Лекція 23. Нелінійна регресія

Нелінійна регресія – окремий випадок регресійного аналізу, в якому розглянутою регресійною моделлю є нелінійна функція, що залежить від параметрів і від однієї або декількох вільних змінних. Відмінність від лінійної регресії полягає тільки в формі зв’язку та методах оцінки параметрів.

23.1. Парна нелінійна регресія

Розрізняють два класи нелінійних регресійних моделей:

1. Регресії, нелінійна за факторами, але лінійна відносно невідомих параметрів (квазілінійна).

До нелінійних регресійних моделей за включеними змінним відносяться такі функції:

поліноми різних ступенів, наприклад, $y = a + bx^3$; рівнобічна гіпербола - $y = a + \frac{b}{x}$;

$y = a + b\ln x$;

$y = a + be^x$;

$y = a + b\sqrt{x}$;

$y = a + bx^2$;

2. Регресії, нелінійна за факторами та параметрами

Для таких регресій не існує єдиного способу зведення їх до лінійних.

До нелінійних регресійних моделей по оцінюваним параметрами відносяться функції:

$y = \frac{1}{a + bx}$, $y = \frac{1}{a + be^{-x}}$

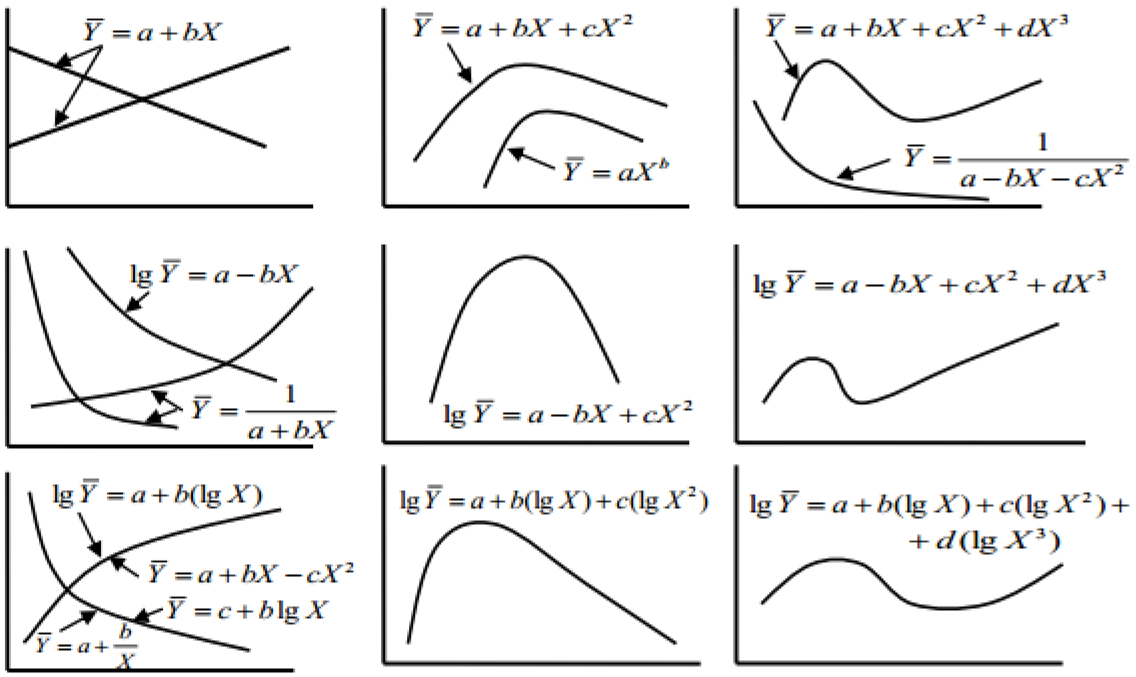
степенева - $y = ax^b$;

$y = ae^{bx}$

показникова - $y = ab^x$;

експоненціальна - $y = e^{a+bx}$.

Дані моделі за своєю природою не є лінійними, тому що вони не можуть бути представленими у вигляді простої регресійної моделі з деякими перетвореннями незалежних змінних.



23.1.1. Побудова та дослідження рівняння нелінійної регресії

Степенева модель

Прологарифмуємо функцію $y = ax^b \cdot \varepsilon$, отримаємо:

$$\ln y = \ln a + b \cdot \ln x + \ln \varepsilon . \qquad (23.1)$$

Введемо позначення:

$$y_1 = \ln y ;\ a_1 = \ln a ;\ x_1 = \ln x ;\ \varepsilon_1 = \ln \varepsilon .$$

Тоді степенева модель приймає лінійний вигляд

$$y_1 = a_1 + bx_1 + \varepsilon_1 \qquad (23.2)$$

і до неї може застосовуватися стандартний МНК.

Після знаходження оцінок a_1 і b, a визначається за формулою $a = e^{a_1}$.

Коефіцієнт b означає еластичність змінної *Y* по змінній *X*. Така модель називається моделлю *з постійною еластичністю*.

Показникова модель

Прологаріфмуємо функцію $y = a \cdot b^x \cdot \varepsilon$.

Отримаємо:

$$\ln y = \ln a + x \cdot \ln b + \ln \varepsilon \tag{23.3}$$

Позначимо:

$$y_1 = \ln y;\ a_1 = \ln a;\ b_1 = \ln b;\ \varepsilon_1 = \ln \varepsilon .$$

Тоді степенева модель приймає лінійний вигляд

$$y_1 = a_1 + b_1 x + \varepsilon_1 \tag{23.4}$$

і до неї може застосовуватися стандартний МНК.

Після знаходження оцінок a_1 і b_1, значення a і b визначається за формулою $a = e^{a_1}$, $b = e^{b_1}$.

Гіпербола

Для приведення моделі до лінійного виду вводять нову змінну $z = \frac{1}{x}$ і модель набирає вигляду

$$y = a + b \cdot z + \varepsilon . \tag{23.5}$$

Подальше застосування МНК дозволяє визначити оцінки a і b.

Парабола

В вихідному рівнянні $y = a + bx + cx^2 + \varepsilon$ виконаємо заміну

$$x_1 = x,\ x_2 = x^2 \tag{23.6}$$

отримаємо лінійне рівняння множинної регресії

$$y = a + bx_1 + cx_2 + \varepsilon \tag{23.7}$$

До нього можна безпосередньо застосовувати МНК.

23.1.2. Дослідження рівняння нелінійної регресії

Рівняння нелінійної регресії доповнюються індексом кореляції:

$$R = \sqrt{1 - \frac{\sum\left(y_i - \hat{y}_{x_i}\right)^2}{\sum\left(y_i - \overline{y}\right)^2}} \qquad (23.8)$$

і середньою помилкою апроксимації

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y_i - \hat{y}_{x_i}}{y_i}\right| \cdot 100\% \qquad (23.9)$$

де, y_i - значення залежної змінної, які отримали в результаті дослідження економічного об‘єкту; $\hat{y}_{x_i}$ - значення залежної змінної, які отримали із відповідного рівняння регресії; $\overline{y}$ - середнє значення залежної змінної

$$\overline{y} = \frac{1}{n}\sum y_i \text{ .}$$

Якість рівняння регресії оцінюється за допомогою індексу детермінації R^2. Для перевірки значимості в цілому рівняння регресії застосовується F-критерій Фішера.

Знаходимо:

$$F_{факт} = \frac{R^2}{1 - R^2}\left(n - 2\right),$$

де n - число спостережень, та порівнюємо його з критичною точкою розподілу Фішера, при рівні значимості δ, числом ступенів свободи $\nu_1 = 1,\ \nu_2 = n - 2$, $F_{крит} = F\left(\delta; \nu_1, \nu_2\right)$.

Якщо $F_{крит} < F_{факт}$, то гіпотеза про випадкову природу оцінюваних параметрів відхиляється, визнається їх статистична значимість і надійність рівняння регресії. Якщо $F_{крит} > F_{факт}$, то приймається гіпотеза про їх статистичну незначущість і ненадійності рівняння регресії.

При аналізі моделі часто застосовують коефіцієнт еластичності. Він характеризує вплив пояснюючого фактора на залежну змінну. Коефіцієнт еластичності обчислюють за формулою:

$$E = \hat{f}'(x)\frac{x}{y}.$$

Він показує на скільки відсотків зміниться залежна змінна y, якщо пояснююча змінна x зміниться на 1%. Значення коефіцієнта еластичності часто залежить від величини x, тому розглядають середній коефіцієнт еластичності

$$\overline{E} = \hat{f}'(\overline{x})\frac{\overline{x}}{y}.$$

Він показує на скільки відсотків в середньому по сукупності зміниться y результат , якщо пояснююча змінна відхилиться від свого середнього значення на 1%.

Наведемо формули для обчислення коефіцієнтів еластичності основних функцій:

Таблиця 23.1

Вид функції $y(x)$	Коефіцієнт еластичності $E = \hat{f}'(x)\frac{x}{y}$
Лінійна $y = a + bx + \varepsilon$	$E = \frac{bx}{a + bx}$
Степенева $y = a \cdot x^b \cdot \varepsilon$	$E = b$
Показникова $y = a \cdot b^x \cdot \varepsilon$	$E = x \cdot \ln b$
Гіпербола $y = a + \frac{b}{x} + \varepsilon$	$E = \frac{-b}{a \cdot x + b}$
Парабола $y = a + b \cdot x + c \cdot x^2 + \varepsilon$	$E = \frac{(2c \cdot x + b) \cdot x}{a + b \cdot x + c \cdot x^2}$

При виборі моделі важливо, щоб вона добре відповідала вихідним даним. Рівняння «тим краще», чим більшу частину розкиду залежної змінної у воно може пояснити. Для оцінки якості нелінійної регресії, аналогічно лінійній залежності, застосовується індекс детермінації R^2 .

Крім того, як показник якості моделі, застосується середня помилка апроксимації:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y_i - \hat{y}_{x_i}}{y_i}\right| \cdot 100\%.$$

Важливим фактором при оцінці якості нелінійної моделі є перевірка значимості коефіцієнтів відповідної лінійної форми моделі за допомогою критерію Стьюдента (t-статистики).

23.2. Множинна нелінійна регресія

Лише деякі природні та економічні процеси можна моделювати за допомогою лінійної моделі. Тому для більш точного й повного опису процесу, необхідно використовувати нелінійну регресійну залежність.

Як для парного так і для багатофакторного регресійного аналізу можна розглядати два типи моделей:

- нелінійна за факторами, але лінійна за невідомими параметрами;
- нелінійна за факторами та параметрами.

Багатофакторні регресійні моделі першого типу

$$Y = \alpha_0 + \alpha_1 F_1(x_1) + \alpha_2 F_2(x_2) + \ldots + \alpha_m F_m(x_m) + l, \tag{23.10}$$

де $F_i(x_i)$, $i = \overline{1,m}$ - будь-які функції.

Заміною змінних $z_i = F_i(x_i)$, $i = \overline{1,m}$, такі регресії зводяться до лінійної моделі

$$Y = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \ldots + \alpha_m z_m + l \tag{23.11}$$

Нелінійність за параметрами є більш серйозною проблемою. Але коли права частина моделі складається з членів вигляду x^{β}, $e^{\beta x}$ помножених між собою, а випадковий член мультиплікативний до добутку цих функцій, то таку модель можна за допомогою логарифмування обох її частин звести до лінійної. Таким прикладом є степенева модель

$$Y = \alpha_0 x_1^{\alpha_1} \cdot x_2^{\alpha_2} \cdot \ldots \cdot x_m^{\alpha_m} \cdot e \tag{23.12}$$

або у вигляді множинної степеневої моделі

$$y_i = \alpha_0^* x_{n1}^{\alpha_0^*} \cdot x_{n2}^{\alpha_1^*} \cdot \ldots \cdot x_{nm}^{\alpha_m^*} \cdot e^{\varepsilon_n} \tag{23.13}$$

яка може описувати деякий виробничий процес, або степенево-показникова регресія (де $m = 4$).

$$Y = \alpha_0 \cdot \alpha_1^{\ x_1} \cdot \alpha_2^{\ x_2} \cdot \ldots \cdot \alpha_m^{\ x_m} \cdot e \tag{23.14}$$

Логарифмуванням обох частин (23.12) та (23.14) можна звести до лінійної (23.11), або для (23.13)

$$\ln y = \alpha_0^* + \alpha_1^* \ln x_{n1} + \alpha_2^* \ln x_{n2} + \ldots + \alpha_m^* \ln x_{nm} + \varepsilon_i \tag{23.15}$$

Здійснивши вибірку обсягу n, дістанемо систему рівнянь, яку у векторно-матричній формі можна записати так:

$$\vec{y}^* = X\vec{\alpha}^* + \vec{\varepsilon}, \tag{23.16}$$

де

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_n \end{pmatrix}, \ X = \begin{pmatrix} 1 & \ln x_{11} & \ln x_{12} & \ldots & \ln x_{1m} \\ 1 & \ln x_{21} & \ln x_{22} & \ldots & \ln x_{2m} \\ 1 & \ln x_{31} & \ln x_{32} & \ldots & \ln x_{3m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \ln x_{n1} & \ln x_{n2} & \ldots & \ln x_{nm} \end{pmatrix}, \ \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \ldots \\ \varepsilon_n \end{pmatrix}.$$

Тоді, за аналогією з попередніми випадками, компоненти вектора визначають із рівняння $\vec{\alpha}^* = \left(X^T X\right)^{-1} X^T \vec{y}^*$.

Якщо досліджувати регресії (23.12) та (23.13), спочатку треба знайти оцінки та надійні інтервали для лінійної регресії, а потім за допомогою зворотних перетворень, знайти оцінки та надійні інтервали і для регресій (23.12) та (23.13).