ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ
КИЇВСЬКОГО НАЦІОНАЛЬНОГО УНІВЕРСИТЕТУ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

О. С. Слабоспицький

ОЦІНЮВАННЯ ПАРАМЕТРІВ СИСТЕМ ПРИ НЕКЛАСИЧНИХ ПРИПУЩЕННЯХ

Конспект лекцій

Київ
Видавництво «Людмила»
2025

УДК 519.2(075.8)
С47

Рецензенти:
чл.-кор. НАН України, д-р техн. наук, проф. М. Ю. Кузнєцов
д-р техн. наук, проф. В. А. Заславський

Рекомендовано до друку
вченою радою факультету комп'ютерних наук та кібернетики
(протокол № 5 від 18 листопада 2025 року)

Ухвалено
науково-методичною комісією факультету комп'ютерних наук та кібернетики
(протокол № 4 від 17 листопада 2025 року)

Слабоспицький О. С.

С47 Оцінювання параметрів систем при некласичних припущеннях : конспект лекцій / О. С. Слабоспицький. – К. : Видавництво «Людмила», 2025. – 46 с.

Розглянуто розв'язання проблеми оцінювання невідомих параметрів регресійної моделі, не тільки у традиційному вигляді, але й коли може бути порушене у постановці задачі якесь з класичних припущень відносно невизначеностей моделі.

Для студентів, які вивчають дисципліни "Теорія оцінювання систем в умовах невизначеності", "Аналіз даних" і споріднені з ними предмети.

УДК 519.2(075.8)

ВСТУП

Розв’язання широкого спектру науково-технічних проблем вимагає знання математичних моделей з потрібною точністю для складних систем, які досліджуються. На жаль, не завжди вони вже наявні. В цьому випадку за доступною інформацією необхідно побудувати або уточнити відповідну математичну модель об’єкта. Фактично це і є основною задачею теорії оцінювання (ідентифікації).

Виділяють два основних класи задач ідентифікації систем. Якщо апріорна інформація про об’єкт бідна або відсутня взагалі, то кажуть, що мають справу з *задачею ідентифікації у широкому розумінні*. Якщо інформація про математичну модель об’єкта доступна з точністю до невідомих параметрів, то виникає потреба у розв’язанні *задачі ідентифікації у вузькому розумінні*. Саме задачам останнього типу і буде приділена основна увага у даній роботі.

У свою чергу самі моделі об’єктів дослідження можна поділити на три типи: чорні скриньки , сірі скриньки та білі скриньки. У випадку, коли внутрішня структура системи повністю невідома, а єдина доступна інформація, яка може бути використана при побудові її моделі, вичерпується тільки знанням пар спостережень вектора вхідних змінних та реакції системи на нього, тобто відповідного вектора вихідних змінних об'єкта, то кажуть що мають справу з *чорною скринькою (black box)*. У ситуації більш багатшої апріорної інформації, коли додатково відома внутрішня структура об’єкта, але не повністю, а тільки з точністю до деяких невідомих параметрів, то кажуть, що стикнулися з *сірою скринькою (grey box)*. Можлива ще робота з моделлю системи з третього типу, *білою скринькою (white box)*, це коли функціонування об’єкта можна описати за допомогою відомих рівнянь, тобто інформація про внутрішню структуру системи доступна повністю. У подальшому основний акцент буде зроблено на роботі з сірими скриньками.

У першому розділі для лінійної регресійної моделі сформульована постановка задачі оптимального оцінювання її невідомих параметрів. Наведено опис усіх класичних припущень, які традиційно при цьому робляться. У ситуації, коли справедливі усі класичні припущення для вищевказаного об'єкта, наводяться розв'язки задач оцінювання вектора невідомих параметрів звичайним методом найменших квадратів та зваженим методом найменших квадратів з довільною ваговою матрицею.

У наступних розділах послідовно припускається, що не є справедливе одне з класичних припущень, і для кожного з цих випадків демонструється яким чином розв'язується необхідна задача оцінювання невідомих параметрів регресійної моделі.

Автор щиро вдячний студентам і співробітникам факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка, які сприяли покращенню цього посібника.

Усі зауваження та побажання щодо посібника будуть із вдячністю сприйняті автором. Їх можна надіслати електронною поштою на адресу: *alexsl@knu.ua*.

1. ОЦІНЮВАННЯ ПАРАМЕТРІВ РЕГРЕСІЙНОЇ МОДЕЛІ ЗА СПРАВЕДЛИВОСТІ КЛАСИЧНИХ ПРИПУЩЕНЬ

У цьому розділі наведено постановку задачі оцінювання вектора невідомих параметрів для регресійної моделі, яка була взята в якості базової. Описані класичні припущення, які вважаються справедливими при цьому. Як основний метод для розв'язання вищевказаної проблеми було обрано звичайний метод найменших квадратів.

Отримані явні формули для оцінки параметрів обраним методом та її похибки оцінювання для лінійної регресійної моделі у випадку справедливості усіх класичних припущень.

1.1. Постановка задачі оцінювання параметрів у разі справедливості класичних припущень

Припустимо, що математична модель об'єкта, який досліджується, задана з точністю до деяких невідомих параметрів. Для практичного використання цієї моделі необхідно вміти розв'язувати задачу знаходження оцінок цих параметрів оптимальних у деякому розумінні.

Нехай потрібно знайти оцінку вектора невідомих параметрів α лінійної регресійної моделі

$$y(k) = x^T(k)\alpha + e(k),\ k = \overline{1, N}\ , \qquad (1.1)$$

де спостереження $y(k)\in\mathbb{R}$, вектор регресорів $x(k)\in\mathbb{R}^p$, $\alpha\in\mathbb{R}^p$, похибка моделі $e(k)\in\mathbb{R}$. Компоненти вектора регресорів представляють собою значення деяких відомих функцій від вектора незалежних змінних.

Необхідно за доступними значеннями $\{y(k)\}_{k=1}^{N}$ та $\{x(k)\}_{k=1}^{N}$ знайти оптимальну у деякому розумінні оцінку вектора невідомих параметрів α.

Спочатку перепишемо регресійну модель (1.1) у матричному вигляді:

$$y = X\alpha + e,$$

де
$$y=\begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix},\quad X=\begin{pmatrix} x^T(1) \\ x^T(2) \\ \vdots \\ x^T(N) \end{pmatrix}\in M_{N,p}(\mathbb{R}),\quad e=\begin{pmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{pmatrix}.$$

Традиційно вважаються справедливими такі класичні припущення класичного регресійного аналізу:

I. $rank(X)=p$;

II. немає ніяких обмежень на вектор невідомих параметрів α, тобто в якості його значення допускається довільне значення з простору $\mathbb{R}^p$;

III. $e\sim\mathcal{N}\left(\theta_N,\sigma^2E_N\right), \sigma^2\in\mathbb{R}_+$,

де θ_N – нульовий вектор розмірності N, E_N – одинична матриця порядку N, $\mathbb{R}_+$ – множина позитивних дійсних чисел;

IV. математична модель для об’єкта, який досліджується, є лінійною за невідомими параметрами.

Зауваження. З цих припущень випливає:

1. матриця X має повний ранг по стовпчикам, а $N\geq p$;

2. пошук оцінки вектора невідомих параметрів α можна здійснювати в усьому просторі $\mathbb{R}^p$, бо на α не накладено ніяких обмежень;
3. $e(k) \sim \mathcal{N}\left(0, \sigma^2\right)$, $k = \overline{1, N}$, тобто $\{e(k)\}_{k=1}^{N}$ є однаково нормально розподіленими та незалежними;
4. в якості апроксимуючої функції для функції регресії у моделі (1.1) була обрана функція, яка є лінійною за вектором невідомих параметрів α.

Пошук оптимальної оцінки $\hat{\alpha}$ вектора невідомих параметрів α регресійної моделі (1.1) проведемо звичайним методом найменших квадратів.

Означення. Оцінка $\hat{\alpha}$ для вектора невідомих параметрів α регресійної моделі (1.1), на якій досягає свого мінімуму функціонал

$$\|e\|^2,$$

називається *оцінкою методу найменших квадратів (оцінкою МНК, estimate of least squares method)*, де $\|e\|^2 = e^T e$.

Її ще називають *оцінкою звичайного методу найменших квадратів*.

Потрібну оцінку $\hat{\alpha}$ знайдемо як частинний випадок більш загальної оцінки.

Означення. Оцінка $\hat{\alpha}_W$ для вектора невідомих параметрів α регресійної моделі (1.1), на якій досягає свого мінімуму функціонал

$$\|e\|_W^2, \ W > 0,$$

називається *оцінкою зваженого методу найменших квадратів (оцінкою ЗМНК, estimate of method of weighted least squares) з ваговою матрицею* W, де $\|e\|_W^2 = e^T W e$.

Спочатку знайдемо оцінку $\hat{\alpha}_W$, а оцінку МНК $\hat{\alpha}$ отримаємо, як частинний випадок оцінки ЗМНК $\hat{\alpha}_W$ з ваговою матрицею $W = E_N$, бо $\hat{\alpha} = \hat{\alpha}_{E_N}$.

1.2. Явні формули для оцінок МНК, ЗМНК та їх похибок оцінювання у випадку справедливості класичних припущень

Знайдемо спочатку явну форму представлення для оцінки $\hat{\alpha}_W$. Розпишемо вираз її критерія якості

$$\|e\|_W^2 = e^T We = \left(y - X\alpha\right)^T W\left(y - X\alpha\right) =$$

$$= \alpha^T X^T WX\alpha - 2\alpha^T X^T Wy + \|y\|_W^2 .$$

Зауваження. Відомо, що

$$grad_\alpha\left(\alpha^T\beta\right) = grad_\alpha\left(\beta^T\alpha\right) = \beta,$$

$$grad_\alpha\left(\alpha^T A\alpha\right) = \left(A + A^T\right)\alpha,$$

де α, β - вектори, а A - матриця відповідних розмірностей.

Тоді маємо можливість записати необхідну умову екстремуму критерія якості

$$\left\{grad_\alpha \|e\|_W^2\right\}\Big|_{\alpha=\hat{\alpha}_W} = \left\{2\left(X^T WX\right)\alpha - 2X^T Wy\right\}\Big|_{\alpha=\hat{\alpha}_W} = \theta_p .$$

Звідси отримуємо так звану *систему нормальних рівнянь для оцінки ЗМНК з ваговою матрицею W*

$$\left(X^T WX\right)\hat{\alpha}_W = X^T Wy.$$

Оскільки матриця $\left(X^T WX\right)$ буде не виродженою тоді і тільки тоді, коли $rank\left(X\right) = p$, то справедливість класичного припущення I дозволяє стверджувати, що оцінка буде єдиною і набуває вигляду

$$\hat{\alpha}_W = \left(X^T W X\right)^{-1} X^T W y.$$

Знайдемо також похибку оцінювання $\Delta\left(\hat{\alpha}_W\right) = \hat{\alpha}_W - \alpha$ для отриманої оцінки $\hat{\alpha}_W$:

$$\Delta\left(\hat{\alpha}_W\right) = \hat{\alpha}_W - \alpha = \left(X^T W X\right)^{-1} X^T W y - \alpha =$$
$$= \left(X^T W X\right)^{-1} X^T W X \alpha + \left(X^T W X\right)^{-1} X^T W e - \alpha =$$
$$= \left(X^T W X\right)^{-1} X^T W e.$$

В результаті похибка оцінювання $\Delta\left(\hat{\alpha}_W\right)$ для оцінки ЗМНК $\hat{\alpha}_W$ може бути обчислена таким чином:

$$\Delta\left(\hat{\alpha}_W\right) = \left(X^T W X\right)^{-1} X^T W e.$$

Останні результати дозволяють констатувати, що справедливе таке твердження.

Теорема. Оцінка ЗМНК $\hat{\alpha}_W$ з ваговою матрицею W, $W > 0$ вектора невідомих параметрів α та її похибка оцінювання $\Delta\left(\hat{\alpha}_W\right)$ для регресійної моделі (1.1) набувають вигляду:

$$\hat{\alpha}_W = \left(X^T W X\right)^{-1} X^T W y, \tag{1.2}$$

$$\Delta\left(\hat{\alpha}_W\right) = \hat{\alpha}_W - \alpha = \left(X^T W X\right)^{-1} X^T W e. \tag{1.3}$$

У частинному випадку, коли $W = E_N$, з (1.2) та (1.3) отримуємо відповідні необхідні результати для оцінки МНК $\hat{\alpha}$ та її похибки оцінювання $\Delta\left(\hat{\alpha}\right)$.

Наслідок. Оцінка МНК та її похибка оцінювання $\Delta\left(\hat{\alpha}\right)$ для регресійної моделі (1.1) набувають вигляду:

$$\hat{\alpha} = \left(X^T X\right)^{-1} X^T y, \tag{1.4}$$

$$\Delta\left(\hat{\alpha}\right) = \hat{\alpha} - \alpha = \left(X^T X\right)^{-1} X^T e. \tag{1.5}$$

Зауваження 1. *Система нормальних рівнянь для оцінки МНК* $\hat{\alpha}$ буде відповідно мати вигляд

$$\left(X^T X\right)\hat{\alpha} = X^T y. \qquad (1.6)$$

Зауваження 2. При знаходженні оцінок МНК та ЗМНК, а також їх відповідних похибок оцінювання була використана справедливість тільки класичних припущень I та II.

У наступних розділах задача оцінювання невідомих параметрів регресійної моделі буде послідовно розглянута та розв'язана, коли не є справедливе якесь з класичних припущень.

Маючи у своєму розпорядженні оцінку МНК $\hat{\alpha}$ для вектора невідомих параметрів α регресійної моделі (1.1) у вигляді

$$\hat{\alpha} = \left(X^T X\right)^{-1} X^T y$$

значення залежної змінної можна оцінити таким чином:

$$\hat{y} = X\hat{\alpha} \qquad \text{або}$$

$$\hat{y}(k) = x^T(k)\hat{\alpha}, \quad k = \overline{1, N} \ .$$

А в якості оцінки σ^2 можна використати незміщену оцінку методу максимальної правдоподібності

$$\hat{\sigma}^2 = \frac{1}{N-p}\left\|y - X\hat{\alpha}\right\|^2 .$$

Фактично у чисельнику останньої оцінки фігурує *залишкова сума квадратів (residual sum of squares, RSS)*, тобто сума квадратів залишків $\hat{e}(k) = y(k) - x^T(k)\hat{\alpha}$ $\left(k = \overline{1, N}\right)$. Дійсно,

$$\sum_{k=1}^{N}\hat{e}^2(k) = \sum_{k=1}^{N}\left(y(k) - x^T(k)\hat{\alpha}\right)^2 = \left\|y - X\hat{\alpha}\right\|^2 .$$

Таким чином, залишкова сума квадратів дорівнює найменшому значенню функціонала $\|e\|^2$, який досягається у точці $\hat{\alpha}$, а саме:

$$\left.\|e\|^2\right|_{\alpha=\hat{\alpha}} = \left\|y - X\hat{\alpha}\right\|^2 = \|\hat{e}\|^2,$$

де $\hat{e} = \begin{pmatrix}\hat{e}(1) & \hat{e}(2) & \cdots & \hat{e}(N)\end{pmatrix}^T$.

1.3. Властивості оцінки МНК і не тільки

Ознайомимося з базовими властивостями оцінки МНК $\hat{\alpha}$ для α, оцінки $\hat{y}$ для y та $\hat{\sigma}^2$ для σ^2.

Т е о р е м а . Для оцінок $\hat{\alpha}, \hat{y}, \hat{\sigma}^2$ справедливі такі властивості:

I. для оцінки $\hat{\alpha}$ та оцінок її компонент $\left\{\hat{\alpha}_i\right\}_{i=1}^{p}$ справедливо:

a. $$\hat{\alpha} \sim \mathcal{N}\left(\alpha, \sigma^2\left(X^T X\right)^{-1}\right); \qquad (1.7)$$

b. $\hat{\alpha}_i \sim \mathcal{N}\left(\alpha_i, \sigma^2 d_i\right),\ i = \overline{1, p}$,

де $d_i = \left\{\left(X^T X\right)^{-1}\right\}_{ii}$, тобто d_i є i-тим діагональним елементом матриці $\left(X^T X\right)^{-1}$;

II. статистика $\dfrac{1}{\sigma^2}(\hat{\alpha}-\alpha)^T\left(X^T X\right)(\hat{\alpha}-\alpha) \sim \chi^2(p)$;

III. для оцінки $\hat{y}$ та оцінок її компонент $\left\{\hat{y}(k)\right\}_{k=1}^{N}$ справедливо:

a. $\hat{y} \sim \mathcal{N}\left(X\alpha, \sigma^2 X\left(X^T X\right)^{-1} X^T\right)$;

b. $\hat{y}(k) \sim \mathcal{N}\left(x^T(k)\alpha, \sigma^2 x^T(k)\left(X^T X\right)^{-1} x(k)\right),\ \ k = \overline{1, N}$;

IV. для оцінки $\hat{\sigma}^2$ справедливо

a. $\dfrac{(N-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(N-p)$;

b. $M\hat{\sigma}^2=\sigma^2,\ D\hat{\sigma}^2=\dfrac{2\sigma^4}{N-p}$;

V. оцінки $\hat{\alpha}$ та $\hat{\sigma}^2$ є незалежними;

VI. оцінки $\hat{\alpha}, \hat{y}$ та $\hat{\sigma}^2$ є ефективними;

VII. *теорема Андерсона – Тейлора*.

Вкажемо явно поточну кількість спостережень N у регресійній моделі

$$y_N = X_N\alpha + e_N,$$

де $y_N, e_N \in \mathbb{R}^N, X_N \in M_{N,p}(\mathbb{R})$.

Позначимо через $\hat{\alpha}(N)=\left(X_N^T X_N\right)^{-1} X_N^T y_N$ оцінку МНК для α у цій моделі.

Тоді оцінка $\hat{\alpha}(N)$ є сильно слушною тоді і тільки тоді, коли

$$\left(X_N^T X_N\right)^{-1} \xrightarrow[N\to\infty]{} \Theta_p,$$

де Θ_p - нульова матриця порядку p.

2. ОЦІНЮВАННЯ ПАРАМЕТРІВ РЕГРЕСІЙНОЇ МОДЕЛІ ЗА УМОВ ПРОБЛЕМ ЗІ СПРАВЕДЛИВІСТЮ ПЕРШОГО КЛАСИЧНОГО ПРИПУЩЕННЯ

Спочатку згадаємо постановку задачі оцінки вектора невідомих параметрів α лінійної регресійної моделі за справедливості всіх класичних припущень. А саме вважаємо, що для її математичної моделі, записаній у матричному вигляді,

$$y = X\alpha + e \tag{2.1}$$

справедливі всі класичні припущення:

I. $rank\left(X\right) = p$;

II. немає ніяких обмежень на вектор невідомих параметрів α, тобто в якості його значення допускається довільне значення з простору $\mathbb{R}^p$;

III. $e \sim \mathcal{N}\left(\theta_N, \sigma^2 E_N\right),\ \ \sigma^2 \in \mathbb{R}_+$;

IV. математична модель для об’єкта, який досліджується, є лінійною за невідомими параметрами.

Тоді відомо, що за справедливості припущень I, II, III, IV, оцінка МНК $\hat{\alpha}$ вектора параметрів α для моделі (2.1) визначається як точка у якій досягає свого мінімуму функціонал

$$\|e\|^2$$

і вона набуває такого вигляду

$$\hat{\alpha} = \left(X^T X\right)^{-1} X^T y\,. \tag{2.2}$$

Почнемо з дослідження задачі оцінювання невідомих параметрів регресійної моделі (2.1) у випадку, коли є проблеми зі справедливістю тільки першого класичного припущення. А саме вважаємо, що ставиться під сумнів повнота рангу по стовпчикам матриці X . Дослідимо цю ситуацію більш детальніше.

Послабимо припущення I, а саме будемо вважати, що матриця X задовольняє умові:

I'. $rank\left(X\right)=r,\quad r<p$.

Припущення I' еквівалентно умові, що стовпчики матриці X лінійно залежні, тобто що

$$\exists a_i \neq \theta_p : \quad Xa_i = \theta_N ,\, i = \overline{1, p-r} . \tag{2.3}$$

Якщо справедливе припущення I', тобто умова (2.3) виконується строго, то кажуть, що знаходяться в умовах *строгої мультиколінеарності*, якщо ж умова (2.3) виконується тільки приблизно, то кажуть, що знаходяться в умовах *мультиколінеарності (multicollinearity).*

Проаналізуємо послідовно обидва випадки.

2.1. Оцінка параметрів в умовах строгої мультиколінеарності

Тобто спочатку розглянемо задачу оцінювання вектора невідомих параметрів α регресійної моделі (2.1) у випадку, коли справедливі припущення I', II, III, IV. Її розв'язок надає таке твердження.

Теорема. *Оцінка МНК* $\hat{\alpha}$ вектора невідомих параметрів α для моделі (2.1) *в умовах строгої мультиколінеарності*, тобто за справедливості припущень I', II, III, IV буде не єдиною і множина усіх цих оцінок буде задаватися згідно

$$\hat{\alpha}\left(c\right)=X^{+}y+\left[E_p-X^{+}X\right]c, c\in\mathbb{R}^p ,$$

а відповідна похибка оцінювання буде обчислюватися таким чином

$$\Delta\left(\hat{\alpha}\left(c\right)\right)=X^{+}e+\left[E_p-X^{+}X\right]\left(c-\alpha\right), c\in\mathbb{R}^p .$$

Найменшу норму серед усіх цих оцінок буде мати оцінка

$$\hat{\alpha} = X^{+} y,$$

причому формула останньої оцінки суттєво спрощується, якщо матриця X має повний ранг по рядках, тобто $rank(X) = N$, і набуває вигляду

$$\hat{\alpha} = X^{T}\left(XX^{T}\right)^{-1} y .$$

Доведення. Проаналізуємо спочатку наш критерій якості

$$\|e\|^2 = \|y - X\alpha\|^2 . \tag{2.4}$$

Так як у нашому випадку, множина усіх оцінок МНК $\hat{\alpha}$ вектора параметрів α для моделі (2.1) визначається як множина точок у яких досягає свого мінімуму функціонал (2.4), то вона співпадає з множиною усіх псевдорозв’язків відносно α такої системи лінійних алгебраїчних рівнянь

$$X\alpha = y .$$

При знаходженні цієї множини скористаємося оператором псевдообернення за Муром-Пенроузом. Дійсно, неважко бачити, що множину усіх псевдорозв'язків останньої системи можна представити у такому вигляді

$$\hat{\alpha}(c) = X^{+} y + \left[E_p - X^{+} X \right] c, c \in \mathbb{R}^p , \tag{2.5}$$

де $(^{+})$ – оператор псевдообернення за Муром-Пенроузом.

Ортогональність доданків з правої частини співвідношення (2.5) дозволяє стверджувати, що найменшу евклідову норму серед усіх цих оцінок буде мати оцінка, яка задається першим доданком, а саме:

$$\hat{\alpha} = X^{+} y . \tag{2.6}$$

У випадку, коли матриця X має повний ранг по рядках, тобто $rank(X) = N$, підрахування її псевдооберненої матриці можна здійснити згідно

$$X^{+} = X^{T}\left(XX^{T}\right)^{-1}.$$

В результаті формула для оцінки (2.6) суттєво спрощується і набуває вигляду

$$\hat{\alpha} = X^{+}y = X^{T}\left(XX^{T}\right)^{-1}y.$$

Залишилося знайти ще вираз для похибки оцінювання $\Delta\left(\hat{\alpha}(c)\right)$ для оцінки (2.5), дійсно

$$\begin{aligned}\Delta\left(\hat{\alpha}(c)\right) &= \hat{\alpha}(c) - \alpha = X^{+}y + \left[E_p - X^{+}X\right]c - \alpha = \\ &= X^{+}X\alpha + X^{+}e + \left[E_p - X^{+}X\right]c - \alpha = \\ &= X^{+}e + \left[E_p - X^{+}X\right]c - \left[E_p - X^{+}X\right]\alpha = \\ &= X^{+}e + \left[E_p - X^{+}X\right](c - \alpha), c \in \mathbb{R}^{p}.\end{aligned}$$

Що і треба було довести. ■

Зауваження 1. При доведенні теореми не було потреби у використанні припущення III.

Зауваження 2. Раніш отримані формули (1.4) - (1.5) для оцінки МНК $\hat{\alpha}$ та її похибка оцінювання $\Delta(\hat{\alpha})$ для випадку, коли справедливі припущення I, II, III, IV, а саме:

$$\hat{\alpha} = \left(X^{T}X\right)^{-1}X^{T}y,$$

$$\Delta(\hat{\alpha}) = \hat{\alpha} - \alpha = \left(X^{T}X\right)^{-1}X^{T}e$$

в умовах строгої мультиколінеарності не працездатні, бо матриця $\left(X^{T}X\right)$ вже є виродженою.

2.2. Оцінка вектора параметрів в умовах мультиколінеарності

Переходимо до знаходження оцінки вектора невідомих параметрів α регресійної моделі (2.1) в умовах мультиколінеарності. У цьому випадку оцінка МНК у вигляді (2.2) теоретично існує, бо матриця $\left(X^{T}X\right)$ є не виродженою, але її практичне використання буде проблематичним. Дійсно наявність мультиколінеарності приводить до того, що існує принаймні одне власне значення матриці $\left(X^T X\right)$, яке попадає в окіл нуля, а це означає, що $\det\left(X^T X\right)$ буде близьким до нуля, тобто матриця $\left(X^T X\right)$ буде погано обумовленою. А це у свою чергу приводить до ряду негативних наслідків для оцінки МНК у представленні (2.2). А саме:

- Оцінка МНК (2.2) буде вже нестійкою. Тобто незначні зміни в елементах матриці X можуть привести до значних змін у оцінці параметрів $\hat{\alpha}=\left(X^T X\right)^{-1} X^T y$. Свій внесок до цього додають також похибки машинного округлення під час обчислення $\hat{\alpha}$. А це у свою чергу приводить до погіршення якості результатів прогнозу згідно побудованої моделі.
- Оцінка МНК (2.2) буде вже мало ефективною, бо згідно властивості (1.7) її характеристика розсіювання має вигляд $\sigma^2\left(X^T X\right)^{-1}$, а тому суттєво зростає.
- А останнє тягне за собою сильну корельованість між собою оцінок компонент вектора невідомих параметрів регресійної моделі, що ускладнює їх наглядну інтерпретацію.

Один з поширених підходів, який використовується для оцінки невідомих параметрів в умовах мультиколінеарності, це використання гребеневої оцінки, яка була запропонована А. Хоерлом (A. Hoerl) у 1962 році.

Означення. *Гребеневою оцінкою* (ridge estimate) називається оцінка вектора невідомих параметрів α регресійної моделі (2.1), яка визначається таким чином:

$$\hat{\alpha}(\varepsilon)=\left(X^T X+\varepsilon E_p\right)^{-1} X^T y, \varepsilon>0, \tag{2.7}$$

де ε – мале позитивне дійсне число.

Зупинимось на запропонованій гребеневій оцінці більш детально.

Ця оцінка дозволяє подолати погано обумовленість матриці $\left(X^T X\right)$. Але з'ясувалося, що гребенева оцінка є зміщеною, і це призвело спочатку до значного відтермінування її широкого використання на практиці. Проте згодом виявилося, що гребенева оцінка при коректному виборі параметра ε дозволяє досягти навіть більшої точності у середньо квадратичному розумінні у порівнянні з звичайною оцінкою МНК $\hat{\alpha} = \left(X^T X\right)^{-1} X^T y$. Саме це відновило втрачений інтерес до використання на практиці гребеневої оцінки.

Введемо позначення $B(\varepsilon) = \left(X^T X + \varepsilon E_p\right)^{-1}$. Це дозволяє гребеневу оцінку записати у такому вигляді

$$\hat{\alpha}(\varepsilon) = B(\varepsilon) X^T y, \varepsilon > 0. \qquad (2.8)$$

Теорема. Для гребеневої оцінки $\hat{\alpha}(\varepsilon)$ справедливі такі властивості:

1) $\hat{\alpha}(\varepsilon) = B(\varepsilon)\left(X^T X\right)\hat{\alpha} = \left[E_p - \varepsilon B(\varepsilon)\right]\hat{\alpha}$,

   де $\hat{\alpha} = \left(X^T X\right)^{-1} X^T y$;

2) $\hat{\alpha}(\varepsilon) \sim \mathcal{N}\left(\alpha + \Delta\left(\hat{\alpha}(\varepsilon)\right), \sigma^2 B(\varepsilon)\left(X^T X\right) B^T(\varepsilon)\right)$,

   де $\Delta\left(\hat{\alpha}(\varepsilon)\right) = -\varepsilon B(\varepsilon)\alpha$;

3) $M\left\|\hat{\alpha}(\varepsilon) - \alpha\right\|^2 = M\left\|\hat{\alpha}(\varepsilon) - M\hat{\alpha}(\varepsilon)\right\|^2 + \left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2$.

Зауваження 1. З цієї теореми випливають такі висновки:

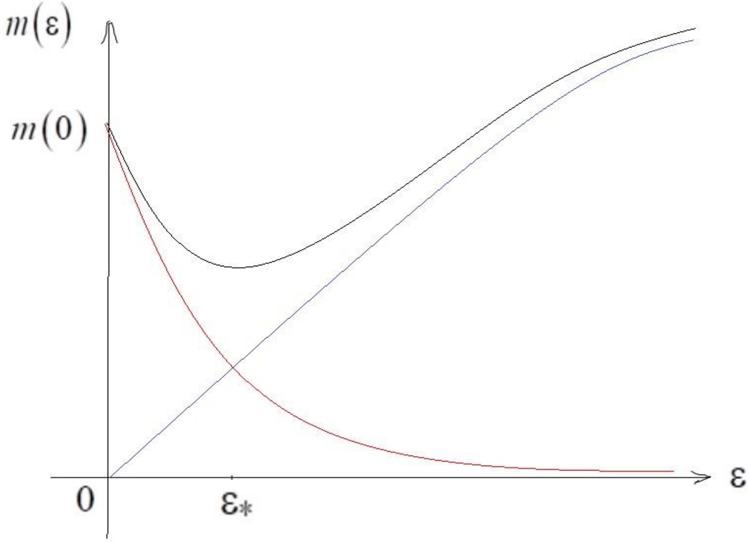
- гребенева оцінка $\hat{\alpha}(\varepsilon)$ є лінійним перетворенням оцінки МНК $\hat{\alpha}$;
- гребенева оцінка $\hat{\alpha}(\varepsilon)$ є зміщеною зі зміщенням $\Delta\left(\hat{\alpha}(\varepsilon)\right)$;
- середньо квадратична похибка оцінювання гребеневої оцінки $M\left\|\hat{\alpha}(\varepsilon) - \alpha\right\|^2$ дорівнює сумі дисперсій гребене-

вих оцінок усіх компонент вектора параметрів α плюс квадрат евклідової норми зміщення $\Delta\left(\hat{\alpha}(\varepsilon)\right)$.

Зауваження 2. Згідно останнього пункту теореми

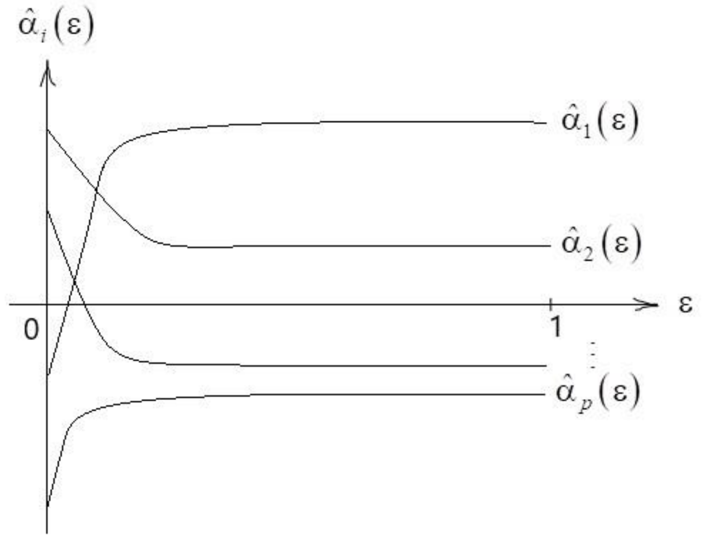
$$m(\varepsilon)=M\left\|\hat{\alpha}(\varepsilon)-\alpha\right\|^2=M\left\|\hat{\alpha}(\varepsilon)-M\hat{\alpha}(\varepsilon)\right\|^2+\left\|\Delta\left(\hat{\alpha}(\varepsilon)\right)\right\|^2.$$

Якщо побудувати графіки кожного з додатків у правій частині останнього рівняння та скласти їх, у підсумку отримаємо графік залежності $m(\varepsilon)$ від ε, який буде мати такий вигляд:



А. Хоерл та Р. Кенард (A. Hoerl, R. Kennard) у 1970 році довели, що завжди існують такі ε_0 при яких $m(\varepsilon_0)<m(0)$, причому ε_* буде найкращим серед таких ε_0. Таким чином, за рахунок незначного зміщення оцінки $\hat{\alpha}(\varepsilon)$ можна суттєво зменшити її середньо квадратичну похибку оцінювання $m(\varepsilon)=M\left\|\hat{\alpha}(\varepsilon)-\alpha\right\|^2$, яка буде меншою навіть $m(0)$, тобто середньо квадратичної похибки оцінювання оцінки звичайного МНК $\hat{\alpha}$.

Зауваження 3. В результаті, залишилося тільки використати гребеневу оцінку $\hat{\alpha}(\varepsilon)$ при $\varepsilon = \varepsilon_*$, або принаймні при ε якомога близькому до ε_*. Знаходження такого ε при якому $m(\varepsilon)$ набуває свого найменшого значення ускладнено тим, що $m(\varepsilon)$ через $\Delta(\hat{\alpha}(\varepsilon)) = -\varepsilon B(\varepsilon)\alpha$ залежить також від вектора невідомих параметрів α. Відповідь на питання, як обирати ε, було дано у публікації А. Хоерла та Р. Кенарда (A. Hoerl, R. Kennard) у 1970 році. Вони в якості практичного підходу по визначенню ε_* запропонували використовувати *гребеневий слід*, тобто об'єднаний графік залежностей гребеневих оцінок $\hat{\alpha}_i(\varepsilon)$, оцінок компонент вектора невідомих параметрів α, від ε, $i = \overline{1, p}$ (як правило, розглядають $\varepsilon \in [0,1]$).



Для цього по цьому графіку визначається таке найменше значення ε_0 при якому усі графіки оцінок $\hat{\alpha}_i(\varepsilon)$ вже стабілізувалися, $i = \overline{1, p}$. Саме це значення ε_0 виявилося близьким до ε_* і його було рекомендовано використовувати на практиці у гребеневій оцінці $\hat{\alpha}(\varepsilon)$. Так як збільшення значення ε буде призводити до зростання не тільки похибки оцінювання $\Delta(\hat{\alpha}(\varepsilon))$, але й згідно теореми до зростання середньо квадратичної похибки оцінювання $m(\varepsilon)$, то стає зрозумілим чому ε потрібно обирати якомога меншим.

3. ОЦІНЮВАННЯ ПАРАМЕТРІВ МОДЕЛІ З УРАХУВАННЯМ ОБЛАСТІ ДОПУСТИМИХ ЗНАЧЕНЬ ДЛЯ НИХ

Познайомимося ще з одним класом задач оцінювання. Принципова відмінність постановки задачі оцінювання параметрів з цього класу задач полягає у тому, що в якості області допустимих значень для вектора невідомих параметрів буде виступати не весь простір $\mathbb{R}^p$, а тільки деяка його підмножина. Для цього випадку потрібно буде знайти відповідну оптимальну оцінку вектора невідомих параметрів.

3.1. Постановка задачі оцінювання параметрів за наявності обмежень

Для наочності проведемо паралель з постановкою задачі оцінювання вектора параметрів α лінійної регресійної моделі у випадку, коли справедливі усі класичні припущення. Нехай її математична модель записана у матричному вигляді

$$y = X\alpha + e \tag{3.1}$$

та мають місце всі класичні припущення

I. $rank(X) = p$;

II. немає ніяких обмежень на вектор невідомих параметрів α, тобто в якості його значення допускається довільне значення з простору $\mathbb{R}^p$;

III. $e \sim \mathcal{N}\left(\theta_N, \sigma^2 E_N\right), \sigma^2 \in \mathbb{R}_+$;

IV. математична модель для об'єкта, який досліджується, є лінійною за невідомими параметрами.

Оцінка $\hat{\alpha}$ вектора параметрів α для моделі (3.1), коли справедливі всі класичні припущення, згідно МНК визначається як точка у якій досягає свого мінімуму функціонал

$$\hat{\alpha} = \arg\min_{\alpha} \|e\|^2$$

і як відомо має такий вигляд

$$\hat{\alpha} = \left(X^T X\right)^{-1} X^T y . \tag{3.2}$$

Змінимо тепер традиційне припущення II. А саме будемо вважати, що в якості області допустимих значень для вектора параметрів α виступає вже не весь простір $\mathbb{R}^p$, а тільки множина усіх α, які задовольняють деякій системі алгебраїчних рівнянь $A\alpha = b$ з відомою матрицею A та відомим вектором b відповідних розмірностей, тобто припущення II трансформується у таке припущення:

II'. $\alpha \in \mathcal{L}$,

де $\mathcal{L} = \left\{\alpha : A\alpha = b, rank\left(A\right) = q\right\}, A \in M_{q,p}\left(\mathbb{R}\right), b \in \mathbb{R}^q$.

Тоді оцінка МНК для об'єкта (3.1) за справедливості припущень I, II', III, IV, тобто за наявності лінійних обмежень, вже є розв'язком такої оптимізаційної задачі

$$\hat{\alpha}_{\mathcal{L}} = \arg\min_{\alpha \in \mathcal{L}} \|e\|^2 .$$

Тобто потрібна оцінка МНК шукається вже не у всьому просторі $\mathbb{R}^p$, а тільки на його підмножині $\mathcal{L}$, непорожність якої гарантується умовою повноти рангу матриці A по рядкам, а саме умовою $rank\left(A\right) = q$.

3.2. Оцінювання параметрів за наявності лінійних обмежень

Залишилось визначити вираз потрібної нам оцінки МНК $\hat{\alpha}_{\mathcal{L}}$ для об'єкта (3.1) за наявності лінійних обмежень, які визначаються припущенням II'.

Вигляд цієї оцінки задається таким твердженням.

Теорема. Оцінка МНК $\hat{\alpha}_{\mathcal{L}}$ для об'єкта (3.1) за справедливості припущень I, II', III, IV, тобто за наявності лінійних обмежень, визначається таким чином

$$\hat{\alpha}_{\mathcal{L}} = \hat{\alpha} - \left(X^T X\right)^{-1} A^T \left[A\left(X^T X\right)^{-1} A^T\right]^{-1} \left[A\hat{\alpha} - b\right], \qquad (3.3)$$

де $\hat{\alpha} = \left(X^T X\right)^{-1} X^T y$.

Наслідок. Похибка оцінювання $\Delta\left(\hat{\alpha}_{\mathcal{L}}\right)$ для оцінки МНК $\hat{\alpha}_{\mathcal{L}}$ для об'єкта (3.1) за наявності лінійних обмежень набуває такого вигляду:

$$\Delta\left(\hat{\alpha}_{\mathcal{L}}\right) = \hat{\alpha}_{\mathcal{L}} - \alpha = U\left(X^T X\right)^{-1} X^T e, \qquad (3.4)$$

де $U = E_p - \left(X^T X\right)^{-1} A^T \left[A\left(X^T X\right)^{-1} A^T\right]^{-1} A$.

Доведення наслідку. Так як згідно (1.5)

$$\Delta\left(\hat{\alpha}\right) = \hat{\alpha} - \alpha = \left(X^T X\right)^{-1} X^T e,$$

то $\hat{\alpha} = \alpha + \left(X^T X\right)^{-1} X^T e$.

Тоді

$$\hat{\alpha}_{\mathcal{L}} = \hat{\alpha} - \left(X^T X\right)^{-1} A^T \left[A\left(X^T X\right)^{-1} A^T\right]^{-1} \left[A\hat{\alpha} - b\right] =$$

$$= \alpha + \left(X^T X\right)^{-1} X^T e - \left(X^T X\right)^{-1} A^T \left[A\left(X^T X\right)^{-1} A^T\right]^{-1} \times$$

$$\times\left[A\alpha + A\left(X^T X\right)^{-1} X^T e - b\right] =$$

$$= \alpha + \left\{E_p - \left(X^T X\right)^{-1} A^T \left[A\left(X^T X\right)^{-1} A^T\right]^{-1} A\right\}\times$$

$$\times\left(X^T X\right)^{-1} X^T e = \alpha + U\left(X^T X\right)^{-1} X^T e.$$

Або остаточно отримуємо

$$\Delta\left(\hat{\alpha}_{\mathcal{L}}\right) = \hat{\alpha}_{\mathcal{L}} - \alpha = U\left(X^T X\right)^{-1} X^T e. \qquad \blacksquare$$

Зауваження. Для моделі (3.1) за справедливості припущень I, II', III, IV оцінки для вектора спостережень y та його компонент $y(k)$ набувають відповідно представлень

$$\hat{y}_{\mathcal{L}} = X\hat{\alpha}_{\mathcal{L}},\ \hat{y}_{\mathcal{L}}(k) = x^T(k)\hat{\alpha}_{\mathcal{L}},\ k = \overline{1, N},$$

а незміщена оцінка $\hat{\sigma}^2_{\mathcal{L}}$ методу максимальної правдоподібності параметра σ^2 має вигляд:

$$\hat{\sigma}^2_{\mathcal{L}} = \frac{1}{N - (p - q)}\left\|y - X\hat{\alpha}_{\mathcal{L}}\right\|^2.$$

Ознайомитися з деякими властивостями наведених оцінок надає можливість таке твердження.

Теорема. Оцінки $\hat{\alpha}_{\mathcal{L}}$, $\hat{\sigma}^2_{\mathcal{L}}$ мають такі властивості:

I. $\hat{\alpha}_{\mathcal{L}} \sim \mathcal{N}\left(\alpha, \sigma^2 U\left(X^T X\right)^{-1}\right)$;

II. $M\hat{\sigma}^2_{\mathcal{L}} = \sigma^2$,

де $U = E_p - \left(X^T X\right)^{-1} A^T \left[A\left(X^T X\right)^{-1} A^T\right]^{-1} A$.

В подальшому, при доведенні теореми, знадобиться така лема, справедливість якої не складно перевірити.

Лема. Мають місце такі матричні тотожності:

- $\left(X^T X\right)^{-1} U^T = U\left(X^T X\right)^{-1}$;
- $U^2 = U$.

Доведення теореми. Спочатку впевнимося у справедливості пункту I цієї теореми.

Так як згідно (3.4) та припущення III класичного регресійного аналізу

$$\begin{cases} \hat{\alpha}_{\mathcal{L}} = \alpha + U\left(X^T X\right)^{-1} X^T e, \\ e \sim \mathcal{N}\left(\theta_N, \sigma^2 E_N\right), \sigma^2 \in \mathbb{R}_+, \end{cases}$$

тоді оцінка $\hat{\alpha}_{\mathcal{L}}$, як лінійне перетворення нормально розподіленого вектора e, теж буде нормально розподілена (див. додаток). Скориставшись теоремою з додатку та допоміжною вищенаведеною лемою, знайдемо параметри цього розподілу

$$\hat{\alpha}_{\mathcal{L}} \sim \mathcal{N}\Big(\alpha + U\left(X^T X\right)^{-1} X^T \theta_N,$$

$$\sigma^2 U\left(X^T X\right)^{-1} X^T E_N X \left(X^T X\right)^{-1} U^T\Big)$$

$$\Rightarrow \hat{\alpha}_{\mathcal{L}} \sim \mathcal{N}\left(\alpha, \sigma^2 U\left(X^T X\right)^{-1} U^T\right)$$

$$\Rightarrow \hat{\alpha}_{\mathcal{L}} \sim \mathcal{N}\left(\alpha, \sigma^2 U U\left(X^T X\right)^{-1}\right)$$

$$\Rightarrow \hat{\alpha}_{\mathcal{L}} \sim \mathcal{N}\left(\alpha, \sigma^2 U\left(X^T X\right)^{-1}\right).$$

Переходимо до доведення пункту II з теореми. Для доведення того, що $M\hat{\sigma}^2_{\mathcal{L}} = \sigma^2$, достатньо впевнитися, що

$$M\left\|y - X\hat{\alpha}_{\mathcal{L}}\right\|^2 = \left[N - (p-q)\right]\sigma^2 .$$

Використання (3.4), а саме що $\hat{\alpha}_{\mathcal{L}} = \alpha + U\left(X^T X\right)^{-1} X^T e$, та допоміжної леми, дозволяє стверджувати, що справедливий такий ланцюг перетворень

$$M\left\|y - X\hat{\alpha}_{\mathcal{L}}\right\|^2 = M\left\|y - X\left(\alpha + U\left(X^T X\right)^{-1} X^T e\right)\right\|^2 =$$

$$= M\left\|y - X\alpha - XU\left(X^T X\right)^{-1} X^T e\right\|^2 =$$

$$= M\left\|e - XU\left(X^T X\right)^{-1} X^T e\right\|^2 =$$

$$= M\left\|\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]e\right\|^2 =$$

$$= Me^T\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]^T\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]e =$$

$$= M\,tr\left\{e^T\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]^T \times\right.$$

$$\left.\times\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]e\right\} =$$

$$= M\,tr\left\{ee^T\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]^T \times\right.$$

$$\left.\times\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]\right\} =$$

$$= tr\left\{M\left(ee^T\right)\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]^T \times\right.$$

$$\left.\times\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]\right\} =$$

$$= \sigma^2 tr\left\{\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]^T\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]\right\} =$$

$$= \sigma^2 tr\left\{\left[E_N - X\left(X^T X\right)^{-1} U^T X^T\right]\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]\right\} =$$

$$= \sigma^2 tr\left\{\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]^2\right\}.$$

А так як матриця $\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]$ є ідемпотентною, тобто

$$\left[E_N - XU\left(X^T X\right)^{-1} X^T\right]^2 = \left[E_N - XU\left(X^T X\right)^{-1} X^T\right],$$

то далі отримуємо таке

$$M\left\|y - X\hat{\alpha}_{\mathcal{L}}\right\|^2 = \sigma^2 tr\left\{E_N - XU\left(X^T X\right)^{-1} X^T\right\} =$$

$$= \sigma^2\left\{N - tr\left[XU\left(X^T X\right)^{-1} X^T\right]\right\} =$$

$$= \sigma^2\left\{N - tr\left[U\left(X^T X\right)^{-1} X^T X\right]\right\} = \sigma^2\left\{N - tr\left[U\right]\right\} =$$

$$= \sigma^2\left\{N - tr\left[E_p - \left(X^T X\right)^{-1} A^T\left[A\left(X^T X\right)^{-1} A^T\right]^{-1} A\right]\right\} =$$

$$= \sigma^2\left\{N - \left[tr\left(E_p\right) - tr\left(\left(X^T X\right)^{-1} A^T\left[A\left(X^T X\right)^{-1} A^T\right]^{-1} A\right)\right]\right\} =$$

$$= \sigma^2\left\{N - \left[p - tr\left(A\left(X^T X\right)^{-1} A^T\left[A\left(X^T X\right)^{-1} A^T\right]^{-1}\right)\right]\right\} =$$

$$= \sigma^2\left\{N - \left[p - tr\left(E_q\right)\right]\right\} = \sigma^2\left\{N - \left[p - q\right]\right\}. \qquad \blacksquare$$

4. ОЦІНЮВАННЯ ПАРАМЕТРІВ РЕГРЕСІЙНОЇ МОДЕЛІ В УМОВАХ ПОРУШЕННЯ ТРЕТЬОГО КЛАСИЧНОГО ПРИПУЩЕННЯ

Переходимо до дослідження задачі оцінювання невідомих параметрів регресійної моделі у випадку, коли може бути порушена справедливість третього класичного припущення. А саме будемо вважати, що похибки математичної моделі можуть вже бути корельованими та неоднорідними.

4.1. Постановка задачі та основні припущення

Знову спочатку апелюємо до постановки задачі оцінювання вектора невідомих параметрів лінійної регресійної моделі у випадку, коли справедливі усі класичні припущення крім третього, яке буде послаблене. Нехай її математична модель записана у матричному вигляді

$$y = X\alpha + e \tag{4.1}$$

та мають місце такі припущення:

I. $rank(X) = p$;

II. немає ніяких обмежень на вектор невідомих параметрів α, тобто в якості його значення допускається довільне значення з простору $\mathbb{R}^p$;

III'. $e \sim \mathcal{N}(\theta_N, V), V > 0$;

IV. математична модель для об’єкта, який досліджується, є лінійною за невідомими параметрами.

Фактично було замінено тільки класичне припущення III на припущення III'. Тоді в якості оцінки вектора параметрів α об'єкта (4.1) за справедливості припущень I, II, III', IV, тобто за наявності неоднорідних та корельованих похибок моделі, пропонується використовувати таку оцінку.

Означення. *Марковською оцінкою (markov estimate)* для вектора невідомих параметрів α регресійної моделі (4.1) за справедливості припущень I, II, III', IV називається оцінка на якій досягає свого мінімуму функціонал

$$\|e\|_{V^{-1}}^2, \ V > 0 .$$

4.2. Оцінювання параметрів регресійної моделі при корельованих і неоднорідних похибках

Яким чином можна прийти саме до оцінки, яка визначена у останньому означенні. Для цього зведемо задачу оцінювання вектора параметрів α об'єкту (4.1) за справедливості припущень I, II, III', IV до задачі оцінювання вектора параметрів α деякої перетвореної моделі, але вже з некорельованими та однорідними похибками моделі, тобто у разі справедливості традиційних припущень I, II, III, IV.

Для цього співвідношення (4.1) помножимо зліва на матрицю $V^{-\frac{1}{2}}$. В результаті отримаємо

$$V^{-\frac{1}{2}}y = V^{-\frac{1}{2}}X\alpha + V^{-\frac{1}{2}}e .$$

Або після введення позначень

$$\tilde{y} = V^{-\frac{1}{2}} y,\ \tilde{X} = V^{-\frac{1}{2}} X,\ \tilde{e} = V^{-\frac{1}{2}} e,$$

маємо

$$\tilde{y} = \tilde{X}\alpha + \tilde{e}. \tag{4.2}$$

Тоді згідно наслідку 1 до теореми з додатку для вектора похибок $\tilde{e}$ перетвореної моделі (4.2) отримаємо, що

$$\tilde{e} \sim \mathcal{N}\left(\theta_N, E_N\right),$$

тобто для $\tilde{e}$ вже справедливо III класичне припущення.

Застосуємо тепер для знаходження оцінки вектора параметрів α об'єкта (4.2) за справедливості класичних припущень I, II, III, IV оцінку звичайного МНК $\hat{\hat{\alpha}}$, в результаті отримаємо

$$\hat{\hat{\alpha}} = \left(\tilde{X}^T \tilde{X}\right)^{-1} \tilde{X}^T \tilde{y} = \left(\left[V^{-\frac{1}{2}} X\right]^T V^{-\frac{1}{2}} X\right)^{-1} \left[V^{-\frac{1}{2}} X\right]^T V^{-\frac{1}{2}} y =$$

$$= \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y = \hat{\alpha}_{V^{-1}}. \tag{4.3}$$

Останній вираз для оцінки $\hat{\hat{\alpha}}$ є не що інше, як оцінка ЗМНК $\hat{\alpha}_{V^{-1}}$ з ваговою матрицею V^{-1}. Але марковська оцінка, згідно її означення, і є оцінкою ЗМНК з ваговою матрицею V^{-1}.

Її похибка оцінювання, у свою чергу згідно (1.3), буде мати вигляд

$$\Delta\left(\hat{\alpha}_{V^{-1}}\right) = \hat{\alpha}_{V^{-1}} - \alpha = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} e.$$

Підсумуємо отримані результати у вигляді такого твердження.

Теорема. Марковська оцінка $\hat{\alpha}_{V^{-1}}$ вектора невідомих параметрів α для регресійної моделі (4.1) за справедливості припущень I, II, III', IV та її похибка оцінювання $\Delta\left(\hat{\alpha}_{V^{-1}}\right)$ визначаються таким чином:

$$\hat{\alpha}_{V^{-1}} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y,$$

$$\Delta\left(\hat{\alpha}_{V^{-1}}\right) = \hat{\alpha}_{V^{-1}} - \alpha = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} e.$$

Проведемо дослідження деяких можливостей марковської оцінки.

Властивості марковської оцінки.

1. $\hat{\alpha}_{V^{-1}} \sim \mathcal{N}\left(\alpha, \left(X^T V^{-1} X \right)^{-1} \right)$.
2. Оцінка $\hat{\alpha}_{V^{-1}}$ є ефективною.
3. Вкажемо явно у регресійній моделі поточну кількість спостережень N:

$$y_N = X_N \alpha + e_N,$$

де $y_N, e_N \in \mathbb{R}^N, X_N \in M_{N,p}(\mathbb{R})$.

Нехай

$$e_N \sim \mathcal{N}(\theta_N, V_N), V_N > 0, \ N \in \mathbb{N}.$$

Позначимо через

$$\hat{\alpha}_{V_N^{-1}}(N) = \left(X_N^T V_N^{-1} X_N \right)^{-1} X_N^T V_N^{-1} y_N$$

марковську оцінку для α у цій моделі.

Тоді марковська оцінка $\hat{\alpha}_{V_N^{-1}}(N)$ є сильно слушною тоді і тільки тоді, коли

$$\left(X_N^T V_N^{-1} X_N \right)^{-1} \xrightarrow[N \to \infty]{} \Theta_p,$$

де Θ_p - нульова матриця порядку p.

Доведення. Так як для марковської оцінки, згідно (4.3), має місце

$$\hat{\alpha}_{V^{-1}} = \hat{\hat{\alpha}} = \left(\tilde{X}^T \tilde{X} \right)^{-1} \tilde{X}^T \tilde{y},$$

то скористаємося властивостями оцінки $\hat{\hat{\alpha}}$, як оцінки МНК для моделі (4.2) за справедливості класичних припущень I, II, III, IV, згідно теореми з розділу 1.3. Послідовно впевнимося у справедливості кожної з властивостей марковської оцінки.

1. Дійсно, згідно (1.7) з вищезгаданої теореми, оцінка $\hat{\hat{\alpha}}$ має такий розподіл

$$\hat{\hat{\alpha}} \sim \mathcal{N}\left(\alpha, \left(\tilde{X}^T \tilde{X}\right)^{-1}\right) \Rightarrow$$

$$\Rightarrow \hat{\hat{\alpha}} \sim \mathcal{N}\left(\alpha, \left(X^T V^{-1} X\right)^{-1}\right) \Rightarrow$$

$$\Rightarrow \hat{\alpha}_{V^{-1}} \sim \mathcal{N}\left(\alpha, \left(X^T V^{-1} X\right)^{-1}\right).$$

2. Так як, згідно вищезгаданої теореми, оцінка $\hat{\hat{\alpha}}$ є ефективною, а $\hat{\alpha}_{V^{-1}} = \hat{\hat{\alpha}}$, то і марковська оцінка $\hat{\alpha}_{V^{-1}}$ є ефективною.

3. Легко бачити що марковська оцінка

$$\hat{\alpha}_{V_N^{-1}}(N) = \hat{\hat{\alpha}}(N) = \left(\tilde{X}_N^T \tilde{X}_N\right)^{-1} \tilde{X}_N^T \tilde{y}_N,$$

де $\hat{\hat{\alpha}}(N)$ є оцінкою МНК для об'єкта

$$\tilde{y}_N = \tilde{X}_N \alpha + \tilde{e}_N$$

за справедливості класичних припущень I, II, III, IV, причому припущення III набуває вигляду

$$\tilde{e}_N \sim \mathcal{N}\left(\theta_N, E_N\right),$$

де

$$\tilde{y}_N = V_N^{-\frac{1}{2}} y_N,\ \tilde{X}_N = V_N^{-\frac{1}{2}} X_N,\ \tilde{e}_N = V_N^{-\frac{1}{2}} e_N.$$

Згідно теореми Андерсона-Тейлора оцінка $\hat{\hat{\alpha}}(N)$ є сильно слушною тоді і тільки тоді, коли

$$\left(\tilde{X}_N^T \tilde{X}_N\right)^{-1} \xrightarrow[N \to \infty]{} \Theta_p.$$

А так як

$$\hat{\hat{\alpha}}(N) = \hat{\alpha}_{V_N^{-1}}(N), \quad \left(\tilde{X}_N^T \tilde{X}_N\right)^{-1} = \left(X_N^T V_N^{-1} X_N\right)^{-1},$$

То і марковська оцінка $\hat{\alpha}_{V_N^{-1}}(N)$, у свою чергу, є сильно слушною тоді і тільки тоді, коли

$$\left(X_N^T V_N^{-1} X_N\right)^{-1} \xrightarrow[N \to \infty]{} \Theta_p. \qquad \blacksquare$$

Зауваження. Практичне використання марковської оцінки ускладнюється обмеженням доступу до інформації про коваріаційну матрицю V для похибок моделі e. Як правило, на практиці доступна тільки деяка її оцінка.

5. ОЦІНЮВАННЯ ПАРАМЕТРІВ НЕЛІНІЙНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ

Звернемось до дослідження задачі оцінювання невідомих параметрів моделі у випадку, коли може бути порушена справедливість четвертого класичного припущення. А саме вважається, що математична модель для об’єкта, який досліджується, є нелінійною за невідомими параметрами. Намітимо шляхи знаходження оцінки цих параметрів.

5.1. Постановка задачі та класифікація нелінійних моделей

Нехай для об'єкта, який досліджується, математична *модель є нелінійною за вектором невідомих параметрів* α і в k -ий момент набуває вигляду

$$y(k) = g\left(x(k), \alpha\right) + e(k), k = \overline{1, N}, \tag{5.1}$$

де $y(k)$ - k -те спостереження залежної скалярної змінної, $x(k)$ - k -те спостереження вектора усіх регресорів, $g\left(x(k), \alpha\right)$ є відомою нелінійною функцією за вектором параметрів α , $\alpha \in \mathbb{R}^p$.

З моделлю такого плану можна стикнутися у ситуації, коли в якості функції, якою наближають функцію регресії η щодо вектора $\vec{\xi}$, виступає деяка задана нелінійна функція за вектором параметрів α .

У моделі (5.1) похибка моделі $e(k)$ є адитивною, але у деяких випадках вона може бути і мультиплікативною.

Необхідно за доступними значеннями $\{y(k)\}_{k=1}^{N}$ та $\{x(k)\}_{k=1}^{N}$ для об'єкта (5.1) знайти оптимальну у деякому розумінні оцінку вектора невідомих параметрів α .

Нелінійні моделі (5.1) поділяються на *внутрішньолінійні* та *внутрішньонелінійні*.

Означення. Нелінійна модель (5.1) називається *внутрішньолінійною*, якщо її шляхом деяких перетворень можна привести до моделі, яка є лінійною за вектором невідомих параметрів, інакше вона називається *внутрішньонелінійною*.

5.2. Знаходження оцінки параметрів нелінійної моделі

Окреслимо шляхи пошуку оцінки вектора невідомих параметрів методом найменших квадратів для вищезгаданих двох класів нелінійних моделей.

І. Внутрішні лінійні моделі. Нехай деяка нелінійна модель є внутрішньолінійною. Значить для неї існує таке перетворення, яке зводить її до моделі, яка є лінійною по вектору невідомих параметрів.

Наведемо декілька прикладів внутрішньолінійних моделей з одним регресором:

- $\dfrac{1}{y(k)} = \alpha_1 + \alpha_2 x(k) + e(k), k = \overline{1, N}$,
- $y(k) = e^{\alpha_1 + \alpha_2 x(k) + e(k)}, k = \overline{1, N}$,
- $y(k) = \alpha_1 x^{\alpha_2}(k) e(k), k = \overline{1, N}$,

де $y(k)$ - k -те спостереження залежної скалярної змінної, $x(k)$ - k -те спостереження єдиного скалярного регресора, α_1, α_2 - скалярні невідомі параметри, в останньому прикладі моделі часто припускають, що мультиплікативні похибки

$$e(k) \sim \mathcal{LN}\left(0,\sigma^2\right), \quad \sigma^2 > 0, \quad k = \overline{1,N} ,$$

тобто логнормально розподілені.

Кожна з цих моделей перепозначенням або/та логарифмуванням зводиться до моделі лінійної за параметрами.

У загальному випадку для внутрішньо лінійної моделі існують такі перетворення $g_1(\bullet),\left\{g_{2j}(\bullet)\right\},\left\{g_{3j}(\bullet)\right\}$, що після їх використання, приходимо до моделі лінійної за деякими параметрами $\left\{\beta_j\right\}$ з похибками моделі $\left\{\tilde{e}(k)\right\}_{k=1}^{N}$:

$$g_1\left(y(k)\right) = \sum_j \beta_j g_{3j}\left(x(k)\right) + \tilde{e}(k), k = \overline{1,N} , \tag{5.2}$$

де

$$\beta_j = g_{2j}(\alpha), \forall j . \tag{5.3}$$

Тоді з системи лінійних рівнянь (5.2), спочатку за допомогою МНК знаходимо оцінки $\left\{\hat{\beta}_j\right\}$.

На наступному кроці, підставивши отримані оцінки $\left\{\hat{\beta}_j\right\}$ у систему рівнянь (5.3), знаходимо з неї оцінку $\hat{\alpha}$ вектора невідомих параметрів α .

Зауваження. Обчислена таким чином оцінка $\hat{\alpha}$, отримується на базі оцінок $\left\{\hat{\beta}_j\right\}$, які мінімізують суму квадратів похибок перетвореної моделі (5.2), а оцінка $\hat{\alpha}$ не гарантує мінімізацію суми квадратів похибок оригінальної моделі (5.1). Тому отриману оцінку $\hat{\alpha}$ можна використовувати тільки в якості деякого наближення до оптимальної оцінки шуканого вектора невідомих параметрів α оригінальної моделі (5.1).

ІІ. Внутрішні нелінійні моделі. Нехай деяка нелінійна математична модель

$$y(k) = g\left(x(k),\alpha\right) + e(k), k = \overline{1,N}$$

є внутрішньонелінійною. Значить для неї не існує або не вдалося знайти таке перетворення, яке зводить її до моделі, яка є лінійною за вектором невідомих параметрів. Яким чином у цьому випадку пропонується шукати оптимальну оцінку вектора невідомих параметрів α ?

У цій ситуації в якості оптимальної оцінки вектора невідомих параметрів α можна взяти розв'язок такої оптимізаційної задачі

$$\hat{\alpha} = \arg\min_{\alpha} Q(\alpha),$$

де $Q(\alpha) = \sum_{k=1}^{N} e^2(k)$.

Для знаходження точки, у якій досягається екстремум цього нелінійного функціонала $Q(\alpha)$ потрібно буде скористатися методами безумовної нелінійної оптимізації в залежності від класу нелінійних функцій з яким стикнулися. Наприклад: градієнтним методом, методом лінеарізації, або компілятом двох останніх – методом Марквартa, методами випадкового пошуку тощо.

ДОДАТОК

Розподіл деяких перетворень нормально розподіленого вектора

Нехай випадковий вектор $\vec{\xi}$ має нормальний розподіл

$$\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n,$$

тоді його щільність набуває вигляду

$$p_{\vec{\xi}}\left(\vec{x}\right) = \frac{1}{\left(2\pi\right)^{\frac{n}{2}}\left(\det\left(V\right)\right)^{\frac{1}{2}}} e^{-\frac{1}{2}\left\|\vec{x}-\vec{m}\right\|_{V^{-1}}^{2}}.$$

Вигляд розподілу лінійного перетворення цього вектора надає таке твердження.

Т е о р е м а . Нехай

- $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n$;
- $\vec{a} \in \mathbb{R}^m$, B – матриця розмірності $m \times n$.

Тоді справедливо

$$\vec{a} + B\vec{\xi} \sim \mathcal{N}\left(\vec{a} + B\vec{m}, BVB^T\right).$$

З останньої теореми випливає справедливість таких висновків.

Наслідок 1. Якщо $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n$, то

$$V^{-\frac{1}{2}}\left(\vec{\xi} - \vec{m}\right) \sim \mathcal{N}\left(\theta_n, E_n\right),$$

де θ_n – нульовий вектор розмірності n, E_n – одинична матриця порядку n.

Наслідок 2. Нехай $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n$, тоді

$$\left(\vec{\xi} - \vec{m}\right)^T V^{-1}\left(\vec{\xi} - \vec{m}\right) \sim \chi^2\left(n\right).$$

Наслідок 3. Якщо $\vec{\xi} \sim \mathcal{N}\left(\vec{m}, V\right), V > 0,\ \vec{\xi} \in \mathbb{R}^n,$ тоді його можна представити як деяке лінійне перетворення стандартно розподіленого нормального вектора $\vec{\xi}_*$, а саме:

$$\vec{\xi} = \vec{m} + V^{\frac{1}{2}}\vec{\xi}_*,$$

де $\vec{\xi}_* \sim \mathcal{N}\left(\theta_n, E_n\right)$.

СПИСОК ЛІТЕРАТУРИ

1. *Чарін, В.С.* Лінійна алгебра / В.С. Чарін. — Київ : Техніка, 2004.
2. *Albert, A.* Regression and the Moore-Penrose Pseudoinverse / A. Albert. — New York : Academic Press, 1972.
3. *Eykhoff, P.* System Identification: Parameter and State Estimation / P. Eykhoff. — Chichester, England : Wiley, 1974.
4. *Graupe, D.* Identification of Systems / D. Graupe. — Huntington, New York : Robert E. Krieger Publishing Company, 1976.
5. *Hansen, P. C.* Least Squares Data Fitting with Applications / P. C. Hansen, V. Pereyra, & G. Scherer. — Baltimore, MD : Johns Hopkins University Press, 2013.
6. Linear Models and Generalizations: Least Squares and Alternatives / C. R. Rao, H. Toutenburg, Shalabh, & C. Heumann. — 3rd Edition. — Berlin : Springer, 2008.
7. *Ljung, L.* System Identification: Theory for the User / L. Ljung. — 2nd Edition. — Upper Saddle River, NJ : Prentice-Hall, 1999.
8. *Weisberg, S.* Applied Linear Regression / S. Weisberg. — 4th Edition. — Hoboken, NJ : Wiley, 2014.

ПРЕДМЕТНИЙ ПОКАЖЧИК

ПОКАЖЧИК ПОЗНАЧЕНЬ

E_N – одинична матриця порядку N,

$\mathbb{R}_+$ – множина позитивних дійсних чисел,

$\|\vec{x}\|^2 = \vec{x}^T \vec{x}$.

$\|\vec{x}\|_Q^2 = \vec{x}^T Q \vec{x}, Q > 0$.

$\hat{y} = X\hat{\alpha}$.

$\hat{y}(k) = x^T(k)\hat{\alpha}, \quad k = \overline{1, N}$.

θ_N – нульовий вектор розмірності N,

Θ_p - нульова матриця порядку p,

$$\hat{\sigma}^2 = \frac{1}{N-p}\|y - X\hat{\alpha}\|^2.$$

$\xi \sim \mathcal{N}(m, \sigma^2)$ – випадкова величина ξ нормально розподілена з параметрами m та σ^2.

$\vec{\xi} \sim \mathcal{N}(\vec{m}, V)$ – випадковий вектор $\vec{\xi}$ нормально розподілений з параметрами $\vec{m}$ та V.

ЗМІСТ

Навчальне видання

СЛАБОСПИЦЬКИЙ Олександр Сергійович

Оцінювання параметрів систем при некласичних припущеннях

Конспект лекцій

Друкується за авторською редакцією

Підписано до друку 21.11.2025. Формат 60×84/16.
Папір офсетний. Гарнітура Таймс. Друк офсетний.
Умов. друк. арк. 3. Наклад 300. Зам. 311.
Надруковано у «Видавництво Людмила»
Свідоцтво про внесення до Державного реєстру
суб’єктів видавничої справи ДК № 5303 від 02.03.2017
«Видавництво Людмила”. 03148, Київ, а/с 115
Тел./факс: +380504697485, 0683408332.
E-mail: lesya3000@ukr.net

Для нотаток