

ТЕМА 8. МЕТОДИ ПРОСТОЇ ЕКСТРАПОЛЯЦІЇ

1. Поняття та класифікація методів простої екстраполяції
2. Парний регресійний аналіз
3. Множинний регресійний аналіз

1. Поняття та класифікація методів простої екстраполяції

Екстраполяція часового ряду – це подовження існуючої динаміки досліджуваного показника на майбутні періоди. Передумовою проведення екстраполяції є припущення про інерційність економічних процесів. Тобто, якщо в минулому спостерігалась певна тенденція розвитку того чи іншого явища, то в майбутньому, з високою ймовірністю, вона також буде зберігатись. При цьому, період прогнозування не повинний перевищувати третину від тривалості ретроспективних даних.

Методи простої екстраполяції відносяться до групи формалізованих (фактографічних) методів прогнозування й призначені для математичного опису трендового компонента, за умови, що всі рівні часового ряду мають однаковий вплив на його формування, незалежно від строку їхньої давності.

До методів простої екстраполяції належать:

1. Методи парного регресійного аналізу – використовуються для встановлення функціонального взаємозв'язку між двома показниками, один з яких виступає фактором, а інший – результатом.

$$y_{\text{розрах}} = f(t), \quad (1)$$

Де t – номер періоду часу; $y_{\text{розрах}}$ – розрахункове значення показника, для якого виконується прогнозування.

В якості факторного показника в даному випадку виступає час t . Серед основних видів парної регресії розрізняють: лінійну, степеневу, логарифмічну, поліноміальну, гіперболічну та експоненційну.

2. Методи множинного регресійного аналізу – використовуються для встановлення функціонального взаємозв'язку між результативним показником й множиною факторів $x_1, x_2, \dots, x_n$, де $n \geq 2$.

$$y_{\text{розрах}} = f(x_1, x_2, \dots, x_n), \quad (2)$$

В свою чергу, для кожного з цих факторів може бути побудований прогноз методами парного регресійного аналізу:

$$x_1 = f_1(t); x_2 = f_2(t); \dots; x_n = f_n(t), \quad (3)$$

В основі побудови регресії (1) та (2) покладений метод найменших квадратів.

2. Парний регресійний аналіз

Процес побудови однофакторної регресійної моделі розпочинається з вибору відповідного типу математичної залежності, яка найкраще відображає характер досліджуваного взаємозв'язку.

Від правильного вибору форми залежності, в кінцевому випадку, залежать результати прогнозування. Для вирішення даної задачі будується діаграма розсіювання, яка може приймати будь-який вигляд, рис.1.

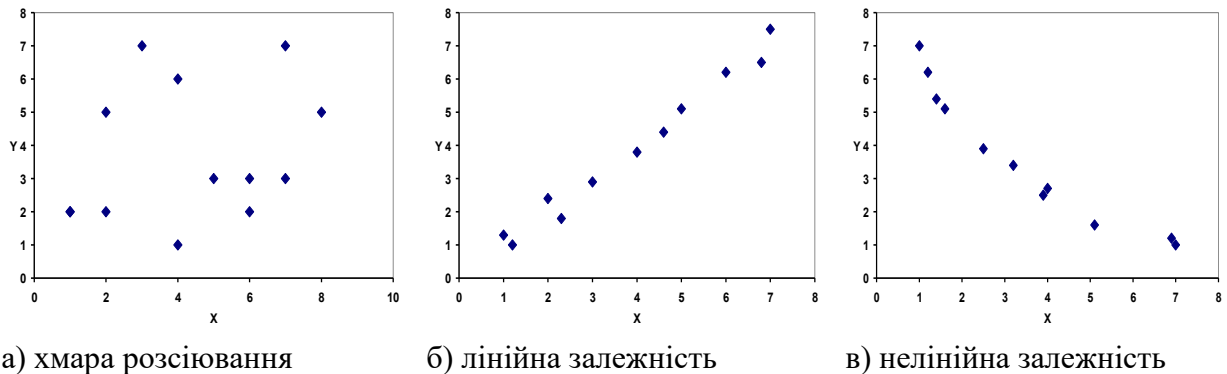


Рис. 1. Діаграми розсіювання

По зовнішньому вигляду скупчення рівнів часового ряду висувається гіпотеза про форму залежності між змінними:

1. Лінійна залежність:

$$y_{\text{розр}} = a_0 + a_1 x;$$

2. Парабола другого та третього порядків, відповідно:

$$y_{\text{розр}} = a_0 + a_1 x + a_2 x^2;$$

$$y_{\text{розр}} = a_0 + a_1 x + a_2 x^2 + a_3 x^3;$$

3. Степенева залежність:

$$y_{\text{розр}} = a_0 x^{a_1};$$

4. Гіперболічна залежність:

$$y_{\text{розр}} = a_0 + \frac{a_1}{x};$$

5. Логарифмічна залежність:

$$y_{\text{розр}} = a_0 + a_1 \ln(x);$$

6. Експоненційна залежність:

$$y_{\text{розр}} = a_0 e^{a_1 x}.$$

В усіх цих прикладах невідомими змінними є параметри рівнянь a_0 , a_1 , ..., a_n , які необхідно знайти. Причому, побудована трендова компонента повинна максимально наближатись до ретроспективних рівнів часового ряду.

Для вирішення даного завдання використовується метод найменших квадратів. Розглянемо приклад його застосування для лінійної залежності, рис. 2.

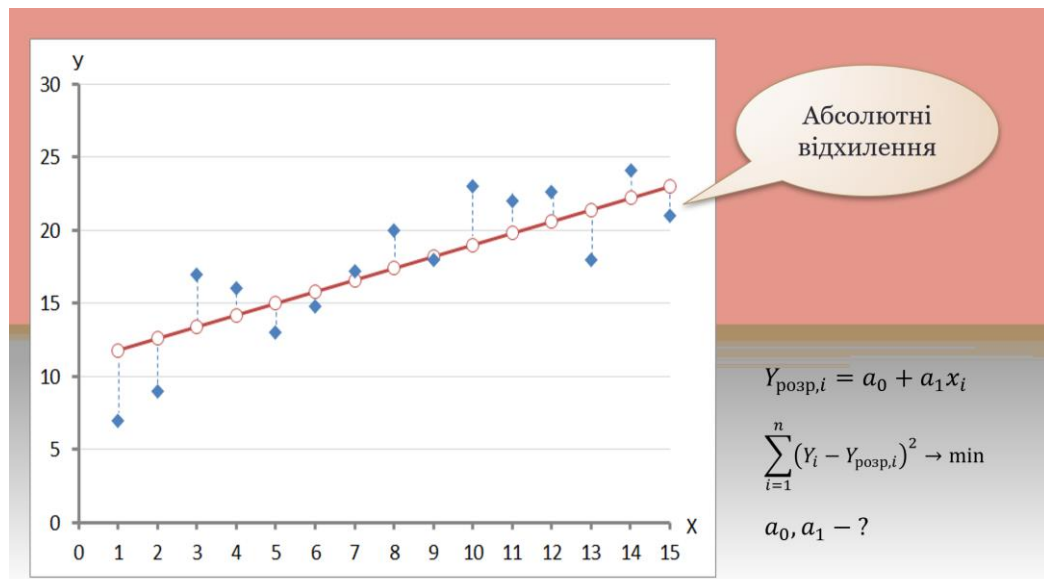


Рис. 2. Метод найменших квадратів

Як видно з рис. 2, метод найменших квадратів виходить із необхідності мінімізації суми квадратів відхилень розрахункових значень досліджуваного показника від вхідних рівнів часового ряду.

Представимо цільову функцію у вигляді:

$$S(a_0, a_1) = \sum (Y_i - a_0 - a_1 X_i)^2 \rightarrow \min$$

Функція S буде мінімізуватись, коли її часткові похідні будуть дорівнювати нулю:

$$\begin{cases} \frac{\partial S}{\partial a_0} = 0 \\ \frac{\partial S}{\partial a_1} = 0 \end{cases} \rightarrow \begin{cases} \frac{\partial S}{\partial a_0} = 2 \sum_{i=1}^n (Y_i - a_0 - a_1 X_i) \times (-1) = 0 \\ \frac{\partial S}{\partial a_1} = 2 \sum_{i=1}^n (Y_i - a_0 - a_1 X_i) \times (-X_i) = 0 \end{cases}$$

Тоді, система рівнянь з двома невідомими набуває вигляду:

$$\begin{cases} \sum_{i=1}^n (Y_i - a_0 - a_1 X_i) = 0 \\ \sum_{i=1}^n (X_i Y_i - a_0 X_i - a_1 X_i^2) = 0 \end{cases} \rightarrow \begin{cases} \sum_{i=1}^n a_0 + a_1 \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i \\ a_0 \sum_{i=1}^n X_i + a_1 \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i \end{cases}$$

Представимо її в наступному вигляді:

$$\begin{cases} n a_0 + a_1 \sum X_i = \sum Y_i \\ a_0 \sum X_i + a_1 \sum X_i^2 = \sum X_i Y_i \end{cases}$$

З останньої системи лінійних алгебраїчних рівнянь знаходимо невідомі змінні a_0 та a_1 :

$$a_0 = \frac{\sum Y_i \sum X_i^2 - \sum X_i \sum X_i Y_i}{n \sum X_i^2 - (\sum X_i)^2}; \quad a_1 = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - (\sum X_i)^2}$$

Для знаходження невідомих параметрів інших форм нелінійних регресій використовують алгебраїчні перетворення для приведення їх до лінійного вигляду. Алгебраїчні перетворення включають логарифмування та заміну змінних. Розглянемо їх більш детально.

2. Парабола другого порядку $y_{\text{розн}} = a_0 + a_1 x + a_2 x^2$.

Введемо заміну змінних: $x_1 = x^2$.

Тоді, рівняння параболи прийме вигляд:

$$y_{\text{розн}} = a_0 + a_1 x + a_2 x_1.$$

Як бачимо, початкова нелінійна залежність з однією змінною x перетворилась на лінійну залежність з двома змінними x та x_1 . Тобто, маємо випадок множинної регресії, який буде розглядатись нижче.

3. Степенева залежність $y_{\text{розн}} = a_0 x^{a_1}$.

Перш ніж провести заміну змінних, виконаємо логарифмування лівої й правої частин нашого рівняння:

$$\ln(y_{\text{розн}}) = \ln(a_0 x^{a_1}), \text{ звідки отримуємо:}$$

$$\ln(y_{\text{розн}}) = \ln(a_0) + \ln(x^{a_1}) = \ln(a_0) + a_1 \times \ln(x).$$

Далі проводимо заміну змінних:

$$\ln(y_{\text{розн}}) = \hat{y};$$

$$\ln(a_0) = \widehat{a_0};$$

$$\ln(x) = \hat{x}.$$

Тоді остаточне рівняння набуває лінійного вигляду:

$$\hat{y} = \widehat{a_0} + a_1 \hat{x}.$$

Після знаходження методом найменших квадратів невідомих параметрів $\widehat{a_0}$ та a_1 , необхідно розрахувати початкове a_0 , згідно попередньої заміни:

$$a_0 = e^{\widehat{a_0}}.$$

4. Гіперболічна залежність $y_{\text{розн}} = a_0 + \frac{a_1}{x}$.

Введемо заміну змінних:

$$\frac{1}{x} = \hat{x}.$$

Тоді, початкове рівняння прийме лінійну форму:

$$y_{\text{розн}} = a_0 + a_1 \hat{x}.$$

5. Логарифмічна залежність $y_{\text{розн}} = a_0 + a_1 \ln(x)$.

Введемо заміну змінних:

$$\ln(x) = \hat{x}.$$

Тоді, початкове рівняння знову приймає лінійну форму:

$$y_{\text{розр}} = a_0 + a_1 \hat{x}.$$

6. Експоненційна залежність $y_{\text{розр}} = a_0 e^{a_1 x}$.

Перш ніж провести заміну змінних, виконаємо логарифмування лівої й правої частин нашого рівняння:

$$\ln(y_{\text{розр}}) = \ln(a_0 e^{a_1 x}), \text{ звідки отримуємо:}$$

$$\ln(y_{\text{розр}}) = \ln(a_0) + \ln(e^{a_1 x}) = \ln(a_0) + a_1 x \times \ln(e) = \ln(a_0) + a_1 x.$$

Далі проводимо заміну змінних:

$$\ln(y_{\text{розр}}) = \hat{y};$$

$$\ln(a_0) = \widehat{a_0}.$$

Тоді остаточне рівняння набуває лінійного вигляду:

$$\hat{y} = \widehat{a_0} + a_1 x.$$

Після знаходження методом невідомих параметрів $\widehat{a_0}$ та a_1 , необхідно розрахувати початкове a_0 :

$$a_0 = e^{\widehat{a_0}}.$$

З наведених залежностей слід обирати ту, яка надає найвищий рівень достовірності за критерієм Фішера, що розглядався вище.

3. Множинний регресійний аналіз

Як зазначалось вище, у випадку парної регресії кількість факторів була обмежена одним фактором, в якості якого виступав час. Множинна регресія, в свою чергу, виходить з двох, або більше факторів.

По аналогії з парним регресійним аналізом, метод найменших квадратів застосовується для визначення невідомих параметрів лише для лінійної множинної регресії. Тому, у випадку нелінійної залежності, вона по аналогії повинна спочатку бути приведена до лінійного вигляду шляхом алгебраїчних перетворень: заміни змінних та логарифмування.

Нехай, лінійна залежність має вигляд:

$$y_{\text{розр}} = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n, \quad (4)$$

Де n – кількість факторів; $b_0, b_1, \dots, b_n$ – невідомі параметри множинної регресії, які необхідно знайти.

Для того, щоб МНК повернув нам всі невідомі параметри, включаючи b_0 , до складу регресії необхідно включити фіктивну змінну $x_0=1$. З урахуванням цього, рівняння (4) запишеться у вигляді:

$$y_{\text{розр}} = b_0 x_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n, \quad (5)$$

Представимо вхідні дані у матричному вигляді:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_m \end{bmatrix}, X = \begin{bmatrix} x_{10} & x_{11} & x_{12} & \dots & x_{1n} \\ x_{20} & x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{m0} & x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1n} \\ 1 & x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}, B = \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_n \end{bmatrix}$$

Тоді, розрахунок невідомих параметрів рівняння (5) буде відбуватись шляхом наступних матричних перетворень:

$$B = [X]^T [X]]^{-1} [X]^T Y, \quad (6)$$

Параметри рівняння регресії (4) у Microsoft Excel можна знайти за допомогою вбудованої функції **=ЛИНЕЙН()**. Дана функція має наступні аргументи:

- I аргумент – вектор-стовпець відомих значень результативного показника Y;
- II аргумент – матриця відомих значень факторів X. Дана матриця **не повинна** містити колонку фіктивних змінних $x_0=1$;
- III, IV аргументи – логічні параметри, повинні приймати значення «істина», або 1.

Після введення даної функції, виділити діапазон комірок, куди повинні повернутись розраховані значення. Розмірність цього діапазону $5 \times (n + 1)$.

Далі слід натиснути <F2> та одночасно <CTRL>+<SHIFT>+<ENTER>. Розраховані значення параметрів регресії, які необхідно було знайти, будуть розташовані у першому рядку:

b_n	...	b_2	b_1	b_0
...	...	...	...	...
...	...	...	...	...

В якості прикладу, розглянемо приведення степеневій множинній регресії до лінійної форми. Нехай, необхідно побудувати регресію виду:

$$y_{\text{розр}} = b_0 x_1^{b_1} x_2^{b_2} \times \dots \times x_n^{b_n}, \quad (7)$$

Логарифмуємо дане рівняння, отримуємо:

$$\ln(y_{\text{розр}}) = \ln(b_0 x_1^{b_1} x_2^{b_2} \times \dots \times x_n^{b_n}), \text{ звідки будемо мати}$$

$$\ln(y_{\text{розр}}) = \ln(b_0) + b_1 \ln(x_1) + b_2 \ln(x_2) + \dots + b_n \ln(x_n).$$

Дали виконуємо заміну змінних:

$$\ln(y_{\text{розр}}) = \hat{y};$$

$$\ln(b_0) = \widehat{b_0};$$

$$\ln(x_i) = \hat{x}_i.$$

Тоді остаточне рівняння набуває лінійного вигляду:

$$\hat{y} = \hat{b}_0 + b_1\hat{x}_1 + b_2\hat{x}_2 + \dots + b_n\hat{x}_n.$$

Аналогічним чином виконуються перетворення для всіх інших форм залежностей.

Результатом застосування методів простої екстраполяції є виділення трендової складової з загальної динаміки рівнів часового ряду. Обов'язково, адекватність отриманих моделей слід перевіряти за критерієм Фішера й обирати для прогнозування ту, яка має найвищий рівень достовірності.