

Тема 2. Визначення основної тенденції у ряді динаміки

2.1 Тренд та сезонні коливання

Рівні часового ряду формуються під впливом багатьох факторів. Основні з цих факторів визначають закономірність, тенденцію розвитку явища, що досліджується. Загальну, основну тенденцію розвитку явища називають **трендом**. Наслідком дії сукупності незначних випадкових факторів є випадкові коливання рівнів ряду.

Часовий ряд як сукупність значень його рівнів містить у собі три основні компоненти:

- 1) основну довготривалу тенденцію зміни рівнів ряду (його тренд);
- 2) короточасні систематичні зміни рівнів ряду (**сезонні коливання**);
- 3) несистематичні випадкові коливання рівнів ряду.

При дослідженні рядів динаміки потрібно виділити ці компоненти та визначити тренд. Для цього використовують спеціальну обробку часового ряду – його **вирівнювання**.

До основних методів вирівнювання (згладжування) часового ряду відносять збільшення інтервалів часу, для яких визначаються рівні ряду, застосування ковзних середніх та аналітичне вирівнювання ряду динаміки.

2.2 Метод збільшення часових інтервалів

Сутність цього методу полягає у перетворенні вихідного часового ряду шляхом заміни його на ряд, рівні якого відносяться до більш тривалих проміжків часу. Отриманий ряд містить збільшені рівні, знайдені шляхом підсумовування рівнів первинного ряду. При цьому коливання рівнів, викликані дією короткотривалих випадкових факторів, зменшуються, тренд часового ряду проявляється більш чітко.

Приклад. Часовий ряд містить дані про обсяги реалізації пральних машин у магазинах міста (таблиця 2.1).

Таблиця 2.1. Часовий ряд для обсягів реалізації пральних машин у магазинах міста

Місяць	1	2	3	4	5	5	7	8	9	10	11	12
Обсяг реалізації, одиниць	366	310	296	380	336	295	380	381	392	444	382	398

Об'єднаємо помісячні рівні ряду у квартальні підсумовуванням відповідних показників обсягів реалізації.

Таблиця 2.2. Перетворений часовий ряд

Квартал	1	2	3	4
Обсяг реалізації, одиниць	972	1011	1153	1224

Після збільшення інтервалів часу у часовому ряді більш помітним став тренд збільшення обсягів реалізації протягом року.

Окремим випадком методу збільшення часових інтервалів є обчислення середніх рівнів для збільшених інтервалів часу, наприклад, у останньому прикладі за рівні ряду можна було прийняти на щоквартальні обсяги реалізації, а середньомісячні обсяги для кожного кварталу.

2.3 Метод ковзної середньої

При застосуванні цього методу фактичні рівні часового ряду замінюють середніми рівнями, розрахованими для послідовно рухомих (ковзних) збільшених інтервалів, що включають у себе m рівнів часового ряду. Наприклад, якщо вибрати $m=3$, то спочатку розраховують середню величину з перших трьох рівнів ряду, вона замінює другий рівень ряду, потім знаходять середню величину для 2-го, 3-го та 4-го рівнів, нею замінюють 3-й рівень ряду і т.д. Кожного разу при розрахунку середнього рівня у сумі трьох рівнів з'являється новий рівень, а два залишаються тими ж самими. У результаті відбувається взаємне знищення випадкових коливань у середніх рівнях. Розраховані за m рівнями ковзні середні відносять до середини кожного інтервалу, що розглядається.

Приклад. У таблиці 2.3 наведено часовий ряд, що відображає динаміку виробництва продукції промисловим підприємством за 12 місяців. Провести згладжування цього часового ряду методом ковзної середньої при $m=3$.

Таблиця 2.3. Часовий ряд динаміки виробництва продукції підприємством за 12 місяців (млн. г.о.)

Місяць	1	2	3	4	5	6	7	8	9	10	11	12
Виробництво продукції, млн. г.о.	82	79	85	80	78	76	72	78	68	70	74	66
Ковзна середня, $m=3$	–	82,0	81,3	81,0	78,0	75,3	75,3	72,7	72,0	70,7	70,0	–

У даному прикладі згладжений ряд ковзних середніх більш чітко у порівнянні з вихідним рядом відображає тенденцію зменшення рівнів ряду.

Згладжування часового ряду методом ковзної середньої найзручніше здійснювати для непарної кількості m рівнів ряду, що використовують при її розрахунку. Якщо m є парним, то отримаємо ковзну середню, що відноситься між часовими періодами. Щоб згладжені рівні ряду відносилися до конкретних періодів часу, з кожної пари суміжних проміжних значень ковзних середніх знаходять середню арифметичну, яку відносять до певного періоду часу.

Основним недоліком згладжування методом ковзної середньої є те, що згладжений ряд є коротшим, ніж вихідний ряд. При непарному m він скорочується на $\frac{m-1}{2}$ періоди часу з кожного кінця, при парному – на $\frac{m}{2}$ періоди з кожного кінця ряду. При застосуванні методу ковзної середньої згладжуються випадкові коливання, проте сезонні коливання у ряді зберігаються. Основним недоліком цього методу є те, що він не дозволяє отримати загальну тенденцію зміни рівнів часового ряду у вигляді математичної моделі (формули, що відображає залежність між рівнями ряду та часом), яку можна використати для прогнозування.

2.4 Метод аналітичного вирівнювання

Сутність методу аналітичного вирівнювання полягає у заміні фактичних рівнів часового ряду $y_i, i=1,2,\dots,n$ теоретичними рівнями \tilde{y}_i , розрахованими за певною математичною моделлю (рівнянням $\tilde{y} = f(t)$), що відображає тренд часового ряду. При цьому кожний фактичний рівень часового ряду розглядають як суму двох складових: $y_i = f(t_i) + \varepsilon_i$, де $f(t)$ – систематична складова часового ряду, що відображає тренд, ε_i – випадкова складова, що відображає випадкові коливання рівнів ряду навколо тренду.

Задача аналітичного вирівнювання зводиться до визначення на основі фактичних даних вигляду функції $\tilde{y} = f(t)$, здатної найбільш адекватно відобразити тренд показника, що досліджується.

При аналітичному вирівнюванні найчастіше використовують наступні функції:

- лінійна функція $\tilde{y} = a_0 + a_1 t$;
- показникова функція $\tilde{y} = a_0 \cdot a_1^t$;
- гіпербола $\tilde{y} = a_0 + \frac{a_1}{t}$;
- парабола $\tilde{y} = a_0 + a_1 t + a_2 t^2$ (може використовуватися і многочлен більш високого степеня);
- тригонометричний многочлен $\tilde{y} = a_0 + \sum_{k=1}^m (a_k \cos kt + b_k \sin kt)$.

Вибір вигляду функції для аналітичного вирівнювання виконується на основі графічного зображення часового ряду, проте існують деякі правила, що дозволяють вибрати вигляд трендової моделі без побудови графіку часового ряду.

Вирівнювання за лінійною функцією $\tilde{y} = a_0 + a_1 t$ звичайно використовують для часових рядів, рівні яких змінюються близько до арифметичної прогресії, тобто коли абсолютні прирости рівнів $\Delta y_i = y_i - y_{i-1}$ приблизно є однаковими. Якщо приблизно однаковими є другі різниці рівнів часового ряду, тобто $\Delta^2 y_i = \Delta y_i - \Delta y_{i-1} \approx \text{const}$, то для моделювання тренду використовують рівняння параболи $\tilde{y} = a_0 + a_1 t + a_2 t^2$.

Якщо при послідовному розташуванні проміжків часу значення рівнів ряду змінюються у геометричній прогресії, тобто ланцюгові темпи зростання у часовому ряді є приблизно сталими, то тренд цього ряду відображається показниковою функцією $\tilde{y} = a_0 \cdot a_1^t$.

Якщо у часовому ряді спостерігається сповільнене зниження рівнів ряду, які зі збільшенням часу не можуть зменшитися до нуля, то для моделювання типу тренду використовують гіперболу $\tilde{y} = a_0 + \frac{a_1}{t}$.

Один і той же часовий ряд можна вирівняти за різними формулами для $\tilde{y} = f(t)$. Щоб визначити, вибір якої формули дає кращий результат, для кожної з них обчислюють суму квадратів відхилень фактичних рівнів часового ряду від теоретичних, тобто суму $\sum_{i=1}^n (y_i - \tilde{y}(t_i))^2$. Та функція $\tilde{y} = f(t)$, для якої ця сума квадратів буде найменшою, найбільш адекватно відображає тренд часового ряду. Проте безпосередньо порівнювати суми квадратів відхилень можливо лише у випадках, коли рівняння трендів $\tilde{y} = f(t)$, що порівнюються, мають однакову кількість параметрів a_0, a_1, \dots, a_m . Якщо кількість m цих параметрів є різною для кожного рівняння, то кожену суму квадратів ділять на величину $(n-m)$ і порівнюють ці відношення між собою.

Для відшукування параметрів a_0, a_1, \dots, a_m моделі тренду $\tilde{y} = f(t)$ найчастіше використовують метод найменших квадратів. При застосуванні цього методу спочатку вибирають тип моделі, потім визначають її параметри a_0, a_1, \dots, a_m з умови мінімуму суми квадратів відхилень фактичних рівнів ряду від їх теоретичних значень (обчислених за рівнянням моделі):

$$\sum_{i=1}^n (y_i - \tilde{y}(t_i))^2 \rightarrow \min$$

На практиці найчастіше використовують модель лінійного тренду $\tilde{y} = a_0 + a_1 t$. Коефіцієнти a_0 та a_1 цієї моделі визначають за формулами:

$$a_1 = \frac{n \sum_{i=1}^n y_i t_i - \sum_{i=1}^n y_i \sum_{i=1}^n t_i}{n \sum_{i=1}^n t_i^2 - \left(\sum_{i=1}^n t_i \right)^2}, a_0 = \bar{y} - a_1 \bar{t},$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \bar{t} = \frac{\sum_{i=1}^n t_i}{n}.$$

Приклад. За даними про щомісячне виробництво продукції на підприємстві, наведеними у таблиці 2.4, виконати вирівнювання ряду динаміки, застосувавши методи збільшення часових інтервалів, ковзної середньої та аналітичного вирівнювання за лінійною функцією.

Таблиця 2.4. Дані про щомісячне виробництво продукції на підприємстві, шт.

Місяць	1	2	3	4	5	6	7	8	9	10	11	12
Обсяг виробництва	240	245	250	248	254	260	258	265	269	270	275	274

Вирівняємо наведений у таблиці ряд динаміки, використавши метод збільшення часових інтервалів. Перейдемо від часового ряду з помісячними даними до ряду з поквартальними даними, об'єднавши обсяги виробництва для кожного кварталу. Отримаємо часовий ряд, наведений у таблиці 2.5.

Таблиця 2.5. Часовий ряд зі збільшеними часовими інтервалами

Квартал	1	2	3	4
Обсяг виробництва, шт.	735	762	792	819
Середньомісячний обсяг виробництва, шт.	245	254	264	273

Часовий ряд зі збільшеними інтервалами свідчить, що існує тенденція до зростання обсягу виробництва продукції.

Вирівняємо часовий ряд, застосувавши метод ковзних середніх. За ковзний інтервал виберемо 5 місяців і знайдемо середнє арифметичне значення

$$\bar{y}_3 = \frac{240 + 245 + 250 + 248 + 254}{5} = 247,4.$$

Знайдене середнє значення відноситься до березня. Зміщуємо ковзний інтервал на один рівень від попереднього і знаходимо нове середнє значення, яке буде відповідати квітню:

$$\bar{y}_4 = \frac{245 + 250 + 248 + 254 + 260}{5} = 251,4.$$

Аналогічно отримуємо ряд ковзних середніх та заміняємо ними фактичні рівні. Результати розрахунків наведені у таблиці 2.6.

Таблиця 2.6. Часовий ряд, вирівняний за методом ковзних середніх

Місяць	Обсяг виробництва, шт.	Ковзні середні
1	240	–
2	245	–
3	250	247,4
4	248	251,4
5	254	254,0
6	260	257,0
7	258	261,2
8	265	264,4
9	269	267,4
10	270	270,6
11	275	–
12	274	–

Ряд ковзних середніх, як і ряд квартальних середніх, вказує на тенденцію до зростання виробництва.

Проведемо аналітичне вирівнювання часового ряду за допомогою лінійного тренду $\tilde{y} = a_0 + a_1 t$. У таблиці 2.7 наведено проміжні розрахунки для обчислення коефіцієнтів тренду.

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{3108}{12} = 259, \bar{t} = \frac{\sum_{i=1}^n t_i}{n} = \frac{78}{12} = 6,5,$$

$$a_1 = \frac{n \sum_{i=1}^n y_i t_i - \sum_{i=1}^n y_i \sum_{i=1}^n t_i}{n \sum_{i=1}^n t_i^2 - \left(\sum_{i=1}^n t_i \right)^2} = \frac{12 \cdot 21152 - 3108 \cdot 78}{12 \cdot 650 - (78)^2} = 6,647,$$

$$a_0 = \bar{y} - a_1 \bar{t} = 259 - 6,647 \cdot 6,5 = 215,795.$$

Таблиця 2.7. Проміжні розрахунки та теоретичні значення рівнів тренду

Фактичне значення, y	t	t^2	ty	Теоретичне значення, \tilde{y}
240	1	1	240	222,442
245	2	4	490	229,089
250	3	9	750	235,736
248	4	16	992	242,383
254	5	25	1270	249,030
260	6	36	1560	255,677
258	7	49	1806	262,324
265	8	64	2120	268,971
269	9	81	2421	275,618
270	10	100	2700	282,265
275	11	121	3025	288,912
274	12	144	3288	295,559
$\sum y = 3108$	$\sum t = 78$	$\sum t^2$	$\sum ty = 21152$	$\sum \tilde{y} = 3108$

Отже, рівняння лінійного тренду має вигляд: $\tilde{y} = 6,647t + 215,795$.

Підставляючи у отримане рівняння значення t , отримуємо теоретичні значення обсягу виробництва. Бачимо, що при аналітичному вирівнюванні чітко відслідковується тенденція зростання обсягу виробництва, проте відхилення для перших та останніх рівнів часового ряду є суттєвими.

2.5 Оцінка якості трендової моделі

Найпростішим способом оцінки якості трендової моделі є використання стандартної похибки цієї моделі:

$$\mu = \sqrt{\frac{\sum_{t=1}^n (y_t - \tilde{y}_t)^2}{n - m}},$$

де y_t фактичне значення рівня ряду y у момент часу t , \tilde{y}_t – значення рівня ряду, розраховане за трендовою моделлю, n – кількість спостережень, m – кількість параметрів у рівнянні тренду.

Для оцінки якості вибраної моделі тренду у ряді динаміки використовують критерій Фішера. Його розраховують за формулою

$$F = \frac{\sum_{t=1}^n (\tilde{y}_t - \bar{y})^2}{m - 1} : \mu^2.$$

У цій формулі \bar{y} – середній рівень ряду.

Чим більшим є значення критерію F , тим кращою є трендова модель у порівнянні з іншими трендовими моделями. Проте навіть краща у порівнянні з іншими трендова модель може виявитися непридатною для практичного використання. Її можна визнати задовільною, якщо розраховане значення критерію Фішера виявиться більшим за деяке табличне значення $F_{крит}$. Значення $F_{крит}$ визначають з таблиці F -розподілу. Ця таблиця має три входи: Q – рівень значущості перевірки, тобто ймовірність визнання адекватної моделі неякісною або ймовірність відхилення вірної гіпотези про наявність зв'язку між динамікою рівнів ряду та часом (на практиці звичайно вибирають $Q = 5\%$), другий вхід – величина $\nu_1 = m - 1$, третій вхід – параметр $\nu_2 = n - m$. При $F < F_{крит}$ практичне застосування побудованої трендової моделі є недоцільним.