УДК 330.43

РЕГРЕСІЙНІ МОДЕЛІ БЕЗ ВІЛЬНОГО ЧЛЕНА

МАЛЯРЕЦЬ Л. М.
доктор економічних наук
Харків
КОЙБІЧУК В. В.
Суми

Найпоширенішим інструментом в дослідженнях об'єктів в економіці є кореляційно-регресійний аналіз. Це обумовлено специфікою стану та розвитку соціально-економічних систем – аналіз їх має ґрунтуватись на причинно-наслідкових механізмах. Розроблення багатофакторних регресійних моделей вимагає встановлення результативної ознаки та факторів, які впливають не неї. Змістовність моделі залежить від дослідника, який моделює взаємозв'язок. Він же стикається з різними проблемами в розробленні регресійної моделі. Насправді цих проблем не так вже й мало. Проблеми оцінювання параметрів множинних регресійних моделей детально аналізувались в авторській роботі [1], де було викладено окремі рекомендації щодо їх вирішення. До цього часу ще не запропоновано ефективних засобів для запобігання мультиколінеарності у тих випадках, коли вона впливає на оцінки параметрів моделі. Відомо, що метод покрокової регресії не оправдався [2, 3]. Було запропоновано *QR*-розклад матриці *X*, за допомогою чого проблема мультиколінеарності в регресійному аналізі розв’язується досить просто [4]. Також в багатофакторному регресійному аналізі при вирішенні реальних економічних задач виникають несподівані проблеми інтерпретації коефіцієнтів регресії [5]. Обчислюючи регресійні моделі для економічного аналізу, належить знати, що може бути отримане рівняння з β-коефіцієнтами, які більші одиниці за модулем. При отриманні такої моделі рекомендується уважно вивчити напрямок причинно-наслідкових зв’язків явищ, що вивчаються, і уточнити специфікацію моделі [6].

Розглянемо одну з проблем множинної регресійної моделі, пов’язану зі змістом вільного члена. Вільний член лінійної багатофакторної моделі $y = b_0 + b_1 x_1 + b_2 x_2$ відіграє ту саму роль, що і в однофакторному випадку – він потрібен в моделі для того, щоб при середніх значеннях пояснюючих змінних $(\bar{x}_1, \bar{x}_2)$ було одержане середнє значення результативної ознаки $(\bar{y})$.

Прикладів випадків регресійних моделей, у яких вільний член дорівнює нулю, багато в економіці: гіпотеза неперервного (постійного) прибутку Мілтона Фрідмана, згідно з якою постійне споживання пропорційне постійному прибутку; у теорії аналізу витрат стверджується, що змінні витрати виробництва пропорційні випуску продукції; у монетаристській теорії рівень зміни цін, тобто рівень інфляції, пропорційний рівневі зміни пропозиції грошей [7].

З позиції математики моделі без вільного члена з'являються у випадках, коли відомо, що лінія регресії обов’язково має проходити через фіксований вузол, найчастіше через початок координат. Наприклад, зображена на *рис. 1* квадратична залежність завжди проходить через початок координат за будь-яких значеннях параметрів: $y = b_1 x + b_2 x^2$.

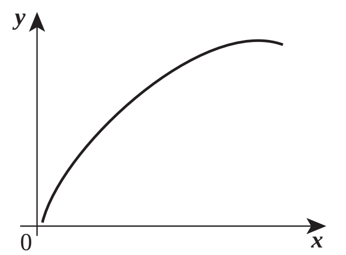

Рис. 1. Квадратична модель

Моделі без вільного члена також з'являються при використанні зваженого метода найменших квадратів, коли для подолання наслідків гетероскедастичності все рівняння регресії попередньо помножується на вагову функцію. Наприклад, нелінійну модель $y = \dfrac{1}{b_0 + b_1 x}$ можна привести до лінійного вигляду (відносно параметрів) функціональним перетворенням залежної змінної $1/y = b_0 + b_1 x$. Але відомо, що при функціональних перетвореннях за-лежної змінної майже завжди з'являється суттєва гетероскедастичність залишків моделі, яка здатна спотворити значення оцінок параметрів моделі. Вагова функція призначена дати більшу вагу надійним спостереженням і меншу нестабільним даним з великою мінливістю. У даному прикладі вагова функція дорівнює $g = y^2$. Помножимо лінійне рівняння (у координатах x, $1/y$) на цю вагову функцію і одержимо модель без вільного члена $y = b_0 y^2 + b_1 x y^2$, яка лінійна відносно параметрів і не має гетероскедастичності. МНК-оцінки параметрів такої моделі – доброякісні і не мають систематичних похибок (зміщення).

Проте відсутність у моделі вільного члена призводить до того, що сума залишків не буде дорівнювати нулю. Усі статистичні характеристики якості моделі – коефіцієнт детермінації, статистики Фішера і Ст'юдента, що обчислені за стандартними формулами, будуть тепер не правильними.

Модель з відсутнім або нульовим перетином можна застосовувати у деяких випадках, проте тут потрібно пам’ятати декілька специфічних рис. По-перше, Σe_i, яка завжди дорівнює нулю в моделі з наявною величиною перетину (загальноприйнята модель), не завжди повинна дорівнювати нулю для регресії, що проходить через початок координат. По-друге, R^2 – коефіцієнт детермінації, що є завжди додатним у загальноприйнятій моделі, у деяких випадках може перетворюватись на від’ємний для регресії, що проходить через початок координат. Цей аномальний результат ми отримуємо тому, що R^2 неоднозначно передбачає, що перетин включається в модель. Тому R^2, який обчислено за правилами, може не відповідати регресійній моделі, що проходить через початок координат.

Саме через ці специфічні риси моделі регресії, що проходить через початок координат, користуватись слід обережно. Якщо немає достатньої попередньої упевненості в її використанні, краще використати загальноприйняту модель із наявним перетином. Це має подвійну перевагу. По-перше, якщо ми включаємо в модель величину перетину, але вона статистично не значима для усіх практичних цілей, ми маємо регресію, що проходить через початок координат. По-друге, і це найважливіше, якщо фактично перетин існує, але ми наполягаємо на тому, щоб застосувати регресію, що проходить через початок координат, то ми припускаємося помилки специфікації, тим самим порушуючи припущення 5-ї класичної моделі лінійної регресії [7, с. 133 – 134].

Розглянемо детально модель без вільного члена. Для оцінки параметрів моделей без вільного члена

$$y = b_1 x_1 + b_2 x_2 + e$$

пропонується додати в метод найменших квадратів обов’язкову умову $\Sigma e = 0$. Маємо задачу на умовний екстремум – треба знайти мінімум суми квадратів залишків моделі за умови рівності нулю їх суми:

$$\begin{cases} \sum e^2 \to \min, \\ \sum e = 0. \end{cases}$$

Складаємо функцію Лагранжа (де множник Лагранжа позначено -2λ)

$$F = \Sigma e^2 - 2\lambda \Sigma e$$

і дорівнюємо нулю її частинні похідні за всіма змінними λ, b_1, b_1 (y, x_1 і x_2 зараз не є змінними, це набори відомих окремих чисел).

$$\begin{cases} \dfrac{\partial F}{\partial \lambda} = 2\sum e = 0 \\ \dfrac{\partial F}{\partial b_1} = 2\sum e \dfrac{\partial e}{\partial b_1} - 2\lambda \sum \dfrac{\partial e}{\partial b_1} = 0 \\ \dfrac{\partial F}{\partial b_2} = 2\sum e \dfrac{\partial e}{\partial b_2} - 2\lambda \sum \dfrac{\partial e}{\partial b_2} = 0 \end{cases}.$$

Врахуємо, що

$$e = y - b_1 x_1 - b_2 x_2,$$

$$\frac{\partial e}{\partial b_1} = -x_1; \quad \frac{\partial e}{\partial b_2} = -x_2,$$

звідки одержуємо таку систему зв'язків на залишки моделі:

$$\begin{cases} \sum e = 0, \\ \sum e x_1 = \lambda \sum x_1, \\ \sum e x_2 - \lambda \sum x_2. \end{cases}$$

Помножимо рівняння $y = b_1x_1 + b_2x_2 + e$ на кожну змінну $(1, x_1, x_2, y, e)$ і просумуємо ці вирази за всіма спостереженнями:

$$\Sigma y = b_1 \Sigma x_1 + b_2 \Sigma x_2 + \Sigma e$$

$$\Sigma yx1 = b_1 \Sigma x_1 x_1 + b_2 \Sigma x_1 x_2 + \Sigma e x_1$$

$$\Sigma yx2 = b_1 \Sigma x_1 x_2 + b_2 \Sigma x2 x_2 + \Sigma e x_2$$

$$\Sigma yy = b_1 \Sigma y x_1 + b_2 \Sigma y x_2 + \Sigma y e$$

$$\Sigma ye = b_1 \Sigma e x_1 + b_2 \Sigma e x_2 + \Sigma ee.$$

Врахуємо зв'язки, що накладені на залишки і з перших трьох сумарних виразів отримуємо таку систему рівнянь для оцінки параметрів моделі без ві-льного члена:

$$\begin{cases} \sum y = \qquad\qquad b_1 \sum x_1 \quad + b_2 \sum x_2, \\ \sum y x_1 = \lambda \sum x_1 + b_1 \sum x_1 x_1 + b_2 \sum x_1 x_2, \\ \sum y x_1 = \lambda \sum x_1 + b_1 \sum x_1 x_2 + b_2 \sum x_2 x_2. \end{cases}$$

Цікаво порівняти отриману систему рівнянь з системою «нормальних рівнянь» для оцінки параметрів моделі з вільним членом $y = b_0 + b_1x_1 + b_2x_2$:

$$\begin{cases} \sum y = \quad n b_0 \qquad b_1 \sum x_1 \quad + b_2 \sum x_2, \\ \sum y x_1 = b_0 \sum x_1 + b_1 \sum x_1 x_1 + b_2 \sum x_1 x_2, \\ \sum y x_1 = b_0 \sum x_1 + b_1 \sum x_1 x_2 + b_2 \sum x_2 x_2. \end{cases}$$

Єдина відмінність нової системи – у першому рівнянні відсутній член nb_0.

З останніх двох сумарних виразів можна отримати:

$$\Sigma y^2 = b_1 \Sigma y x_1 + b_2 \Sigma y x_2 + b_1 \Sigma e x_1 + b_2 \Sigma e x_2 + \Sigma e_2 =$$

$$= b_1 \Sigma y x_1 + b_2 \Sigma y x_2 + \lambda(b_1 \Sigma x_1 + b_2 \Sigma x_2) + \Sigma e_2 =$$

$$= \lambda \Sigma y + b_1 \Sigma y x_1 + b_2 \Sigma y x_2 + \Sigma e_2,$$

що дуже нагадує аналогічний вираз для моделі з вільним членом (з заміною λ на b_0).

Згадаємо приклад з історії науки. У 1927 р. економіст Пол Дуглас, розглядаючи діаграми логарифмів капітальних витрат (K), обсягів випуску про-дукції (Y) і витрат труда (L), виявив, що відстані від точок графіка ln$Y$ до точок графіків ln$K$, ln$L$ складають однакову пропорцію для всіх спостережень з 1900 по 1922 рр. На підставі цього факту математик Чарльз Кобб показав, що така особливість має місце для залежності $Y = K^b L^{1-b}$, яка зараз має назву модель Кобба – Дугласа.

Своє відкриття П. Дуглас зробив лише тому, що для 1989 року він прийняв значення всіх показників K, Y, L за 1 (100%), тобто задав фіксований вузол, де перетинаються графіки залежностей ln$K$, ln$Y$, ln$L$ за часом ($x = t - 1989$). Фактично він помітив лінійність трендів логарифмів показників (*рис. 2*), і цього факту має бути досить, що зробити всі подальші висновки.

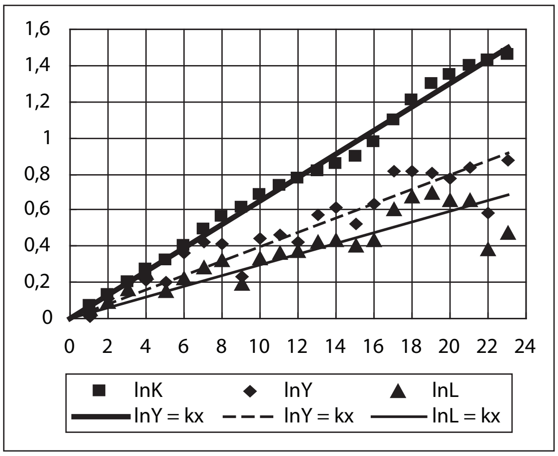

Рис. 2. Спостереження П. Дугласа

Розглянемо детально апроксимацію ln$L = f(x)$. Дані П. Дугласа наведені в *табл. 1*.

Якщо апроксимувати залежність звичайною лінійною моделлю

$$\ln L = b_0 + b_1 x + e,$$

то будуть одержані такі МНК-оцінки параметрів b_0 = = 0,08299; b_1 = 0,02435. Сума залишків моделі дорівнює нулю $\Sigma e = 0$, сума квадратів залишків $\Sigma e^2 = 0{,}1829$. Виконується співвідношення $s_y^2 = s_p^2 + s_e^2$, де позначено $y = \ln L$, $yp = b_0 + b_1 x$, s_p^2 – дисперсія розрахункових значень. У числах маємо розклад загальної дисперсії на дві складові 0,03404 = 0,02608 + 0,00796, звідки знаходимо коефіцієнт детермінації $R^2 = 0{,}7664$. Через фіксований вузол лінія регресії не проходить (*рис. 3*).

Рис. 3. Апроксимація часового тренду для ln*L*

Таблиця 1

Вихідні дані про відносні значення *K*, *Y*, *L*

№	t	K	Y	L	lnK	lnY	lnL
1	1900	1,07	1,01	1,05	0,0677	0,0100	0,0488
2	1901	1,14	1,12	1,10	0,1310	0,1133	0,0953
3	1902	1,22	1,22	1,18	0,1989	0,1989	0,1655
4	1903	1,31	1,24	1,29	0,2700	0,2151	0,2546
5	1904	1,38	1,22	1,16	0,3221	0,1989	0,1484
6	1905	1,49	1,43	1,25	0,3988	0,3577	0,2231
7	1906	1,63	1,52	1,33	0,4886	0,4187	0,2852
8	1907	1,76	1,51	1,38	0,5653	0,4121	0,3221
9	1908	1,85	1,26	1,21	0,6152	0,2311	0,1906
10	1909	1,98	1,55	1,40	0,6831	0,4383	0,3365
11	1910	2,08	1,59	1,44	0,7324	0,4637	0,3646
12	1911	2,16	1,53	1,45	0,7701	0,4253	0,3716
13	1912	2,26	1,77	1,52	0,8154	0,5710	0,4187
14	1913	2,36	1,84	1,54	0,8587	0,6098	0,4318
15	1914	2,44	1,69	1,49	0,8920	0,5247	0,3988
16	1915	2,66	1,89	1,54	0,9783	0,6366	0,4318
17	1916	2,98	2,25	1,82	1,0919	0,8109	0,5988
18	1917	3,35	2,27	1,96	1,2090	0,8198	0,6729
19	1918	3,66	2,23	2,00	1,2975	0,8020	0,6931
20	1919	3,87	2,18	1,93	1,3533	0,7793	0,6575
21	1920	4,07	2,31	1,93	1,4036	0,8372	0,6575
22	1921	4,17	1,79	1,47	1,4279	0,5822	0,3853
23	1922	4,31	2,40	1,61	1,4609	0,8755	0,4762

Якщо ми приймемо модель без вільного члена $\ln L = kx + e$, то одержимо $k = 0{,}02965$. Графік залежності проходить через фіксований вузол (див. рис. 3).

Сума залишків цієї моделі не дорівнює нулю $\Sigma e = 0{,}4467$, сума квадратів залишків $\Sigma e^2 = 0{,}2200$ (це більше в порівняння з попередньою моделлю). Не виконується співвідношення $s_y^2 = s_p^2 + s_e^2,$ тому не можна обчислити коефіцієнт детермінації (який має показувати частку мінливості, що пояснюється моделлю). Також не можна обчислити інші статистичні характеристики.

Оцінімо параметри моделі без вільного члена новим методом. Складаємо систему рівнянь для визначення k і λ:

$$\begin{cases} \sum y = \quad k\sum x, \\ \sum yx = \lambda\sum x + k\sum x^2, \end{cases}$$
$$\sum y^2 = \lambda\sum y + k\sum yx + \sum e^2.$$

З першого рівняння цієї системи (яке еквівалентно умові $\Sigma e = 0$) відразу знаходимо нову оцінку параметра k:

$$k = \frac{\sum y}{\sum x} = \frac{8,6289}{276} = 0,03126.$$

Для однофакторної моделі нова методика пропонує визначати єдиний параметр моделі k з обов'язкової умови $\Sigma e = 0$, не з $\Sigma e^2 \to \min$. З другого рівняння додатково знаходимо $\lambda = -0{,}02536$. Третє співвідношення може застосовуватися для контролю всіх обчислень.

Сума залишків для цієї моделі дорівнює нулю $\Sigma e = 0$, сума квадратів залишків $\Sigma e^2 = 0{,}2313$ (це більше в порівняння з попередніми моделями). Співвідношення не виконується. Графіки всіх трьох моделей наведено на рис. 3.

Отже, удосконалений метод має переваги порівняно зі стандартним: виконання умови рівності нулю суми залишків моделі. Моделі без вільного члена можуть мати в аналізі соціально-економічних систем, тому слід розробляти їх та інтерпретувати з урахуванням особливостей. ■

ЛІТЕРАТУРА

1. **Егоршин А. А., Малярец Л. М.** Проблемы эконометрического оценивания. Коммунальное хозяйство городов. Научно-технический сборник. Выпуск 61. Серия: Экономические науки.– Киев : Техніка, 2005.– 304 с. (С. 267 – 273)

2. **Егоршин А. А.** Корреляционно-регрессионный анализ: [учеб. пособ. для студ. экон. спец. вузов] / А. А. Егоршин, Л. М. Малярец.– Х. : Основа, 1998.– 208 с.

3. **Егоршин А. А.** Практикум по эконометрии в Excel: [пособ. для студ. высш. учебн. завед.] / А. А. Егоршин, Л. М. Малярец.– Х. : ИД «ИНЖЭК», 2005.– 100 с.

4. Малярець Л. М., Рижих І. Ю. Застосування QR-розкладу прямокутних матриць Хаусхолдеровими відображеннями в регресійному аналізі // Економіка розвитку.– Харків : ХНЕУ.– 2009.– № 1(49).– С. 16 – 20.

5. Малярец Л. М., Никольский И. А. Об одной математической проблеме многомерного регрессионного анализа // Бизнес Информ.– Харьков : ИД «ІНЖЕК».– 2009.– № 8.– С. 81 – 87.

6. Малярец Л. М., Азаренков Г. Ф. Особенности параметров линейных регрессионных моделей в реальних экономических задачах / У монографії: «Лібермановські читання: економічна спадщина та сучасні проблеми / Під заг. ред. д. е. н., проф. Пономаренка В. С., Кизима М. О., к. е. н., доц. Зими О. Г.– Харків : ФОП Лібуркіна Л. М.; ВД «ІНЖЕК», 2009.– 296 с.

7. Лук'яненко І. Г. Економетрика : підручник. / І. Г. Лук'яненко, Л. І. Краснікова.– К. : Тов. «Знання», КОО, 1998.– 494 с.