Пістунов І.М.

Економетрика: навч. наоч. посіб. Дніпро : НТУ «ДП», 2024. 36 с.

В посібнику подано приклад і наведено порядок підготовки даних, їх аналізу на достовірність та достатність, перед розрахунком коефіцієнтів лінійних та квазілінійних моделей.
Наведено поняття похибки прогнозування та моделювання, статистичної достовірності коефіцієнтів моделей.
Наведено приклади розрахунків, надано пояснення щодо структури даних.
Призначено для студентів спеціальності 051 «Економіка»

Рецензенти:
Васильєва Н.К., завідувач каф. інформаційних систем ДДАЕУ, проф.
Пилипенко Ю.І., зав каф. Економічної теорії та міжнародних економічних відносин НТУ “ДП”, проф.

Назва «економетрика» вперше було введено в 1926 р норвезьким економістом і статистом Рагнаром Фришем (Frisch). Тому, можна сказати, що економетрія - порівняно молода галузь науки. 29 грудня 1930 за ініціативою І. Фішера (1867-1947), Р. Фріша, Я. Тінберген (1903-1995), Й. Шумпетера, О.Андерсона (1887-1960) та інших вчених на засіданні Американської асоціації розвитку науки (США, Клівленд, штат Огайо) було створено економетричне суспільство, на якому норвезький учений Р.Фріш дав новій науці назву «економетрика». Створене Економетричне товариство позначило себе так: «Міжнародне товариство для розвитку економетричної теорії і її зв'язок зі статистикою і математикою».

Визначення економетрики: «Економетрика - це не те ж саме, що економічна статистика. Вона не ідентична і тому, що ми називаємо економічної теорією, хоча значна частина цієї теорії носить кількісний характер. Економетрика не є синонімом додатків математики до економіки. Як показує досвід, кожна з трьох відправних точок - статистика, економічна теорія та математика - необхідна, але не достатня умова для розуміння кількісних співвідношень в сучасному економічному житті. Це єдність всіх трьох складових. І це єдність утворює економетрику »

Об'єкт курсу «Економетрика» - сукупність різних соціально-економічних процесів, що протікають в економічній системі.

Предмет курсу «Економетрика» - економетричні методи і моделі, що дозволяють визначити і вивчити кількісні взаємозв'язку між соціально - економічними процесами і явищами.

Мета курсу «Економетрика» - познайомитися з основними економетричними завданнями, навчитися формувати економетричні моделі різних рівнів.

Завдання курсу «Економетрика» - побудова і аналіз основних економетричних моделей, оцінка параметрів, перевірка значущості моделі, прогнозування на основі побудованої моделі.

Курс «Економетрика» тісно пов'язаний з іншими дисциплінами і базується на фундаменті математичних і економічних знань. Навчальний курс «Економетрика» спирається на курси «Мікроекономіка», «Макроекономіка», «Статистика», «Теорія ймовірностей і математична статистика», багатовимірні статистичні методи і т.ін.

Завдання, які вирішуються за допомогою економетрії можна класифікувати за трьома параметрами: за

За кінцевим прикладним цілям виділяють:

а) прогноз економічних і соціально-економічних показників, які характеризують стан і поведінку економічної системи;

б) імітація можливих сценаріїв соціально-економічного розвитку аналізованої системи.

За рівнем ієрархії аналізованої економічний системи виділяють:

а) макрорівень, тобто країни в цілому, аналіз між країнами;

б) мезорівень - регіони всередині країни, галузі, корпорації і т.д .;

в) мікрорівень - підприємства, фірми, сім'ї.

За профілем економетричного моделювання: дослідження може бути спрямоване на вирішення проблем попиту і пропозиції, ціноутворення, інвестиційної, фінансової або соціальної політики або на рішення певного комплексу проблем.

Поняття економетричної моделі

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t\text{-}1} + \varepsilon_{1t},$$
$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t\text{-}1} + \varepsilon_{2t},$$
$$T_t = \gamma_0 + \gamma_1 Y_t + \varepsilon_{3t},$$
$$Y_t = C_t + I_t + G_t,$$

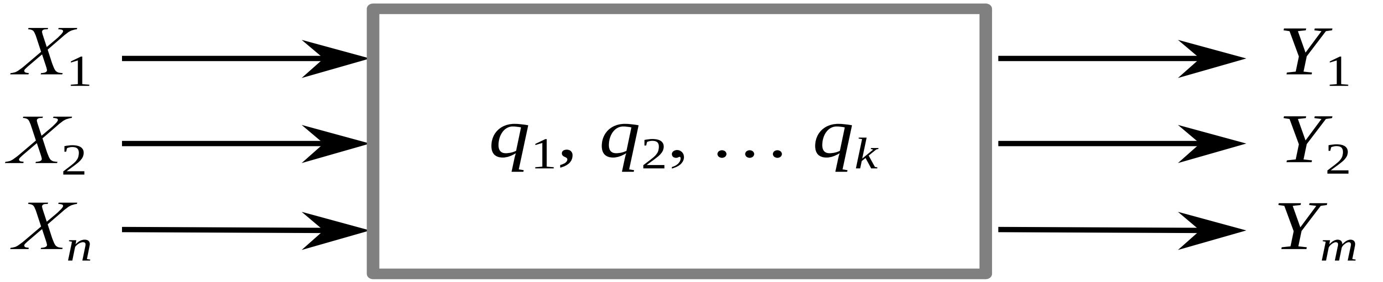
де C_t - споживання; *Yt* - дохід; I_t - інвестиції; T_t - податки; G_t- витрати на уряд; ε_t= 1,2, 3 - відшкодування.

Екзогенні змінні - змінні, які задаються як би ззовні і в певній мірі керовані. У економетричної моделі вони є пояснюють або незалежними змінними.
Ендогенні змінні - це змінні, значення яких формуються всередині соціально-економічної системи, яка аналізується, під впливом екзогенних змінних і у взаємодії один з одним. У економетричної моделі вони є пояснюється або залежними змінними. **Зумовлені змінні** - це пояснюючі змінні, які формуються з екзогенних змінних і *лагових ендогенних змінних*. Значення лагових ендогенних змінних були виміряні в минулі по відношенню до поточного моменту часу і тому можуть бути розглядатися як відомі, задані.

- **Лаґ часови́й** (рос. *временной лаг*; англ. *time lag*; нім. *temporales Log n*) — економічний показник, що відображає відставання або випередження у часі одного економічного явища (причини) порівняно з іншим, пов'язаним з ним явищем (наслідком).
- Часовий лаг становить значну перешкоду для економічного планування, оскільки період, протягом якого одне економічне явище (причина) призведе до остаточного виникнення іншого економічного явища (наслідку), може становити навіть до 18 місяців

(з ВІКІПЕДІЇ)

При побудова економетричної моделі реалізується метод моделювання за принципом «чорного ящика»: задаючи вхідні характеристики системи отримують значення «виходу». Внутрішня структура при цьому не відома.



$$Y = f(X, \varepsilon)$$

де *Y* - ендогенні змінні; *X*- екзогенні (або зумовлені) змінні;
ε - випадкова або стохастична складова.

Класифікація економетричних моделей

- Однофакторна лінійна $y = \alpha + \beta x + \varepsilon$:
- Однофакторна нелінійна: $y = \alpha e^{\beta x} + \varepsilon$:
- Багатофакторна лінійна $y = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n + \varepsilon$
- Багатофакторна нелінійна: $y = \alpha x_1^{\beta_1} x_2^{\beta_2} + \varepsilon$:

- Система відносно незалежних змінних:

$$y_1 = f_1(x_1, x_2, \dots x_m)$$
$$y_2 = f_2(x_1, x_2, \dots x_m)$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$y_n = f_n(x_1, x_2, \dots x_m)$$

- Рекурсивна система

$$y_1 = f_1(x_1, x_2, \dots x_m)$$
$$y_2 = f_2(x_1, x_2, \dots x_m, y_1)$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$y_n = f_n(x_1, x_2, \dots x_m, y_1, y_2, \dots, y_{n-1})$$

Обчислення основних статистичних характеристик вибірки

1. Середнє
$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \, .$$

2. Розмах вибірки R — це різниця між найбільшим і найменшим значенням випадкової величини у вибірці.

3. Мода вибірки M_O — те значення випадкової величини, що зустрічається у вибірці найчастіше.

4. Медіана вибірки M_e — серединне значення ранжованої вибірки.

5. Незміщена вибіркова дисперсія
$$D^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} .$$

6. Вибіркове середньоквадратичне відхилення (математичний стандарт)
$$D = \sqrt{D^2}$$

7. Ексцес характеризує згладжений або загостре-ний розподіл ознаки в сукупності у порівнянні з нормальним розподілом.

$$\hat{E}=\left(\frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{D}\right)^4\right)-\frac{3(n-1)^2}{(n-2)(n-3)}.$$

8. Асиметрія визначає зміщення ознаки в сукупності відносно її середньої величини. Додатна асиметрія – це зрушення розподілу у бік Позитивних відхилень, від'ємна, у бік негативних.

$$\hat{A}=\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{D}\right)^3.$$

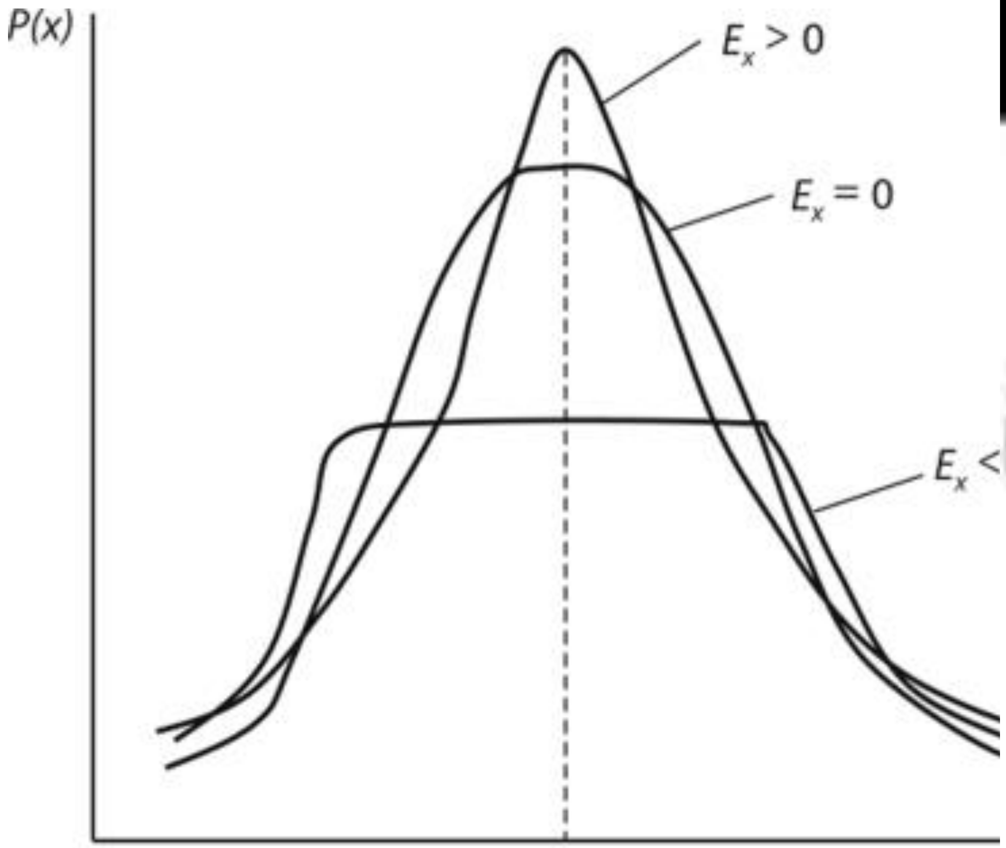
9. Дисперсії ексцесу та асиметрії

$$D_{\hat{E}}^2=\frac{24n(n-1)^2}{(n+5)(n+3)(n-2)(n-3)};\quad D_{\hat{A}}^2=\frac{6n(n-1)}{(n+3)(n+1)(n-2)}.$$

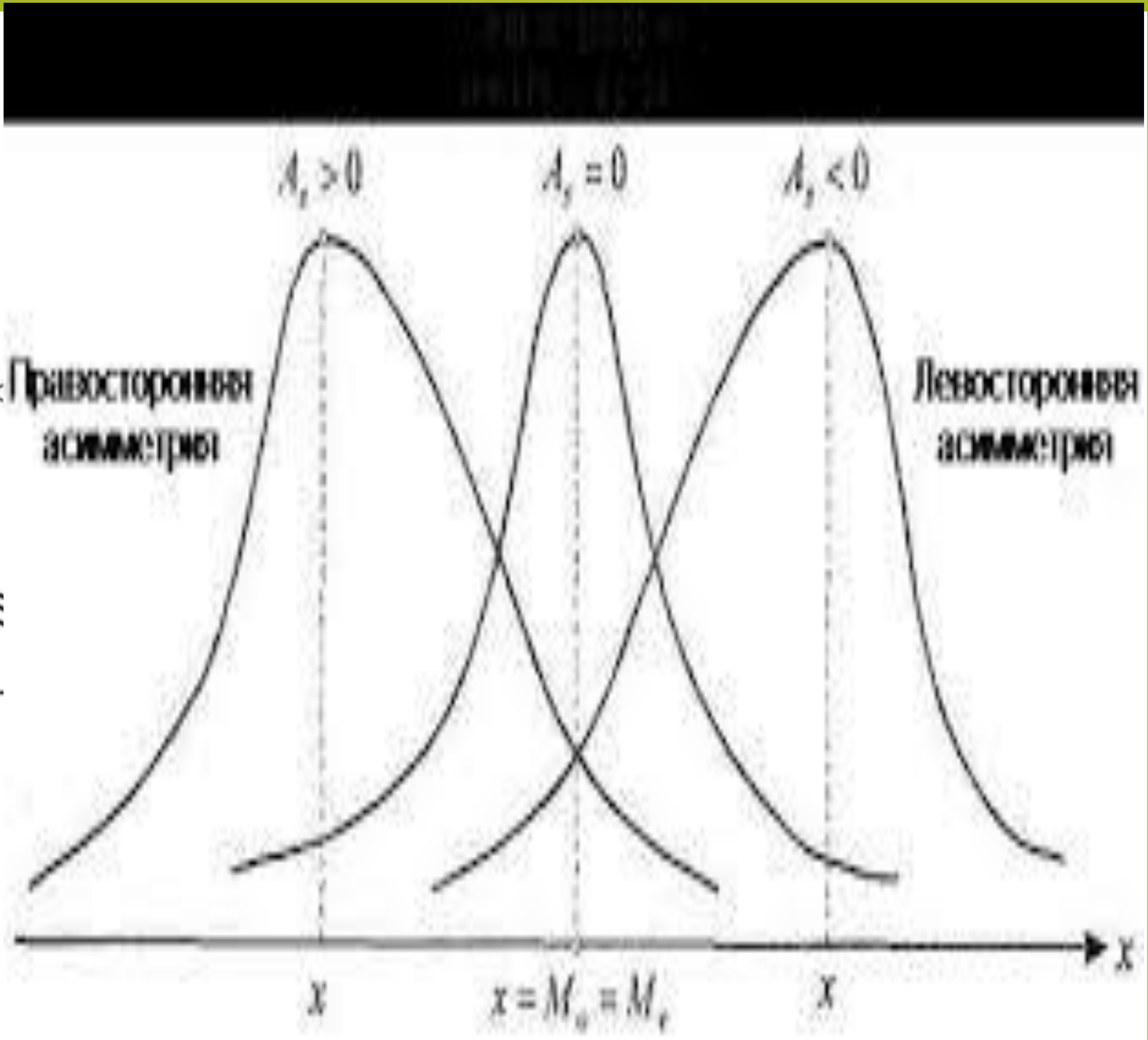
10. Критерій згоди для цих показників.

$$\left|\hat{E}\right|\le 2D_{\hat{E}};\quad \left|\hat{A}\right|\le 2D_{\hat{A}}.$$

<u>Якщо його дотримано, значить вибірка є достатньою для подальшого використання</u>.


P(x)
$E_x > 0$
$E_x = 0$
$E_x <$

Ексцес


$A_s > 0$
$A_s = 0$
$A_s < 0$
Правосторонняя асимметрия
Левосторонняя асимметрия
x
$x = M_o = M_e$
x
x

11. Середня помилка вибірки

$$S = \sqrt{\frac{D^2}{n}}.$$

Кореляційний аналіз. Основне завдання кореляційного аналізу полягає у виявленні взаємозв'язку між випадковими змінними шляхом оцінки коефіцієнтів кореляції і детермінації, а також перевірки значущості отриманих значень. У економетриці кореляційний аналіз застосовується для відбору факторів, що викликають найбільший вплив на досліджуваний показник і оцінки якості побудованих економетричних моделей. Мірою взаємозв'язку між двома змінними *v* і *w* є вибіркова коваріація, що обчислюється за правилом:

$$Cov(v, w) = \frac{1}{n-1}\sum_{i=1}^{n}\left[(v_i - v_{cp})(w_i - w_{cp})\right].$$

У практичних розрахунках зазвичай використовується вибірковий парний коефіцієнт парної кореляції, який визначається за наявного набору фактичних даних

$$r(v, w) = \frac{\sum_{i=1}^{n}\left[(v_i - v_{cp}) * (w_i - w_{cp})\right]}{\sqrt{\sum_{i=1}^{n}(v_i - v_{cp})^2 * \sum_{i=1}^{n}(w_i - w_{cp})^2}} = \frac{Cov(v, w)}{S_v * S_w}.$$

Довірчий інтервал

$$P(|\Theta[X] - \ddot{\Theta}[X]| < \varepsilon) = \beta \qquad \ddot{\Theta}[X] - \varepsilon < \Theta[X] < \ddot{\Theta}[X] + \varepsilon$$

$$\varepsilon_m = \ddot{\sigma}_x Л(\beta) \qquad \varepsilon_D = \ddot{D}_x Л(\beta)\sqrt{\frac{0,8N+1,2}{N(N-1)}}$$

$$\varepsilon_{p_i} = Л(\beta)\sqrt{\frac{k_i(1-k_i)}{N}} \qquad \ddot{\sigma}_m = \sqrt{\frac{D_x}{N}}$$

Де $Л(\beta)$– зворотне значення функції Лапласа для квантиля таблиці z= $\dfrac{x - m_x}{\sigma_x}$

Парний коефіцієнт кореляції має такі властивості:

1) приймає значення в інтервалі [-1; 1], тобто $\left| r(v,w) \right| \leq 1$;

2) не залежить від вибору початку відліку і одиниці виміру,

де a, b, c, d - постійні величини, причому a і c - позитивні; $r(av+b; cw+d) = r(v,w)$

3) Якщо $r(v,w) > 0$, то між змінними є прямий зв'язок, тобто при зростанні (убуванні) однієї з них інша також зростає (спадає); якщо, то зв'язок є зворотнім, тобто при зростанні однієї змінної інша убуває;

4) Якщо $r(v,w) = \pm 1$, то між змінними є функціональна лінійна залежність, а якщо, то лінійний зв'язок між змінними відсутній; відповідного, чим ближче модуль коефіцієнта парної кореляції до одиниці, тим тісніше зв'язок між змінними.

Оцінка значущості вибіркового коефіцієнта парної кореляції

Для оцінки значущості вибіркового коефіцієнта парної кореляції застосовується t-критерій Стьюдента. При цьому фактичне значення цього критерію визначається за формулою

$$t_{набл} = \sqrt{\frac{r^2}{1-r^2}(n-2)} \,,$$

Отримане значення порівнюється з табличним критичним значенням, яке залежить від рівня значущості *α* і числа ступенів свободи $\nu = n - 2$. Критичне значення може бути знайдено за відповідними таблицями, а при використанні табличного процесора Excel - за допомогою функції СТЬЮДРАСПОБР (*α*; *ν*).При отриманні $t_{набл} > t_{кр}$ значення коефіцієнта кореляції *r* визнається значимим, тобто між змінними є лінійна кореляційна залежність.

Етапи побудови і аналізу економетричних моделей

- • Якісний аналіз (постановка мети аналізу, визначення сукупності, визначення результативних і факторних ознак, вибір періоду, за який проводиться аналіз, вибір методу аналізу).
- • Попередній аналіз модельованої сукупності (перевірка однорідності сукупності, виключення аномальних спостережень, уточнення необхідного обсягу ознак, встановлення законів розподілу ознак).
- • Побудова економетричної моделі (встановлення переліку чинників, розрахунок оцінок параметрів рівнянь регресії, перебір конкуруючих варіантів моделі).
- • Оцінка адекватності моделі (перевірка статистичної суттєвості рівняння залежності в цілому і його окремих параметрів; перевірка відповідності формальних властивостей оцінок завданням дослідження).
- • Економічна інтерпретація і практичне використання моделі.

ЛІНІЙНА РЕГРЕСІЯ

- Розглянемо дві змінні *x* та *y*, де *y* - залежна змінна (регресанта, ендогенна змінна), *x* - незалежна змінна (регресор, екзогенна змінна). Змінна *y* характеризує результат або ефективність функціонування аналізованої економічної системи і її знання формуються всередині системи під впливом ряду інших змінних або факторів. За своєю природою залежна змінна завжди стохастична (випадкова). Змінна *x* - це змінна, що за своєю природою може бути як випадковою так і невипадковою. Ця змінна піддається управлінню і планування, її називають незалежною змінної. Якщо змінну можна задавати «ззовні» аналізованої системи, то її називають екзогенної.

- Функція $y = f(X^*)$ називається функцією регресії *y* по *x* (регресією Y по X), якщо вона описує зміни умовного середнього значення результуючої змінної *y* (за умови, що значення пояснюють зміни X зафіксовані на деякому рівні X*) в залежності від значення змінних X. Співвідношення між змінними будемо позначати:

$$f(X^*) = E(y|\, X = X^*) = E(y|\, X^*)$$

$$f(X) = E(y|\, X).$$

Лінійна парна регресія

Якщо f (x) є лінійною функцією, то ми маємо загальний вигляд моделі парної лінійної регресії:

$$y_p = a + b * x$$

де *a* - постійна величина (або вільний член рівняння), *b* - коефіцієнт регресії, що визначає нахил лінії, уздовж якої розсіяні спостереження. Коефіцієнт регресії характеризує зміну змінної y при зміні значення x на одиницю. Якщо $b > 0$, то змінні позитивно коррельовані, якщо $b < 0$ - негативно корельовані. Фактичне значення досліджуваної змінної y тоді може бути представлено у вигляді:

$$y = a + b * x + \varepsilon$$

де *ε* - різниця між фактичним значенням (результатом спостереження) і значенням, розрахованим за рівнянням моделі. Якщо модель адекватно описує досліджуваний процес, то *ε* - незалежна нормально розподілена випадкова величина з нульовим математичним сподіванням (M*ε* = 0) і постійної дисперсією ($D\varepsilon = \sigma^2$). Наявність випадкової компоненти *ε* відображає той факт, що присутні інші чинники, що впливають на досліджувану змінну і не враховані в моделі.

Визначення параметрів лінійної парної моделі методом МНК

Для оцінки параметрів *a* і *b* лінійної парної регресії з використанням наявного набору результатів спостережень найбільш часто використовують метод найменших квадратів (МНК), який мінімізує суму квадратів ε_i - відхилення результатів спостережень y_i від розрахованих за лінійною моделлю значень y_{pi}:

$$b = \frac{\sum_{i=1}^{n}\left[(x_i - x_{cp}) * (y_i - y_{cp})\right]}{\sum_{i=1}^{n}(x_i - x_{cp})^2},$$

$$a = y_{cp} - b * x_{cp}$$

Таке рішення може існувати тільки при виконанні умови $\sum_{i=1}^{n}(x_i - x_{cp})^2 \neq 0$, тобто коли не всі спостереження проводилися при одному і тому ж значенні факторної змінної (сума квадратів дорівнює нулю, якщо кожний доданок дорівнює нулю). Ця умова називається умовою ідентифікованих моделі.

Перевірка значущості параметрів парної лінійної моделі

Для оцінки надійності отриманих значень *a* і *b* проводиться перевірка їх значущості з використанням стандартної помилки оцінки, яка, в свою чергу, визначається за значеннями ряду залишків ε_i:

$$S_{cm} = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n-m-1}}$$

де *n* - кількість спостережень, *m* - кількість факторних змінних в моделі.

На першому етапі обчислюються t-статистика:

$$t_a = \frac{|a|}{S_a}, \quad t_b = \frac{|b|}{S_b}$$

$$S_a = S_{cm} * \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - x_{cp})^2}}, \quad S_b = S_{cm} * \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - x_{cp})^2}}$$

На другому етапі визначається критичне значення $t_{кр}(\alpha; n\text{-}m\text{-}1)$ за таблицями або за допомогою функції СТЮДРАСПОБР в Excel. Рівень значущості α задається, а число ступенів свободи обчислюється за кількістю спостережень *n* і числа факторів *m* (в парній моделі фактор x єдиний).

Нарешті, на третьому етапі обчислені значення t-статистик порівнюються з критичним значеннями $t_{кр}$. Якщо розрахункове значення більше табличного, то відповідний параметр (коефіцієнт рівняння) вважається значущим. В іншому випадку коефіцієнт значущим не є, тобто його можна покласти рівним нулю.

Перевірка виконання передумов МНК

1. Перевірка рівності математичного сподівання рівнів ряду залишків нулю здійснюється в ході перевірки, відповідної $H_0: |\varepsilon| = 0$ **З цією метою будується t-статистика**

$$\hat{t} = \frac{|\bar{\varepsilon}|}{S_\Sigma}\sqrt{n}, \qquad S_\Sigma = \sqrt{\frac{\sum_{t=1}^{n}(\varepsilon_t - \bar{\varepsilon})^2}{n-1}},$$

де $\bar{\varepsilon}$ **- середнє арифметичне значення рівнів ряду залишків** ε_t**,** S_Σ **- середньоквадратичне відхилення для цієї послідовності. На рівні значущості *α* гіпотеза відхиляється, якщо**

$\hat{t} > t_{\alpha,\nu}$ **, де** $t_{\alpha,\nu}$ **- критерій розподілу Стьюдента з довірчою ймовірністю (1-*α*) і ступенями свободи** $\gamma = n-1$**.**

2. Для перевірки умови випадковості виникнення окремих відхилень від тренду часто використовується критерій поворотних точок. Значення випадкової змінної вважається поворотною точкою, якщо воно одночасно більше (або одночасно менше) значень попереднього і наступного члена. Якщо залишки випадкові, то поворотна точка припадає в середньому приблизно на кожні 1,5 спостереження.

Існує певна залежність між середньою арифметичною, дисперсією кількості поворотних точок в ряді залишків *p* і числом членів вихідного ряду спостережень *n*. З використанням цієї залежності, критерій випадковості відхилень від тренду з довірчою ймовірністю 0,95 можна представити у вигляді:

$$p > \left[\frac{2}{3}(n-2) - 1{,}96\sqrt{\frac{16n-29}{90}} \right]$$

тут квадратні дужки означають округлення до цілого.

Якщо нерівність не виконується, то ряд залишків не можна назвати випадковим (тобто він містить регулярну компоненту) і, отже, модель не є адекватною.

3. Наявність (відсутність) автокорреляції в відхиленнях фактичних значень від моделі зростання найпростіше перевірити за допомогою критерію Дарбіна-Уотсона. З цією метою будується статистика Дарбіна-Уотсона (*d* - статистика), в основі якої лежить розрахункова формула.

$$d = \frac{\sum_{t=1}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2}$$

Для формулювання висновку про наявність (відсутність) автокорреляции отримане значення необхідно порівняти з критичними значеннями (нижнє) і (верхнє), які визначаються за спеціальними таблицями для трьох рівнів значимості (α= 0,01; α = 0,025; α = 0,05).

4. Відповідність ряду залишків нормальному закону розподілу перевіримо за допомогою R / S - критерію

$$R / S = \frac{(\varepsilon_{\max} - \varepsilon_{\min})}{S_{\Sigma}}$$

Оцінка якості рівняння регресії

Коефіцієнт детермінації показує, яка частка зміни досліджуваної ознаки врахована в моделі. Іншими словами коефіцієнт детермінації показує, яка частина зміни досліджуваної змінної може бути обчислена, виходячи із змін включених в модель факторних змінних за допомогою обраного типу функції, що зв'язує факторні змінні і досліджуваний ознака в рівнянні моделі.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \varepsilon_i^2}{\sum_{i=1}^{n} (y_i - y_{cp})^2} = \frac{\sum_{i=1}^{n} (y_{pi} - y_{cp})^2}{\sum_{i=1}^{n} (y_i - y_{cp})^2}$$

Коефіцієнт детермінації R^2 може приймати значення від 0 до 1. Чим ближче коефіцієнт детермінації R^2 до одиниці, тим краще якість моделі.

Індекс кореляції $R = \sqrt{R^2}$ характеризує тісноту обраного при побудові моделі типу зв'язку між врахованими в моделі факторами і досліджуваної змінної. У разі лінійної парної регресії його значення за абсолютною величиною збігається з коефіцієнтом парної кореляції r (x, y), який ми розглянули раніше, і характеризує тісноту лінійного зв'язку між x і y. Значення індексу кореляції, очевидно, також лежать в інтервалі від 0 до 1. Чим ближче величина R до одиниці, тим тісніше обраний вид функції пов'язує між собою факторні змінні і досліджуваний ознака, тим краще якість моделі.

Середня відносна помилка апроксимації виражається у відсотках і характеризує точність моделі. Прийнятним точність моделі при

$$E_{отн.ср} = 100 * \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\varepsilon_i}{y_i} \right|$$

вирішенні практичних завдань може визначатися, виходячи з міркувань економічної доцільності з урахуванням конкретної ситуації. Широко застосовується критерій, відповідно до якого точність вважається задовільною, якщо середня відносна похибка менше 15%. Якщо $E_{отн.ср.}$ менше 5%, то кажуть, що модель має високу точність. Не рекомендується застосовувати для аналізу і прогнозу моделі з незадовільною точністю, тобто, коли $E_{отн.ср.}$ більше 15%.

Критерій Фішера

$$F = \frac{(n - m - 1) * R^2}{m * (1 - R^2)}$$

Критичне значення F-критерію визначається за таблицями при заданому рівні значущості α і ступенях свободи $\nu_1 = m, \quad \nu_2 = n - m - 1$ (можна використовувати функцію FРАСПОБР в Excel). Тут, як і раніше, m - число факторів, врахованих в моделі, n - кількість спостережень. Якщо розрахункове значення більше критичного, то рівняння моделі визнається значущим. Чим більше розрахункове значення F-критерію, тим краще якість моделі.

Нелінійні моделі парної регресії

Парабола другої степені

$$y_p = a + bx + cx^2$$

Гіперболічна модель

$$y_p = a + \frac{b}{x}$$

Степенева модель

$$y_p = a * x^b$$

Показова модель

$$y_p = a * b^x$$

Прогнозування із застосуванням парного рівняння регресії

Точковий прогноз обчислюємо шляхом підстановки в рівняння прогнозного значення факторної змінної.

$$y_{прогн}^{точ} = a_0 + a_1 * x_{прогн}$$

Імовірність реалізації точкового прогнозу практично дорівнює нулю. Тому на додаток до точкового прогнозу розраховується середня помилка прогнозу або

$$L = S_{ст} * t_{\alpha, n-m-1} * \sqrt{1 + \frac{1}{n} + \frac{(x_{прогн} - x_{ср})^2}{\sum_{i=1}^{n}(x_i - x_{ср})^2}}$$

довірчий інтервал прогнозу з досить великою надійністю. Розмах прогнозного інтервалу L залежить від стандартної помилки , віддалення $x_{прогн}$ від свого середнього значення в ряді спостережень $x_{ср}$, кількості спостережень n і рівня значущості прогнозу α

Тоді фактичні значення досліджуваної ознаки з ймовірністю (1-α) потраплять в інтервал

$$y_{прогн} \in \left[y_{прогн}^{точеч} - L;\ y_{прогн}^{точеч} + L\right]$$

$$S_{ст} = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n - m - 1}}$$

Модель множинної регресії

Загальний вигляд моделі

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m + \varepsilon$$

Кожен з коефіцієнтів рівняння має наступну економічну інтерпретацію: він показує, наскільки зміниться значення досліджуваної ознаки при зміні відповідного фактору на 1 при незмінних інших факторних змінних.

Для адекватності моделі необхідно, щоб випадкова величина *ε*, що є різницею між фактичними і розрахунковими значеннями, мала нормальний закон розподілу з математичним сподіванням рівним нулю і постійної дисперсією σ^2.

Матричний вигляд моделі

$$Y_в = X_в A + E$$

де *Yв* - вектор вибіркових даних спостережень

досліджуваної змінної (n елементів),

Xв - матриця вибіркових даних спостережень

факторних змінних (елементів), *A* – вектор

параметрів рівняння (*m* + 1 елементів), а

E - вектор випадкових відхилень (*n* елементів):

$$Y_в = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, \quad X_в = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix}, \quad A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{pmatrix}, \quad E = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_m \end{pmatrix}.$$

Оцінка параметрів моделі за допомогою МНК. Відбір факторів

Формула для обчислення параметрів регресійного рівняння методом найменших квадратів (МНК) за даними спостережень

$$A = (X_{в}^{T} X_{в})^{-1} X_{в}^{T} Y_{в}$$

Знаходження параметрів за допомогою цього рівняння можливо лише тоді, коли між різними стовпцями і різними рядками матриці вихідних даних X відсутня сувора лінійна залежність (інакше не існує зворотна матриця). Ця умова не виконується, якщо існує лінійна або близька до неї зв'язок між результатами двох різних спостережень, або ж якщо такий зв'язок існує між двома різними факторними змінними. Лінійних або близька до неї зв'язок між факторами називається мультиколеніарностью. Щоб позбутися від мультиколеніарності, в модель включають один з лінійно пов’язаних між собою факторів, причому той, який більшою мірою пов'язаний з досліджуваною змінної.

На практиці щоб позбутися від мультиколеніарності ми будемо перевіряти

для кожної пари факторних змінних виконання наступних умов.

$$\begin{cases} \left| r_{x_i x_j} \right| < 0{,}8 \\ \left| r_{x_i x_j} \right| < \left| r_{y x_i} \right| \\ \left| r_{x_i x_j} \right| < \left| r_{y x_j} \right| \end{cases}$$

Якщо хоча б одна з умов (не виконується, то в модель включають тільки один

з цих двох чинників, а саме, той, у якого модуль коефіцієнта кореляції з *Y* більше.

```
. regress lnv lnk lnl

      Source |       SS       df       MS              Number of obs =      25
-------------+------------------------------           F(  2,    22) =  397.54
       Model |  44.1727741     2  22.086387            Prob > F      =  0.0000
    Residual |  1.22225984    22  .055557265           R-squared     =  0.9731
-------------+------------------------------           Adj R-squared =  0.9706
       Total |  45.3950339    24  1.89145975           Root MSE      =  .23571

------------------------------------------------------------------------------
         lnv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         lnk |   .2454281   .1068574     2.30   0.032     .0238193    .4670368
         lnl |    .805183   .1263336     6.37   0.000     .5431831    1.067183
       _cons |   1.844416   .2335928     7.90   0.000     1.359974    2.328858
------------------------------------------------------------------------------
```

1 2 3 4 5 6 7 8 9 10 11 12

F - тест для виявлення мультиколінеарності (алгоритм Феррара - Глобера)

1. Обчислимо матрицю коефіцієнтів парної кореляції.

2. Для кожної змінної побудуємо регресійну залежність від інших факторів. Спочатку, наприклад, будуємо лінійну регресію виду (наприклад для 3-х факторів)

$$y_1 = c_0 + c_1 x_1 + c_2 x_2$$

3. Для кожної моделі знаходимо залишки ε_i і за ними

– величину R^2 для кожної моделі

$$F = \frac{\frac{R^2}{k}}{\frac{1-R^2}{n-k-1}}$$

4. Розраховуємо F-статистику Фішера за формулою

і порівнюємо їх із $F_{табл}$

$$r_{xy.z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}.$$

5. Розраховуємо часткові коефіцієнти кореляції за формулою

(для прикладу подано випадок із 3-х факторів)

$$t_{xy} = \frac{r_{xy.z}\sqrt{n-m}}{\sqrt{1-r_{xz.z}^2}}.$$

6. Розрахуємо критерій Стьюдента і порівнюємо його з табличним.

Якщо табличне більше розрахованого – мультиколінеарності немає.

Оцінка впливу окремих факторів на досліджувану змінну

Важливу роль при оцінці впливу окремих факторів грають коефіцієнти регресійної моделі a_j. Однак безпосередньо з їх допомогою не можна зіставити фак-тори за ступенем їх впливу на залежну змінну через відмінності одиниць виміру і різного масштабу коливань.

Коефіцієнти еластичності показує, на скільки відсотків змінюється досліджувана змінна при зміні факторної змінної на 1 відсоток.

$$Э_j = a_j * \frac{x_{jcp}}{y_{cp}}$$

Бета – коефіцієнт

$$\beta_j = a_j * \frac{S_{x_j}}{S_y} \quad S_{x_j} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij} - x_{jcp})^2}, \quad S_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - y_{cp})^2}$$

показує, на яку частину величини середньоквадратичного відхилення зміниться змінна y зі зміною відповідної незалежної змінної xj на величину свого середньоквадратичного відхилення при фіксованому рівні значень інших факторних змінних.

Дельта – коефіцієнт *показує частку впливу фактору в сумарному впливі всіх факторів*

$$\Delta_j = r(x_j, y) * \frac{\beta_j}{R^2}$$

Побудова прогнозів на основі моделі множинної лінійної регресії

Точковий прогноз $y^{точ}_{прогн} = a_0 + a_1 x_{прогн1} + a_2 x_{прогн2} + \dots + a_m x_{прогнm}$ $\qquad Y^{точ}_{прогн} = X_{прогн} A$

Інтервальний прогноз $V = X_{прогн} \times (X^T_{в} \times X_{в})^{-1} \times X^T_{прогн}$

$$L_i = S_{cm} * t_{\alpha, n-m-1} * \sqrt{1 + v_{ii}}$$

$$v_{ii} = \frac{m \prod_{j=1}^{m} (x_{ijпрогн} - x_{jcp})^2}{\prod_{j=1}^{m} \sum_{i=1}^{n} (x_{ij} - x_{jcp})^2}$$

$$S_{cm} = \sqrt{\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n - m - 1}}$$

$$y_{i\,прогн} \in \left[y^{точеч}_{i\,прогн} - L_i ;\ \ y^{точеч}_{i\,прогн} + L_i \right]$$

Системи лінійних одночасних рівнянь

Система лінійних незалежних рівнянь

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1 \\ y_2 = a_{20} + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2 \\ \dots\dots\dots\dots \\ y_k = a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{km}x_m + \varepsilon_k \end{cases}$$

Система лінійних рекурсивних рівнянь

$$\begin{cases} y_1 = a_{10} + a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m + \varepsilon_1 \\ y_2 = b_{21}y_1 + a_{20} + a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m + \varepsilon_2 \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{30} + a_{31}x_1 + a_{32}x_2 + \cdots + a_{3m}x_m + \varepsilon_3 \\ \dots\dots\dots \\ y_k = b_{k1}y_1 + b_{k2}y_2 + \cdots + b_{kk-1}y_{k-1} + a_{k0} + a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{km}x_m + \varepsilon_k \end{cases}$$

Система одночасних (взаємозалежних) рівнянь

$$\begin{cases} y_1 = a_{10} + b_{12}y_2 + b_{13}y_3 + \dots + b_{1k}y_k + a_{11}x_1 + a_{12}x_2 + \dots + a_{1m}x_m + \varepsilon_1 \\ y_2 = a_{20} + b_{21}y_1 + b_{23}y_3 + \dots + b_{2k}y_k + a_{21}x_1 + a_{22}x_2 + \dots + a_{2m}x_m + \varepsilon_2 \\ y_3 = a_{30} + b_{31}y_1 + b_{32}y_2 + \dots + b_{3k}y_k + a_{31}x_1 + a_{32}x_2 + \dots + a_{3m}x_m + \varepsilon_3 \\ \dots\dots\dots \\ y_k = a_{k0} + b_{k1}y_1 + b_{k2}y_2 + \dots + b_{kk-1}y_{k-1} + a_{k1}x_1 + a_{k2}x_2 + \dots + a_{km}x_m + \varepsilon_k \end{cases}$$

Приведена форма моделі формі моделі

$$\begin{cases} y_1 = \delta_{10} + \delta_{11}x_1 + \delta_{12}x_2 + \dots + \delta_{1m}x_m + \varepsilon_1 \\ y_2 = \delta_{20} + \delta_{21}x_1 + \delta_{22}x_2 + \dots + \delta_{2m}x_m + \varepsilon_2 \\ \dots\dots\dots \\ y_k = \delta_{k0} + \delta_{k1}x_1 + \delta_{k2}x_2 + \dots + \delta_{km}x_m + \varepsilon_k \end{cases}$$

Проблема ідентифікації: Для того, щоб модель була такою, що ідентифікується, необхідно, щоб кожне рівняння моделі було ідентифіковане. Якщо хоча б одне рівняння СФМ неідентифіковане, то вся модель вважається неідентифікованою.

Необхідною умовою ідентифікації окремого рівняння моделі є рахункове правило. Якщо позначити через *H* число досліджуваних змінних *yi*, присутніх в *i*-му рівнянні, а через *D* позначити число факторних змінних x_j, відсутніх в *i*-му рівнянні, то рахункове правило має вигляд:

- якщо $D + 1 < H$, то рівняння неідентифіковані;
- якщо $D + 1 = H$, то рівняння ідентифіковані;
- якщо $D + 1 > H$, то рівняння над ідентифіковані .

- Достатня умова ідентифікованих окремого рівняння моделі виконується, якщо визначник матриці, складеної з коефіцієнтів в інших рівнняннях при змінних (як досліджуваних *y*, так і факторних *x*), відсутніх в даному *i*-му рівнянні не дорівнює нулю, а ранг цієї матриці , водночас, не менше, ніж кількість всіх досліджуваних змінних в системі рівнянь за виключенням одного.