показати, що розглянуті вище процеси (випадковий крок плюс шум та випадковий крок плюс дрейф і шум) представляють собою окремі випадки ЛЛТ.

Знайдемо розв'язок для $y(k)$. Спочатку запишемо його для $\lambda(k)$:

$$\lambda(k) = \lambda_0 + \sum_{i=1}^k v(i),$$

а також для $\mu(k)$:

$$\mu(k) = \mu(k-1) + \lambda_0 + \sum_{i=1}^k v(i) + \varepsilon(k),$$

або

$$\mu(k) = \mu_0 + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + (k-1)v(2) + (k-2)v(3) + \dots + v(k).$$

Оскільки $y_0 = \mu_0 + \eta_0$, то розв'язок для $y(k)$ має вигляд:

$$y(k) = y_0 + [\eta(k) - \eta_0] + \sum_{i=1}^k \varepsilon(i) + k(\lambda_0 + v(1)) + (k-1)v(2) + \dots + v(k).$$

В даному 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 ЛЛТ нерегулярна компонента визначатиметься членом $A(L)\eta(k)$. Детермінований тренд залежить у даному 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 у 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 на s кроків для ЛЛТ:

$$\begin{aligned}y(k+s) &= y_0 + [\eta(k+s) - \eta_0] + \sum_{i=1}^{k+s} \varepsilon(i) + (k+s)(\lambda_0 + v(1)) + \\ &\quad + (k+s-1)v(2) + (k+s-2)v(3) + \dots + v(k+s)\end{aligned}$$

або

$$y(k+s) = y(k) + [\eta(k+s) - \eta(k)] + \sum_{i=1}^k \varepsilon(i) + s[\lambda_0 + v(1) + v(2) + \dots + v(k)] + \sum_{i=1}^s (s+1-i)v(k+i).$$

Умовне математичне сподівання:

$$E_k [y(k+s)] = [y(k) - \eta(k)] + s[\lambda_0 + v(1) + v(2) + \dots + v(k)].$$

Нахил тренду визначається сумою $s[\lambda_0 + v(1) + v(2) + v(3) + \dots + v(k)]$.

10.3 Методи вилучення тренду

При моделюванні процесів з трендом, можна описати тренд за допомогою вибраної детермінованої чи випадкової функції або вилучити його зовсім із 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);
- *опис тренду* за допомогою вибраної детермінованої функції з подальшим використанням залишків для опису стаціонарних коливань іншою моделлю, наприклад, АРКС.

Якщо процес має один одиничний корінь, то перші різниці будуть стаціонарними, а якщо більше, то необхідно обчислювати різниці вищих порядків.

Розглянемо спочатку процес випадкового кроку + шум. Його розв'язок має вигляд:

$$y(k) = y(0) + a_0 k + \sum_{i=1}^k \varepsilon(i).$$

Перша різниця:

$$\begin{aligned} \Delta y(k) &= y(k) - y(k-1) = y_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) - \left(y_0 + a_0(k-1) + \sum_{i=1}^{k-1} \varepsilon(i) \right) = \\ &= a_0 + \varepsilon(k). \end{aligned}$$

Очевидно, що послідовність $\{\Delta y(k)\}$ буде стаціонарною, оскільки вона містить тільки константу і білий шум. Математичне сподівання для цього процесу

$$E[\Delta y(k)] = E[a_0 + \varepsilon(k)] = a_0,$$

$$\text{Var}[\Delta y(k)] \equiv E[\Delta y(k) - a_0]^2 = E[\varepsilon(k)]^2 = \sigma_\varepsilon^2,$$

і коваріація

$$\text{Cov}[\Delta y(k), \Delta y(k-s)] \equiv E[(\Delta y(k) - a_0)(\Delta y(k-s) - a_0)] =$$

$$= E[\varepsilon(k)\varepsilon(k-s)] = 0.$$

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:

$$E[\Delta y(k)] = E[\varepsilon(k) + \Delta \eta(k)] = 0,$$

$$\text{Var}[\Delta y(k)] \equiv E[\Delta y(k)]^2 = E\{[\varepsilon(k) + \Delta \eta(k)]^2\} =$$

$$= E\{[\varepsilon(k)]^2 + 2\varepsilon(k)\Delta \eta(k) + [\Delta \eta(k)]^2\} =$$

$$= \sigma_\varepsilon^2 + 2E[\varepsilon(k)\Delta \eta(k)] + E\{[\eta(k)]^2 - 2\eta(k)\eta(k-1) + [\eta(k-1)]^2\} =$$

$$= \sigma_\varepsilon^2 + 2\sigma_\eta^2,$$

$$\text{Cov}[\Delta y(k), \Delta y(k-1)] = E\{[\varepsilon(k) + \eta(k) - \eta(k-1)][\varepsilon(k-1) + \eta(k-1) - \eta(k-2)]\} =$$

$$= -\sigma_\eta^2,$$

і коваріація

$$\text{Cov}[\Delta y(k), \Delta y(k-s)] =$$

$$= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)][\varepsilon(k-s) + \eta(k-s) - \eta(k-s-1)]\} = 0, \quad \forall s \geq 1.$$

При $s = 1$ коефіцієнт кореляції між $\Delta y(k)$ і $\Delta y(k-1)$ дорівнює

$$\rho(1) = \frac{\text{Cov}[\Delta y(k), \Delta y(k-1)]}{\text{Var}[\Delta y(k)]} = -\frac{\sigma_\eta^2}{\sigma_\varepsilon^2 + 2\sigma_\eta^2}.$$

Легко встановити, що $-0,5 \leq \rho(1) \leq 0$, а всі інші коефіцієнти кореляції є нульовими. Оскільки перші різниці $\Delta y(k)$ у даному випадку представляють собою процес КС(1), то процес випадкового кроку + шум можна формально визначити як АРІКС(0, 1, 1). Зважаючи на те, що константа не впливає на

кореляційну функцію, то додатково можна сказати, що розв'язок (10.1.17) 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:

$$\Delta y(k) = \Delta \mu(k) + \Delta \eta(k) = a(k) + \varepsilon(k) + \Delta \eta(k)$$

i

$$\begin{aligned} \Delta^2 y(k) &= \Delta a(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k) \\ &= \delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k) \end{aligned}$$

Оскільки процес $a(k)$ нестационарний, то можна легко показати, що перші різниці процесу $y(k)$ також нестационарні. Математичне сподівання, дисперсія і коваріація для других різниць $\Delta^2 y(k)$:

$$E[\Delta^2 y(k)] = E[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)] = 0,$$

$$\begin{aligned} \text{Var}[\Delta^2 y(k)] &= E[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)]^2 = \\ &= E\{[\delta(k)]^2 + [\Delta \varepsilon(k)]^2 + [\Delta^2 \eta(k)]^2 + \\ &\quad + 2\delta(k)\Delta \varepsilon(k) + 2\delta(k)\Delta^2 \eta(k) + 2\Delta \varepsilon(k)\Delta^2 \eta(k)\} = \\ &= \sigma_\delta^2 + 2\sigma_\varepsilon^2 + 6\sigma_\eta^2, \end{aligned}$$

$$\begin{aligned} \text{Cov}[\Delta^2 y(k), \Delta^2 y(k-1)] &= E\{[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)] \times \\ &\quad [\delta(k-1) + \Delta \varepsilon(k-1) + \Delta^2 \eta(k-1)]\} = \\ &= E\{[\delta(k) + \varepsilon(k) - \varepsilon(k-1) + \eta(k) - 2\eta(k-1) + \eta(k-2)] \times \\ &\quad [\delta(k-1) + \varepsilon(k-1) - \varepsilon(k-2) + \eta(k-1) \\ &\quad - 2\eta(k-2) + \eta(k-3)]\} = \\ &= -\sigma_\varepsilon^2 - 4\sigma_\eta^2, \end{aligned}$$

$$\begin{aligned} \text{Cov}[\Delta^2 y(k), \Delta^2 y(k-2)] &= E\{[\delta(k) + \Delta \varepsilon(k) + \Delta^2 \eta(k)] \times \\ &\quad \times [\delta(k-2) + \Delta \varepsilon(k-2) + \Delta^2 \eta(k-2)]\} = \\ &= E\{[\delta(k) + \varepsilon(k) - \varepsilon(k-1) + \eta(k) - 2\eta(k-1) + \eta(k-2)] \times \\ &\quad \times [\delta(k-2) + \varepsilon(k-2) - \varepsilon(k-3) + \eta(k-2) \\ &\quad - 2\eta(k-3) + \eta(k-4)]\} = \\ &= \sigma_\eta^2. \end{aligned}$$

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 у 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):

$$A(L)y(k) = B(L)\varepsilon(k), \quad (10.3.1)$$

де $A(L)$ і $B(L)$ – поліноми порядку p і q від 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 на комплексній площині. Тепер рівняння (10.3.1) можна записати як

$$(1-L)A^*(L)y(k) = B(L)\varepsilon(k).$$

Позначимо $y^*(k) = \Delta y(k)$ і перейдемо до запису:

$$A^*(L)y^*(k) = B(L)\varepsilon(k). \quad (10.3.2)$$

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 показано, що шляхом диференціювання у дискретній формі модель нестаціонарного процесу можна перетворити у 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:

$$y(k) = y(0) + \alpha_1 k + \varepsilon(k).$$

Перша різниця цього процесу $\Delta y(k)$ описується рівнянням:

$$\Delta y(k) = \alpha_1 + \varepsilon(k) - \varepsilon(k-1),$$

яке не 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 регресії

$$y(k) = \alpha_0 + \alpha_1 k + \varepsilon(k).$$

Залишок цієї моделі $e(k) = y(k) - \hat{y}(k)$ буде стаціонарним. У більш загальному випадку ряд має поліноміальний тренд

$$y(k) = \alpha_0 + \alpha_1 k + \alpha_2 k^2 + \alpha_3 k^3 + \dots + \alpha_n k^n + e(k), \quad (10.3.3)$$

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. У випадку коректної моделі (10.3.3) її залишки будуть представляти собою стаціонарний процес, який можна описати моделлю АРКС.

Моделі, стаціонарні стосовно різниць та тренду

Вище ми 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 у 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:

$$A(L)y(k) = \alpha_0 + \alpha_1 k + e(k),$$

характеристичне рівняння якого $A(L)$ має корені, що лежать за межами одиничного кола (стаціонарний процес, оскільки ми користуємось поняттям зворотного характеристичного рівняння). 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 (10.3.2), тобто,

$$A(L)y^*(k) = \alpha_1 + (1-L)B(L)\varepsilon(k),$$

то бачимо, що перехід до перших різниць процесу ССТ призвів 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, який перетворюється в стаціонарний в результаті переходу до 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.

У 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 у даному випадку лише статистичними артефактами в умовах наявності нестаціонарних 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% [13]. Згідно з деякими 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:

$$y(k) = 0,8479 - 0,0064 \cdot k + e(k).$$

(-14,16)

В дужках наведено 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 з процесу детермінованого тренду не приводить до стаціонарності ряду залишків.

Для 70 спостережень значення $2N^{-1/2} = 0,239$ майже точно відповідає значенню коефіцієнта часткової автокореляції для лагу 5. Результати оцінювання двох альтернативних моделей для курсу йена/долар наведено у таблиці 10.1. Значення константи a_0 є статистично незначимим для обох моделей. Значення Q -статистики для АКФ (до 17-го значення, оскільки $N/4 \approx 17$) показують, що їх можна розглядати як групу з нульовою значимістю. Однак, за критерієм Байєса-Шварца з невеликою перевагою можна вибрати модель КС(1). Очевидним фактом є той, що будь-яка з наведених у таблиці моделей є набагато кращою за модель з детермінованим трендом.

Таблиця 10.1 Моделі курсу обміну валют йена/долар

	<i>Оцінки</i>	$Q(17)$ - статистика	<i>BSC</i>
AR(1)	a_0 : -0,0104 (0,0095)	$Q(17) = 19,06$ (0,3249)	-114,359
	a_1 : 0,3684 (0,1148)		
MA(1)	a_0 : -0,0116 (0,0082)	$Q(17) = 19,22$ (0,2573)	-114,932
	β_1 : 0,3686 (0,1123)		

В дужках наведено стандартні похибки.

10.4 Бізнес-цикли

Традиційно, при виконанні досліджень бізнес-циклів макроекономічні змінні розкладають на детермінований віковий (часовий) тренд, циклічну та нерегулярну компоненту. Типовим прикладом вікового тренду є виробництво ВВП у розвинених країнах після Другої світової війни. Нахил тренду визначається такими факторами довгострокового впливу, як технологічне зростання, підвищення врожайності та досягненнями в освіті. Однією з причин наявності відхилень від тренду є хвилеподібні зміни в галузі економічної активності, які називають *бізнес-циклами*. Зважаючи на численні історичні факти розвитку більшості країн світу, періодів процвітання та занепаду уникнути неможливо.

Амплітуду циклічних змін можна регулювати (зменшувати) шляхом реалізації належної монетарної та фіскальної політики. З точки зору викладеного вище матеріалу тренд представляє собою стаціонарну компоненту зростання, а циклічна та нерегулярна компоненти є нестаціонарними.

Хоча існують періоди спаду і процвітання, досвід економічного розвитку в період після Другої світової війни свідчить про те, що бізнес-цикли не мають постійного періоду. Існує також розповсюджена думка про те, що на довгих часових інтервалах макроекономічні змінні відбуваються з постійною швидкістю, а будь-які відхилення від тренду незмінно вирівнюються “невидимою рукою”.

Віра в те, що тренд не змінюється в часі, призводить до загальноприйнятої практики “вилучення тренду” з макроекономічних часових рядів шляхом його опису лінійною функцією або поліноміальною регресією.

В результаті застосування такого підходу отримують, наприклад, рівняння тренду для ВВП у вигляді: $y(k) = c_0 + c_1k + \varepsilon(k)$, а залишки представляють собою стаціонарну частину ряду. Якщо залишки дійсно стаціонарні, то циклічну та нерегулярну компоненти можна описати за допомогою відомих рівнянь.

Однак, проблема із застосування аналізу такого типу полягає у тому, що тренд може не бути детермінованим. Як показано вище, некоректно віднімати тренд від ряду, який він є стаціонарним після обчислення різниць. З економічної точки зору дуже важливо знати – будуть макроекономічні змінні стаціонарними стосовно різниць чи детермінованого тренду. Якщо змінна є стаціонарною стосовно детермінованого тренду, то поточні збурення будь-якого типу не будуть створювати для неї довгострокових ефектів (проблем).

Наприклад, розглянемо функцію прогнозування для моделі, стаціонарної стосовно тренду: $y(k) = \alpha_0 + \alpha_1k + \varepsilon(k)$. Якщо $\varepsilon(k)$ – процес білого шуму, то прогноз $y(k+s)$ визначається як $\alpha_0 + \alpha_1(k+s)$, тобто на майбутні значення y не впливають ні поточні ні минулі значення. Більше того, дисперсія похибки прогнозу залишається незмінною при відомих постійних значеннях коефіцієнтів α_0 і α_1 . Похибкою прогнозу для довільного значення s завжди буде $\varepsilon(k+s)$, а дисперсією похибки прогнозу буде $\text{var}[\varepsilon(k+s)]$. Навіть якщо послідовність $\{\varepsilon(k)\}$ є корельованою, довгострокові прогнози будуть поступово залежати тільки від значень α_0 , α_1 і горизонту прогнозування s .

Однак, ситуація суттєво змінюється у випадку, коли послідовність $\{y(k)\}$ має стохастичний тренд. Розглянемо, для прикладу, модель випадкового кроку + шум:

$$\begin{aligned} y(k) &= \mu(k) + \eta(k), \\ \mu(k) &= \mu(k-1) + \varepsilon(k). \end{aligned}$$

Розв'язок цього рівняння для моменту $k + s$ має такий вигляд:

$$y(k + s) = y(k) + \sum_{i=1}^s \varepsilon(k + i) + \eta(k + s) - \eta(k).$$

Зазначимо, що при довгостроковому прогнозуванні дисперсія похибки прогнозу буде необмеженою. Оцінка прогнозу на s кроків визначається за виразом:

$$E_k[y(k + s)] = y(k) + \eta(k),$$

і дисперсія похибки прогнозу

$$\begin{aligned} \text{Var} \{y(k + s) - E_k[y(k + s)]\} &= \text{var} \left[\sum_{i=1}^s \varepsilon(k + i) + \eta(k + s) \right] = \\ &= s \sigma_{\varepsilon}^2 + \sigma_{\eta}^2. \end{aligned}$$

Із розширенням горизонту прогнозування (збільшення кількості кроків прогнозування) довірчий інтервал для прогнозів буде зростати і при $s \rightarrow \infty$ дисперсія похибки прогнозу буде нескінченною.

Коректно виконані дослідження свідчать, що існує ряд макроекономічних змінних, які є процесами, стаціонарними стосовно різниць, а не стосовно тренду [12, 55]. Згідно з літературними джерелами, виконано аналіз багатьох макроекономічних змінних (для США) – реальний ВВП, номінальний ВВП, промислове виробництво, зайнятість, рівень безробіття, дефлятор ВВП, індекс споживчих цін, заробітна плата, об'єм грошей в обороті, доходи від цінних паперів та інші. Вибірка даних для споживчих цін починалась в 1860, а для ВВП в 1909 році. Всі вибірки закінчувались в 1970 році. Деякі результати цього дослідження наведено в таблиці 10.2.

Таблиця 10.2 Вибрані значення коефіцієнтів кореляції

	$\rho(1)$	$\rho(2)$	$r(1)$	$r(2)$	$d(1)$	$d(2)$
Реальний ВВП	0,95	0,90	0,34	0,04	0,87	0,66
Номінальний ВВП	0,95	0,89	0,44	0,08	0,93	0,79
Промислове виробництво	0,97	0,94	0,03	-0,11	0,84	0,67
Рівень безробіття	0,75	0,47	0,09	-0,29	0,75	0,46

У перших двох стовпчиках наведено значення коефіцієнтів кореляції першого і другого порядку для реального і номінального ВВП, промислового

виробництва і рівня безробіття. Зазначимо, що коефіцієнти автокореляції для перших трьох рядів свідчать про наявність процесів з одиничними коренями. Хоча для рівня безробіття $\rho(1) = 0,75$, автокореляція другого порядку $\rho(2) = 0,47$, тобто, є меншою 0,5.

Коефіцієнти $r(1)$ і $r(2)$ – це значення автокореляції першого і другого порядку для перших різниць. Їх значення свідчать про стаціонарність цих процесів, а також про те, що дані згенеровані процесами, стаціонарними стосовно різниць. Додатні значення коефіцієнтів автокореляції для перших різниць реального та номінального ВВП при значенні лагу 1 припускають існування процесу $КС(1)$. Коефіцієнти $d(1)$ і $d(2)$ – автокореляційні коефіцієнти для процесів з вилученими трендами, тобто для випадку моделювання тренду поліномами. Подальшою підтримкою аргументу щодо стаціонарності вибраних процесів стосовно різниць є те, що диференціювання процесів, стаціонарних стосовно тренду, призводить до виникнення неінвертованих ковзних процесів. Разом з тим, жодний з процесів, що розглядались (диференціювались), не мав одиничного кореня у ковзному середньому.

Результати моделювання лінійного тренду та обчислення коефіцієнтів автокореляції для відповідних залишків наведено у двох останніх стовпчиках таблиці 10.2. Цікаво зазначити, що коефіцієнти вибіркової автокореляційної функції для даних з вилученим трендом мають досить великі значення. *Це підтверджує той факт, що вилучення тренду з часового ряду, стаціонарного стосовно різниць, шляхом моделювання (опису тренду) не приводить до його стаціонарності.* З таблиці також видно, що вилучення тренду з ряду, що характеризує зайнятість, не впливає на значення коефіцієнтів автокореляції.

Звичайно, що аналіз автокореляційних функцій – це не єдиний підхід до визначення факту, чи буде ряд стаціонарним стосовно різниць. Для цього існують формальні статистичні тести. Однак, процедура тестування також не може бути простою і прямолінійною, тому що класичні тести призначені для аналізу стаціонарних процесів. Відносно розглянутого прикладу можна сказати, що у даному випадку неможливо відхилити нуль-гіпотезу, що дані є стаціонарними відносно різниць. Якщо це так, то макроекономічні змінні не зростають монотонно і гладко з постійною швидкістю. Однак, деякі макроекономічні впливи (удари) мають тривалу природу і ефекти від їх виникнення ніколи не зникає. Бажано, щоб подібні впливи були позитивними, але аналіз реальних процесів свідчить про те, що це далеко не завжди так.

10.5 Стохастичні тренди та декомпозиція ряду

З прикладу, наведеного у попередньому параграфі, 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. Віднімаючи детерміновану складову від 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 стохастичний, 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 у *середньому*. Наприклад, розглянемо модель випадкового кроку + шум:

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

Оскільки $E[\varepsilon(k)] = 0$, то *середня* зміна значення $y(k)$ визначається константою a_0 . Звичайно, що у 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 [13] 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):

$$y(k) = y(k-1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2). \quad (10.5.1)$$

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:

$$y(k) = a_0 k + y_0 + \sum_{i=1}^k e(i). \quad (10.5.2)$$

Для моменту $k + s$ отримаємо:

$$y(k+s) = a_0(k+s) + y_0 + \sum_{i=1}^{k+s} e(i). \quad (10.5.3)$$

Підставимо (10.5.2) в (10.5.3) для того щоб виключити y_0 :

$$y(k+s) = a_0 s + y(k) + \sum_{i=1}^s e(k+i). \quad (10.5.4)$$

Тепер повернемося до $\varepsilon(k)$:

$$\sum_{i=1}^s e(k+i) = \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i), \quad (10.5.5)$$

$$y(k+s) = a_0 s + y(k) + \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i). \quad (10.5.6)$$

Запишемо функції прогнозування для різних значень s . Оскільки $E_k[\varepsilon(k+i)] = 0, \forall i > 0$, то отримаємо:

$$\begin{aligned} E_k[y(k+1)] &= a_0 + y(k) + \beta_1 \varepsilon(k) + \beta_2 \varepsilon(k-1), \\ E_k[y(k+2)] &= 2a_0 + y(k) + (\beta_1 + \beta_2) \varepsilon(k) + \beta_2 \varepsilon(k-1), \\ &\vdots \\ E_k[y(k+s)] &= s a_0 + y(k) + (\beta_1 + \beta_2) \varepsilon(k) + \beta_2 \varepsilon(k-1). \end{aligned} \quad (10.5.7)$$

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_0s . Отриманий результат дає можливість зробити такі висновки:

1. Вказаний тренд визначається як умовне математичне сподівання граничного значення функції прогнозування. У нестрогому смислі він є “довгостроковим” прогнозом. Значення цього прогнозу буде відрізнятись в кожний момент часу при надходженні нових значень послідовності $\{\varepsilon(k)\}$. В будь-який момент k нерегулярна компонента ряду представляє собою різницю між $y(k)$ і трендом $\mu(k)$. Таким чином, використовуючи (10.5.7), знайдемо, що нерегулярна (коливальна) компонента ряду визначається за виразом:

$$y(k) - E_k[y(k+s)] + sa_0 = -(\beta_1 + \beta_2)\varepsilon(k) - \beta_2\varepsilon(k-1)$$

або

$$y(k) - \mu(k) = -(\beta_1 + \beta_2)\varepsilon(k) - \beta_2\varepsilon(k-1). \quad (10.5.8)$$

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

$$\begin{aligned} \Delta\mu(k) &= \Delta y(k) + (\beta_1 + \beta_2)\Delta\varepsilon(k) + \beta_2\Delta\varepsilon(k-1) = \\ &= [y(k) - y(k-1)] + (\beta_1 + \beta_2)\varepsilon(k) - \beta_1\varepsilon(k-1) - \beta_2\varepsilon(k-2). \end{aligned}$$

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 у момент k складається з дрейфу a_0 та інновації у вигляді білого шуму $(1 + \beta_1 + \beta_2)\varepsilon(k)$.

В роботі [40] 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 (10.5.8), отримаємо значення нерегулярної компоненти. Повторюючи розглянуту обчислювальну процедуру для кожного моменту часу k , отримаємо всю нерегулярну послідовність. З рівняння (10.5.8) видно, що ця нерегулярна компонента є різницею між $y(k)$ і трендом, тобто тепер можна легко отримати складову тренду.

10.6 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:

$$y(k) = a_0 + y(k-1) + \varepsilon(k) + \beta_1\varepsilon(k-2) + \beta_2\varepsilon(k-2) + \dots \quad (10.6.1)$$

Введемо позначення

$$e(k) = \varepsilon(k) + \beta_1\varepsilon(k-1) + \beta_2\varepsilon(k-2) + \beta_3\varepsilon(k-3) + \dots \quad (10.6.2)$$

і, по аналогії з розглянутими вище прикладами, запишемо розв'язок рівняння (10.6.1) у вигляді:

$$y(k+s) = y(k) + a_0s + \sum_{i=1}^s e(k+i). \quad (10.6.3)$$

У відповідності з рівнянням (10.6.2) останню складову у рівнянні (10.6.3) представимо у вигляді:

$$\sum_{i=1}^s e(k+i) = \sum_{i=1}^s \varepsilon(k+i) + \beta_1 \sum_{i=1}^s \varepsilon(k-1+i) + \beta_2 \sum_{i=1}^s \varepsilon(k-2+i) + \beta_3 \sum_{i=1}^s \varepsilon(k-3+i) + \dots \quad (10.6.4)$$

Використовуючи можливість розкладання сум, наявних в (10.6.4), можна записати функцію прогнозу на s кроків:

$$E_k[y(k+s)] = y(k) + a_0s + \left(\sum_{i=1}^s \beta_i\right)\varepsilon(k) + \left(\sum_{i=2}^{s+1} \beta_i\right)\varepsilon(k-1) + \left(\sum_{i=3}^{s+2} \beta_i\right)\varepsilon(k-2) + \dots \quad (10.6.5)$$

Для того щоб визначити структуру стохастичного тренду покладемо $s \rightarrow \infty$:

$$\lim_{s \rightarrow \infty} E_k[y(k+s)] = y(k) + a_0s + \left(\sum_{i=1}^{\infty} \beta_i\right)\varepsilon(k) + \left(\sum_{i=2}^{\infty} \beta_i\right)\varepsilon(k-1) + \left(\sum_{i=3}^{\infty} \beta_i\right)\varepsilon(k-2) + \dots \quad (10.6.6)$$

З іншого боку, прогноз на s кроків можна представити у вигляді тотожності:

$$y(k+s) = \Delta y(k+s) + \Delta y(k+s-1) + \Delta y(k+s-2) + \dots + \Delta y(k+1) + y(k). \quad (10.6.7)$$

$$\begin{aligned} \text{Наприклад, } y(k+2) &= \Delta y(k+2) + \Delta y(k+1) + y(k) = \\ &= y(k+2) - y(k+1) + y(k+1) - y(k) + y(k) = y(k+2). \end{aligned}$$

Тепер прогноз на основі умовного математичного сподівання

$$\begin{aligned} \lim_{s \rightarrow \infty} E_k[y(k+s)] &= \lim_{s \rightarrow \infty} E_k \{ [y(k+s) - y(k+s-1)] + [y(k+s-1) - y(k+s-2)] + \dots \\ &\quad + [y(k+2) - y(k+1)] + [y(k+1) - y(k)] + y(k) \} = \\ &= \lim_{s \rightarrow \infty} E_k [\Delta y(k+s) + \Delta y(k+s-1) + \dots + \Delta y(k+2) + \Delta y(k+1)] + y(k). \end{aligned} \quad (10.6.8)$$

Таким чином, для того щоб побудувати прогноз за виразом (10.6.8), необхідно знайти значення $y(k+1), \dots, y(k+s)$. Останній вираз і (10.6.3) дають можливість знайти нерегулярну компоненту (тобто, стохастичний тренд) як

$$\begin{aligned} Irt &= y(k) - \left[\lim_{s \rightarrow \infty} [E_k(y(k+s))] + a_0s \right] = \\ &= - \lim_{s \rightarrow \infty} E_k [\Delta y(k+s) + \Delta y(k+s-1) + \dots + \Delta y(k+2) + \Delta y(k+1)] - s a_0. \end{aligned} \quad (10.6.9)$$

На основі наведених викладок можна сформулювати наступну 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:

$$E_k[\Delta y(k + s) + \Delta y(k + s - 1) + \dots + \Delta y(k + 1)] + y(k).$$

Наприклад, при $s = 10$ для $k = 1$ необхідно обчислити суму:

$$\zeta_1 = E_1[\Delta y(11) + \Delta y(10) + \dots + \Delta y(2)] + y(1).$$

Також, у відповідності до (10.6.8), можна записати:

$$y(1) + E_1[\Delta y(11) + \Delta y(10) + \dots + \Delta y(2)] = E_1[y(11)].$$

Звідси можна записати, що стохастична частина тренду буде представлена в момент $k = 1$ наступним виразом:

$$E_1[y(11)] - a_0 s,$$

де $a_0 s$ – детермінована частина тренду. Аналогічно можна записати суму для моменту $k = 2$:

$$\zeta_2 = E_2[\Delta y(12) + \Delta y(11) + \dots + \Delta y(3)] + y(2).$$

Для ряду, що складається з N спостережень, компонента випадкового тренду в момент часу $k = N$ визначається сумою:

$$\zeta_N = E_N[\Delta y(N + 10) + \Delta y(N + 9) + \dots + \Delta y(N + 1)] + y(N).$$

В результаті отримаємо послідовність значень сум:

$$\{\zeta(k)\} = [\zeta_1, \zeta_2, \dots, \zeta_N].$$

Крок 3. Формування нерегулярної компоненти ряду:

$$Irt(k) = -E_k[\Delta y(k + 10) + \Delta y(k + 9) + \dots + \Delta y(k + 1)].$$

Для більшості рядів значення s можна брати невеликим, наприклад, $s = 2$ для процесу, що описується рівнянням:

$$y(k) = y(k-1) + a_0 + \varepsilon(k) + \beta_1 \varepsilon(k-1) + \beta_2 \varepsilon(k-2),$$

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 10.1. Розглянемо модель курсу обміну валют

$$\Delta y(k) = -0,0113 + \varepsilon(k) + 0,2973\varepsilon(k-1).$$

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

$$E_k[\Delta y(k+100) + \Delta y(k+99) + \dots + \Delta y(k+1)],$$

можна замінити на $-100 \cdot 0,0113 + 0,2973\varepsilon(k)$ при $s = 100$. Так, для першого використаного спостереження стохастичну частину тренду можна 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.

10.7 Розкладання тренду для моделі випадкового кроку + шум + дрейф

Як було показано вище, модель випадкового кроку + шум + дрейф має вигляд:

$$y(k) = \mu(k) + \eta(k), \quad (10.7.1)$$

$$\mu(k) = a_0 + \mu(k-1) + \varepsilon(k), \quad (10.7.2)$$

$$E[\eta(k)\varepsilon(k)] = 0,$$

де $\eta(k)$, $\varepsilon(k)$ – некорельовані процеси білого шуму. Запишемо розв'язок для моментів k і $(k+s)$:

$$y(k) = \mu_0 + a_0 k + \sum_{i=1}^k \varepsilon(i) + \eta(k), \quad (10.7.3)$$

$$y(k+s) = \mu_0 + a_0(k+s) + \sum_{i=1}^{k+s} \varepsilon(i) + \eta(k+s). \quad (10.7.4)$$

З рівняння (10.7.3) знайдемо:

$$\mu_0 = y(k) - a_0 k - \sum_{i=1}^k \varepsilon(i) - \eta(k) \quad (10.7.5)$$

і підставимо цей вираз в (10.7.4):

$$y(k+s) = y(k) + a_0 s + \sum_{i=1}^s \varepsilon(k+i) + \eta(k+s) - \eta(k). \quad (10.7.6)$$

Запишемо рівняння (10.7.3) для моменту $y(k+1)$:

$$y(k+1) = \mu_0 + a_0(k+1) + \sum_{i=1}^{k+1} \varepsilon(i) + \eta(k+1) \quad (10.7.7)$$

і знайдемо першу різницю:

$$\begin{aligned} \Delta y(k+1) &= y(k+1) - y(k) = \\ &= \mu_0 + a_0 k + a_0 + \sum_{i=1}^k \varepsilon(i) + \varepsilon(k+1) + \eta(k+1) - \\ &\quad - \mu_0 - a_0 k - \sum_{i=1}^k \varepsilon(i) - \eta(k) = \\ &= a_0 + \varepsilon(k+1) + \eta(k+1) - \eta(k). \end{aligned} \quad (10.7.8)$$

Знайдемо математичне сподівання, дисперсію і коваріацію для $\Delta y(k)$:

$$E[\Delta y(k)] = E[a_0 + \varepsilon(k) + \eta(k) - \eta(k-1)] = a_0. \quad (10.7.9)$$

$$\begin{aligned} \text{var}[\Delta y(k)] &= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)][\varepsilon(k) + \eta(k) - \eta(k-1)]\} = \\ &= \sigma_\varepsilon^2 + 2\sigma_\eta^2 = \text{const}. \end{aligned} \quad (10.7.10)$$

$$\begin{aligned} \text{cov}[\Delta y(k), \Delta y(k-1)] &= \\ &= E\{[\varepsilon(k) + \eta(k) - \eta(k-1)][\varepsilon(k-1) + \eta(k-1) - \eta(k-2)]\} = \\ &= -\sigma_\eta^2. \end{aligned} \quad (10.7.11)$$

$$\text{cov}[\Delta y(k), \Delta y(k+s)] = 0, \quad k \geq 2.$$

Знайдені характеристики дозволяють представити послідовність $\{\Delta y(k)\}$ через компоненти білого шуму як

$$\Delta y(k) = a_0 + e(k) + \beta_1 e(k-1), \quad (10.7.12)$$

де $e(k)$ – послідовність, яка включає в себе обидві компоненти: $\{\varepsilon(k)\}$ і $\{\eta(k)\}$, які необхідно розділити.

Якщо скористатись рівнянням (10.7.12) для описання вихідного процесу, то у відповідності до (10.7.11) тільки $\text{cov}[\Delta y(k), \Delta y(k-1)] \neq 0$. Таким чином, основні статистичні характеристики процесу (10.7.12) мають вигляд:

$$\begin{aligned}
\text{Var} [\Delta y(k)] &= \text{Var} [e(k) + \beta_1 e(k-1)] = \\
&= E \left\{ [e^2(k) + 2\beta_1 e(k)e(k-1) + \beta_1^2 e^2(k-1)] \right\} = \\
&= (1 + \beta_1^2) \text{Var} [e(k)] = (1 + \beta_1^2) \sigma_\varepsilon^2. \quad (10.7.13)
\end{aligned}$$

$$\begin{aligned}
\text{Cov} [\Delta y(k), \Delta y(k-1)] &= E \{ [e(k) + \beta_1 e(k-1)] [e(k-1) + \beta_1 e(k-2)] \} = \\
&= E [\beta_1 e(k-1) e(k-1)] = \beta_1 \sigma_\varepsilon^2. \quad (10.7.14)
\end{aligned}$$

Тепер можна порівняти значення статистичних характеристик процесу, отримані раніше:

$$\sigma_\varepsilon^2 + 2\sigma_\eta^2 = (1 + \beta_1^2) \sigma_\varepsilon^2 \quad \text{і} \quad -\sigma_\eta^2 = \beta_1 \sigma_\varepsilon^2. \quad (10.7.15)$$

Звідси випливає, що

$$\sigma_\eta^2 = -\beta_1 \sigma_\varepsilon^2, \quad (10.7.16)$$

$$\begin{aligned}
\sigma_\varepsilon^2 &= (1 + \beta_1^2) \sigma_\varepsilon^2 - 2\sigma_\eta^2 = (1 + \beta_1^2) \sigma_\varepsilon^2 + 2\beta_1 \sigma_\varepsilon^2 = \\
&= \sigma_\varepsilon^2 (1 + \beta_1)^2. \quad (10.7.17)
\end{aligned}$$

Тепер вилучимо самі послідовності $\{\varepsilon(k)\}$ і $\{\eta(k)\}$. З рівняння (10.7.6) запишемо прогноз на один крок:

$$E_k [y(k+1)] = y(k) + a_0 - \eta(k) \quad (10.7.18)$$

і для (10.7.12):

$$E_k [\Delta y(k+1)] = a_0 + \beta_1 e(k). \quad (10.7.19)$$

Умовний прогноз для $y(k+1) = y(k) + a_0 + e(k+1) + \beta_1 e(k)$ має вигляд:

$$E_k [y(k+1)] = y(k) + a_0 + \beta_1 e(k). \quad (10.7.20)$$

Оскільки функції прогнозування (10.7.18) і (10.7.20) повинні бути еквівалентними, то $y(k) + a_0 - \eta(k) = y(k) + a_0 + \beta_1 e(k)$

$$\text{або} \quad \eta(k) = -\beta_1 e(k). \quad (10.7.21)$$

З рівняння для перших різниць $\Delta y(k) = a_0 + \varepsilon(k) + \Delta \eta(k)$ знайдемо, що

$$\varepsilon(k) = \Delta y(k) - a_0 + \beta_1 \Delta \eta(k), \quad (10.7.22)$$

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 (10.7.21) і (10.7.22) можна розкласти ряд, у якому кореляція між трендом і нерегулярною компонентою дорівнює нулю.

Наведений приклад є ілюстрацією такого 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:

$$E[\varepsilon(k)\eta(k)]/\sigma_\varepsilon\sigma_\eta = 1,$$

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 у 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 (10.6.8) 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. Зазначимо, що це питання не має нічого спільного з коректною формою моделі АРІКС; задача тут полягає у 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 у даному випадку в тому, що немає можливості ідентифікувати “істинну” модель на основі вибіркового даних.

Таким чином, у 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:

$$\Delta y(k) = 0,0043 + 0,395 \Delta y(k-1) + \varepsilon(k), \quad \sigma_\varepsilon = 0,0108.$$

Припускаючи, що інновації тренду та нерегулярної компоненти є некорельованими, знайдемо рівняння:

$$y(k) = \mu(k) + A(L)\eta(k),$$

де $\Delta \mu(k) = 0,0076 + \varepsilon(k)$, $\sigma_\varepsilon = 0,0061$,

$$A(L)\eta(k) = (1 - 1,487L + 0,569L^2)\eta(k), \quad \sigma_\eta = 0,0079.$$

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

$$\Delta y(k) = \frac{0,0043 + \varepsilon(k)}{1 - 0,395L},$$

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 десятого розділу

1. Поясніть, якими функціями можна описати детермінований тренд? Які функції, на вашу думку, необхідно застосовувати в першу чергу і чому?
2. Запишіть математичну модель, яка відповідає наступній 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?
7. Який розв'язок різницевого рівняння називають невластим? 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 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; $E[\varepsilon(k)\eta(k-s)] = 0, \forall k, s$. Знайдіть 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 описання реального випадкового процесу, 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, при умові, 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? Поясніть результати на прикладі.
 12. Дайте пояснення скороченню: АРІКС(p, d, q)? Яким чином можна знайти оцінку d ?
 13. Запропонуйте послідовність побудови моделі процесу із стохастичним трендом.

Р о з д і л 11

ПРОГНОЗУВАННЯ КОІНТЕГРОВАНИХ ПРОЦЕСІВ

11.1 Задача моделювання нестационарних коінтегрованих процесів

У 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 вважається, що тренд вказує на відносно довгострокові зміни процесу.

Під трендом будемо розуміти змінне у часі поточне середнє значення процесу, яке може бути отримане, наприклад, за допомогою процедури оцінювання поточного середнього, фільтрації, описане за допомогою полінома або іншої 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 на відносно коротких часових інтервалах.

У загальному випадку задача коректного математичного опису тренду може бути не дуже простою. Якщо для процесу з трендом побудувати модель авторегресії з ковзним середнім, то вона може виявитись коректною з точки зору параметрів, що використовуються для встановлення її адекватності

(коефіцієнт детермінації наближається до 1, статистика Дарбіна-Уотсона близька до 2, коефіцієнти 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 констатує наявність спільного тренду.*

Від нестационарного ряду можна перейти до стаціонарного шляхом знаходження перших різниць або різниць вищого порядку. Але при переході до часового ряду у вигляді різниць у ньому залишається лише та інформація, що відповідає короткостроковим змінам (коливанням) процесу навколо тренду. Тобто *втрачається інформація щодо довгострокових змін процесу, яка міститься у тих складових змінної, що вилучаються при переході до різниць.* Задача вилучення тренду за допомогою різниць чи детермінованої функції розв'язується в кожному випадку аналізу даних окремо, оскільки наперед невідомо, який саме тренд міститься у процесі – детермінований чи випадковий. Остаточний вибір методу побудови моделі відбувається на етапі аналізу моделі для залишків, які ми отримуємо після опису (вилучення) тренду.

Тому у загальному випадку існує проблема коректного моделювання процесів (часових рядів), що містять позитивні (зростаючі) чи негативні (спадаючі) тренди, а також інші складові, які приводять до відхилення від стаціонарності.

Одним із підходів до коректного математичного опису нестационарних рядів з трендом є так званий *коінтеграційний підхід* (аналіз), який дає можливість побудувати модель для кількох нестационарних процесів, що розглядаються в рамках однієї постановки задачі. *Якщо ряди нестационарні, але коінтегровані, то їх лінійна комбінація може бути стаціонарним процесом.* Зокрема, цей підхід дозволяє будувати моделі корегування похибок (МКП), що містять механізм корегування похибок (математичного опису) відносно довгострокових ефектів процесів, що моделюються.

Концепція коінтегрованості змінних передбачає існування довгострокового зв'язку між значеннями змінних. Тобто припускається існування спільної врівноваженої траєкторії руху цих змінних, від якої вони можуть відхилитися на коротких проміжках часу, але економічні механізми в цілому діють таким чином, що рівновага відтворюється і зберігається на довгих часових інтервалах шляхом корегування відповідних відхилень від врівноваженого стану. Існування коінтегрованості змінних дає можливість коректно використовувати частину з них у якості керуючих впливів.

Якщо процеси, що розглядаються спільно, коінтегровані, то можна побудувати відповідну модель корегування похибки, що має наступні характеристики [24]:

- вона одночасно відображає короткострокові та довгострокові аспекти динаміки досліджуваних процесів;
- 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 шляхом обчислення різниць з метою видалення тренду. Проте задача ускладнюється у випадку аналізу багатовимірних часових рядів, що містять нестационарні змінні. У цьому випадку для того щоб описати нестационарні змінні будують лінійні комбінації кількох нестационарних процесів, які називають коінтегрованими.

Наявність коінтеграційних співвідношень характерно для багатьох процесів у економіці, особливо для процесів економіки перехідного періоду (ЕПП). Багато процесів ЕПП відрізняються високою динамікою, нестационарністю, наявністю інтегруючих компонент. У цілому це позитивний ефект, оскільки коінтеграція може свідчити про існування причинно-наслідкових зв'язків між змінними, які можна об'єднувати у одній математичній моделі.

Для ілюстрації концепції коінтеграційного підходу розглянемо рівняння, що описує попит на грошову масу:

$$m(k) = \beta_0 + \beta_1 p(k) + \beta_2 y(k) + \beta_3 r(k) + e(k), \quad (11.1.1)$$

де $m(k)$ – довгостроковий попит на грошову масу; $p(k)$ – рівень споживчих цін; $y(k)$ – фактичний прибуток; $r(k)$ – процентна ставка центрального банку; $e(k)$ – збурення у вигляді 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, фактичний прибуток (для його вимірювання можна скористатися фактичним внутрішнім валовим продуктом (ВВП)), а також відповідні значення процентних ставок центрального банку.

З логічних міркувань видно, що $\beta_1 = 0$, $\beta_2 > 0$, $\beta_3 < 0$, що повинно бути підтверджено експериментальними даними. Очевидно також, що послідовність $\{e(k)\}$ повинна включати ті складові, що не увійшли 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 *нестационарними інтегрованими процесами типу I(1)* (лінійний тренд). А це означає, що значення кожної із згаданих змінних може істотно коливатися і не повертатися до очікуваного довгострокового (асимптотичного) рівня. Теорія коінтеграційного аналізу ґрунтується на тому, що існує така комбінація нестационарних змінних, яка є стаціонарним процесом. Запишемо рівняння (11.1.1) у такий спосіб:

$$e(k) = m(k) - \beta_0 - \beta_1 p(k) - \beta_2 y(k) - \beta_3 r(k). \quad (11.1.2)$$

Оскільки послідовність $\{e(k)\}$ повинна бути стаціонарною, то лінійна комбінація у правій частині (11.1.2) також повинна бути стаціонарною. Цей приклад ілюструє основну ідею коінтеграційного методу: *рівновага процесів, що включають нестационарні змінні, потребує існування стаціонарної комбінації цих змінних*. Наведений приклад із попитом на грошову масу є тільки однією ілюстрацією стаціонарної комбінації нестационарних змінних.

Формально *коінтеграційний аналіз* починається із розгляду множини економічних змінних, що знаходяться у певній довгостроковій рівновазі:

$$\beta_1 x_1(k) + \beta_2 x_2(k) + \dots + \beta_n x_n(k) = 0. \quad (11.1.3)$$

Якщо позначити вектори параметрів і вимірів через β і \mathbf{x} , відповідно, то рівняння (11.1.3) можна представити у векторному вигляді:

$$\mathbf{e}(k) = \beta \mathbf{x}(k), \quad \beta = [\beta_1, \beta_2, \dots, \beta_n]^T, \quad \mathbf{x} = [x_1(k), x_2(k), \dots, x_n(k)]^T$$

Введемо наступне *визначення коінтеграції*:

Компоненти вектора $\mathbf{x}(k)$ є *коінтегрованими* порядку d, b (що позначається як $\mathbf{x}(k) \sim CI(d, b)$), якщо виконуються такі умови:

- усі компоненти вектора $\mathbf{x}(k)$ є інтегрованими процесами порядку d ;
- існує такий вектор $\beta = [\beta_1, \beta_2, \dots, \beta_n]^T$, 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.

Вектор β називають коінтеграційним вектором. Якщо у рівнянні (11.1.1) рівень цін, фактичний прибуток і процентна ставка є інтегрованими процесами порядку $I(1)$, а лінійна комбінація змінних

$$e(k) = m(k) - \beta_0 - \beta_1 p(k) - \beta_2 y(k) - \beta_3 r(k)$$

стаціонарна, то ці змінні є інтегрованими порядку $(1,1)$. Вектори вимірів і параметрів мають вигляд:

$$\mathbf{x}^T(k) = [m(k) \ 1 \ p(k) \ y(k) \ r(k)], \quad \beta^T = [1 - \beta_0 - \beta_1 - \beta_2 - \beta_3].$$

Відхилення від довгострокової ринкової рівноваги визначається значеннями випадкової складової $e(k)$. Оскільки $\{e(k)\}$ – стаціонарна послідовність, то відхилення від рівноваги носять короткочасний характер.

Чотири важливих зауваження стосовно означення коінтеграції

1. Поняття коінтеграції відноситься до лінійної комбінації нестаціонарних змінних. Теоретично припускається також існування довгострокової нелінійної залежності між елементами множини інтегрованих змінних. Проте на сьогодні теорія коінтеграційного аналізу не дає можливості встановлювати нелінійні зв'язки. Слід також відзначити, 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 x_1 обирають $\lambda = 1/\beta_1$.

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 інтегрованості $I(d)$, не є коінтегрованою. Відсутність коінтеграції означає, що відсутня довгострокова рівновага між змінними, тобто характер їхньої зміни може істотно різнитися у часі. Якщо змінні інтегровані, але мають різний порядок інтегрованості, то вони не можуть бути коінтегровані. 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» не зовсім вдалий, оскільки економісти-теоретики й економісти-математики вкладають у нього різний зміст. Економісти-теоретики зазвичай розуміють під рівновагою рівність між

необхідними і фактичними 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) підтримувалася за рахунок ринкових впливів або за рахунок визначених правил поведіння індивідуумів. У більшості випадків *під рівновагою розуміють причинні, поведінкові або просто співвідношення 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 у наведеному 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:

$$\begin{aligned} m(k) &= \gamma_0 - \gamma_1[y(k) + p(k)] + e_1(k) = \\ &= \gamma_0 - \gamma_1 y(k) - \gamma_1 p(k) + e_1(k) \end{aligned} \quad (11.1.4)$$

де $\{e_1(k)\}$ – стаціонарний процес залишків моделі, що описує випускання в оборот грошової маси.

Для моделі потреби у грошовій масі (11.1.1) існує два коінтеграційних вектори для таких змінних як пропозиція грошової маси, рівень цін, фактичний прибуток і процентна ставка. Їх можна представити у вигляді коінтеграційної матриці, що складає з двох векторів:

$$\mathbf{B} = \begin{bmatrix} 1 & -\beta_0 & -\beta_1 & -\beta_2 & -\beta_3 \\ 1 & -\gamma_0 & \gamma_1 & \gamma_1 & 0 \end{bmatrix}.$$

Обидві лінійні комбінації, що задаються добутком $\mathbf{B}\mathbf{x}(k)$, є стаціонарними. Таким чином, *ранг коінтеграції* для $\mathbf{x}(k)$ дорівнює двом. У випадку наявності множини коінтеграційних векторів може виявитися неможливим ідентифікувати скорочену форму взаємозв'язків.

4. У переважній більшості випадків *публікації з коінтеграції фокусуються на багатовимірних процесах*, кожний 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 дуже обмежене число економічних змінних, порядок інтегрованості яких перевищує одиницю. Хоча для економіки перехідного періоду досить характерним є другий порядок для більшості процесів.

У випадку, коли не виникає невизначеностей, багато авторів використовують термін «коінтеграція» якщо змінні мають порядок коінтегрованості $CI(1,1)$. Очевидно, що існує також ряд інших можливостей. 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 $I(1)$.

Приклади коінтегрованих процесів

Розглянемо приклади коінтегрованих і некоінтегрованих процесів. Послідовності $\{y(k)\}, \{z(k)\}$ отримані в результаті використання моделі випадкового кроку + шум, тобто моделі типу $x(k) = x(k-1) + \varepsilon(k)$, де $\varepsilon(k)$ – псевдовипадкова послідовність нормально розподілених величин із нульовим середнім і скінченною дисперсією $Var[\varepsilon(k)] = \sigma_\varepsilon^2$. Фактично тут використано моделі типу $y(k) = \mu(k) + \varepsilon_y(k)$, де $\mu(k) = \mu(k-1) + v(k)$, $\varepsilon_y(k), v(k)$ – некорельовані процеси білого шуму.

Як свідчить моделювання послідовності $\{y(k)\}$, на перших двадцяти кроках для обох послідовностей спостерігається спадаючий часовий тренд (хоча це зовсім необов'язково, він може бути також позитивним або відсутнім взагалі). У будь-якому випадку немає тенденції цих процесів до деякого сталого стану, тобто вони є нестационарними, і нуль-гіпотеза про наявність одиничного кореня може підтвердитися, якщо застосувати відповідний тест.

Різниця між цими процесами, $[y(k) - z(k)]$, виявляється стаціонарною; «похибка рівноваги» $e(k) = y(k) - z(k)$ має нульове середнє і постійну дисперсію. В результаті тестування встановлено, що цей процес є процесом $I(0)$, тобто неінтегрованим процесом.

Також встановлено, що між трьома процесами типу випадковий крок + шум існує коінтеграція. Жодний із згенерованих процесів не має тенденції до встановлення на деякому постійному рівні, а формальне застосування тесту Дікі-Фулера на наявність одиничних коренів свідчить про їхню наявність у всіх процесах. На відміну від попереднього випадку, тут немає двох процесів, що були б коінтегровані між собою, оскільки кожний із рядів має свою власну тенденцію розвитку. Проте для розглянутих процесів існує лінійна комбінація:

$$e(k) = y(k) + z(k) - w(k), \quad (11.1.5)$$

яка є стаціонарною. Таким чином, можна зробити висновок, що принаймні одна із змінних у розглянутій системі повинна бути обмежена («скомпенсована») значеннями двох інших змінних.

На рис.11.1 наведено *графік розсіювання* змінних $y(k)$, $z(k)$ одна відносно іншої, при цьому очевидно, що всі двадцять відліків утворюють упорядковані пари $[y(i), z(i)]$, $i=1,2,\dots,20$. З рис. 11.1 випливає, що малі значення послідовності $\{y(k)\}$ асоціюються з малими значеннями послідовності $\{z(k)\}$. Оскільки обидва ряди мають однакову тенденцію до зміни у часі, то між ними існує позитивна кореляція. Графік цієї залежності у вигляді прямої лінії, отриманої в результаті застосування

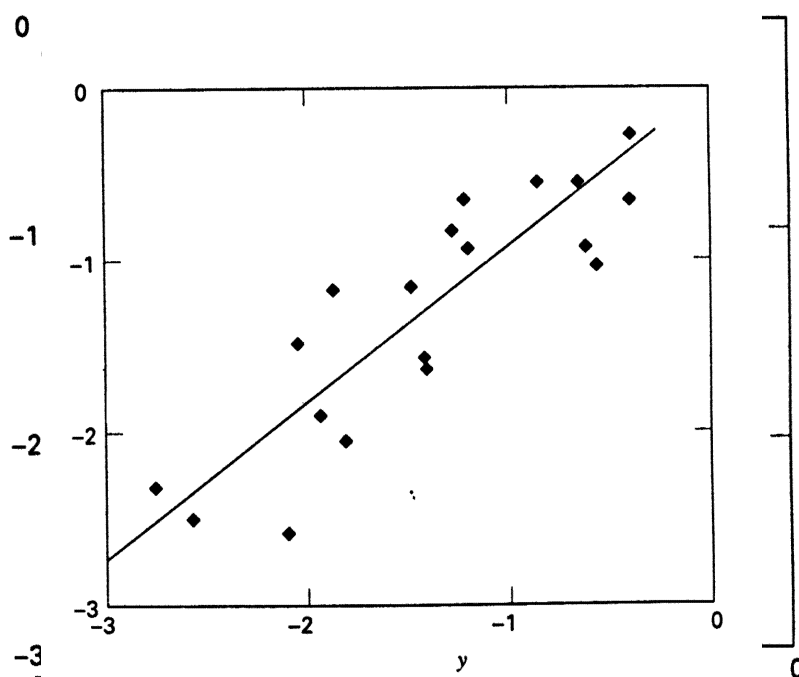


Рис.11.1 Графік розсіювання змінних

МНК до обох рядів, наведений на рис. 11.1. Ця лінія підтверджує наявність сильної позитивної залежності і представляє собою «довгостроковий» рівноважний взаємозв'язок між рядами. Відхилення («похибки») – це стаціонарні відхилення від довгострокової рівноваги.

11.2 Моделі коінтеграції

11.2.1 Коінтеграція і тренди

У процесі аналізу коінтегрованих змінних встановлено, що вони мають однакові стохастичні тренди. Це полегшує розуміння сутності коінтегрованості. З метою 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:

$$y(k) = \mu_y(k) + \varepsilon_y(k), \quad (11.2.1)$$

$$z(k) = \mu_z(k) + \varepsilon_z(k), \quad (11.2.2)$$

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:

$$\begin{aligned} \beta_1 y(k) + \beta_2 z(k) &= \beta_1 [\mu_y(k) + \varepsilon_y(k)] + \beta_2 [\mu_z(k) + \varepsilon_z(k)] = \\ &= [\beta_1 \mu_y(k) + \beta_2 \mu_z(k)] + [\beta_1 \varepsilon_y(k) + \beta_2 \varepsilon_z(k)]. \end{aligned} \quad (11.2.3)$$

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 (11.2.3) 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

$$\beta_1 \mu_y(k) + \beta_2 \mu_z(k) = 0. \quad (11.2.4)$$

Очевидно, що значення змінних $\mu_x(k)$, $\mu_y(k)$ будуть безупинно змінюватися в часі. Оскільки передбачається, що обидва коефіцієнти β_1 , β_2 повинні відрізнятися від нуля, то (11.2.4) буде виконуватися тільки у тому випадку, якщо

$$\mu_y(k) = -\mu_z(k) \frac{\beta_2}{\beta_1}.$$

Єдиною можливістю забезпечення існування цієї рівності при ненульових значеннях коефіцієнтів β_1 , β_2 є рівність стохастичних трендів. Таким чином, якщо два стохастичні 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 наведеному вище прикладі послідовності $\{y(k)\}$, $\{z(k)\}$ задовольняли моделі:

$$y(k) = \mu(k) + \varepsilon_y(k); \quad z(k) = \mu(k) + \varepsilon_z(k),$$

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:

$$y(k) - z(k) = [\mu(k) + \varepsilon_y(k)] - [\mu(k) + \varepsilon_z(k)] = \varepsilon_y(k) - \varepsilon_z(k).$$

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 $\varepsilon(k) = \varepsilon_y(k) - \varepsilon_z(k)$. Додатковий аналіз свідчить, що характер зміни цієї похибки відповідає стаціонарному процесу. 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 $\beta_3 / \beta_4 = \beta_1 / \beta_2$.

У випадку наявності коінтеграції між процесами випадкового кроку + шум кожний процес є інтегрованим порядку $I(1)$, що підтверджується тестом Дікі-Фулера. Встановлено, що попарні комбінації змінних не є коінтегрованими. На відміну від першого випадку характери зміни рядів не близькі один до другого. Проте за побудовою тренд змінної $w(k)$ є простою сумою трендів змінних $y(k)$, $z(k)$, тобто:

$$\mu_w(k) = \mu_y(k) + \mu_z(k).$$

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 наступне:

$$\begin{aligned} y(k) + z(k) - w(k) &= [\mu_y(k) + \varepsilon_{y(k)}] + [\mu_z(k) + \varepsilon_z(k)] - [\mu_w(k) + \varepsilon_w(k)] = \\ &= \varepsilon_y(k) + \varepsilon_z(k) - \varepsilon_w(k). \end{aligned}$$

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

$$\mathbf{x}(k) = \mu(k) + \varepsilon(k), \quad (11.2.5)$$

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 вимірності. Якщо необхідно виразити тренд однієї із змінних у вигляді лінійної комбінації інших трендів, то повинен існувати такий вектор β , що

$$\beta_1 \mu_1(k) + \beta_2 \mu_2(k) + \dots + \beta_n \mu_n(k) = 0.$$

Якщо помножити (11.2.5) на вектор β , то отримаємо

$$\beta^T \mathbf{x}(k) = \beta^T \mu(k) + \beta^T \varepsilon(k).$$

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

$$\mathbf{V} \mu(k) = 0,$$

де $\mathbf{V}[r \times n]$ – матриця, що складається з елементів β_{ij} . Наприклад, якщо для n змінних існує два незалежних коінтеграційних вектори, то матриця \mathbf{V} має вигляд:

$$\mathbf{V} = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1n} \\ \beta_{21} & \beta_{22} & \dots & \beta_{2n} \end{bmatrix}.$$

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.

11.2.2 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 у порівнянні 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 і/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 і те, що на короткочасну динаміку повинно впливати відхилення від довгострокового взаємозв'язку.

Модель динаміки, що логічно випливає з наведених міркувань, називають *моделлю корегування похибок*. У моделі корекції 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:

$$\Delta r_S(k) = \alpha_S [r_L(k-1) - \beta r_S(k-1)] + \varepsilon_S(k), \quad \alpha_S > 0; \quad (11.2.7)$$

$$\Delta r_L(k) = -\alpha_L [r_L(k-1) - \beta r_S(k-1)] + \varepsilon_L(k), \quad \alpha_L > 0, \quad (11.2.8)$$

де $r_L(k)$, $r_S(k)$ – довгострокова і короткострокова процентна ставка, відповідно; $\varepsilon_S(k)$, $\varepsilon_L(k)$ – збурення у вигляді процесів білого шуму, що можуть бути корельовані між собою; α_S , α_L , β – 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 $r_L(k) = \beta r_S(k)$.

Наведена модель ілюструє взаємозв'язок між МКП і коінтегрованими змінними. За припущенням, різниця $\Delta r_S(k)$ є стаціонарною, що забезпечує порядок інтегрованості $I(0)$ для лівої частини рівняння (11.2.7). Очевидно, що ліва частина (11.2.8) також повинна мати ступінь інтегрованості $I(0)$.

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 (11.2.8). 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:

$$\Delta r_S(k) = a_{10} + \alpha_S [r_L(k-1) - \beta r_S(k-1)] + \sum a_{11}^i \Delta r_S(k-1) + \sum a_{12}^i \Delta r_L(k-1) + \varepsilon_S(k) \quad (11.2.9)$$

$$\Delta r_L(k) = a_{20} - \alpha_L [r_L(k-1) - \beta r_S(k-1)] + \sum a_{21}^i \Delta r_S(k-1) + \sum a_{22}^i \Delta r_L(k-1) + \varepsilon_L(k) \quad (11.2.10)$$

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 (11.2.9) і (11.2.10) 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 моделлю корекції похибки.

Отриманий результат легко узагальнюється 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:

$$\Delta \mathbf{x}(k) = \mathbf{f}_0 + \mathbf{F} \mathbf{x}(k-1) + \mathbf{F}_1 \Delta \mathbf{x}(k-1) + \mathbf{F}_2 \Delta \mathbf{x}(k-2) + \dots + \mathbf{F}_p \Delta \mathbf{x}(k-p) + \varepsilon(k), \quad (11.2.11)$$

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 вектора $\mathbf{x}(k)$ мають порядок інтегрованості $I(1)$. Якщо для цих змінних існує представлення у вигляді МКП типу (11.2.11), то обов'язково буде мати місце стаціонарна лінійна комбінація змінних із порядком інтегрованості $I(0)$. Розв'язок рівняння (11.2.11) відносно $\mathbf{F} \mathbf{x}(k-1)$ має вигляд:

$$\mathbf{F} \mathbf{x}(k-1) = \Delta \mathbf{x}(k) - \mathbf{f}_0 - \sum \mathbf{F}_i \Delta \mathbf{x}(k-i) - \varepsilon(k).$$

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:

$$f_{11}x_1(k-1) + f_{12}x_2(k-1) + \dots + f_{1n}x_n(k-1).$$

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 (11.2.11) є наявність у ньому матриці \mathbf{F} . Необхідно зазначити таке:

- Якщо всі елементи матриці \mathbf{F} рівні нулю, то рівняння (11.2.11) 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 опису процесу, якщо $\mathbf{x}(k)$ представляється у вигляді МКП типу (11.2.11).

Для того щоб дослідити взаємозв'язок між коінтеграцією і корекцією похибок, необхідно розглянути властивості простої векторної авторегресійної моделі типу:

$$x(k) = a_{11}x(k-1) + a_{22}y(k-1) + \varepsilon_x(k), \quad (11.2.12)$$

$$y(k) = a_{21}x(k-1) + a_{22}y(k-1) + \varepsilon_y(k), \quad (11.2.13)$$

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:

$$\begin{aligned} (1 - a_{11}L)x(k) - a_{12}Ly(k) &= \varepsilon_x(k), \\ -a_{21}Lx(k) + (1 - a_{22}L)y(k) &= \varepsilon_y(k). \end{aligned}$$

Наступним кроком є знаходження розв'язку цих рівнянь відносно $x(k)$, $y(k)$. Перепишемо останню систему у матричному вигляді:

$$\begin{bmatrix} (1 - a_{11}L) & -a_{12}L \\ -a_{21}L & (1 - a_{22}L) \end{bmatrix} \begin{bmatrix} x(k) \\ y(k) \end{bmatrix} = \begin{bmatrix} \varepsilon_x(k) \\ \varepsilon_y(k) \end{bmatrix}.$$

Використовуючи правило Крамера для знаходження оберненої матриці, запишемо рішення для $x(k)$, $y(k)$:

$$x(k) = \frac{(1 - a_{22})\varepsilon_x(k) + a_{12}L\varepsilon_y(k)}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}, \quad (11.2.14)$$

$$y(k) = \frac{a_{21}L\varepsilon_x(k) + (1 - a_{11}L)\varepsilon_y(k)}{(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2}. \quad (11.2.15)$$

Таким чином, система рівнянь (11.2.12), (11.2.13) перетворена в два різницевих рівняння другого порядку, що мають однакові зворотні характеристичні рівняння виду:

$$(1 - a_{11}L)(1 - a_{22}L) - a_{12}a_{21}L^2 = 0.$$

Розв'язуючи це рівняння відносно L , одержимо два корені, які можна перетворити у корені звичайного характеристичного рівняння таким чином: $\lambda = 1/L$. Тобто «пряме» характеристичне рівняння має вигляд:

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0. \quad (11.2.16)$$

Оскільки обидва рівняння (11.2.12) і (11.2.13) мають одне і те ж характеристичне рівняння, то рівняння (11.2.16) 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 (11.2.14), (11.2.15) 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 впливає з (11.2.12), (11.2.13), 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 (11.2.14) 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

$$(1 - L)y(k) = \Delta y(k) = \frac{1 - a_{22}L}{1 - \lambda_2 L} \varepsilon_x(k) + \frac{a_{12}L}{1 - \lambda_2 L} \varepsilon_y(k),$$

який є стаціонарним при $|\lambda_2| < 1$.

Таким чином, для того щоб забезпечити умову коінтегрованості змінних, тобто $CI(1,1)$, 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 (11.2.16)):

$$0,5(a_{11} + a_{22}) + 0,5 \sqrt{(a_{11}^2 + a_{22}^2) - 2a_{11}a_{22} + 4a_{12}a_{21}} = 1.$$

Після спрощення дане співвідношення приймає вигляд:

$$a_{11} = [(1 - a_{22}) - a_{12}a_{21}] / (1 - a_{22}). \quad (11.2.17)$$

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

$$a_{22} > -1, \quad (11.2.18)$$

і

$$a_{12}a_{21} + (a_{22})^2 < 1. \quad (11.2.19)$$

Умови (11.2.17) – (11.2.18) представляють собою обмеження, які необхідно накласти на коефіцієнти рівнянь (11.2.12) – (11.2.13) для досягнення коінтегрованості розглянутих змінних порядку $CI(1,1)$. Для того щоб визначити вплив цих обмежень на характер розв'язку рівнянь, перепишемо (11.2.12) – (11.2.13) у вигляді:

$$\begin{bmatrix} \Delta x(k) \\ \Delta y(k) \end{bmatrix} = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} x(k-1) \\ y(k-1) \end{bmatrix} + \begin{bmatrix} \varepsilon_x(k) \\ \varepsilon_y(k) \end{bmatrix}. \quad (11.2.20)$$

З (11.2.17) випливає, що $a_{11} - 1 = -a_{12}a_{21} / (1 - a_{22})$. Тому після деяких перетворень представимо (11.2.20) у формі:

$$\Delta x(k) = -[a_{12}a_{21} / (1 - a_{22})]x(k-1) + a_{12}y(k-1) + \varepsilon_x(k), \quad (11.2.21)$$

$$\Delta y(k) = a_{21}x(k-1) - (1 - a_{22})y(k-1) + \varepsilon_y(k). \quad (11.2.22)$$

Рівняння (11.2.21), (11.2.22) 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:

$$\begin{aligned}\Delta x(k) &= \alpha_x [x(k-1) - \beta y(k-1)] + \varepsilon_x(k), \\ \Delta y(k) &= \alpha_y [x(k-1) - \beta y(k-1)] + \varepsilon_y(k),\end{aligned}$$

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 (11.2.18), (11.2.19) 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 (11.2.7), (11.2.8), 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:

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.

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 (11.2.20) можна представити у вигляді:

$$\Delta \mathbf{z}(k) = \mathbf{F} \mathbf{z}(k-1) + \varepsilon(k). \quad (11.2.23)$$

Очевидно, що оцінити векторну авторегресійну модель для коінтегрованих змінних, використовуючи тільки перші різниці, неможливо. Якщо у рівнянні (11.2.23) виключити з розгляду $\mathbf{F} \mathbf{z}(k-1)$, 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} дорівнює нулю, а змінні $x(k)$, $y(k)$ мають представлення у вигляді корекції похибки у вигляді (11.2.21), (11.2.22).

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 (11.2.16). 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 (11.2.12), (11.2.13), 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.

У 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 дорівнює 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.* 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 (11.2.9),

(11.2.10) послідовність $\{r_L(k)\}$ не впливає за Грейнджером на $\{r_S(k)\}$, якщо усі коефіцієнти $a_{12}(i) = 0$ і $\alpha_S = 0$.

11.3 Методика побудови моделі коінтегрованих процесів

Побудова моделі коінтегрованих процесів складається з декількох 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 визначенням 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:

$$\begin{aligned} \Delta^2 y(k) &= \Delta y(k) - \Delta y(k-1) = y(k) - y(k-1) - [y(k-1) - y(k-2)] = \\ &= y(k) - 2y(k-1) + y(k-2), \end{aligned}$$

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 визначення існування нестационарності (існування одиничних коренів) запропоновано ряд тестів.

Перевірка 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

$$y(k) = a_1 y(k-1) + \varepsilon(k),$$

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 ступінь 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'язку задачі оцінювання коефіцієнтів моделі часового ряду передбачає скінченність дисперсії залишків (похибок) $e(k)$ моделі. Наявність нестационарності призводить до порушення цього припущення. Наприклад, розглянемо рівняння

$$\begin{aligned} y(k) &= y(k-1) + e(k) = [y(k-2) - e(k-1)] + e(k) = \dots \\ &= y(0) + e(k) + e(k-1) + e(k-2) + \dots + e(1). \end{aligned}$$

Оскільки залишки $e(k)$ незалежні і мають постійну дисперсію, то дисперсія $y(k)$ зростає до нескінченності при $k \rightarrow \infty$. В такому випадку для описання динаміки ряду можна скористатись рівнянням

$$y(k) - y(k-1) = a_1 y(k-1) - y(k-1) + e(k)$$

або

$$\Delta y(k) = b y(k-1) + e(k),$$

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 описання тренда. Якщо включити середнє, то воно приймає вигляд:

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k)$$

або

$$\Delta y(k) = a_0 + a_1 y(k-1) - y(k-1) + \varepsilon(k) = a_0 + b y(k-1) + \varepsilon(k).$$

Із врахуванням тренду останнє рівняння приймає вигляд:

$$y(k) = a_0 + a_1 k + a_2 y(k-1) + \varepsilon(k),$$

де k – дискретний час. Це рівняння можна записати для першої різниці

$$y(k) - y(k-1) = a_0 + a_1 k + a_1 y(k-1) - y(k-1) + \varepsilon(k)$$

або

$$\Delta y(k-1) = a_0 + a_1 k + b y(k-1) + \varepsilon(k).$$

Для такої моделі було б некоректно використовувати 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, тобто 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:

$$\Delta y(k) = a_0 + by(k-1) + c_1 \Delta y(k-1) + c_2 \Delta y(k-2) + \dots + c_n \Delta y(k-n) + \varepsilon(k).$$

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

$$\Delta y(k) = b y(k-1) + c_1 \Delta y(k-1) + c_2 \Delta y(k-2) + \dots + c_n \Delta y(k-n) + \varepsilon(k),$$

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 [12, 67]. 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

$$\tau_\infty + \frac{\tau_1}{N} + \frac{\tau_2}{N^2},$$

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 Маккінноном в 1991 [67]).

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 наступних значеннях τ :

$\tau_{\infty} = -3,43$ ($\alpha = 1\%$) або $\tau_{\infty} = -2,86$ ($\alpha = 5\%$);

$\tau_1 = -6,00$ ($\alpha = 1\%$) або $\tau_1 = -2,74$ ($\alpha = 5\%$);

$\tau_2 = -29,25$ ($\alpha = 1\%$) або $\tau_2 = -8,36$ ($\alpha = 5\%$).

Нульова гіпотеза при наявності середнього та тренду

В даному випадку застосовується така ж процедура, що і вище, але при наступних значеннях τ :

$\tau_{\infty} = -3,96$ ($\alpha = 1\%$) або $\tau_{\infty} = -3,41$ ($\alpha = 5\%$);

$\tau_1 = -8,35$ ($\alpha = 1\%$) або $\tau_1 = -4,04$ ($\alpha = 5\%$);

$\tau_2 = -47,44$ ($\alpha = 1\%$) або $\tau_2 = -17,83$ ($\alpha = 5\%$).

Розширений тест Дікі-Фуллера

Для того щоб скористатись розширеним тестом ДФ, необхідно побудувати наступне рівняння регресії:

$$\Delta y(k) = a_0 + a_1 k + b y(k-1) + \sum_{i=1}^p c_i \Delta y(k-i) + \varepsilon(k), \quad (11.3.1)$$

де a_0, a_1, b, c_i – невідомі коефіцієнти регресії. Якщо всі коефіцієнти $c_i = 0, i = 1, 2, \dots, p$, то рівнянням (11.3.1) 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 [24] пропонується наступне просте евристичне правило: якщо візуально з графіка **не** можна зробити висновок про наявність тренду, то в модель (11.3.1) необхідно включати тільки константу (перетин) a_0 , навіть якщо значення коливаються навколо нуля. Якщо візуальний аналіз ряду свідчить про наявність тренду, то в модель (11.3.1) необхідно ввести a_0 і $a_1 k$.

За допомогою базової моделі (11.3.1) тестуються такі гіпотези:

$H_0 : b = 0$, або часовий ряд нестационарний: $\{y(k)\} \sim I(\text{int}), \text{int} > 0$;

$H_1 : b < 0$, або часовий ряд стаціонарний: $\{y(k)\} \sim I(0), \text{int} = 0$.

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:

$$|\tau| = \left| \frac{\hat{b}}{SE_{\hat{b}}} \right| \geq |\tau_{\text{крит}}|$$

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:

$$\Delta^2 y(k-1) = a_0 + b \Delta y(k-1) + \sum_{i=1}^p c_i \Delta^2 y(k-i) + \varepsilon(k). \quad (11.3.2)$$

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 (11.3.2) тестуються наступні гіпотези:

- H_0 : $b = 0$, або часовий ряд перших різниць нестационарний:
 $\{\Delta y(k)\} \sim I(\text{int})$, $\text{int} > 0$;
- 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:

$$e(k) = y(k) - \beta x(k),$$

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 декількох 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

$$[x(k) - \alpha - \beta y(k)] \sim I(0), \quad (11.3.3)$$

$$[y(k) - \alpha' - \beta' x(k)] \sim I(0), \quad (11.3.4)$$

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 (11.3.3) та (11.3.4):

$$e_y(k) = y(k) - \alpha - \beta x(k), \quad (11.3.5)$$

$$e_x(k) = x(k) - \alpha' - \beta' x(k). \quad (11.3.6)$$

Нехай обидва ряди $x(k)$ і $y(k)$ мають порядок інтегрованості 1. Базовим для виконання тесту Інгла-Грейнджера є регресійне рівняння виду:

$$\Delta e(k) = be(k-1) + \sum_{i=1}^p c_i \Delta e(k-i) + \varepsilon(k), \quad (11.3.7)$$

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 (11.3.5) або (11.3.6).

За допомогою моделі (11.3.7) тестуються наступні гіпотези:

– $H_0 : \hat{b} = 0$, або часові ряди $x(k)$ і $y(k)$ не коінтегровані; при цьому

$$e(k) \sim I(1);$$

– $H_1 : \hat{b} < 0$, або часові ряди $x(k)$ і $y(k)$ коінтегровані; при цьому

$$e(k) \sim I(0).$$

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:

$$\Delta x(k) = a_{10} + \sum_{i=1}^p b_{1i} \Delta x(k-i) + \sum_{i=1}^p c_{1i} \Delta y(k-i) + \lambda_1 e(k-1) + \varepsilon_1(k), \quad (11.3.8)$$

$$\Delta y(k) = a_{20} + \sum_{i=1}^p b_{2i} \Delta y(k-i) + \sum_{i=1}^p c_{2i} \Delta x(k-i) + \lambda_2 e_x(k-1) + \varepsilon_2(k), \quad (11.3.9)$$

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 $\lambda_1 < 0, \lambda_2 < 0$) корегується в інших періодах. Необхідними умовами забезпечення динамічної стійкості моделей (11.3.8) та (11.3.9) є наступні: $-1 \leq \lambda_1 \leq 0, -1 \leq \lambda_2 \leq 0$.

Модель (11.3.8), (11.3.9) завжди коректна завдяки тому, що попередні етапи її побудови забезпечують виконання припущень щодо стаціонарності її змінних, тобто

$$\Delta x(k) \sim I(0), \quad \Delta y(k) \sim I(0), \quad e_x(k-1) \sim I(0), \quad e_y(k-1) \sim I(0), \\ \varepsilon_1(k) \sim I(0), \quad \varepsilon_2(k) \sim I(0) \text{ за припущенням.}$$

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%. У окремих випадках суб’єкти господарської діяльності можуть звільнитись 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 реальні 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 реальні 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 одинадцятого розділу

1. Які процеси називають інтегрованими? Що означає порядок інтегрованості? Яке формально позначають інтегрованість та коінтегрованість?
2. Чи є відмінності в термінах: „інтегровані процеси”, „процеси з трендом”, нестационарні процеси” та „процеси з одиничними коренями”? Якщо є, то поясніть, які конкретно?
3. Якими методами можна вилучити тренд з процесу? Чи є ці методи повністю рівнозначними за кінцевим результатом?
4. Знайдіть другі різниці для 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 від початкової моделі?
5. Що означає коінтегрованість процесів? Яким умовам повинні задовольняти коінтегровані процеси? Наведіть приклад макроекономічних процесів (або процесів іншої природи), які можуть бути коінтегрованими?
6. Якщо процеси мають однаковий порядок інтегрованості, чи обов'язково вони будуть коінтегрованими?
7. Поясніть суть тестування на інтегрованість за Дікі-Фулерам? В чому полягає відмінність розширеного тесту Дікі-Фулера від звичайного?
8. Яку статистику використовують при тестуванні на інтегрованість? Чи можна використати для цієї мети t – статистику Стьюдента?
9. Поясніть послідовність процедури побудови моделі коінтегрованих процесів? Яку роль відіграє модель корегування похибок? Конкретизуйте, про які похибки йде мова?
10. Яка мета побудови моделі коінтегрованих процесів, які складові включає ця модель? Поясніть на прикладі.
11. Поясніть, яким чином можна використати модель коінтегрованих процесів для створення системи керування процесами, які моделюються?

12. 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 $x(k)$?
13. Сформулюйте задачу імітаційного моделювання коінтегрованих процесів за допомогою лінійної комбінації випадкових процесів? Поясніть існування можливості побудови такої моделі?
14. В чому полягає мета побудови моделі корегування похибок? Яка її принципова відмінність від звичайної регресії?
15. Поясніть на прикладі можливість використання рангу матриці F для встановлення факту коінтегрованості інтегрованих послідовностей $x(k)$, $y(k)$.
16. Побудуйте модель коінтеграції для валового внутрішнього продукту України та іншого вибраного вами макроекономічного процесу (скористайтесь щомісячними даними за період 01.1995 – 12.2005). Поясніть важливість наявності коінтегрованості між макроекономічними процесами перехідного періоду.
17. Які елементи повинна включати система підтримки прийняття рішень для моделювання коінтегрованих процесів?

Р о з д і л 12

ВСТУП ДО ЙМОВІРНІСНОГО АНАЛІЗУ ДАНИХ

Теорія ймовірностей і прикладна статистика – це мистецтво і наука збору, обробки даних і аналізу отриманих результатів з метою прийняття коректних обґрунтованих рішень технічного, ділового, політичного персонального або іншого характеру.

Ймовірнісно-статистичні методи необхідно розглядати як важливу трудомістку частину процесу прийняття рішень, яка дозволяє приймати обґрунтовані тактичні та стратегічні рішення. Ці методи ґрунтуються на глибокому аналізі наявної інформації та інтуїції фахівців.

Коректне використання ймовірнісно-статистичних методів надає значні переваги стосовно прийняття коректних рішень практично в усіх напрямках людської діяльності: в конкуренції за підвищення якості та продаж нової продукції, дає можливість отримати високоякісні оцінки прогнозів, обґрунтувати фінансові, макроекономічні, а також рішення стосовно раціонального ведення домашнього господарства і розв'язати багато інших задач.

12.1 Задачі, які розв'язують ймовірнісно-статистичними методами

- Виявлення причинно-наслідкових зв'язків у функціонуванні процесів різної природи з метою побудови коректних математичних моделей.

- Аналіз і діагностика стану об'єктів різної природи і прийняття рішення стосовно коректності їх функціонування. Технічні системи, технологічні процеси, медична діагностика.
- Прогнозування і управління соціально-економічними процесами і системами (регіональний, галузевий та державний рівні). Виявлення циклів в економіці та бізнесі. – Динамічні БМ.
- Формування сценаріїв розвитку соціально-економічних процесів і систем. Використання побудованих сценаріїв для планування розвитку виробництва та різних видів бізнесу.
- Виявлення катастрофічних явищ у природі, технічних системах, економіці.
- Автоматичне керування технологічними процесами і технічними системами: – побудова адекватних моделей, застосування методів статистичного керування, аналіз якості керування за допомогою статистичних параметрів.
- Ймовірно-статистична підтримка прийняття експертних рішень.
- Підтримка прийняття особистих рішень та рішень стосовно ведення домашнього господарства.
- Контроль стану навколишнього середовища: три типи моделей: мережні моделі (нейромережі та мережі Байєса); моделі у вигляді диференціальних рівнянь в частинних похідних; рівняння AP, APКС, множинна регресія.
- Аналіз даних і прийняття рішень в генетиці, біології (існує напрям досліджень – біостатистика), психології (наприклад, видається сучасний журнал: *Journal of Experimental Psychology* в США).
- Розпізнавання образів.

Ймовірно-статистичні дослідження корисні особливо тоді, коли є великі групи людей, підприємств або інших об'єктів чи процесів (*генеральна сукупність*), які необхідно дослідити, але повне дослідження провести неможливо. Для того щоб отримати неідеальне, але практично корисне розуміння ситуації стосовно генеральної сукупності, можна відібрати невелику групу (*вибірку*) даних, яка складається із деяких, але не всіх, об'єктів генеральної сукупності.

Процес узагальнення результатів дослідження на всю генеральну сукупність називають імовірнісним або статистичним висновком або рішенням.

Випадкова вибірка з генеральної сукупності є одним із кращих способів накопичення даних, оскільки генеральна сукупність, як правило, занадто велика, щоб її вивчати повністю.

Основою для прийняття рішення у ймовірно-статистичних

дослідженнях є математична модель досліджуваного процесу, яка характеризується своєю структурою. Коректна оцінка структури дає можливість прийняти коректне рішення на основі такої моделі.

Поняття структури математичної моделі включає в себе такі елементи:

- вимірність моделі;
- порядок моделі;
- нелінійність та її тип (нелінійність стосовно і не лінійність стосовно параметрів);
- час запізнення та його оцінка;
- збурення та його тип (детерміноване або випадкове);
- обмеження на змінні і параметри моделі.

12.2 Можливі помилки при використанні ймовірнісно-статистичних методів

Ймовірнісно-статистичні методи – *надзвичайно потужний інструмент аналізу даних різної природи і їх силу можна використати коректно або некоректно*. Є багато прикладів некоректного застосування статистичних методів, що проявилось у формулюванні висновків, протилежних до дійсності.

Наприклад, у 1936 році в США була допущена помилка журналом “*Literary Digest*”, який помилково визначив вибірку даних. Журнал виконав телефонне *опитування своїх передплатників* і спрогнозував, що *Альфред Лендон* легко перемає свого суперника *Франкліна Рузвельта* на виборах президента США. Однак, Рузвельт переміг Лендона в 46 із 48 штатів. Справа в тому, що 1936 рік – це один з років економічної депресії і дозволити собі телефон і передплату журналу могли тільки люди з хорошим матеріальним станом. Тобто вибірка, визначена журналом, не представляла всіх прошарків виборців у США – вона *не була представницькою*. Передплатники журналу у своїй більшості збирались голосувати за Лендона, в той час як основна частина генеральної сукупності була за Рузвельта. Після виборів журнал “*Literary Digest*” швидко втратив свою популярність і досить швидко перестав існувати.

Звичайно, що можна навести багато інших прикладів некоректного застосування ймовірнісно-статистичних методів і не тільки в минулому, а й сьогодні. На жаль, бувають випадки, коли некоректне застосування цих методів зумовлене не тільки поганими знаннями, але й цілеспрямованим навмисним формуванням некоректних висновків на основі зібраних даних.

Очевидно, що майже всі статистичні процедури можна виконати некоректно як в результаті незнання, так і навмисно. Однак, як свідчить практика, навмисне спотворення методів статистичного висновку приводить тільки до негативних результатів. Якщо ми будемо стверджувати, що зростання ВВП України складає 13% на рік, а фактичне зростання складає 5%, то добробут населення і довір'я до влади від цього не підвищаться.

Зазначимо, що невірні дані і некоректно виконаний ймовірно-статистичний аналіз можуть призвести до прийняття принципово неправильних тактичних і стратегічних рішень з катастрофічними наслідками.

Випадки неправильного використання звичайних статистичних методів можна знайти в американській книзі “*Як обманювати за допомогою статистики*”.

Іноді говорять, що за допомогою статистики можна довести, що завгодно. Однак, якраз це “що завгодно” пов’язане з випадками некоректного застосування ймовірно-статистичних методів. Деякі компанії виконують дослідження, за допомогою яких доводять, що їхній продукт переважає за своїми якостями аналогічні продукти, що випускаються конкуруючими фірмами. Очевидно, що висновки будь-якого дослідження можна підігнати під бажаний результат за рахунок зміни техніки збору даних або неправильної інтерпретації встановлених залежностей. У повсякденному житті кожний з нас зустрічається із статистичними даними та висновками, що знаходяться у протиріччі між собою. Тому немає нічого дивного в тому, що багато людей висловлюють недовіру до статистики.

Однак, у випадках зловживання ймовірно-статистичними методами *винні не цифри і не використані методи, а ті виконавці*, які мають схильність обманювати суспільство за їх допомогою. Чесні наміри і знання статистичних методів завжди дають корисні змістовні результати.

Іноді можна зустріти заперечення проти застосування ймовірно-статистичних методів у зв’язку з тим, що отримувані за ними висновки носять узагальнений характер і не відображають індивідуальних якостей людей. Так, висновок стосовно того, що індивідууми з вищою освітою матимуть деякий рівень середньої зарплати, не сприймається деякими людьми. Вони вважають, що їх це не стосується і що висновок невірний. Так само як і у випадку голосування за деякого визначеного кандидата в президенти не всі представники однієї соціальної групи голосуватимуть за одного кандидата.

З цього приводу можна сказати, що ймовірно-статистичні висновки є узагальнюючими, але вони *не є абсолютними*. Загальне судження ніколи не залишається незмінним. Якщо ми будемо наводити приклади, які будуть знаходитись у протиріччі з висновком, то тим самим будемо надавати

висновку *форму абсолютного судження*. Розв'язання цієї проблеми полягає в тому, що кожний індивідуум повинен сприймати справедливість ймовірнісно-статистичного судження, але в той же час повинен розуміти, що ніхто не забирає в нього свободи вибору за допомогою цього судження.

Загалом можна сказати, *що основною метою застосування ймовірнісно-статистичного аналізу є обґрунтоване прийняття рішень на основі аналізу даних та експертних оцінок*. Якщо ми хочемо, щоб рішення були об'єктивними і якісними, то необхідно знати і вміти користуватись методами ймовірнісно-статистичного аналізу та іншими методами прийняття рішень, що доповнюють їх. Доповнюючими методами є *методи оптимізації, чітка і нечітка логіка, нейромережі* та інші методи. Інтегроване (спільне) застосування декількох методів прийняття рішень значно підвищує ймовірність отримання і вибір самих якісних рішень з множини можливих альтернатив.

12.3 Ймовірнісні та детерміновані моделі

Моделювання процесів різної природи пов'язане із невизначеностями структурного, параметричного і статистичного характеру. Велика перевага ймовірнісно-статистичного підходу до аналізу даних і прийняття рішень полягає у тому, що він дає можливість коректно боротись із такими невизначеностями, тобто враховувати їх при побудові моделей та формуванні альтернативних розв'язків задач.

Структурні невизначеності пов'язані з неможливістю (у багатьох випадках) знайти прийнятну за точністю оцінку структури моделі внаслідок відсутності необхідної інформації. Тобто оцінити з належною точністю ті елементи, які входять у поняття структури математичної моделі.

Параметричні невизначеності зумовлені труднощами отримання незміщених, консистентних і ефективних оцінок параметрів (коефіцієнтів) моделі в умовах низької інформативності даних, наявності пропусків, екстремальних значень та похибок (шумів) вимірів.

Невизначеності статистичного характеру зумовлені відсутністю належної інформації (статистичних характеристик) стосовно збурюючих впливів на стан об'єкта, шумів вимірів, відсутністю належної інформації стосовно розподілів інших випадкових змінних, які є у моделі.

Очевидно, що всі три зазначені типи невизначеностей взаємопов'язані між собою. Так, в усіх трьох випадках велику роль стосовно величини похибок оцінок структури і параметрів моделі відіграє неповна інформація

стосовно початкового і поточного стану досліджуваного процесу. Коректний підхід до побудови моделі потребує мінімізації впливу невизначеностей усіх типів на якість результуючої моделі.

У подальшому викладенні матеріалу будемо характеризувати досліджувані об'єкти і процеси множиною *кількісних* (числових) або *якісних* (символьних, лінгвістичних) змінних:

$$\mathbf{x} = \{ X_1, X_2, \dots, X_n \},$$

де кожна змінна X_j має свою область визначення D_j .

Детермінована модель представляє собою детерміновану функцію або множину правил стосовно взаємного зв'язку між змінними $X_i, i=1, \dots, n$. Наприклад, моделлю ідеального гармонічного коливання може бути така:

$$y_1(t) = A_1 \sin(\omega_1 t + \varphi_1), \quad (12.3.1)$$

або

$$y_2(t) = A_2 \cos(\omega_2 t + \varphi_2). \quad (12.3.2)$$

Прикладом детермінованого правила може бути таке:

$$\text{якщо } X_1 = \text{"людина"}, \text{ то } X_2 = \text{"розумна"}, \quad (12.3.3)$$

де X_1 – змінна, яка відноситься до множини живих організмів; X_2 – змінна, яка характеризує когнітивні можливості деякого організму (тобто його здатність до мислення).

Якщо кількість змінних і кількість значень, які вони можуть приймати, великі, то об'єкт може перебувати у одному із множини його можливих станів.

Ймовірнісне моделювання передбачає, що кожний стан (або множина станів) об'єкта може наступити із деякою ймовірністю. При використанні такого підходу кожний стан характеризується спільним розподілом ймовірностей (СРЙ) множини змінних

$$\mathbf{x} = \{ X_1, X_2, \dots, X_n \},$$

який позначимо так: $P(X_1, X_2, \dots, X_n)$. Областю визначення функції спільного розподілу ймовірностей є Декартів добуток:

$$D_1 \times D_2 \times \dots \times D_n,$$

а значення, які може приймати СРЙ, знаходяться в інтервалі $[0; 1]$.

Переваги ймовірнісного моделювання

1. Ймовірнісні моделі точніше представляють функціонування об'єктів і процесів у будь-яких галузях нашої діяльності. Стосовно можливостей опису станів досліджуваного об'єкта ймовірнісне моделювання набагато потужніше ніж детерміноване. Детерміновану модель можна розглядати як окремий випадок ймовірнісної. Наприклад, можна стверджувати, що

$$\text{«Миші сірі»}, \quad (12.3.4)$$

або

$$\text{«}x\% \text{ мишей сірі»}, \quad (12.3.5)$$

де $x \rightarrow 100$. Вербальна модель (12.3.4) – це окремий випадок моделі (12.3.5), оскільки *остання децю складніша і має ймовірнісний характер*. Модель (12.3.5) можна записати у формалізованому вигляді з використанням умовної ймовірності так:

$$p(x_2 = \text{«сірий»} / x_1 = \text{«мишка»}) = y. \quad (12.3.6)$$

У даному випадку імовірнісна модель точніше представляє колір мишей, оскільки у природі є білі миші, але їх кількість дуже незначна.

Іншим прикладом може бути детермінована і стохастична модель у просторі станів, яка широко використовується у проектуванні систем керування:

$$\mathbf{x}(k) = \mathbf{A}(k, k-1)\mathbf{x}(k-1) + \mathbf{B}(k, k-1)\mathbf{u}(k-1) + \mathbf{w}(k-1), \quad (12.3.7)$$

де $\mathbf{x}(k)$ – n -вимірний вектор станів системи; $\mathbf{u}(k-1)$ – m -вимірний вектор детермінованих вхідних величин (сигнали керування); $\mathbf{w}(k-1)$ – n -вимірний вектор випадкових зовнішніх збурень; $\mathbf{A}(k, k-1)$ – $(n \times n)$ матриця динаміки системи (вона містить коефіцієнти, що характеризують динаміку, тобто швидкість зміни станів у часі); $\mathbf{B}(k, k-1)$ – $(n \times m)$ матриця коефіцієнтів керування. Подвійний часовий аргумент у вигляді $(k, k-1)$ означає, що величина з цим аргументом використовується в момент k , але її значення ґрунтується на попередніх даних, які відомі на момент $k-1$ включно. Далі будемо записувати для простоти матриці \mathbf{A} і \mathbf{B} з одним аргументом, тобто $\mathbf{A}(k)$ та $\mathbf{B}(k)$. Очевидно, що стаціонарна система описується матрицями з постійними коефіцієнтами, які записують просто \mathbf{A} і \mathbf{B} .

У класичній постановці задачі оптимального оцінювання і керування

послідовність зовнішніх збурень $\mathbf{w}(k)$ задовольняє властивостям білого гаусового шуму з нульовим середнім значенням і коваріаційною матрицею \mathbf{Q} , тобто статистики шуму мають вигляд:

$$\begin{aligned} E[\mathbf{w}(k)] &= 0, \quad \forall k, \\ E[\mathbf{w}(k)\mathbf{w}^T(j)] &= \mathbf{Q}(k)\delta_{kj}, \end{aligned} \quad (12.3.8)$$

де δ_{kj} – дельта-функція Кронекера, що визначається як $\delta_{kj} = \begin{cases} 0 & \text{для } k \neq j \\ 1 & \text{для } k = j \end{cases}$;

$\mathbf{Q}(k)$ – додатно визначена коваріаційна матриця зовнішніх збурень розмірності $(n \times n)$. Діагональні елементи матриці представляють собою дисперсії компонент вектора збурень $\mathbf{w}(k)$. Очевидно, що для застосування моделі (12.3.7) необхідно знати параметри нормального розподілу процесу $\mathbf{w}(k)$.

Окремим випадком моделі (12.3.7) є спрощена детермінована модель у просторі станів:

$$\mathbf{x}(k) = \mathbf{A}(k, k-1)\mathbf{x}(k-1) + \mathbf{B}(k, k-1)\mathbf{u}(k-1). \quad (12.3.9)$$

Вона простіша, але і грубіша одночасно, оскільки не враховує важливі зовнішні впливи на процес.

2. Друга перевага імовірнісної моделі полягає у тому, що її побудова вимагає систематичного дослідження усіх можливих значень кожної змінної, тобто кожної підмножини D_j , а також усіх можливих комбінацій змінних (конфігурацій об'єкта). Тобто необхідно розглядати кожний елемент з декартового добутку:

$$D_1 \times D_2 \times \dots \times D_n.$$

Такий підхід зменшує вплив на процес побудови моделі факторів, пов'язаних із культурними і психологічними особливостями індивідуума, а також зменшується ризик залишити поза увагою деякі аспекти функціонування об'єкта.

Таким чином, імовірнісний підхід до моделювання надає можливість максимально точного представлення функцій об'єкта, оскільки при цьому розглядаються всі його теоретично можливі стани (конфігурації). Кожному стану ставиться у відповідність один елемент із нескінченної множини значень в інтервалі $[0; 1]$.

3. Третім позитивним фактором використання ймовірнісних моделей є те, що їх застосування досить часто є простішим ніж детермінованих.

Приклад 12.1. При виборі майбутньої спеціальності випускникам шкіл бажано мати інформацію стосовно ймовірності подальшого працевлаштування. У такому випадку не обов'язково потрібно займатись збором відповідної статистики, що вимагає багато часу. Можна звернутись до декількох випускників вибраної спеціальності і отримати потрібну інформацію. Вона може бути отримана у вигляді:

„Близько 80% випускників даної спеціальності знаходять роботу відразу після закінчення університету.”

Звичайно, що це наближена інформація, але її може виявитись цілком достатньо для прийняття рішення абітурієнтом.

Недоліки імовірнісного моделювання

1. У багатьох випадках **потрібно мати апріорну інформацію про процес** у вигляді оцінок початкового стану, типів розподілів змінних, попередніх оцінок параметрів розподілів і т. ін. Очевидно, що ця задача не завжди розв'язується просто і швидко.

2. Основні труднощі виникають із **побудовою та використанням спільних розподілів ймовірностей** для змінних, що характеризують поведінку процесу. При цьому виникають такі проблеми:

- Графічне представлення СРЙ можливе тільки у тих випадках, коли кількість змінних не перевищує 2. Приклад представлення СРЙ у самому простому випадку ($n=2; m=2$, де n – кількість змінних; m – кількість значень, які можуть приймати змінні X_1 і X_2) наведено у табл. 12.1.
- Використання СРЙ призводить до виникнення проблеми комбінаторного „вибуху” – великої кількості комбінацій значень змінних. Так, якщо кожна змінна приймає m різних значень ($m \geq 1$), то спільний розподіл ймовірностей описується m^n потенціально можливими конфігураціями (станами) об'єкта. Наприклад, якщо $n=m=10$, то отримаємо 10^{10} (10 мільярдів) значень.

Спільний розподіл ймовірностей для $n=m=2$

X_1, X_2	$x_{2,1}$	$x_{2,2}$
$x_{1,1}$	$p(X_1 = x_{1,1}; X_2 = x_{2,1})$	$p(X_1 = x_{1,1}; X_2 = x_{2,2})$
$x_{1,2}$	$p(X_1 = x_{1,2}; X_2 = x_{2,1})$	$p(X_1 = x_{1,2}; X_2 = x_{2,2})$

12.4 Безумовна незалежність подій

Згідно з пропозицією Декарта, множину n змінних можна розділити на декілька підмножин меншого розміру і аналізувати окремо.

Припустимо, наприклад, що множину n змінних можна розділити на дві підмножини розмірністю j і $(n-j)$ змінних, відповідно:

$$p(X_1, X_2, \dots, X_n) = p(X_1, \dots, X_j) \cdot p(X_{j+1}, \dots, X_n). \quad (12.4.1)$$

Тепер задача моделювання спрощується. Спочатку можна обчислити спільну ймовірність для підмножини змінних X_1, \dots, X_j , а потім для підмножини X_{j+1}, \dots, X_n . Остаточний результат отримаємо за виразом (12.4.1).

Рівність (12.4.1) означає, що підмножини X_1, \dots, X_j і X_{j+1}, \dots, X_n – незалежні або *безумовно незалежні*. Це означає, що будь-яка інформація стосовно підмножини X_1, \dots, X_j , наприклад, $X_1 = 7$ або $X_1 + X_2 > 5$, не змінить розподіл ймовірностей другої підмножини X_{j+1}, \dots, X_n .

Однак, для реальних об'єктів і процесів безумовна незалежність цих двох підмножин малоймовірна. Якщо така ситуація має місце, то це означає, що початкова постановка задачі стосовно моделювання такого об'єкта була некоректною. У такому випадку є сенс будувати дві моделі, тобто описувати математично кожену множину окремо.

12.5 Правила обчислення ймовірностей

Ймовірність – це ступінь нашої віри у те, що деяка подія відбудеться.

Існує також поняття „свідчення” стосовно попереднього твердження; свідчення проявляється (знаходить своє вираження) через інші твердження. Приклади розглянемо нижче.

Вибірковий простір експерименту

Множину можливих результатів деякого експерименту (або послідовності подій) називають *вибірковим простором* цього експерименту. Будемо припускати, що вибіркового простору досліджуваного експерименту містить всі можливі результати цього експерименту і що кожна пара результатів представляє собою взаємовиключні події. Такі припущення гарантують, що експеримент закінчиться точно одним із визначених результатів у вибіркового простору.

Наприклад, при підкиданні костей вибіркового простору буде такий:

$$S = \{1, 2, 3, 4, 5, 6\},$$

а для опису результату футбольного матчу можна скористатись таким простором:

$$S = \{winner, loser\}.$$

Підмножина вибіркового простору складається із однієї або декількох *подій*. Наприклад, подію, пов'язану з тим, що ми отримаємо значення три («3») або вище при використанні 6-гранних костей, відповідає такій підмножині:

$$\{3, 4, 5, 6\} \subseteq \{1, 2, 3, 4, 5, 6\}.$$

Така подія матиме місце якщо в результаті підкидання кості отримаємо елемент, який належить вказаній підмножині.

В загальному випадку ми говоримо, що подія $A = \text{''істина''}$, тобто мала місце у деякому експерименті, якщо результат експерименту є елементом події A . Якщо подія містить тільки один елемент, то його також вважають результатом.

Для того щоб виміряти ступінь невизначеності стосовно результату експерименту використовують ймовірність події $p(A)$, $\forall A \subseteq S$. Ці ймовірності повинні задовольняти трьом аксіомам.

Аксіома 1. *Ймовірність того, що результат деякого експерименту буде належати його вибіркового простору дорівнює одиниці: $p(S) = 1$.*

Іншими словами: ступінь віри (довіри) у те, що деяка подія A відбудеться, характеризується дійсним числом в інтервалі від нуля до одиниці:

$$0 \leq p(A) \leq 1. \quad (12.5.1)$$

Аксіома 2. Будь-яка подія A повинна мати невід'ємну ймовірність: *Для всіх $A \subseteq S$ виконується умова, що $p(A) \geq 0$.*

Аксиома 3. Якщо дві події A і B не перетинаються (рис. 12.1а), тобто не залежать одна від одної, то ймовірність комбінованої події визначається сумою ймовірностей двох окремих подій:

Якщо $A \subseteq S$, $B \subseteq S$ і $A \cap B = \emptyset$, то $p(A \cup B) = p(A) + p(B)$.

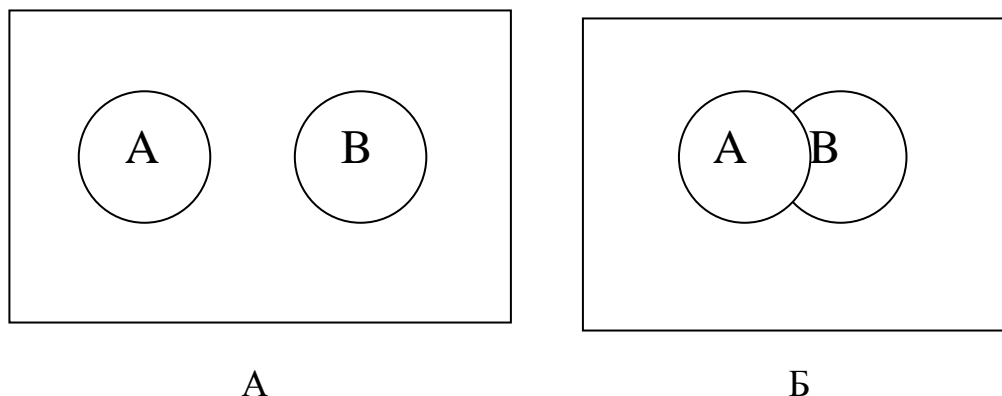


Рис. 12.1 Незв'язані події (А) і зв'язані події (Б)

Наприклад, якщо подія B стосується числа 3 на грані кості, тобто $B = \{3\}$; а подія A пов'язана з тим, що випаде парне число, тобто: $A = \{2, 4, 6\}$, то події A і B – не перетинаються (незалежні). А тому ймовірність настання однієї з цих подій буде такою:

$$p(A \cup B) = p(A) + p(B) = \frac{1}{6} + \frac{3}{6} = \frac{4}{6}.$$

Якщо події перетинаються (рис. 12.1б), можна легко показати, що

$$p(A \cup B) = p(A) + p(B) - p(A \cap B), \quad (12.5.2)$$

де $A \cap B$ – перетин подій A і B , який представляє подію, яка означає *появу обох подій* A і B .

Приклад 12.2. Розглянемо приклад з гральними картами (52 карти у колоді). Позначимо через A подію, яка означає, що із колоди витягли будь-яку карту *пікової* масті, а через B – подію, яка означає, що із колоди витягли короля (будь-якої масті). Очевидно, що ці події залежні (пов'язані між собою), оскільки їх перетин означає, що ми витягли *пікового короля*, тобто

$$A \cap B = \{\text{піковий король}\}.$$

Таким чином, ймовірність того, що ми витягнемо з колоди короля або будь-яку карту *пікової* масті складе:

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}.$$

(Пояснення: у колоді з 52 карт міститься 13 карт пікової масті (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, валет, дама, король і туз), 4 королі і 1 піковий король.)

Примітка: Іноді ми будемо підкреслювати, що значення ймовірності знайдено частотним методом (а не експертним суб'єктивним); таке значення ймовірності будемо позначати через $p^\#$. Якщо подія A описується тільки одним результатом a , то ми будемо використовувати позначення $p(a)$, а у інших випадках $p(\{a\})$.

Правило додавання ймовірностей можна поширити на випадок будь-якої кількості *взаємовиключних* подій A_1, A_2, \dots, A_n :

$$p(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_n) = p(A_1) + p(A_2) + \dots + p(A_n). \quad (12.5.3)$$

Взаємовиключними є подія A та її логічне заперечення \bar{A} , які утворюють повну множину станів для деякого об'єкта із двома станами:

$$p(\bar{A}) = 1 - p(A).$$

Наприклад, це справедливо для підкидання монети із двома можливими результатами: *орел і решка*. Третій стан – балансування на боці – практично неможливий. Інший приклад утворюють стани електричної лампочки: *робочий* (нормальний), коли лампочка може давати світло під дією електричного струму; і *неробочий* (спіраль перегоріла).

При оцінюванні ймовірностей подій завжди використовують додаткову інформацію, навіть якщо вона не задана у явному вигляді. Наприклад, нехай відомо, що деяка монета завжди падає орлом вгору після підкидання. Очевидно, що у даному випадку існує додаткова інформація (або *свідчення*) про те, що монета незбалансована, а тому завжди падає на один і той же бік.

Необхідно розрізнити „свідчення” і „додаткову (так звану *фонову*) інформацію”, оскільки деякі твердження необхідно розглядати як свідчення, а деякі – як додаткову інформацію.

Наприклад, нехай виконується аналіз дезоксирибонуклеїнової кислоти (ДНК). Припустимо, що методика аналізу відповідає теоретичним вимогам і її повністю дотримано у процесі аналізу. Всі ці *припущення складають додаткову інформацію* або її частину, оскільки це обов'язкові елементи виконання аналізу, яких необхідно притримуватись.

Свідчення, які мають відношення до виконання аналізу, можуть стосуватися лише опису результатів аналізу, а також статистичних даних, які характеризують надійність свідчень.

Додаткову інформацію можна ввести у означення умовної ймовірності; так, позначення $p(A|I)$ означає:

– „ймовірність виникнення події A (тобто $A=істина$) за умови наявності додаткової інформації I ”.

А позначення $p(A|B,I)$ означає:

– „ймовірність виникнення події A за умови, що відбулась подія B і за наявності додаткової інформації I ”.

При цьому A не залежить від B тільки у тому випадку, якщо

$$p(A|B,I) \neq p(A|I). \quad (12.5.4)$$

Поняття ймовірнісної залежності

Існує також поняття „ймовірнісної залежності” між подіями та „ймовірнісної незалежності”. Це практично неформалізовані поняття і наявність того чи іншого випадку необхідно розглядати у кожному випадку окремо.

Наприклад, нехай монетка підкидається два рази. Чи будуть пов’язані між собою результати першого і другого підкидання? Іншими словами, чи впливає на визначення ймовірності результату другого підкидання знання стосовно результату першого підкидання?

Очевидно, що у випадку гарно збалансованої монетки відповіддю буде „ні”. Однак, якщо невідомо, що монета добре збалансована, то відповіддю буде „так”.

Звичайно, що одного підкидання недостатньо для встановлення взаємної залежності (чи незалежності) результатів. Однак, якщо розглянути множину дослідів, то можна встановити різницю між двома ситуаціями – збалансована і незбалансована монета. Якщо невідомо, чи монета добре збалансована, то можна виконати ряд дослідів для того щоб встановити наскільки якісно виготовлена монета. Результати підкидання не будуть у даному випадку незалежними.

12.6 Умовні ймовірності, теорема Байєса

Коли ми висловлюємо твердження стосовно ймовірності події A , тобто $p(A)$, то неявно висловлюємо умови, пов’язані з іншими відомими факторами. Наприклад, „ймовірність того, що при підкиданні кості випаде число 6 дорівнює $\frac{1}{6}$ ”, як правило, передбачає, що кості виготовлені правильно (збалансовані). Тобто попереднє твердження має починатись так: „за умови, що кості збалансовані, ймовірність того ...”. Ймовірності такого

типу називають умовними, а загальна форма твердження у таких випадках має вигляд:

– „За умови відомої (ймовірності) події B , ймовірність події A складає p .”

Твердження такого типу формально позначаються так: $p(A|B)=c$. Однак, необхідно зазначити, що запис $p(A|B)=c$ не означає, що у всіх випадках, коли $B=$ „істина”, то ймовірність A дорівнює c . Це означає тільки те, що коли B настає і більше нічого не впливає на A , тоді ймовірність події A дорівнює c .

Властивість 12.1 (умовна ймовірність). Якщо настає (відома) подія B , то результат (експерименту) визначається елементами підмножини B , а результати, які відповідають події A , визначаються перетином: $A \cap B$. Тобто, якщо відома інформація про B , то необхідно визначити ймовірність $A \cap B$. Це дає можливість сформулювати властивість умовної ймовірності таким чином:

$$p(A|B) = \frac{p(A \cap B)}{p(B)}, \quad p(B) > 0.$$

Наприклад, умовна ймовірність того, що при підкиданні кості ми отримаємо 4, за умови, що ми отримуємо парне число, визначається так:

$$p(A=\{4\}|B=\{2,4,6\}) = \frac{p(\{4\})}{p(\{2,4,6\})} = \frac{1/6}{3/6} = \frac{1}{3},$$

за умови, що кості виготовлені якісно (збалансовані).

Очевидно, що у розрахунках умовних ймовірностей умова може стосуватись більше однієї події, наприклад:

$$p(A|B \cap C) = \frac{p(A \cap B \cap C)}{p(B \cap C)}.$$

Останній вираз можна переписати так, щоб отримати фундаментальне правило обчислення (умовних) ймовірностей.

Теорема 12.1 (фундаментальне правило):

$$p(A|B) \cdot p(B) = p(A \cap B). \quad (12.6.1)$$

Ця формула інтерпретується так: ймовірність того, що настануть обидві події A і B , визначається ймовірністю A за умови B і ймовірністю B .

Введення умови стосовно ще однієї події приводить до виразу:

$$p(A|B \cap C) \cdot p(B|C) = p(A \cap B|C).$$

Оскільки $p(A \cap B) = p(B \cap A)$, а також $p(A \cap B|C) = p(B \cap A|C)$, то на основі (12.6.1) можна записати:

$$p(A|B) \cdot p(B) = p(A \cap B) = p(B|A) \cdot p(A).$$

Це дає можливість записати правило (теорему) Байєса.

Теорема 12.2 (правило Байєса):

$$p(A|B) = \frac{p(B|A) p(A)}{p(B)}. \quad (12.6.2)$$

Правило Байєса – це метод оновлення ймовірності події A за умови, що наявна інформація стосовно події B . З цієї причини введено такі пояснення:

- $p(A)$ – називають апіорною ймовірністю події A ;
- $p(A|B)$ – апостеріорна ймовірність події A за умови, що мала місце подія B ;
- ймовірність $p(B|A)$ називають *правдоподібністю події A за умови (настання) події B* (пояснюючий приклад стосовно цієї назви наведемо нижче).

Таким чином, теорему Байєса можна інтерпретувати так:

$$\text{Апостеріорна ймовірн.} \propto \text{Правдоподібність} \times \text{Апіорна ймовірн.},$$

де символ „ \propto ” означає пропорційність.

Правило Байєса у контексті наявності ще однієї події C можна записати так:

$$p(A|B, C) = \frac{p(B|A, C) p(A|C)}{p(B|C)}.$$

Приклад 12.3. Нехай дві хвороби a_1 і a_2 мають один і той же симптом b ; і нехай $p(b|a_1)=0,9$, а $p(b|a_2)=0,3$. Припустимо, що апіорні ймовірності для обох хвороб однакові: $p(a_1)=p(a_2)$.

Якщо з'являється симптом b , то за правилом Байєса:

$$p(a_1|b) = \frac{p(b|a_1) p(a_1)}{p(b)} = 0,9 \frac{p(a_1)}{p(b)};$$

$$p(a_2|b) = \frac{p(b|a_2) p(a_2)}{p(b)} = 0,3 \frac{p(a_2)}{p(b)}.$$

Незважаючи на те, що ми не можемо завершити обчислення апостеріорних ймовірностей, можна зробити висновок, що ймовірність хвороби a_1 у три рази перевищує ймовірність a_2 , за умови наявності симптому b .

Якщо a_1 і a_2 – це єдині причини виникнення симптому b , то, припускаючи, що ймовірність захворіти двома хворобами одночасно дорівнює нулю, можна записати, що:

$$p(a_1|b) + p(a_2|b) = 1,$$

а звідси:

$$0,9 \frac{p(a_1|b)}{p(b)} + 0,3 \frac{p(a_2|b)}{p(b)} = 1.$$

Тепер, якщо апіорні ймовірності захворювання першою і другою хворобами однакові, тобто: $p(a_1)=p(a_2)$, то

$$\frac{p(a_1|b)}{p(b)} = \frac{p(a_2|b)}{p(b)} \Rightarrow \frac{p(a_1|b)}{p(b)} (0,9 + 0,3) = 1,$$

а звідси маємо:

$$\frac{p(a_1|b)}{p(b)} = \frac{p(a_2|b)}{p(b)} = \frac{1}{0,9 + 0,3} = \frac{1}{1,2},$$

що дає можливість розрахувати ймовірності захворювання кожною хворобою за умови наявності симптому b :

$$p(a_1|b) = \frac{p(b|a_1) p(a_1)}{p(b)} = 0,9 \frac{p(a_1)}{p(b)} \approx 0,9 \frac{p(a_1|b)}{p(b)} = 0,9 \cdot \frac{1}{1,2} = 0,75;$$

$$p(a_2|b) = \frac{p(b|a_2) p(a_2)}{p(b)} = 0,3 \frac{p(a_2)}{p(b)} \approx 0,3 \frac{p(a_2|b)}{p(b)} = 0,3 \cdot \frac{1}{1,2} = 0,25.$$

Таким чином, ми виконали тут розрахунки без інформації стосовно $p(B)$; ця ймовірність відіграє тільки роль нормуючої величини.

Умовна незалежність

Якщо інформація стосовно події B не змінює ступінь нашої впевненості стосовно появи події A , то говорять, події A і B – незалежні.

Означення 12.1 (незалежність). Події A і B незалежні, якщо

$$p(A|B) = p(A).$$

Наприклад, якщо ми підкидаємо одна за одною дві добре збалансованих (виготовлених) кості, то випадання числа 2 на одній з них не впливає на число, яке випадає на іншій.

Коли дві події незалежні одна від одної, то *фундаментальне правило обчислення ймовірностей* можна записати так:

$$p(A \cap B) = p(A|B) p(B) = p(A) \cdot p(B).$$

Тобто ймовірність того, що обидві події матимуть місце, визначається добутком ймовірностей окремих подій. (Вище, фундаментальне правило (12.6.1) ми записали так: $p(A|B) \cdot p(B) = p(A \cap B)$.)

Симетричність незалежності

Поняття незалежності симетричне, тобто, якщо A не залежить від B , то B не залежить від A :

$$p(B|A) = \frac{p(A \cap B)}{p(A)} = \frac{p(A|B) p(B)}{p(A)} = \frac{p(A) p(B)}{p(A)} = p(B),$$

$$p(A) > 0.$$

З іншого боку, якщо подія A неможлива, то ми її просто не розглядаємо, тобто умова $p(A) > 0$ не має значення.

Концепція незалежності у випадку, коли умова стосується декількох подій.

Так, якщо інформація стосовно події B не змінює нашу ступінь упевненості у настанні події A , але відома інформація про подію C (яка може

впливати на A), то говорять, що A і B умовно незалежні при наявності інформації стосовно події C .

Означення 12.2 (умовна незалежність). Дві події A і B умовно незалежні за наявності інформації стосовно події C , якщо:

$$p(A|B \cap C) = p(A|C).$$

Твердження стосовно умовної незалежності (так само як і твердження стосовно умовної залежності) симетричне. Якщо A умовно не залежить від B за умови наявності інформації стосовно події C , то B умовно не залежить від A за умови наявності інформації стосовно події C :

$$\begin{aligned} p(B|A \cap C) &= \frac{p(A \cap B|C) p(C)}{p(A|C) p(C)} = \frac{p(A|B \cap C) p(B|C)}{p(A|C)} = \\ &= \frac{p(A|C) p(B|C)}{p(A|C)} = p(B|C). \end{aligned}$$

Ймовірність виникнення обох умовно незалежних подій можна обчислити за допомогою розглянутого вище правила множення:

$$p(A \cap B|C) = p(A|C) \cdot p(B|C).$$

Зазначимо, що випадок, коли події незалежні, – це окремий випадок умовної незалежності при $C = \emptyset$.

Приклади умовної незалежності

Стосовно реальних об'єктів концепція умовної незалежності є реалістичнішою. Розглянемо її на простому прикладі.

Приклад 12.4. Припустимо, що студенту необхідно у понеділок здавати іспит. Для того щоб визначити ступінь ризику не здати іспит внаслідок поганого самопочуття розглянемо такі ситуації:

- 1 – перед екзаменом студент не менше 7 годин і почувається добре;
- 2 – перед початком екзамену студент відчуває значну втому.

У цьому прикладі існує причинний зв'язок між тривалістю сну студента перед екзаменом, ступенем його втоми і ризиком не здати екзамен успішно. Припустимо, що перед іспитом студент стомлений, почуває себе погано. Для

оцінювання ризику не має значення, яка причина його втомі – неспокійний сон перед іспитом чи інша причина.

Якщо ж студент почувається перед початком іспиту добре, то якість його сну у ніч з неділі на понеділок не має значення для оцінювання ризику не здати іспит. Такий попередній аналіз дає можливість стверджувати, що ризик не здати іспит є умовно незалежним від якості сну студента перед рейсом, за умови, що він почувається добре.

Для формального опису ситуації введемо три бінарні змінні:

- X_1 – характеризує якість сну студента перед іспитом;
- X_2 – стан студента перед іспитом (добре або погане самопочуття);
- X_3 – студент здасть іспит чи ні.

Тобто X_3 не залежить від X_1 при будь-якому значенні X_2 (для встановлення значення X_3 досить знати X_2), що можна записати так:

$$p(X_3 | X_1 \text{ і } X_2) = p(X_3 | X_1 \cap X_2) = p(X_3 | X_2). \quad (12.6.3)$$

Таким чином, для того щоб визначити ймовірність успішної здачі іспиту студентом досить знати значення X_2 .

У даному випадку немає необхідності вводити функцію трьох змінних:

$$p(X_1, X_2, X_3) = p(X_1) \cdot p(X_2 | X_1) \cdot p(X_3 | X_2), \quad (12.6.4)$$

яка послідовно описує процес оцінювання ймовірності того, що студент успішно здасть екзамен. Тобто якість сну студента впливає на самопочуття, а самопочуття визначає ступінь ризику не здати екзамен:

$$\text{якість сну} \rightarrow \text{самопочуття} \rightarrow \text{ступінь ризику не здати іспит}. \quad (12.6.5)$$

Приклад 12.5. Розглянемо тестування спортсменів на допінг. Як правило, для цього виконується аналіз крові, A , і аналіз сечі, B . Ці два аналізи виконуються незалежно один від другого у різних лабораторіях.

Очевидно, що результати цих аналізів характеризуються залежними змінними. Якщо аналіз A – позитивний, то аналіз B також, скоріше всього, буде позитивним.

Ці аналізи можна розглядати як незалежні події, оскільки їх роблять у різних лабораторіях за різними методиками. Тобто результати обох аналізів є умовно незалежними для будь-якого стану спортсмена.

Для формального опису ситуації із аналізами можна ввести три бінарні змінні:

- X_1 – характеризує стан спортсмена (приймав допінг чи ні);
- X_2 – результат аналізу крові (позитивний чи негативний);
- X_3 – результат аналізу сечі (позитивний або негативний).

Для цих змінних можна записати:

$$p(X_2 | X_1, X_3) = p(X_2 | X_1). \quad (12.6.6)$$

Для того щоб оцінити ймовірність того, що аналіз сечі буде позитивним, досить знати, що спортсмен прийняв допінг. А рівність (12.6.6) має такий смисл: „для того щоб оцінити ймовірність того, що аналіз крові буде позитивним, досить знати, що спортсмен прийняв допінг”.

Розширення правила множення на довільну кількість подій

Закон множення також можна поширити на будь-яку кількість незалежних подій:

$$p(A_1, A_2, \dots, A_n | I) = p(A_1 | I) \cdot p(A_2 | A_1, I) \cdot \dots \cdot p(A_n | A_1, A_2, \dots, A_{n-1}, I),$$

де I – додаткова інформація, наприклад, стосовно дотримання умов експерименту.

Нехай A_1, A_2, \dots, A_n – така послідовність незалежних подій, пов’язаних із підкиданням монети:

- результат першого підкидання – „орел”;
- результат другого підкидання – „орел”;
- результат третього підкидання – „решка”;
- ...

Якщо монета добре збалансована, то результати підкидань від 1-го до $(n-1)$ -го не будуть впливати на майбутні результати (майбутні результати не залежать від попередніх), в тому числі і на результат n -го підкидання, що можна формально записати так:

$$p(A_1, A_2, \dots, A_n) = p(A_1 | I) \cdot p(A_2 | I) \cdot \dots \cdot p(A_n | I).$$

12.7 Розширення правила додавання (другого правила обчислення ймовірностей)

Для будь-яких пропозицій A і B ступінь довіри до того, що A – істинна, за наявності додаткової інформації I , визначається сумою (правилом) таких ймовірностей:

$$p(A|I) = p(A, B|I) + p(A, \bar{B}|I), \quad (12.7.1)$$

де B – подія з двома станами.

Важливість цього правила підтверджується тим фактом, що деякій пропозиції легше присвоїти ступінь довіри, якщо вона розглядається у контексті інших пропозицій, які мають стосунок до неї.

Застосовуючи до наведеного правила закон множення ймовірностей, отримаємо:

$$p(A|I) = p(A|B, I) \cdot p(B|I) + p(A|\bar{B}, I) \cdot p(\bar{B}|I). \quad (12.7.2)$$

Вираз (12.7.2) – це **розширення другого правила (закону) (додавання ймовірностей взаємовиключних подій)**, яке мало вигляд (12.5.3):

$$p(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_n) = p(A_1) + p(A_2) + \dots + p(A_n).$$

Повернемося до прикладу з підкиданням монети. Нехай A – це така пропозиція: „результатом першого підкидання монети є орел”.

Загальну ступінь довіри до A (ймовірність настання A) можна розширити можна легко визначити за допомогою виразу (12.7.2). Правило (12.7.1) легко розширюється на будь-яку кількість пропозицій, за умови що вони є взаємовиключними і утворюють повну множину. У прикладі з монетою існує тільки три пропозиції:

- B_1 = „монета якісна”;
- B_2 = „монета має два орли (герби)”;
- B_3 = „монета має дві решки” (лицеві сторони).

Таким чином, ймовірність того, що результатом першого підкидання буде орел, можна обчислити за виразом:

$$p(A|I) = p(A|B_1, I) \cdot p(B_1|I) + p(A|B_2, I) \cdot p(B_2|I) + p(A|B_3, I) \cdot p(B_3|I)$$

Отриманий вираз можна узагальнити на n взаємовиключних пропозицій (подій) (це знаменник теорему Байєса):

$$p(A|I) = \sum_{i=1}^n p(A|B_i, I) \cdot p(B_i|I). \quad (12.7.3)$$

Насамкінець, на основі виразу (12.7.2) можна стверджувати, що події A і B зв'язані між собою, якщо

$$p(A|B,I) \neq p(A|\bar{B},I). \quad (12.7.4)$$

Дійсно, якщо подія B не має відношення до A , то

$$p(A|B,I) = p(A|\bar{B},I).$$

Оскільки $p(\bar{B}|I) = 1 - p(B|I)$, то права частина виразу (12.7.2) тепер тотожно дорівнює лівій частині незалежно від значення $p(B|I)$:

$$\begin{aligned} p(A|I) &= p(A|B,I) \cdot p(B|I) + p(A|\bar{B},I) \cdot [1 - p(B|I)] = \\ &= p(A|B,I) \cdot p(B|I) + p(A|\bar{B},I) - p(A|\bar{B},I) \cdot p(B|I) = \\ &= p(A|B,I), \end{aligned}$$

оскільки $p(A|B,I) = p(A|\bar{B},I)$.

Поняття апіорної умовної ймовірності

Отримання нової інформації (свідчення) стосовно деякої події можна розглядати як двокроковий процес у часі – моменту t_0 і моменту t_1 . У момент часу t_0 починається пошук свідчення B , яке пов'язане з подією A . В цей момент ймовірність того, що $A = \text{”істина”}$, за умови, що $B = \text{”істина”}$, можна обчислити за виразом (12.7.1). Якщо позначити через $p_0(\cdot)$ – ймовірність події у момент t_0 , то правило (12.4.1) можна переписати у вигляді:

$$p_0(A|B,I) = \frac{p_0(B|A,I) \cdot p_0(A|I)}{p_0(B|A,I) \cdot p_0(A|I) + p_0(B|\bar{A},I) \cdot p_0(\bar{A}|I)}, \quad (12.7.5)$$

що випливає із третього правила (закону) обчислення ймовірностей (закон множення) та розширення другого закону – виразу (12.7.2). У виразі (12.7.5) $p_0(A|I)$, $p_0(\bar{A}|I)$ і $p_0(A|B,I)$ – апіорні (початкові) ймовірності відповідних подій.

Ймовірність $p_0(B|A,I)$ називають *правдоподібністю* A за умови B (у момент t_0). Аналогічно, ймовірність $p_0(B|\bar{A},I)$ – це правдоподібність \bar{A} за умови B (у момент t_0). Поняття „*правдоподібності*” – це ще один із елементів процедури формулювання гіпотез з подальшою метою прийняття рішення.

Оцінювання правдоподібності пов'язане із відповіддю на запитання:

„якщо гіпотеза A вірна, (або невірна), то з якою ступінню довіри буде вірною гіпотеза B ?

Нехай у момент часу t_1 виявляється, що гіпотеза B – вірна. Якою буде ймовірність того, що A – також вірна у цей же момент t_1 , тобто $p_1(A|I) = ?$

Очевидно, що необхідно врахувати відомий факт, що B – вірна; це частина додаткової інформації у момент t_1 . Тому загальна ймовірність події A у момент t_1 повинна враховувати те, що мала місце подія B , тобто:

$$p_1(A|I) = p_0(A|B, I). \quad (12.7.6)$$

Ситуація не зміниться якщо у момент t_1 з'являється інформація стосовно гіпотези B (B – вірна), але до цього моменту подія B не розглядалась. Встановлення відношення події B до A , а також ступеня впливу цього відношення на ймовірності подій вимагає розгляду правдоподібностей і початкових ймовірностей у момент t_0 . Тобто, яка ймовірність того, що подія B – вірна за наявності деякої додаткової інформації і за припущення, що гіпотеза A – вірна (або невірна)? Якою була б апіорна ймовірність того, що A – вірна за наявності деякої додаткової інформації, не враховуючи те, що B – вірна?

12.8 Правдоподібність і ймовірність

Правдоподібність і ймовірність відіграють значну роль у імовірнісних дослідженнях, а тому важливо знати різницю між правдоподібністю події A при даному (відомому) B і ймовірністю A , яке залежить від B .

Розглянемо такий *приклад*: випадковий свідок бачив, що Іванов втік від будинку, де було скоєно злочин. Нехай пропозиція B звучить так:

B : „Іванов втік від місця злочину у той час, коли він був скоєний”;

а пропозицію A сформульовано так:

A : „Іванов скоїв злочин”.

Цілком логічно припустити, що ймовірність $p(B|A, I)$ є високою, але ймовірність $p(A|B, I)$ не обов'язково повинна бути високою.

Якщо Іванов дійсно скоїв злочин, то можна очікувати, що він постарасться швидше покинути місце злочину. Гіпотеза стосовно того, що він винен – це хороше пояснення до свідчення (тобто втікав). Однак, той факт, що він біг від будинку ще не дає високої ймовірності того, що саме він скоїв злочин. Цьому факту можна дати також інше пояснення.

Якщо між правдоподібністю гіпотези A , за наявності свідчення B , і ймовірністю цієї ж гіпотези за умови B (тобто при існуванні залежності між A і B) ставлять знак рівності, то це називають *софізмом переміни умов* (fallacy of the transposed conditions).

Він виникає у тому випадку, коли із факту: якщо мало місце A , то B виникне з великою ймовірністю, помилково роблять такий висновок: якщо мало місце B , то A виникне із високою ймовірністю. Найкращим засобом боротьби з таким софізмом є теорема Байєса у формі (12.6.2):

$$p(A|B,I) = \frac{p(B|A,I) \cdot p(A|I)}{p(B|A,I) \cdot p(A|I) + p(B|\bar{A},I) \cdot p(\bar{A}|I)}.$$

Члени у знаменнику не обов'язково близькі один до одного. У другій складовій знаменника правдоподібність \bar{A} за наявності B також може бути великою, тобто може існувати інше хороше пояснення для свідчення B , навіть якщо A = "невірно". Більше того, необхідно прийняти до уваги ступінь довіри до A у випадку, якщо відома тільки додаткова (другорядна) інформація.

Очевидно, що наведену вище теорему Байєса можна розширити на випадок n взаємовиключних пропозицій A_i , $i=1, \dots, n$, які утворюють повну множину подій і представляють собою альтернативні пояснення для B :

$$p(A_i|B,I) = \frac{p(B|A_i,I) \cdot p(A_i|I)}{\sum_{i=1}^n p(B|A_i,I) \cdot p(A_i|I)}.$$

Очевидно, що ТБ можна застосовувати повторно для розв'язання однієї і тієї ж задачі. У процесі дослідження поставленої задачі може з'явитись нова інформація у вигляді експертних оцінок та статистичних даних (вимірів), які дадуть можливість уточнити прийняте рішення. У такому випадку раніше отриманим результатом можна скористатись як апріорною інформацією.

12.9 Обчислення ймовірностей для змінних

До цих пір ми розглядали обчислення ймовірностей для простих змінних та результатів настання подій у деякому вибірковому просторі. Тепер

розширимо розглянуті вище концепції стосовно обчислення ймовірностей на *змінні* (набори вибірових просторів). Результатом кожного експерименту є *стан*, який характеризується значенням змінної. Множину станів, пов'язаних із змінною A , будемо позначати через

$$sp(A) = (a_1, a_2, \dots, a_n).$$

За аналогією до вибірового простору подій ці стани повинні бути *взаємно виключними* та утворювати *скінченну повну множину*. Останнє припущення гарантує, що змінна (процес) перебуває у одному із своїх станів (хоча ми можемо не знати у якому), а перше припущення означає, що у кожний момент часу змінна перебуває тільки у одному стані.

Наприклад, якщо позначити через D змінну, яка характеризує результат підкидання костей, то її простір станів буде таким:

$$sp(D) = (1, 2, 3, 4, 5, 6).$$

У подальшому змінні будемо позначати великими (прописними) літерами, а їх стани – малими.

Невизначеність стосовно станів (a_1, a_2, \dots, a_n) змінної A будемо виражати через розподіл ймовірностей $P(A)$ цих станів:

$$P(A) = (x_1, \dots, x_n); \quad x_i \geq 0; \quad \sum_{i=1}^n x_i = x_1 + \dots + x_n = 1,$$

де x_i – ймовірності перебування змінної A у стані a_i . Розподіл називають *рівномірним (uniform, even)*, якщо всі ймовірності однакові.

Примітка. У загальному випадку ймовірність перебування A у стані a_i позначають так: $p(A = a_i)$, або просто $p(a_i)$, якщо змінна, що аналізується, очевидна із контексту.

Так само як і для звичайних подій, розглянемо *умовні ймовірності* для змінних. Якщо A має стани (a_1, \dots, a_m) а змінна B має стани (b_1, b_2, \dots, b_m) , то $P(A|B)$ містить $n \times t$ умовних ймовірностей $p(a_i | b_j)$, які визначають ймовірність стану a_i за умови b_j . Тобто умовна ймовірність для змінної за умови наявності інформації про іншу змінну, представляє собою множину ймовірностей (які представляють, як правило, у табличному вигляді). При цьому кожне число таблиці – це ймовірність деякої конфігурації (комбінації) станів змінних (табл. 12.1).

Крім того, оскільки $P(A|B)$ визначає розподіл ймовірностей для кожної події $B = b_j$, то згідно з аксіомою 1 сума ймовірностей по всіх значеннях змінної A повинна дорівнювати 1 для кожного стану змінної B :

$$\sum_{i=1}^n p(A = a_i | B = b_j) = 1, \text{ для кожного } b_j, j = 1, \dots, m. \quad (12.9.1)$$

Таблиця 12.1

Приклад таблиці умовних ймовірностей $P(A|B)$ для бінарної змінної A за умови тернарної змінної B . Для кожного стану змінної B сума ймовірностей станів A дорівнює 1.

	b_1	b_2	b_3
a_1	0,4	0,3	0,6
a_2	0,6	0,7	0,4

Спільний розподіл ймовірностей

Ймовірність появи декількох станів можна виразити *спільною ймовірністю* для двох або більше змінних. Для кожної конфігурації змінних (a_i, b_j) змінних A і B , $P(A, B)$ визначає ймовірність настання обох станів: $A = a_i$ і $B = b_j$. Таким чином розподіл $P(A, B)$ складається з $n \times m$ чисел (приклад у табл. 12.2).

Зазначимо, що оскільки простори станів обох змінних A і B є взаємно виключними і повними множинами, то всі комбінації їх станів (Декартів добуток) також будуть взаємно виключними і повними множинами, а тому їх можна розглядати як вибіркового простір (за аналогією з простими подіями). Таким чином, згідно з аксіомою 1, маємо:

$$P(A, B) = \sum_{i=1}^n \sum_{j=1}^m p(A = a_i, B = b_j) = 1.$$

Таблиця 12.2

Приклад таблиці розподілу спільних ймовірностей $P(A, B)$ для бінарної змінної A за умови тернарної змінної B . Зазначимо, що сума всіх значень ймовірностей у таблиці дорівнює 1.

	b_1	b_2	b_3
a_1	0,16	0,12	0,12
a_2	0,24	0,28	0,08

Коли фундаментальне правило $p(A|B) \cdot p(B) = p(A \cap B)$ застосовується до змінних A і B , то таку процедуру обчислення ймовірностей необхідно застосувати до кожної пари (a_i, b_j) , а всього таких пар маємо $n \times m$:

$$p(a_i | b_j) p(b_j) = p(a_i, b_j).$$

Це означає, що для побудови таблиці $P(A|B)$ кожну ймовірність з множини $P(A|b_j)$ необхідно помножити на $p(b_j)$, що дасть нам $P(A, b_j)$. Виконуючи такі обчислення для кожного b_j , отримаємо $P(A, B)$.

Таблицю 12.2 можна отримати в результаті застосування фундаментального правила до таблиці 12.1 при $P(B) = (0,4; 0,4; 0,2)$ (відповідна ілюстрація наведена у табл. 12.3).

Таблиця 12.3

Таблиця 12.2 спільних ймовірностей $P(A, B)$ отримана як добуток $P(B) = (0,4; 0,4; 0,2)$ на значення $P(A|B)$ з таблиці 12.1.

	b_1	b_2	b_3
a_1	$0,4 \cdot 0,4$	$0,3 \cdot 0,4$	$0,6 \cdot 0,2$
a_2	$0,6 \cdot 0,4$	$0,7 \cdot 0,4$	$0,4 \cdot 0,2$

У випадку застосування до змінних фундаментальне правило визначається теоремою 12.3.

Теорема 12.3 (фундаментальне правило для змінних) – обчислення ймовірностей одночасної появи відповідних станів змінних A і B :

$$P(A, B) = P(A|B) \cdot P(B),$$

а після введення умови стосовно третьої змінної отримаємо:

$$P(A, B|C) = P(A|B, C) \cdot P(B|C).$$

Обчислення маргінальних ймовірностей (маргіналізація, marginalization)

Використовуючи таблицю спільних ймовірностей, можна розрахувати розподіл значень $P(A)$, розглядаючи ті значення B , які можуть мати місце

разом із кожним станом a_i змінної A . Тобто для кожного стану a_i змінної A необхідно розрахувати m різних взаємно виключних значень: $(a_i, b_1), \dots, (a_i, b_m)$.

Таким чином, згідно з аксіомою 3 маємо:

$$p(a_i) = \sum_{j=1}^m p(a_i, b_j).$$

Такі обчислення називають *маргіналізацією*, тобто говорять, що змінна B маргіналізується з розподілу $P(A, B)$ (що приводить до $P(A)$). Процедура маргіналізації позначається так:

$$P(A) = \sum_B P(A, B) = \text{mar}_B P(A, B).$$

Маргіналізуючи B з таблиці розподілу спільних ймовірностей 12.2, отримаємо:

$$P(A) = (0,16 + 0,12 + 0,12; 0,24 + 0,28 + 0,08) = (0,4; 0,6).$$

Тобто $P(A) = (0,4; 0,6)$ – це і є розподіл, який нас цікавить.

Маргіналізація A з таблиці 12.2 дає такий результат для ймовірностей станів B :

$$P(B) = (0,16 + 0,24; 0,12 + 0,28; 0,12 + 0,08) = (0,4; 0,4; 0,2).$$

Таким чином, процедура маргіналізації дає можливість видалити змінні із спільного розподілу ймовірностей і знайти таким чином остаточний розподіл, який нас цікавить.

Теорема Байєса для змінних

Теорема 12.4 (правило Байєса для змінних):

$$P(B|A) = \frac{P(A|B) P(B)}{P(A)} = \frac{P(A, B)}{\sum_B P(A, B)},$$

а після введення додаткової умови за змінною C отримаємо:

$$P(B|A, C) = \frac{P(A|B, C) P(B|C)}{P(A|C)} = \frac{P(A, B|C)}{\sum_B P(A, B|C)}.$$

Дві наведені рівності можливі завдяки (1) фундаментальному правилу обчислення ймовірностей і (2) описаній вище процедурі маргіналізації.

У таблиці 12.4а наведено приклад обчислення розподілу $P(B|A)$, використовуючи $P(A)$, $P(B)$ і $P(A|B)$ з наведених вище таблиць.

Таблиця 12.4а

Умовні ймовірності $P(B|A)$, обчислені шляхом застосування правила Байєса до $P(A|B)$ у таблиці 12.1: $P(A)=(0,4; 0,6)$ і $P(B)=(0,4; 0,4; 0,2)$. Зазначимо, що сума ймовірностей по B дорівнює 1 для кожного стану A .

	a_1	a_2
b_1	$(p(a_1 b_1) \cdot p(b_1)) / p(a_1)$	$(p(a_2 b_1) \cdot p(b_1)) / p(a_2)$
b_2	$(p(a_1 b_2) \cdot p(b_2)) / p(a_1)$	$(p(a_2 b_2) \cdot p(b_2)) / p(a_2)$
b_3	$(p(a_1 b_3) \cdot p(b_3)) / p(a_1)$	$(p(a_2 b_3) \cdot p(b_3)) / p(a_2)$

Таблиця 12.4б містить результати підстановки числових значень у таблицю 12.4а.

Таблиця 12.4б

Умовні ймовірності $P(B|A)$, обчислені шляхом застосування правила Байєса до $P(A|B)$ у таблиці 12.1, $P(A)=(0,4; 0,6)$ і $P(B)=(0,4; 0,4; 0,2)$. Зазначимо, що сума ймовірностей по B дорівнює 1 для кожного стану A .

	a_1	a_2
b_1	$(0,4 \cdot 0,4) / 0,4$	$(0,6 \cdot 0,4) / 0,6$
b_2	$(0,3 \cdot 0,4) / 0,4$	$(0,7 \cdot 0,4) / 0,6$
b_3	$(0,6 \cdot 0,2) / 0,4$	$(0,4 \cdot 0,2) / 0,6$

Результати обчислень за таблицею 12.4б наведені у таблиці 12.4в. Значення $p(a_1|b_1) \dots p(a_i|b_i)$ взяті з таблиці 12.1.

Таблиця 12.4в

Результати обчислень за таблицею 12.4а, тобто значення $p(b_i|a_j)$

	a_1	a_2
b_1	0,4	0,4
b_2	0,3	0,47
b_3	0,3	0,13
	$\sum p(b_i a_1)=1$	$\sum p(b_i a_2)=1$

Концепція умовної незалежності для змінних

Означення 12.3 (умовна незалежність для змінних). Дві змінні A і C умовно незалежні при відомій змінній B , якщо:

$$p(a_i | c_k, b_j) = p(a_i | b_j)$$

для кожного $a_i \in sp(A)$, $b_j \in sp(B)$ і $c_k \in sp(C)$.

У скороченому записі: $P(A|C,B) = P(A|B)$.

Це означає, що за відомого стану змінної B ніяка інформація стосовно стану змінної C не змінить ймовірності A . Зазначимо, що твердження незалежності виконується для кожного стану B ; якщо множина, яка задає умову (тобто B), порожня, то говорять, що A і C маргінально незалежні або просто незалежні ($P(A|C) = P(A)$).

Якщо дві змінні A і C умовно незалежні за наявності третьої (зв'язаної з ними) змінної B , то фундаментальне правило (теорема 12.3) спрощується до вигляду:

$$P(A,C|B) = P(A|B,C) \cdot P(C|B) = P(A|B) \cdot P(C|B).$$

Згідно з цим виразом необхідно перемножити дві таблиці умовних ймовірностей, які відносяться до своїх областей визначення. Таке множення є безпосереднім (прямим) розширенням тих операцій, які розглянуто у попередніх прикладах:

$$p(a_i, c_k | b_j) = p(a_i | b_j) \cdot p(c_k | b_j).$$

Наприклад, перемножуючи $P(A|B)$ на $P(C|B)$, які задані таблицями 12.1 і 12.5, відповідно, отримаємо таблицю 12.6. Тобто спочатку знаходимо таблицю 12.5 для $P(C|B)$ (у даному випадку ми її просто задаємо, а не розраховуємо, а розрахунок можна зробити за формулою Байєса:

$$P(C|B) = \frac{P(B|C) P(C)}{P(B)} = \frac{P(C,B)}{\sum_C P(C,B)},$$

а потім використовуємо для обчислень таблиці 12.1 і 12.5.

Таблиця 12.5

Приклад таблиці умовних ймовірностей $P(C|B)$ для тернарної змінної C за умови тернарної змінної B . Для кожного стану змінної B сума ймовірностей станів C дорівнює 1.

	b_1	b_2	b_3
c_1	0,2	0,9	0,3
c_2	0,05	0,05	0,2
c_3	0,75	0,05	0,5

Використовуючи наведені нижче вирази, обчислимо елементи таблиці умовних ймовірностей (таблиця 12.6а, 12.6б і 12.6в) для умовно незалежних змінних за наявності третьої (зв'язаної з ними) змінної B , тобто:

$$P(A,C|B) = P(A|B,C) \cdot P(C|B) = P(A|B) \cdot P(C|B)$$

$$p(a_i, c_k | b_j) = p(a_i | b_j) \cdot p(c_k | b_j) =$$

Таблиця 12.6а

Обчислення умовних ймовірностей у символній формі

	b_1	b_2	b_3
c_1	$(p(a_1 b_1) \cdot p(c_1 b_1);$ $p(a_2 b_1) \cdot p(c_1 b_1))$	$(p(a_1 b_2) \cdot p(c_1 b_2);$ $p(a_2 b_2) \cdot p(c_1 b_2))$	$(p(a_1 b_3) \cdot p(c_1 b_3);$ $p(a_2 b_3) \cdot p(c_1 b_3))$
c_2	$(p(a_1 b_1) \cdot p(c_2 b_1);$ $p(a_2 b_1) \cdot p(c_2 b_1))$	$(p(a_1 b_2) \cdot p(c_2 b_2);$ $p(a_2 b_2) \cdot p(c_2 b_2))$	$(p(a_1 b_3) \cdot p(c_2 b_3);$ $p(a_2 b_3) \cdot p(c_2 b_3))$
c_3	$(p(a_1 b_1) \cdot p(c_3 b_1);$ $p(a_2 b_1) \cdot p(c_3 b_1))$	$(p(a_1 b_2) \cdot p(c_3 b_2);$ $p(a_2 b_2) \cdot p(c_3 b_2))$	$(p(a_1 b_3) \cdot p(c_3 b_3);$ $p(a_2 b_3) \cdot p(c_3 b_3))$

Таблиця 12.6б

Підстановка числових значень умовних ймовірностей; наприклад,

$p(c_1 | b_1) = 0,2$ – взято з табл. 12.5, а $p(a_1 | b_1) = 0,4$ – з табл. 12.1.

	b_1	b_2	b_3
c_1	(0,2·0,4; 0,2·0,6)	(0,9·0,3; 0,9·0,7)	(0,3·0,6; 0,3·0,4)
c_2	(0,05·0,4; 0,05·0,6)	(0,05·0,3; 0,05·0,7)	(0,2·0,6; 0,2·0,4)
c_3	(0,75·0,4; 0,75·0,6)	(0,05·0,3; 0,05·0,7)	(0,5·0,6; 0,5·0,4)

Таблиця 12.6в

Результати обчислень умовних ймовірностей
 $P(A, C | B) = P(A | B, C) \cdot P(C | B) = P(A | B) \cdot P(C | B)$

	b_1	b_2	b_3
c_1	(0,08; 0,12)	(0,27; 0,63)	(0,18; 0,12)
c_2	(0,02; 0,03)	(0,015; 0,035)	(0,12; 0,08)
c_3	(0,3; 0,45)	(0,015; 0,035)	(0,3; 0,2)

Суми усіх значень умовних ймовірностей для кожного стовпчика дорівнюють 1.

Приклад виконання обчислень значень таблиць умовних ймовірностей за наявності додаткових свідчень (тобто наявності значень окремих станів)

Для ілюстрації розглянутих вище теорем розглянемо процедуру обчислення умовних ймовірностей для трьох змінних A , B і C . Для цих змінних дана таблиця 12.7 числових значень спільного розподілу ймовірностей. До цієї відомої таблиці додатково поступають свідчення $A = a_2$ і $C = c_1$.

Таблиця 12.7

Таблиця спільного розподілу ймовірностей для змінних A , B і C . Три числа, які дані у кожній клітинці таблиці, відповідають станам c_1 , c_2 і c_3 .

	b_1	b_2	b_3
a_1	(0; 0,05; 0,05)	(0,05; 0,05; 0)	(0,05; 0,05; 0,05)
a_2	(0,1; 0,1; 0)	(0,1; 0; 0,1)	(0,2; 0; 0,05)

Задача полягає у тому, щоб обчислити значення таблиці умовних ймовірностей $P(B | a_2, c_1)$.

Спочатку розглянемо ту частину таблиці 12.7, яка відповідає $A = a_2$ і $C = c_1$; аналізуючи таблицю, отримаємо:

$$P(a_2, B, c_1) = (0,1; 0,1; 0,2), \quad (12.9.2)$$

тобто це перші елементи множин значень у клітинках другого рядка. Перші – тому що $C = c_1$, а другого рядка, тому що $A = a_2$.

Для того щоб обчислити значення $P(B|a_2, c_1)$, скористаємось теоремою 12.4:

$$P(B|a_2, c_1) = \frac{P(a_2, B, c_1)}{P(a_2, c_1)} = \frac{P(a_2, B, c_1)}{\sum_B P(a_2, B, c_1)}. \quad (12.9.3)$$

Маргіналізуючи B з рівняння (12.9.2) (маргіналізація у даному випадку означає обчислення ймовірності спільної появи станів (a_2, c_1) , тобто це знаменник (12.9.3)), отримаємо:

$$P(a_2, c_1) = \sum_B P(a_2, B, c_1) = 0,1 + 0,1 + 0,2 = 0,4.$$

Тепер виконаємо операцію ділення (нормування) у рівнянні (12.9.3):

$$P(B|a_2, c_1) = \left(\frac{0,1}{0,4}; \frac{0,1}{0,4}; \frac{0,2}{0,4} \right) = (0,25; 0,25; 0,5).$$

Іншими словами можна сказати, що ми ставили задачу перетворення $P(a_2, B, c_1)$ у розподіл ймовірностей. Оскільки сума значень ймовірностей в (12.9.2) не дорівнює 1, то ми нормуємо розподіл шляхом ділення кожного числа на суму їх усіх значень.

Припустимо тепер, що ми отримали тільки свідоцтво $A = a_2$ і необхідно обчислити $P(B|a_2, C)$. Обчислення елементів цієї таблиці ймовірностей виконується так само, як і у попередньому випадку, за винятком того, що тепер обчислення будуть ґрунтуватись на таблицях.

Як і у попередньому випадку, зосередимось на тій частині таблиці ймовірностей 12.7 $P(A, B, C)$, яка відповідає $A = a_2$; в результаті отримаємо таблицю 12.8.

Таблиця 12.8

Таблиця ймовірностей $P(a_2, B, C)$, яка відповідає тій частині таблиці 12.7, що обмежена значенням $A = a_2$.

	b_1	b_2	b_3
c_1	0,1	0,1	0,2
c_2	0,1	0	0
c_3	0	0,1	0,05

Тобто для побудови цієї таблиці вибрані елементи нижнього рядка (який відповідає a_2) і трьом станам змінних B і C .

Для того щоб обчислити розподіл $P(B|a_2, C)$, скористаємось виразом:

$$P(B|a_2, C) = \frac{P(a_2, B, C)}{P(a_2, C)} = \frac{P(a_2, B, C)}{\sum_B P(a_2, B, C)}. \quad (12.9.4)$$

Ймовірність $P(a_2, C)$ обчислимо шляхом маргіналізації B із таблиці 12.8:

$$P(a_2, C) = (0,1 + 0,1 + 0,2; 0,1 + 0 + 0; 0 + 0,1 + 0,05) = (0,4; 0,1; 0,15), \quad (12.9.5)$$

а отриманий результат підставимо у рівняння (12.9.4). Остаточний результат наведено у таблицях 12.9а і 12.9б.

Таблиця 12.9

$$P(B|a_2, C) =$$

	b_1	b_2	b_3
c_1	0,1/0,4	0,1/0,4	0,2/0,4
c_2	0,1/0,1	0/0,1	0/0,1
c_3	0/0,15	0,1/0,15	0,05/0,15

Таблиця 12.9б

Результати обчислень $P(B|a_2, C)$, використовуючи $P(a_2, B, C)$

(табл. 12.8) і $P(a_2, C)$ (рівняння (12.9.5)).

	b_1	b_2	b_3
c_1	0,25	0,25	0,5
c_2	1	0	0
c_3	0	2/3	1/3

12.10 Вступ до алгебри потенціалів

Розглянемо властивості алгебри, яка стосується множення і маргіналізації таблиць. Такі таблиці не обов'язково повинні бути таблицями (умовних) ймовірностей, а в загальному випадку їх називають *потенціалами*.

Потенціал ϕ – це дійсна функція, визначена над областю скінченних значень змінних χ :

$$\phi : sp(\chi) \rightarrow \mathfrak{R}.$$

Область визначення потенціалів (*домен*) позначається так: $dom(\phi)$. Наприклад, областю визначення потенціалу $P(A, B | C)$ є $dom(P(A, B | C)) = \{A, B, C\}$.

Потенціали можна *перемножувати*; ця операція позначається крапкою (яка може не наводитись у явному вигляді). Множення потенціалів має такі властивості:

1. Область визначення добутку потенціалів:

$$dom(\phi_1 \phi_2) = dom(\phi_1) \cup dom(\phi_2).$$

2. Комутативний закон для добутку потенціалів: $\phi_1 \cdot \phi_2 = \phi_2 \cdot \phi_1$.

3. Асоціативний закон для добутку потенціалів:

$$(\phi_1 \cdot \phi_2) \phi_3 = \phi_1 (\phi_2 \cdot \phi_3).$$

4. Існування одиничного потенціалу: одиничним називають потенціал, який містить тільки одиниці („1”) і який визначений на будь-якій області, такій, що $\mathbf{1} \cdot \phi = \phi$ для всіх потенціалів ϕ .

Визначений вище оператор маргіналізації (підрозділ 12.9) можна узагальнити на потенціали таким чином: $\sum_A \phi$ – це потенціал над областю $dom(\phi) \setminus \{A\}$.

Далі: операція маргіналізації *комутативна*:

$$\sum_A \sum_B \phi = \sum_B \sum_A \phi.$$

5. Для потенціалів типу $P(A | \mathfrak{G})$, де \mathfrak{G} – множина змінних, маємо:

властивість одиничного потенціалу: $\sum_A P(A | \mathfrak{G}) = 1$.

Для маргіналізації добутку виконується співвідношення:

6. Розподільний закон: якщо $A \notin \text{dom}(\phi_1)$, то $\sum_A \phi_1 \phi_2 = \phi_1 \sum_A \phi_2$.

Розподільний закон відомий у формі: $ab + ac = a(b + c)$, а попередня формула (у пункті 6) – це такий же закон, але у застосуванні до таблиць. Для перевірки його дії розглянемо обчислення, наведені у таблицях 12.10 – 12.14. Бачимо, що таблиці 12.12 і 12.14 – однакові і відповідають лівій і правій частинам закону розподілу.

Таблиця 12.10

Потенціали $\phi_1(A, B)$ і $\phi_2(C, B)$

$B \setminus A$	a_1	a_2		$B \setminus C$	c_1	c_2
b_1	x_1	x_2		b_1	y_1	y_2
b_2	x_3	x_4		b_2	y_3	y_4

Таблиця 12.11

Результат множення потенціалів $\phi_1(A, B)$ і $\phi_2(C, B)$.

Добутки $(x_i y_j, x_i y_l)$ відповідають станам c_1 і c_2 .

$B \setminus A$	a_1	a_2
b_1	$(x_1 y_1, x_1 y_2)$	$(x_2 y_1, x_2 y_2)$
b_2	$(x_3 y_3, x_3 y_4)$	$(x_4 y_3, x_4 y_4)$

Таблиця 12.12

Маргіналізація по C таблиці 12.11.

Продовження ілюстрації розподільного закону $\sum_C \phi_1(A, B) \cdot \phi_2(C, B)$

$B \setminus A$	a_1	a_2
b_1	$x_1 y_1 + x_1 y_2$	$x_2 y_1 + x_2 y_2$
b_2	$x_3 y_3 + x_3 y_4$	$x_4 y_3 + x_4 y_4$

Таблиця 12.13

Маргіналізація по C потенціалу $\phi_2(C, B)$.

Продовження ілюстрації розподільного закону: $\sum_C \phi_2(C, B)$

B	
b_1	$y_1 + y_2$
b_2	$y_3 + y_4$

Процедуру маргіналізації будемо називати також *проекцією*. Наприклад, якщо A і B маргіналізуються з $\phi(A, B, C)$, то можна сказати, що ϕ проектується на C ; ця операція позначається так: $\phi^{\downarrow C}$. Використовуючи це позначення, властивості маргіналізації можна представити таким чином (тут V і W – множини змінних), як представлено у табл. 12.14.

Таблиця 12.14

Продовження ілюстрації розподільного закону: $\phi_1(A, B) \sum_C \phi_2(C, B)$

$B \setminus A$	a_1	a_2
b_1	$x_1 (y_1 + y_2)$	$x_2 (y_1 + y_2)$
b_2	$x_3 (y_3 + y_4)$	$x_4 (y_3 + y_4)$

7. Комутативний закон для проекції: $(\phi^{\downarrow V})^{\downarrow W} = (\phi^{\downarrow W})^{\downarrow V}$.

8. Дистрибутивний закон для проекції:

якщо $dom(\phi_1) \subseteq V$, то $(\phi_1, \phi_2)^{\downarrow V} = \phi_1(\phi_2^{\downarrow V})$.

12.11 Випадкові змінні

Розглянемо вибіркового простір S . *Випадкова змінна* визначається дійсною функцією $\mathbf{V}: S \rightarrow \mathfrak{R}$. Наприклад, якщо при підкиданні костей вважається, що випадання значень $\{1, 2, 3\}$ означає програш однієї гривні, а

випадання $\{4, 5, 6\}$ означає виграш однієї гривні, то випадкова змінна є функцією із значеннями:

"-1" на області $\{1, 2, 3\}$ і "+1" на області $\{4, 5, 6\}$.

Середнє значення μ випадкової величини V у вибірковому просторі S визначається так:

$$\mu(V) = \sum_{s \in S} V(s) p(s). \quad (12.11.1)$$

Наприклад, для збалансованого грального кубика середнє значення випадкової змінної, пов'язаної із виграшем або програшем однієї гривні, складає:

$$(-1) \cdot \frac{1}{6} + (-1) \cdot \frac{1}{6} + (-1) \cdot \frac{1}{6} + (+1) \cdot \frac{1}{6} + (+1) \cdot \frac{1}{6} + (+1) \cdot \frac{1}{6} = 0.$$

Середнє називають ще *очікуваним значенням* (або сподіванням).

Мірою розсіювання випадкової змінної є дисперсія σ^2 , яка визначається як середнє суми квадратів відхилень значень змінної від її середнього:

$$\sigma^2(X) = \sum_{s \in S} [X(s) - \mu(X)]^2 p(s). \quad (12.11.2)$$

Для наведеного вище прикладу:

$$\sigma^2 = 3(-1-0)^2 \frac{1}{6} + 3(1-0)^2 \frac{1}{6} = 1.$$

Вступ до неперервних розподілів

Розглянемо експеримент із стрільбою стріл у мішень одиничного розміру: $[1,0 \times 1,0]$. Результатами стрільби будуть попадання у точки мішені з координатами: (x, y) на одиничній площі. Оскільки для будь-якого конкретного результату (координат) ймовірність попадання складає 0, то розподіл ймовірностей присвоюється до підмножин одиничної площі. Цю процедуру (оцінювання розподілу ймовірностей) можна розглядати як розподіл "1"-ї ймовірності на вибірковому просторі. Наприклад, можна

присвоїти ймовірність попадання стріли у квадратик (елемент) площею $[x, x+\varepsilon] \times [y, y+\varepsilon]$. Для більшої систематизації цього процесу виберемо деяке натуральне число n і розділимо одиничну площу мішені на невеликі квадратики:

$$\left[\frac{i}{n}, \frac{i+1}{n} \right] \times \left[\frac{j}{n}, \frac{j+1}{n} \right].$$

Тепер присвоїмо ймовірності попадання стріли в ці квадратики:

$$P \left(\left[\frac{i}{n}, \frac{i+1}{n} \right] \times \left[\frac{j}{n}, \frac{j+1}{n} \right] \right),$$

які мають площу $\frac{1}{n^2}$. Якщо $P \left(\left[\frac{i}{n}, \frac{i+1}{n} \right] \times \left[\frac{j}{n}, \frac{j+1}{n} \right] \right) = x$, то значення ймовірності x розподілене на невеликій площі із середньою густиною $x/(1/n^2) = n^2 x$.

Це дає можливість визначити *функцію густини (щільності)* розподілу (її називають також *частотною функцією*) $f(x, y)$ у вигляді:

$$f(x, y) = \lim_{n \rightarrow \infty} n^2 P \left(\left[x, x + \frac{1}{n} \right] \times \left[y, y + \frac{1}{n} \right] \right).$$

Якщо S – неперервний вибірковий простір, то функція щільності – це невід’ємна дійсна функція f на просторі S , така, що для будь-якої підмножини $A \subseteq S$ виконується рівність:

$$\int_A f(s) ds = p(A).$$

Зокрема,

$$\int_S f(s) ds = 1.$$

Якщо S визначена на інтервалі $[a, b]$ (можливо нескінченному) і результатами експерименту є дійсні числа (наприклад, висота або вага), а нас цікавить середнє (середня висота або вага), його можна визначити так:

$$\mu = \int_a^b x f(x) dx,$$

а дисперсія визначається за виразом:

$$\sigma^2 = \int_a^b (\mu - x)^2 f(x) dx.$$

Математично середнє і дисперсія – це середнє і дисперсія одиничної функції $I(x) = x$, але ми будемо використовувати термін „середнє і дисперсія розподілу”.

Р о з д і л 13

ДЕЯКІ ВАРІАНТИ ТЕОРЕМИ БАЙЄСА

13.1 Загальна форма теореми Байєса для довільних подій

Теорема Байєса (ТБ) стосується умовних ймовірностей подій загального характеру. Нехай $\mathbf{A} = [A_1, \dots, A_n]$ – множина взаємовиключних подій, які повністю описують деякий процес. Розглянемо деяку подію A_j , виникнення якої зв'язане з подією B . Теорема Байєса дає можливість визначити умовну ймовірність настання події A_j при умові, що мала місце подія B , використовуючи при цьому інформацію щодо умовної ймовірності настання події B (при умові, що мала місце подія A_j). Виходячи з наведеного формулювання призначення теореми Байєса, можна сказати, що вона стосується так званої “інверсної ймовірності”. Теорема Байєса для довільних подій записується так:

$$p(A_j | B) = \frac{p(B | A_j) p(A_j)}{\sum_{j=1}^n p(B | A_j) p(A_j)}, \quad p(B) \neq 0. \quad (13.1.1)$$

Зазначимо, що інтерпретація події A_j має персональний характер. Тобто, ймовірність настання події A_j (до появи B , яка може вплинути на A) визначається апіорно і називається *апіорною ймовірністю* події A_j . Аналогічно, $p(A_j | B)$ в рівнянні (13.1.1) є *апостеріорною ймовірністю* події

A_j , обчисленою при умові, що мала місце подія B . Таким чином, можливою інтерпретацією теореми (13.1.1) є наступна: вона представляє собою *механізм для оновлення ступеня упевненості в тому, що відбудеться A_j у світлі нової інформації B* , яка має відношення до A_j . Первинна (апріорна) інформація $p(A_j)$ замінюється уточненою $p(A_j|B)$. З цієї точки зору теорему можна розглядати як правило навчання на основі минулого досвіду. Оновлення інформації відбувається від початкового ступеня довіри (або ймовірності події) $p(A_j)$ до оновленої ймовірності $p(A_j|B)$.

13.1.1 Теорема Байєса для взаємодоповнюючих подій

Окремим випадком теореми (13.1.1) є її формулювання для взаємодоповнюючих подій. Позначимо через \bar{C} подію, яка є доповненням для C . Тобто, C і \bar{C} – взаємовиключні події, які повністю описують деякий процес. Для будь-якої іншої події B ($p(B) \neq 0$) теорема Байєса для таких подій має вигляд:

$$p(C|B) = \frac{p(B|C)p(C)}{p(B|C)p(C) + p(B|\bar{C})p(\bar{C})}. \quad (13.1.2)$$

13.1.2 Роль апріорних ймовірностей

Апріорна ймовірність $p(A_j)$ необхідна для теореми Байєса для подій (13.1.1). Апріорні ймовірності – це ступені віри (довіри), які доступні аналітику до появи та аналізу даних, які мають відношення до проблеми. У тих випадках, коли подальші дані щодо проблеми отримати неможливо, то апріорними ймовірностями можна скористатись для формулювання висновку (прийняття рішень). Тобто, ТБ в таких випадках не використовується. Наприклад, у політиці більшість рішень приймаються без чисельних даних, вони приймаються на основі суджень із врахуванням наявної (різномірної) інформації. Все що є у розпорядженні аналітика в даному випадку – це апріорна інформація. З цього випливає, що до визначення апріорних ймовірностей потрібно підходити дуже виважено.

В науці, бізнесі, праві, техніці і медицині висновки та прийняття рішень щодо невідомих величин приймаються на основі навчання, попереднього досвіду та розуміння теорії, яка лежить в основі явищ та процесів. Якщо вдається зібрати досить великий об'єм інформації щодо проблеми, то апріорні ймовірності, як правило, опускаються з розгляду і для подальшого аналізу використовуються тільки зібрані дані. Однак, якщо об'єм зібраної інформації

невеликий, то при формулюванні висновку апіорна інформація може відігравати значну роль.

13.1.3 Апостеріорні ймовірності

Апостеріорні ймовірності – це ймовірності, які ми отримуємо в результаті застосування теореми Байєса. Для того щоб розв’язок задачі був коректним, сума апостеріорних ймовірностей взаємовиключних подій, які повністю описують деякий процес, повинна дорівнювати 1. Якщо після виконання аналізу даних за допомогою ТБ появились нові дані, то апостеріорні ймовірності, отримані на основі попередніх даних, необхідно використати як апіорні ймовірності для нових даних.

13.1.4 Відношення шансів подій

Відношення доповнюючих ймовірностей – це альтернативне представлення апіорних або апостеріорних ймовірностей. Так, якщо, p – ймовірність події E , то шанс *проти* появи події E визначається відношенням: $\frac{1-p}{p}$, а шанс *на користь* появи події E визначається відношенням $\frac{p}{1-p}$.

Приклад 13.1. Застосування теореми Байєса для подій: розпізнавання відбитків пальців. Знімання відбитків пальців запропоновано Френсісом Гальтоном в 1890-х роках. Цей метод швидко знайшов широке застосування для розпізнавання (ідентифікації) людей за відбитками їх пальців, які на змінюються на протязі всього життя. Існує також метод ідентифікації особи, який ґрунтується на визначенні її групи крові. Наприклад, до групи A відносяться особи, які мають гени типу A ; до групи B – ті, що мають гени типу B ; до групи AB – особи, які мають гени обох типів – A і B і до групи O відносять тих, що не мають генів жодної з груп – A і B . Однак, виникає запитання, що робити у випадках, коли аналіз крові виконати неможливо?

Пізніше було встановлено, що люди сильно відрізняються один від одного за будовою ланцюжків дезоксирибонуклеїнової кислоти (ДНК), які можна знайти у кожній клітині тіла. Тобто, в окремих випадках, для ідентифікації особи можна скористатись дослідженням клітин тіла замість відбитків пальців або аналізів крові. Ця методика була розроблена в Англії професором А. Джеффри в університеті Лісестера в 1984 році. Згідно з цією методикою, для ідентифікації особи необхідно порівняти кілька різних характеристик ДНК для окремо взятих зразків. Розглянемо основні принципи розпізнавання за цією методикою.

Нехай на місці злочину знайдено матеріал, який містить інформацію щодо ДНК *злочинця* (наприклад, краплі крові, пучок волосся, сім'я, зідрану шкіру і т.ін.). Припустимо, слідчі коректно вважають, що (*підозрювана*) особа, ДНК якої виявлено на місці злочину, дійсно вчинила цей злочин. Таким чином, ДНК підозрюваного необхідно співставити з ДНК дійсного злочинця. Застосуємо для розв'язку цієї задачі теорему Байєса.

Введемо в розгляд такі події:

- C – підозрюваний вчинив злочин;
- \bar{C} – підозрюваний не вчинив злочин;
- M – ДНК підозрюваного та ДНК, визначене за допомогою знайдених частинок тіла, співпадають.

Нехай дослідження підозрюваного показало, що його ДНК збігається з ДНК, встановленого за допомогою частинок тіла, знайдених на місці злочину. В результаті декларується збіжність двох ДНК. Чи дійсно в такому випадку підозрюваний вчинив злочин? За теоремою Байєса (13.1.2) *апостеріорна ймовірність* того, що підозрюваний вчинив злочин, при умові збіжності типів ДНК:

$$p(C|M) = \frac{p(M|C)p(C)}{p(M|C)p(C) + p(M|\bar{C})p(\bar{C})}. \quad (13.1.3)$$

Введемо позначення: p_{ca} – апіорна ймовірність події C , тобто, апіорну ймовірність події $p(C)$ і запишемо теорему у вигляді:

$$p(C|M) = \frac{p(M|C)p_{ca}}{p(M|C)p_{ca} + p(M|\bar{C})(1-p_{ca})} = \frac{p(M|C)p_{ca}}{p(M)}.$$

Оскільки

$$p(\bar{C}|M) = \frac{p(M|\bar{C})(1-p_{ca})}{p(M|\bar{C})(1-p_{ca}) + p(M|C)p_{ca}}, \quad (13.1.4)$$

то *апостеріорний шанс* того, що підозрюваний не вчинив злочин має вигляд:

$$\frac{p(\bar{C}|M)}{p(C|M)} = \frac{p(M|\bar{C})}{p(M|C)} \cdot \frac{(1-p_{ca})}{p_{ca}}, \quad (13.1.5)$$

де $[(1-p_{ca})/p_{ca}]$ – апіорний шанс того, що підозрюваний не вчинив злочин. З цього прикладу видно, що перевагою використання відношення шансів є те, що немає необхідності обчислювати безумовну ймовірність $p(M)$, тобто, немає необхідності обчислювати *константу пропорційності* (знаменник) в теоремі Байєса.

Оскільки $p(M | C) = 1 - p(\bar{M} | C)$, можна записати:

$$r = \frac{p(M | C)}{p(M | \bar{C})} = \frac{p(M | C)}{1 - p(\bar{M} | C)}, \quad (13.1.6)$$

а із врахуванням (13.1.5):

$$\frac{p(C | M)}{p(\bar{C} | M)} = r \left[\frac{p_{ca}}{1 - p_{ca}} \right], \quad (13.1.7)$$

де r – *фактор Байєса*, який представляє собою відношення апостеріорного шансу (на користь того, що підозрюваний скоїв злочин) до апріорного шансу, який оправдовує підозрюваного. Таким чином, незалежно від значення p_{ca} , чим більшим буде фактор Байєса, тим більшим буде шанс на користь події C (підозрюваний скоїв злочин). Зазначимо, що значення r залежить тільки від даних. В подальшому фактор Байєса буде використовуватись при перевірці гіпотез.

Підозрюваного бачили біля місця злочину приблизно в той же час, коли було скоєно злочин. Для простоти обчислень прийmemo, що апріорна ймовірність $p_{ca} = 0,5$.

В даному прикладі є ймовірності двох помилок:

– $p(\bar{M} | C)$ – ймовірність висновку, що немає співпадання ДНК, якщо воно повинне бути;

– $p(M | \bar{C})$ – ймовірність того, що є співпадання ДНК, коли його не має бути (припустимо, що підозрюваний – випадкова особа, яке не є злочинцем).

Ці ймовірності визначаються поточним рівнем розвитку технології визначення ДНК, припустимо (за статистичними даними), що $p(M | \bar{C}) = 0,02$, а $p(\bar{M} | C) = 0,001$.

Обчислення цих значень досить складне і залежить від результатів порівняння ДНК підозрюваного з ДНК осіб, які мають таке ж етнічне походження (з відповідної бази даних), що і підозрюваний. Воно залежить також від результатів порівняння декількох незалежних характеристик цих двох зразків ДНК (оцінених за різними зразками).

Використовуючи наведені вище значення для $p(M | \bar{C})$, $p(\bar{M} | C)$, а також те, що $p(M | C) = 1 - p(\bar{M} | C) = 1 - 0,001 = 0,999$, обчислимо ймовірність того, що підозрюваний вчинив злочин:

$$p(C | M) = \frac{p(M | C) p_{ca}}{p(M | C) p_{ca} + p(M | \bar{C})(1 - p_{ca})} =$$

$$= \frac{0,999 \cdot 0,5}{0,999 \cdot 0,5 + 0,02 \cdot 0,5} = 0,9804,$$

або біля 98%. З (13.1.8) знайдемо, що фактор Байєса $r = \frac{0,9804}{1 - 0,9804} = 50,02$.

Апостеріорний шанс того, що підозрюваний не скоїв злочин складає: $r(1-p)/p = 0,02002$, або апостеріорне свідчення на користь того, що підозрюваний скоїв злочин складає біля 50:1.

Таким чином, збіжність ДНК є, в даному випадку, переконливим доведенням того, що підозрюваний скоїв злочин. У випадку, коли апіорний шанс того, що підозрюваний скоїв злочин, відрізняється від $|1|$, то необхідно пожити його на r для того щоб отримати новий фактор Байєса.

13.2 Теорема Байєса для дискретних даних та дискретних параметрів

Нехай X_1, \dots, X_n – взаємно незалежні однаково розподілені спостережувані випадкові змінні з функціями густини розподілу $f(x | \Theta = \theta)$, де Θ – невідома і неспостережувана дискретна випадкова змінна, можливі значення якої $\theta_1, \dots, \theta_p$ визначають конкретний розподіл $(X | \Theta = \theta)$. Необхідно визначити ймовірність (шанс) того, що $\Theta = \theta$. Теорема Байєса для даного випадку:

$$p\{\Theta = \theta | x_1, \dots, x_n\} \equiv h(\theta | \text{дані}) = \frac{f(x_1 | \theta) \dots f(x_n | \theta) g(\theta)}{\sum_{i=1}^m f(x_1 | \theta_i) \dots f(x_n | \theta_i) g(\theta_i)}, \quad (13.2.1)$$

де $h(\cdot)$ – апостеріорна щільність розподілу ймовірностей для Θ ; $g(\cdot)$ – апіорна щільність розподілу ймовірностей для Θ . Очевидно, що знаменник виразу (13.2.1) не повинен дорівнювати нулю.

Інтерпретація теореми Байєса для дискретних даних і дискретних параметрів

Зазначимо, що X і Θ мають дискретні розподіли ймовірностей. При цьому значення X_i , отримані в результаті конкретного експерименту, повинні нести інформацію щодо ймовірностей тих чи інших значень Θ . Знаменник

виразу (13.2.1) – це константа, яка залежить тільки від даних X і не залежить від θ . Таким чином, цей знаменник є константою по відношенню до Θ . Це дає можливість записати рівняння (13.2.1) в еквівалентній формі:

$$h(\theta | x_1, \dots, x_n) = c \cdot f(x_1 | \theta) \dots f(x_n | \theta) g(\theta) \propto f(x_1 | \theta) \dots f(x_n | \theta) g(\theta), \quad (13.2.2)$$

де c – коефіцієнт пропорційності; \propto – символ пропорційності. Константу c завжди можна відновити, оскільки її необхідно визначати через

$$\sum_{i=1}^m p\{\Theta = \theta_i | x_1, \dots, x_n\} = \sum_{i=1}^n h(\theta_i | x_1, \dots, x_n) = 1. \quad (13.2.3)$$

Вираз $f(x_1 | \theta) \dots f(x_n | \theta)$ в (13.2.2) зазвичай називають функцією правдоподібності для даних. Таким чином, з теореми Байєса для дискретних даних і дискретних параметрів у формі (13.2.2) випливає, що апостеріорний умовний розподіл (щільність) ймовірностей для Θ є пропорційним добутку функції правдоподібності для даних і апіорної щільності ймовірностей. Даний результат можна є справедливим як для одновимірних, так і для багатовимірних даних. При цьому елементи даних X_i не повинні бути взаємно незалежними; вони можуть бути корельованими, що, як правило, має місце у стохастичних процесах. Однак, в такому випадку в рівнянні (13.2.2) необхідно використовувати правдоподібність у формі: $f(x_1, \dots, x_n | \theta)$.

Приклад 13.2. Аналіз якості на виробництві: дискретні дані і дискретні параметри (висновок щодо пропорційності).

Компанія *МММ* є виробником плат оперативної пам'яті (оперативні запам'ятовуючі пристрої або ОЗП) для персональних комп'ютерів. Виробництво модулів пам'яті сягає сотень тисяч одиниць на місяць. Позначимо через θ – ймовірність виготовлення дефектного модуля пам'яті для серії з n одиниць. Для простоти приймемо, що θ може приймати три можливих значення:

- 0,25 – хороший результат;
- 0,50 – прийнятний результат;
- 0,75 – поганий результат.

Статистика виробничого процесу компанії *МММ* свідчить, що на протязі п'яти попередніх років виробництва ОЗП утримувались наступні показники якості процесу:

– на протязі 60% часу вироблялись пристрої з ймовірністю дефекту $\theta = 0,25$; – на протязі 30% часу вироблялись пристрої з ймовірністю дефекту $\theta = 0,50$;

– на протязі 10% часу з попередніх 5 років вироблялись пристрої з ймовірністю дефекту $\theta = 0,75$.

Компанія вирішила скористатись цими результатами як апіорними ймовірностями для того щоб спрогнозувати рівень якості продукції в майбутньому. Значення цих апіорних ймовірностей наведені в таблиці 13.1.

Таблиця 13.1. Апіорні ймовірності

	Якість продукції		
	Хороша	Прийнятна	Погана
Ймовірність дефекту (θ)	0,25	0,50	0,75
Щільність ймовірностей для $\theta(p(\theta))$	0,60	0,30	0,10

Після виробництва 10000 пристроїв оперативної пам'яті компанія вирішила перевірити цю партію для того щоб встановити – чи підтримується якість на попередньому рівні?

Виявилось, що з трьох пристроїв пам'яті, випадково вибраних для перевірки, два пристрої мають дефекти. Який висновок щодо якості продукції необхідно зробити, тобто, яким є апостеріорний розподіл для θ ?

В даному випадку дані мають дискретний характер. Припустимо, що вони мають біноміальний розподіл з параметром θ .

– кількість успішних подій $r = 2$ (успішними вважаємо події, пов'язані з випуском дефектних пристроїв) з трьох можливих $n = 3$.

Таким чином, функція правдоподібності має вигляд:

$$L(\theta) = C_2^3 \theta^2 (1 - \theta),$$

де $\theta = [0,25; 0,5; 0,75]$. Рівняння (13.2.2) в даному випадку має вигляд:

$$h(\theta | r, n) \propto L(r | \theta, n) g(\theta).$$

Обчислимо правдоподібність $L(\theta)$:

$$\text{якщо } \theta = 0,25, \text{ то } L(\theta) = \frac{1 \cdot 2 \cdot 3}{1 \cdot 2} (0,25)^2 (1 - 0,25) = 0,140625;$$

$$\text{якщо } \theta = 0,50, \text{ то } L(\theta) = \frac{1 \cdot 2 \cdot 3}{1 \cdot 2} (0,50)^2 (1 - 0,50) = 0,375000;$$

$$\text{якщо } \theta = 0,75, \text{ то } L(\theta) = \frac{1 \cdot 2 \cdot 3}{1 \cdot 2} (0,75)^2 (1 - 0,75) = 0,421875.$$

Апіорні та апостеріорні щільності ймовірності для h наведені в таблиці 13.2.

Таблиця 13.2. Апостеріорні ймовірності

θ	Апріорні ймов. для θ	Правдоподібність $L(\theta)$	(апріорна ймов.) \times (правдоподібн.)	$h =$ апостер. щільн. Для θ
0,25	0,60	0,140625	0,0843750	0,35294
0,50	0,30	0,375000	0,1125000	0,47059
0,75	0,10	0,421875	0,0421875	0,17647
Загальна сума:			0,2390625	0,99999

Елементи в стовпчику h знайдені в результаті ділення значень третього стовпчика ((апріорна ймов.) \times (правдоподібн.)) на загальну суму для цього (передостаннього) стовпчика з метою нормування остаточного результату до 1, тобто:

$$p(\Theta = 0,25 | r, n) \equiv h(0,25 | 2,3) = \frac{0,60 \cdot 0,140625}{0,6 \cdot 0,14625 + 0,3 \cdot 0,375 + 0,1 \cdot 0,421875} =$$

$$= \frac{0,084375}{0,2390625} = 0,35294;$$

$$p(\Theta = 0,50 | r, n) \equiv h(0,50 | 2,3) = \frac{0,30 \cdot 0,375}{0,6 \cdot 0,14625 + 0,3 \cdot 0,375 + 0,1 \cdot 0,421875} =$$

$$= \frac{0,084375}{0,2390625} = 0,47059;$$

$$p(\Theta = 0,75 | r, n) \equiv h(0,75 | 2,3) = \frac{0,10 \cdot 0,421875}{0,6 \cdot 0,14625 + 0,3 \cdot 0,375 + 0,1 \cdot 0,421875} =$$

$$= \frac{0,0421875}{0,2390625} = 0,17647.$$

З таблиці 13.2 видно, що найбільш ймовірним значенням θ , на основі аналізу якості останньої партії пристроїв пам'яті, є $\theta = 0,5$, тобто, виробництво знаходиться на "прийнятному" рівні щодо якості готової продукції.

Очевидно, що перед тим як зробити висновок щодо модифікації виробничого процесу з метою підвищення якості продукції, компанія *МММ* повинна тестувати більше число блоків пам'яті ніж $n = 3$; наприклад, $n = 100$, 500 або 1000.

Теорема Байєса для дискретних даних і дискретних моделей

Припустимо, що ставиться задача порівняти декілька моделей M_1, M_2, \dots, M_k з метою визначення – яка з них є найкращою (правдоподібною). В даному випадку ситуація схожа на ту, що мала місце в прикладі 2.2, де було три альтернативи для ступеня дефектності θ пристроїв оперативної пам'яті. Тобто ми можемо розглядати можливі значення θ як показники якості моделей, кожна з яких є претендентом на пояснення деякого явища. Можна мати деяке уявлення щодо ймовірності наближення кожної з моделей до явища, що розглядається, але для цього потрібно знайти обґрунтування. Для обґрунтування можна скористатись експериментальними даними з метою порівняння апостеріорних ймовірностей, що характеризують наближення моделей до цих даних. Теорему Байєса для дискретного числа моделей можна записати у вигляді:

$$p(M_i | X = x) = \frac{f(x | M_i) p(M_i)}{\sum_{j=1}^k f(x | M_j) p(M_j)}, \quad (13.2.4)$$

де $i = 1, \dots, k$; $f(x | M_i)$ – правдоподібність дискретних даних щодо моделі M_i , яка визначається через щільність розподілу.

13.3 Теорема Байєса для неперервних даних і дискретного параметра

Припустимо, що X – спостережувана, абсолютно неперервна змінна з функцією розподілу $f(x | \theta)$. Позначимо через Θ неспостережуваний невідомий дискретний параметр розподілу змінної X , а через $g(\theta)$ – апіорну щільність розподілу Θ . В даному контексті теорема Байєса для апостеріорної щільності розподілу θ визначається виразом:

$$h(\theta | x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n | \theta) g(\theta)}{\sum_{i=1}^k L(x_1, \dots, x_n | \theta_i) g(\theta_i)}, \quad (13.3.1)$$

де $L(x_1, \dots, x_n | \theta)$ – функція правдоподібності для θ . Форма (13.3.1) еквівалентна наступній:

$$h(\theta | x) \propto L(x_1, \dots, x_n | \theta) g(\theta).$$

Інтерпретація теореми Байєса для неперервних даних і дискретного параметра

Фактичні дані можуть бути багатовимірними і можуть бути взаємно залежними; але вони повинні бути неперервними, а знаменник (13.3.1) повинен бути відмінним від нуля.

Приклад 13.3. Група студентів складається з трьох підгруп. Для тестування необхідно вибрати студента із однієї з цих трьох підгруп. (Неперервні дані і дискретний параметр.)

Таким чином, для тестування необхідно вибрати одного студента з групи, яка складається з трьох підгруп. Задача полягає в тому, щоб визначити: з якої підгрупи вибрали студента для тестування?

Спочатку розглянемо яка апіорна інформація може бути використана. Припустимо, що в минулому році всі три підгрупи навчалися з тими ж викладачами, що і поточного року; екзамен був таким же, а результати тестування студентів в деякий момент часу були такими, як наведені в таблиці 13.3. З цієї таблиці видно, що оцінки студентів мають нормальний розподіл із середніми значеннями m_i та дисперсією $S_i^2 = 225$, $i = 1, 2, 3$.

Таблиця 13.3. Моделі минулорічних оцінок студентів

Модель	Середнє	Модель	Апіорна ймовірн. $p\{M\}$
M_1	$\theta_1 = 74$	$N(74, 225)$	0,25
M_2	$\theta_2 = 81$	$N(81, 225)$	0,50
M_3	$\theta_3 = 68$	$N(68, 225)$	0,25

Для того щоб допомогти передбачити оцінки, які будуть отримані поточного року, необхідно присвоїти апіорні ймовірності оцінкам поточного року на основі інформації щодо минулорічних результатів та поточної інформації щодо студентів. Випадково виберемо студента A , який має хороші поточні оцінки. Також відомо, що кращою серед трьох підгруп є друга підгрупа, яка характеризується моделлю M_2 . На основі наявної інформації ми присвоюємо апіорні ймовірності трьом моделям так, як показано в 4-у стовпчику таблиці 13.3.

Правдоподібність (щільність ймовірностей) для нормального розподілу визначається за виразом:

$$L(x|\theta_i) = \frac{1}{15\sqrt{2\pi}} \exp\left\{-0,5\left(\frac{x-\theta_i}{15}\right)^2\right\},$$

$$i=1,2,3.$$

Оскільки в поточному році студент A набрав 76% від максимальної оцінки, то для обчислення функцій правдоподібності скористаємось оцінкою $x=0,76$. Знайдемо z -оцінки (нормовані або стандартизовані випадкові змінні) за виразом:

$$z_i = \frac{x - \theta_i}{15}.$$

Оскільки $z_i \sim N(0, 1)$ із стандартною нормальною щільністю розподілу

$$\Phi(z_i) \equiv \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_i^2},$$

то значення правдоподібності можна визначити за допомогою таблиці площ стандартного нормального розподілу. Відповідні результати наведені в таблиці 13.4.

Таблиця 13.4. Апостеріорні ймовірності для моделей оцінок студентів

Модел ь	z_i	$\Phi(z_i)$	Правдопод. $\Phi(z_i)/15$	Апріорн і ймовірн.	(Правдопо д.) \times (апріорні)	Апост еріор ні
M_1	0,133	0,395	0,02633	0,25	0,0065825	0,264
M_2	-0,333	0,377	0,02513	0,50	0,0125650	0,504
M_3	0,533	0,346	0,02307	0,25	0,0057675	0,231
Загальна сума					0,0249150	0,999

З таблиці 13.4 видно, що модель M_2 має найвищу апостеріорну ймовірність з трьох моделей. Таким чином, робимо висновок, що випадково вибраний для тестування студент навчається в третій групі.

Зазначимо, що в даному прикладі ми скористались тільки одним експериментальним значенням. Можна було б поставити задачу визначення підгрупи для кількох студентів, які належать до однієї підгрупи. Це збільшило б об'єм експериментальних даних. Більше того, оскільки вибірка складалась тільки з одного значення (оцінка студента A), то це мало несуттєвий вплив на зміну апріорних ймовірностей, тобто, апріорні та апостеріорні ймовірності майже однакові. Якби вибірка даних була більшою, то апріорні ймовірності не мали б такого значного впливу на результат.

13.4 Теорема Байєса для дискретних даних і неперервного параметра

Теорема Байєса для дискретних даних і неперервного параметра θ з щільністю розподілу $g(\theta)$ має вигляд:

$$h(\theta | x_1, \dots, x_n) = \frac{L(x_1, \dots, x_n | \theta) g(\theta)}{\int_{\theta} L(x_1, \dots, x_n | \theta) g(\theta) d\theta}, \quad (13.4.1)$$

де інтегрування виконується по усіх значеннях θ ; $L(x_1, \dots, x_n | \theta)$ – функція правдоподібності, а $g(\theta)$ – щільність ймовірності. Апостеріорний розподіл неперервний, тобто, h – щільність розподілу.

Приклад 13.4. Аналіз якості на виробництві: дискретні дані і неперервний параметр.

Розглянемо ще раз приклад 13.2, пов'язаний з аналізом якості на виробництві. Однак тепер θ – неперервний параметр (ймовірність виробництва пристрою пам'яті з дефектом), який має обмеження: $0 \leq \theta \leq 1$. В прикладі 13.2 з трьох випадково вибраних пристроїв два виявились з дефектом. В даному випадку також скористаємось біноміальним розподілом (правдоподібністю), але θ вже не обмежується трьома значеннями. Необхідно вибрати апріорний розподілу для θ .

Випадок 1: рівномірний розподіл. Припустимо, що фірма-виробник не має статистики щодо дефектних пристроїв пам'яті за минулі роки і на момент виконання статистичного аналізу немає інформації – чи всі партії виробів зовсім без дефектів, чи більшість виробів – дефектні, чи якість посередня. В подібній ситуації для θ необхідно вибрати рівномірний апріорний розподіл на одиничному інтервалі:

$$g(\theta) = \begin{cases} 1, & 0 < \theta < 1, \\ 0, & \text{інакше} \end{cases}$$

Апостеріорна щільність розподілу для θ (по відношенню до міри Лебега) визначається виразом:

$$h(\theta | r, n) = \frac{L(r | \theta, n) g(\theta)}{\int_0^1 L(r | \theta, n) g(\theta) d\theta}.$$

Для конкретних значень $r = 2$, $n = 3$ маємо:

$$h(\theta | r = 2, n = 3) = \frac{C_2^3 \theta^2 (1 - \theta)}{\int_0^1 C_2^3 \theta^2 (1 - \theta) d\theta} = \frac{\theta^2 (1 - \theta)}{\int_0^1 \theta^2 (1 - \theta) d\theta} =$$

$$= \frac{\theta^2 (1-\theta)}{B(3,2) \int_0^1 \frac{1}{B(3,2)} \theta^{3-1} (1-\theta)^{2-1} d\theta} = \frac{1}{B(3,2)} \theta^{3-1} (1-\theta)^{2-1},$$

де

$$B(a,b) \equiv \int_0^1 y^{a-1} (1-y)^{b-1} dy; \quad (13.4.2)$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}; \quad (13.4.3)$$

$\Gamma(a)$ – гамма-функція аргумента a . Бета-розподіл випадкової змінної X з параметрами a і b ($a > 0, b > 0$) визначається за виразом:

$$f(x|a,b) = \begin{cases} \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1}, & 0 < x < 1, \\ 0, & \text{інакше} \end{cases} \quad (13.4.4)$$

Таким чином, видно, що апостеріорною щільністю розподілу для θ є бета-розподіл з параметрами 3 і 2. Рівномірний апріорний та апостеріорний бета-розподіл для θ показані на рис. 13.1.

Випадок 2: за апріорний вибираємо бета-розподіл. Тепер вважаємо, що є більше інформації щодо θ , тобто, приймаємо, що апріорним розподілом для θ є бета-розподіл. Апостеріорна щільність розподілу для θ визначається по аналогії з першим випадком, але тепер $g(\theta)$ – бета-розподіл.

Відповідно, вибираємо наступну функцію бета-розподілу:

$$g(\theta) = \begin{cases} \frac{1}{B(a,b)} \theta^{a-1} (1-\theta)^{b-1}, & 0 < \theta < 1, \\ 0, & \text{інакше} \end{cases} \quad (13.4.5)$$

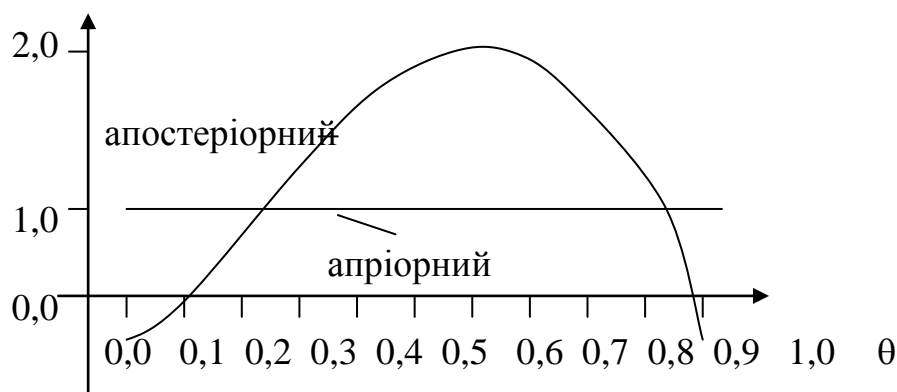


Рис. 13.1. Рівномірна апріорна щільність і апостеріорна щільність бета (3,2): θ – ймовірність виготовлення дефектного блока пам'яті.

для належних значень параметрів a і b , які конкретизують вигляд розподілу. Визначимо апостеріорну щільність розподілу для θ :

$$h(\theta | r = 2, n = 3) = \frac{[C_2^3 \theta^2 (1 - \theta)] \left[\frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \right]}{\int_0^1 [C_2^3 \theta^2 (1 - \theta)] \left[\frac{1}{B(a, b)} \theta^{a-1} (1 - \theta)^{b-1} \right] d\theta}.$$

Отриманий вираз можна спростити:

$$\begin{aligned} h(\theta | r = 2, n = 3) &= \frac{\theta^{a-1} (1 - \theta)^b}{\int_0^1 \theta^{a+1} (1 - \theta)^b d\theta} = \\ &= \frac{\theta^{(a+2)-1} (1 - \theta)^{(b+1)-1}}{B(a+2, b+1) \int_0^1 \frac{\theta^{(a+2)-1} (1 - \theta)^{(b+1)-1}}{B(a+2, b+1)} d\theta} \end{aligned}$$

і оскільки інтеграл в знаменнику останнього виразу повинен дорівнювати 1, то апостеріорну щільність для θ можна записати у вигляді:

$$h(\theta | r = 2, n = 3) = \frac{1}{B(a+2, b+1)} \theta^{(a+2)-1} (1 - \theta)^{(b+1)-1},$$

$$0 < \theta < 1.$$

Зазначимо, що так само як і в першому випадку апостеріорна щільність розподілу знову належить до сімейства бета-розподілу, але з іншими параметрами. Використовуючи апріорну інформацію, необхідно знайти параметри a і b .

Відомо, що статистичні параметри бета-розподілу (13.4.5) визначаються за виразами:

$$E(\theta) = \frac{a}{a+b}; \quad \text{var}(\theta) = \frac{ab}{(a+b)^2 (a+b+1)}.$$

Припустимо що наявна інформація щодо виробничої діяльності компанії дала можливість визначити, що $E(\theta) = 0,2$ і $\text{var}(\theta) = 0,01$. Розв'язок наведених виразів щодо a і b дає наступний результат: $a = 3$; $b = 12$. Таким чином, можемо конкретизувати апріорну щільність розподілу:

$$g(\theta) = \frac{1}{B(3, 12)} \theta^{3-1} (1 - \theta)^{12-1},$$

а підстановка значень параметрів в апостеріорну щільність приводить до виразу:

$$h(\theta | r = 2, n = 3) = \frac{1}{B(5,13)} \theta^{5-1} (1-\theta)^{13-1},$$

$$0 < \theta < 1.$$

Зазначимо, що апіорна та апостеріорна щільності розподілів цілком схожі між собою (рис. 13.5). Це можна пояснити незначним числом експериментальних даних, $n = 3$, які не змінили суттєво апіорну інформацію.

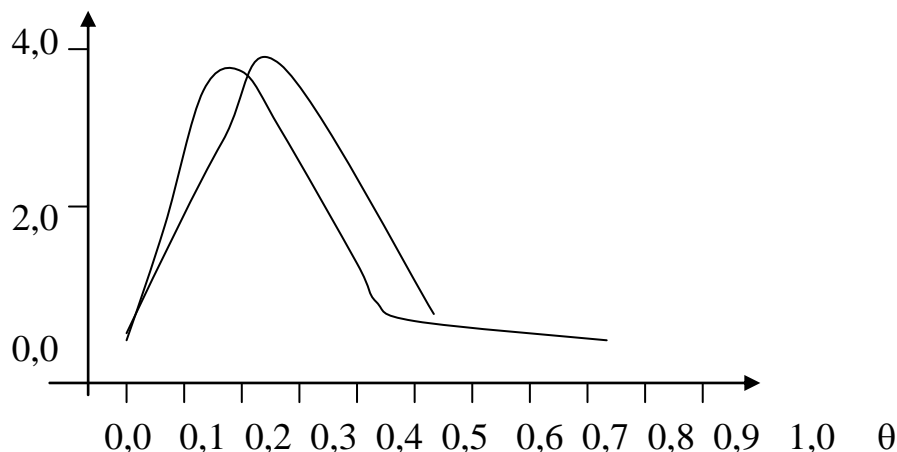


Рис. 13.5. Апіорна щільність бета (3,12) і апостеріорна щільність бета (5,13): θ – ймовірність виготовлення дефектного блока пам'яті.

Наведені ілюстративні приклади застосування теореми Байєса свідчать про те, що вона представляє собою потужний інструмент ймовірнісного аналізу даних. При цьому існує широка можливість використання експертних оцінок для оцінювання апіорних ймовірностей та аналізу ситуацій.

Подальший розвиток цього інструментарію знайшов своє вираження у байєсівських мережах (БМ) – графічних моделях, які відображають наявні причинно-наслідкові зв'язки між змінними досліджуваних об'єктів. Байєсівська мережа – це універсальний інструмент аналізу складних об'єктів з таких причин: – такі моделі можуть бути надзвичайно високої розмірності; – вони можуть поєднувати числові та якісні змінні; – для оцінювання структури і параметрів БМ існує множина методів, які взаємно доповнюють один одного; – для формування ймовірнісного висновку існують точні та наближені методи, які надають можливість знайти розв'язок у складних ситуаціях; – мережа може бути статистичною або динамічною залежно від того, як формулюється постановка задачі. Дослідженню можливостей застосування таких моделей будуть присвячені наступні публікації авторів.

Перелік посилань

Основна література

1. *Бард Й.* Нелинейное оценивание параметров. – Москва: Статистика, 1979. – 346 с.
2. *Бідюк П.І.* Системний підхід до прогнозування на основі моделей часових рядів // Системні дослідження та інформаційні технології, 2003, № 3, с. 88-110.
3. *Бідюк П.І.* Моделювання і прогнозування гетероскедастичних процесів // Системні дослідження та інформаційні технології, 2004, № 1, с. 115-134.
4. *Бідюк П.І., Савенков О.І., Баклан І.В.* Часові ряди: моделювання і прогнозування. – Київ: ЕКМО, 2003. – 144 с.
5. *Бідюк П.І., Литвиненко В.І.* Моделювання і прогнозування гетероскедастичних процесів // Збірник наукових праць Миколаївського гуманітарного університету ім. Петра Могили, 2004, Вип. 22, с.24-39.
6. *Бідюк П.І., Половцев О.В.* Аналіз і моделювання економічних процесів перехідного періоду. – Київ: ПЛАБ-75, 1999. – 140 с.
7. *Бокс Дж., Дженкінс Г.* Анализ временных рядов / в 2-х томах. – Москва: Мир, 1974. – 406 с.
8. *Грін В. Г.* Економетричний аналіз. – Київ: Основи, 2005. – 1198 с.
9. *Гроуп Д.* Методы идентификации систем. – Москва: Мир, 1979. – 304 с.
10. *Дэннис Дж., Шнабель Р.* Численные методы безусловной оптимизации и решения нелинейных уравнений. – Москва: Мир, 1988. – 440 с.
11. *Зайченко Ю.П.* Основи проектування інтелектуальних систем. – Київ: Слово, 2003. – 351 с.
12. *Згуровский М.З., Подладчиков В.Н.* Аналитические методы калмановской фильтрации. – Київ: Наукова думка, 1995. – 285 с.
13. *Згуровский М.З., Бідюк П.І.* Анализ и управление большими космическими конструкциями. – Київ: Наукова думка, 1997. – 450 с.
14. *Зельнер А.* Байесовские методы в эконометрии. – Москва: Статистика, 1980. – 438 с.
15. *Ивахненко А.Г., Лапа В.Г.* Предсказание случайных процессов. – К.: Наукова думка, 1971. – 416 с.
16. *Кашьяп Р.Л., Рао А.Р.* Построение динамических стохастических моделей по экспериментальным данным. – Москва: Наука, 1983. – 385с.
17. *Киричков В.Н.* Идентификация объектов систем управления технологическими процессами. – Київ: Вища школа, 1990. – 263 с.
18. *Кроптя А.В., Бідюк П.І.* Аналіз і методи оцінювання екстремальних значень // Наукові вісті НТУУ «КПІ», 2005, № 4, с. 34 – 47.
19. *Кузовков Н.Т.* Модальное управление и наблюдающие устройства. – М.: Машиностроение, 1976. – 184 с.
20. *Лук'яненко І.Г., Городніченко Ю.О.* Сучасні економетричні методи у фінансах. – Київ: Літера, 2002. – 350 с.

21. *Марпл С.Л.* Цифровой спектральный анализ.– Москва: Мир, 1990. – 584 с.
22. *Моррис У.* Наука об управлении – байесовский подход. – Москва: Мир, 1971. – 304 с.
23. *Райбман Н.С., Капитоненко В.В., Овсепян Ф.А., Варлаки П.М.* Дисперсионная идентификация. – Москва: Наука, 1981. – 336 с.
24. *Саркисян С.А., Каспин В.И., Лисичкин В.А., Минаев Э.С., Пасечник Г.С.* Теория прогнозирования и принятия решений. – Москва: Высшая школа, 1977. – 351 с.
25. *Себер Дж.* Линейный регрессионный анализ. – Москва: Мир, 1980. – 256 с.
26. *Терентьев А.Н., Бидюк П.И.* Эвристический метод построения байесовских сетей // Математичні машини і системи, 2006, № 3, с. 12-23.
27. *Феллер В.* Введение в теорию вероятностей и ее приложения. – Москва: Мир, 1984. – 527 с. (т.1); 751 с. (т.2).
28. *Хей Дж.* Введение в методы байесовского статистического вывода. – Москва: Финансы и статистика, 1987. – 335 с.
29. *Химмельблау Д.* Анализ процессов статистическими методами. – Москва: Мир, 1973. – 957 с.
30. *Четыркин Е.М., Калихман И.Л.* Вероятность и статистика. – Москва: Финансы и статистика, 1982. – 319 с.
31. *Эллиотт Р.* Стохастический анализ и его приложения. – Москва: Мир, 1986. – 352 с.
32. *Ющенко В., Лисицький В.* Гроші: розвиток попиту та пропозиції в Україні. – Київ: Скарби, 1998. – 285 с.
33. *Andreasen 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.
34. *Andrews D.W.K.* Asymptotics for semiparametric econometric models via stochastic equicontinuity // Econometrica, 1994, v. 62, № 1, pp. 43-72.
35. *Bernardo J.M., Smith A.F.M.* Bayesian theory. – New York: John Wiley & Sons, Ltd., 2000. – 586 p.
36. *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.
37. *Casella G., Berger R.L.* Statistical inference. – Pacific Grove, CA: Duxbury, 2002. – 660 p.
38. *Chow C.K., Liu C.N.* Approximating discrete probability distributions with dependence trees // IEEE Transactions on Information Theory, 1968, v.14. – pp. 462-467.
39. *Dahl C.M., Levine M.* Nonparametric estimation of volatility models with serially dependent innovations / TR-09-04, Purdue University, 2004, 31 p.

40. *Dechter R.* Constraint processing. – San Francisco, CA: Morgan Kaufmann Publishers, 2003. – 479 p.
41. *Enders W.* Applied Econometric Time Series. – New York: John Wiley & Sons, Inc., 1995. – 450 p.
42. *Gilks W.R., Richardson S., Spiegelhalter D.J.* Markov Chain Monte Carlo in Practice. – London: Chapman & Hall, 1996. – 486 p.
43. *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.
44. *Johnston J., DiNardo J.* Econometric methods. - New York: McGraw-Hill, Inc., 1997. – 525 p.
45. *Lewis D.C.* Industrial and business forecasting methods. – London: Butterworth Scientific, 1982. – 145 p.
46. *Tsay R.S.* Analysis of financial time series. – New York: J. Wiley & Sons, Inc., 2002. – 455 p.
47. *West J.C.* Nonlinear Signal Distortion Correlation // Int. J. Control, 1965, No.2, pp. 529-538.
48. *Vapnik V.* Statistical Learning Theory. – New York: Wiley, 1998.
49. *Vapnik V.* An Overview of Statistical Learning Theory // IEEE Transactions on Neural Networks, September 1999, 10(5): pp. 988-999.

Додаткова література

1. Айвазян С.А., Бежаева З.И., Староверов О.В. Классификация многомерных наблюдений. – М.: Статистика, 1974. – 240 с.
2. Андерсон Т. Статистический анализ временных рядов. – М.: Мир, 1976. – 755 с.
3. Бідюк П.І. Проектування інформаційних СППР. – Київ: НТУУ «КПІ», 2005. – 121 с.
4. Бриллинджер Д. Временные ряды: обработка данных и теория. – М.: Мир, 1980. – 536 с.
5. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в науке и технике. – М.: Мир, 1980. – 610 с.
6. Довгий С.А., Савенков А.И., Бідюк П.И. Моделирование и прогнозирование экономических процессов переходного периода. – Киев: Укртелеком, 2003. – 240 с.
7. Дорохина Е.Ю., Преснякова Л.Ф., Тихомиров Н.П. Сборник задач по эконометрике. – М.: Экзамен, 2003. – 224 с.
8. Дрейпер Н.Р., Смит Г. Прикладной регрессионный анализ. – Киев: Диалектика, 2007. – 912 с.
9. Дюк В., Самойленко А. Data Mining: учебный курс. СПб: Питер, 2001. – 368 с.

10. Зайченко Ю.П. Нечеткие модели и методы в интеллектуальных системах. – К.: Издательский дом «Слово», 2008. – 344 с.
11. Закс Л. Статистическое оценивание. – М.: Статистика, 1976. – 598 с.
12. Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д. Принятие решений на основе самоорганизации. – Москва, Сов. Радио: 1976. – 363 с.
13. Кельберт М.Я., Сухов Ю.М. Основные понятия теории вероятностей и математической статистики. – М.: МЦНМО, 2007. – 456 с.
14. Кендалл М.Дж., Стьюарт А. Теория распределений. – М.: Наука, 1966. – 588 с.
15. Кобзарь А.И. Прикладная математическая статистика. – М.: Физматлит, 2006. – 816 с.
16. Кокс Д., Хинкли Д. Теоретическая статистика. – Москва: Мир, 1978. – 560 с.
17. Ликеш И., Ляга Й. Основные таблицы математической статистики. – М.: Финансы и статистика, 1985. – 358 с.
18. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. – М.: Финансы и статистика, 2003. – 414 с.
19. Терентьев А.Н., Бидюк П.И. Эвристический метод построения байесовских сетей / Міжнародна НТК „Інтелектуальні системи підтримки прийняття рішень та прикладі аспекти сучасних інформаційних технологій. – Євпаторія, травень 2006, т. 1, с. 401-403.
20. Тихомиров Н.П., Дорохина Е.Ю. Эконометрика. – М.: Экзамен, 2007. – 512 с.
21. Уилкс С. Математическая статистика. – Москва: Наука, 1967. – 632 с.
22. Хей Дж. Введение в методы байесовского статистического вывода. – М.: Финансы и статистика, 1987. – 335 с.
23. Эддоус М., Стэнсфилд Р. Методы принятия решений. – М.: ЮНИТИ, 1997. – 590 с.
24. Casella G., Berger R.L. Statistical Inference. – Pacific Grove, CA: DUXBURY, 2002. – 660 p.
25. Chow C.K., Liu C.N. Approximating discrete probability distributions with dependence trees. // IEE Transactions on information theory, Vol. IT-14, No. 3, May 1968, 6 pp.
26. Chung K.L., AitSahlia F. Elementary Probability Theory. – New York: Springer, 2003. – 470 p.
27. Cuthbert D., Wood F.S. Fitting Equations to Data. – New York: John Wiley & Sons, 1980. – 458 p.
28. Dean T., Kanazava K. A model for reasoning about persistence and causation // Computational Intelligence, 1989, v. 5, No. 3, pp. 142-150.
29. Garg S. Controls and health management technologies for intelligent aerospace propulsion systems / NASA-TM, 2004 – 212915. – 28 p.
30. Ghahramani Z., On structural variational inference / Technical Report: CRG-TR-97-1, University of Toronto, CSD, 1997. – 67 p.

31. 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.
32. *Handbook of Human-Computer Interaction* / Helander M. (ed.). – Amsterdam: Elsevier Science Publishers B.V. (North Holland), 1988. – 1150 p.
33. Kalman R.E., Bucy R.S. New results in linear filtering and prediction // *Journal of Basic Engineering (ASME)*, 1961, D83, pp. 95-108.
34. Magnus J.R., Neudecker H. *Matrix Differential Calculus with Applications*. – New York: John Wiley & Sons, 1999. – 496 p.
35. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition // *IEEE, Trans. ASSP*, 1989, v. 77, No.2, pp. 257-285.
36. Rauch H.E. Solutions to the linear smoothing problem // *IEEE Trans. Automatic Control*, 1963, AC-8(4), pp. 371-372.
37. Rossi P.E., Allenby G.M. Bayesian statistics and marketing // *Marketing Science*, 2003, v. 22, № 3, pp. 304-328.
38. Roweis S., Ghahramani Z. A unifying review of linear Gaussian models // *Neural Computation*, 1999, v. 11, N0.2, pp. 305-345.
39. Stair R.M. *Principles of Information Systems*. – Washington: ITPC, 1996. – 656 p.
40. Strait P.T. *Probability and Statistics with Applications*. – New York: Harcourt Brace Jovanovich, Inc., 1983. – 581 p.
41. Welch G., Bishop G. An introduction to the Kalman filter / *Technical report: TR95-041*. – University of North Carolina, Chapel Hill, 1995. – 16 p.
42. Whitten J.L., Bentley L.D., Barlow V.M. *Systems Analysis & Design Methods*. – Boston: Irwin, 1989. – 797 p.

ДЛЯ НОТАТОК

ДЛЯ НОТАТОК

ДЛЯ НОТАТОК

Наукове видання

**Бідюк Петро Іванович
Гожий Олександр Петрович**

**ЙМОВІРНІСНО-СТАТИСТИЧНІ
МЕТОДИ МОДЕЛЮВАННЯ
І ПРОГНОЗУВАННЯ**

Монографія

Друкується у авторських редакції.
Комп'ютерна верстка, дизайн обкладинки *А. Іщенко*.
Друк *О. Гончарова*. Фальцювальні-палітурні роботи *Ю. Шаповалова*.

Підп. до друку 20.10.2014 р.
Формат 60x84¹/₁₆. Папір офсет.
Гарнітура «Times New Roman». Друк ризограф.
Ум. друк. арк. 51,15. Обл.-вид. арк. 15,5.
Тираж 300 пр. Зам. № 4515.

Видавець і виготовлювач: ЧДУ ім. Петра Могили.
54003, м. Миколаїв, вул. 68 Десантників, 10.
Тел.: 8 (0512) 50-03-32, 8 (0512) 76-55-81, e-mail: rector@chdu.edu.ua.
Свідоцтво суб'єкта видавничої справи ДК № 3460 від 10.04.2009 р.