УДК 517.5
DOI: 10.32626/2308-5878.2019-19.125-131

М. Ю. Савкіна, канд. фіз.-мат. наук,

Інститут математики НАН України, м. Київ

РІВНІСТЬ ОЦІНОК МНК ТА ЕЙТКЕНА МОДЕЛІ ЛІНІЙНОЇ РЕГРЕСІЇ У ВИПАДКУ ГЕТЕРОСКЕДАСТИЧНИХ ВІДХИЛЕНЬ

В роботі у випадку гетероскедастичних незалежних відхилень досліджується модель лінійної регресії, функція якої має вигляд $f(t) = at + b,$ де a та b — невідомі параметри. Наближені значення (спостереження) функцій $f(t)$ реєструються в рівновіддалених точках відрізка $[0,1].$ Сформульовано теорему 1, яка дає умови на дисперсії відхилень, при яких оцінка Ейткена параметра a збігається з його оцінкою МНК. При цих умовах оцінки Ейткена та МНК параметра $M_2 = \max\left\{\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}\right\}.$ не будуть збігатися. Також сформульовано теорему 2, яка дає умови для збігу оцінки Ейткена та оцінки МНК параметра $b.$ На підставі теорем 1 та 2 в даній роботі досліджено властивості дисперсій відхилень, які надають рівність цим оцінкам окремо для параметра a та для параметра $b.$ Показано, для рівності оцінок Ейткена та МНК параметра a відхилення будуть мати найбільшу та найменшу дисперсію в двох сусідніх точках спостереження, розташованих в середині відрізка $[0,1],$ для рівності оцінок параметра b — в околі точки 2/3. Знайдено асимптотичні значення дисперсій всіх відхилень, якщо відношення найбільшої до найменшої дисперсії прямує до нескінченності. Доведено, що в цьому випадку дисперсії всіх відхилень будуть не більше найменшої дисперсії ніж у 3 рази для параметра a та не більше ніж у 5 разів для параметра $b.$

Ключові слова: *метод найменших квадратів, регресійна модель, оцінка Ейткена.*

Вступ. У класичній регресії передбачається, що відхилення в регресійній моделі гомоскедастичні та не корелюють одне з іншим. Це доволі жорстка умова, яка досить часто не виконується. В зв'язку з цим дослідження моделі, в якій відхилення корельовані, або принаймні, гетероскедастичні, має великий інтерес. Проте в тому випадку, коли коваріаційна матриця відхилень не є одиничною, оцінка звичайного методу найменших квадратів (МНК) невідомих параметрів моделі хоч і буде незміщеною та спроможною, вже не буде ефективною. Так виник зважений МНК [1, с. 78]; оцінка невідомих параметрів моделі, яка отримується за допомогою зваженого МНК, називається оцінкою Ейткена.

Однак використання ефективної оцінки Ейткена передбачає знання коваріаційної матриці відхилень, яка на практиці як правило невідома. Тому доводиться користуватися оцінкою МНК. Отже знаходження випадків, коли ці оцінки збігаються, має велике значення.

У [2, с. 610] доведено теорему, яка дає необхідну і достатню умову на матрицю плану [3, с. 49] для збігу оцінки МНК та оцінки Ейткена. Зауважимо, що в цій теоремі оцінки збігаються для всіх параметрів одночасно.

Розглянемо таку модель лінійної регресії

$$y_i = at_i + b + \varepsilon_i, i = 0,1,\dots,n, \tag{1}$$

де $t_i = \dfrac{i}{n}, i = 0,1,\dots,n,$ — точки спостереження, $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_n$ — незалежні у сукупності випадкові величини з $E\varepsilon_i = 0$ та $D\varepsilon_i = \gamma_i \sigma^2, i = 0,1,\dots,n,$ $m = \gamma_{\frac{2n+2}{3}}$, та b — невідомі параметри, які підлягають оцінюванню.

Поставимо задачу знайти умови на дисперсії відхилень, при яких збігаються оцінка МНК та оцінка Ейткена кожного параметру моделі (1) окремо (одночасно для обох параметрів ці оцінки будуть збігатися тільки у випадку всіх рівних дисперсій). Виходячи з загальних формул для оцінки МНК та оцінки Ейткена параметрів лінійної регресійної моделі, в роботі [4] доведено наступні теореми для відповідних оцінок параметрів a та $(n+1)$ моделі (1).

Теорема 1. Якщо в моделі (1) умова i не виконується, то оцінка МНК та оцінка Ейткена параметра a збігаються тоді і тільки тоді, коли:

1) у випадку парного n

$$(n+1)$$

$\gamma_{\frac{n}{2}}, \gamma_{\frac{n}{2}+1}$ — будь-які, але $\gamma_{\frac{n}{2}} \neq \gamma_{\frac{n}{2}+1}$;

2) у випадку непарного n

$$\gamma_i = \frac{(2i-n)\gamma_{\frac{n-1}{2}}\gamma_{\frac{n+1}{2}}}{\left(i - \dfrac{n-1}{2}\right)\gamma_{\frac{n-1}{2}} + \left(i - \dfrac{n+1}{2}\right)\gamma_{\frac{n+1}{2}}}, \tag{2}$$

$$i = 0,1,\dots,\frac{n-3}{2},\frac{n+3}{2},\dots,n,$$

$\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}$ — будь-які, але $\gamma_{\frac{n-1}{2}} \neq \gamma_{\frac{n+1}{2}}$.

Використовуючи формулу (2), дослідимо властивості γ_i, $i = 0, 1, \ldots, \frac{n-3}{2}, \frac{n+3}{2}, \ldots, n.$

Позначимо $m = \min\left\{\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}\right\}$, $M = \max\left\{\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}\right\}$. Маємо

$$M = \alpha m,\ \alpha \geq 1.$$

Нехай $m = \gamma_{\frac{n-1}{2}}$. Представимо γ_i у вигляді

$$\gamma_i = \frac{2i-n}{\left(i - \frac{n-1}{2}\right)\alpha^{-1} + \left(i - \frac{n+1}{2}\right)} \cdot m. \tag{3}$$

Оскільки $0 < \alpha^{-1} < 1$, з (3) маємо

$$m < \gamma_i < \min\left\{2\left(1 + \frac{1}{2i-n-1}\right)m, M\right\}, i = 0, 1, \ldots, \frac{n-3}{2}, \frac{n+3}{2}, \ldots, n.$$

Далі,

$$\lim_{\alpha\to\infty} \gamma_i = 2\left(1 + \frac{1}{2i-n-1}\right)m < 2m,\ i = 0, 1, \ldots, \frac{n-3}{2},$$

$$\lim_{\alpha\to\infty} \gamma_i = 2\left(1 + \frac{1}{2i-n-1}\right)m < 3m,\ i = \frac{n+3}{2}, \ldots, n,$$

тобто якщо $m \ll M$, то

$$m < \gamma_i < 3m,\ i = 0, 1, \ldots, \frac{n-3}{2}, \frac{n+3}{2}, \ldots, n. \tag{4}$$

Якщо $m = \gamma_{\frac{n+1}{2}}$, маємо

$$\gamma_i = \frac{2i-n}{\left(i - \frac{n-1}{2}\right) + \left(i - \frac{n+1}{2}\right)\alpha^{-1}} \cdot m.$$

У випадку $m \ll M$ також отримуємо (4), тобто жодне з γ_i крім $\gamma_{\frac{n-1}{2}}$, $\gamma_{\frac{n+1}{2}}$ не може бути більше найменшого з $\gamma_{\frac{n-1}{2}}, \gamma_{\frac{n+1}{2}}$ більше ніж у 3 рази і не може бути менше його.

Крім того,

$$\gamma_i > \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{n-1}{2}},$$

$$\gamma_i < \gamma_{i+1}, \text{ якщо } m = \gamma_{\frac{n+1}{2}},$$

$$i = 0,1,\dots,\frac{n-3}{2},\frac{n+1}{2},\dots,n .$$

Теорема 2. Якщо в моделі (1) умова $\gamma_i = \gamma, i = 0,1,\dots,n,$ не виконується, то оцінка МНК та оцінка Ейткена параметра b збігаються тоді і тільки тоді, коли:

1) у випадку $n = 3k+1, k = 1,2,\dots$

$$\gamma_i = \gamma_{\frac{2n-2}{3}}, i = 0,1,\dots,\frac{2n-5}{3},\frac{2n+4}{3},\dots,n,$$

$\gamma_{\frac{2n+1}{3}}, \gamma_{\frac{2n-2}{3}}$ — будь-які, але $\gamma_{\frac{2n+1}{3}} \neq \gamma_{\frac{2n-2}{3}}$;

2) у випадку $n = 3k,\ k = 1,2,\dots$

$$\gamma_{\frac{2n+3l}{3}} = \frac{(3l-1)\gamma_{\frac{2n+3}{3}}\gamma_{\frac{2n}{3}}}{(l-1)\gamma_{\frac{2n+3}{3}} + 2l\gamma_{\frac{2n}{3}}}, l = -\frac{2n}{3},\dots,-2,-1,2,\dots,\frac{n}{3}, \tag{5}$$

$\gamma_{\frac{2n}{3}}, \gamma_{\frac{2n+3}{3}}$ — будь-які, але $\gamma_{\frac{2n-3}{3}} \neq \gamma_{\frac{2n}{3}}$;

3) у випадку $n = 3k+2,\ k = 1,2,\dots$

$$\gamma_{\frac{2n+2+3l}{3}} = \frac{(3l+1)\gamma_{\frac{2n-1}{3}}\gamma_{\frac{2n+2}{3}}}{(l+1)\gamma_{\frac{2n-1}{3}} + 2l\gamma_{\frac{2n+2}{3}}}, l = -\frac{2n+2}{3},\dots,-2,1,2,\dots,\frac{n-2}{3}, \tag{6}$$

$\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}$ — будь-які, але $\gamma_{\frac{2n-1}{3}} \neq \gamma_{\frac{2n+2}{3}}$.

У випадку $n = 3k,\ k = 1,2,\dots,$ використовуючи формулу (5), дослідимо властивості $\gamma_i, i = 0,1,\dots,\frac{2n-3}{3}, \frac{2n+6}{3},\dots,n.$

Позначимо $m_1 = \min\left\{\gamma_{\frac{2n}{3}}, \gamma_{\frac{2n+3}{3}}\right\}$, $M_1 = \max\left\{\gamma_{\frac{2n}{3}}, \gamma_{\frac{2n+3}{3}}\right\}$. Маємо

$$M_1 = \beta m_1,\ \beta \geq 1.$$

Нехай $m_1 = \gamma_{\frac{2n}{3}}$. Представимо $\gamma_{\frac{2n+3l}{3}}$ у вигляді

$$\gamma_{\frac{2n+3l}{3}} = \frac{3l-1}{(l-1)+2l\beta^{-1}} \cdot m_1. \tag{7}$$

Оскільки $0<\beta^{-1}<1$, з (7) маємо

$$m_1<\gamma_i<\min\left\{\left(3+\frac{2}{l-1}\right)m_1\ ,M_1\right\}, l=-\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}\ .$$

Далі,

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}}=\left(3+\frac{2}{l-1}\right)m_1<3m_1,\ l=-\frac{2n}{3},\ldots,-2,-1\ ,$$

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}}=\left(3+\frac{2}{l-1}\right)m_1<5m_1,\ l=2,\ldots,\frac{n}{3},$$

тобто, якщо $m_1\ll M_1$, то $m_1<\gamma_{\gamma_{\frac{2n+3l}{3}}}<5m_1,\ l=-\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}.$

Якщо $m_1=\gamma_{\frac{2n+3}{3}}$, маємо

$$\gamma_{\frac{2n+3l}{3}}=\frac{3l-1}{(l-1)\beta^{-1}+2l}\cdot m_1. \tag{8}$$

Оскільки $0<\beta^{-1}<1$, з (8) маємо

$$m_1<\gamma_{\frac{2n+3l}{3}}<\min\left\{\left(\frac{3}{2}-\frac{1}{2l}\right)m_1,M_1\right\}, l=-\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}.$$

Далі,

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}}=\left(\frac{3}{2}-\frac{1}{2l}\right)m_1<2m_1,\ l=-\frac{2n}{3},\ldots,-2,-1\ ,$$

$$\lim_{\alpha\to\infty}\gamma_{\frac{2n+3l}{3}}=\left(\frac{3}{2}-\frac{1}{2l}\right)m_1<\frac{3}{2}m_1,\ l=2,\ldots,\frac{n}{3},$$

тобто, якщо $m_1\ll M_1$, то $m_1<\gamma_{\gamma_{\frac{2n+3l}{3}}}<2m_1,\ l=-\frac{2n}{3},\ldots,-2,-1,2,\ldots,\frac{n}{3}.$

Крім того,

$$\gamma_i>\gamma_{i+1},\ \text{якщо}\ m=\gamma_{\frac{2n}{3}}\ ,$$

$$\gamma_i<\gamma_{i+1},\ \text{якщо}\ m=\gamma_{\frac{2n+3}{3}}\ ,$$

$$i=0,1,\ldots,\frac{2n-3}{3},\ \frac{2n+6}{3},\ldots,n.$$

У випадку $n=3k+2$, $k=1,2,\ldots$, використовуючи формулу (6), дослідимо властивості $\gamma_i, i=0,1,\ldots,\frac{2n-4}{3},\ \frac{2n+5}{3},\ldots,n.$

Далі, позначимо

$$m_2 = \min\left\{\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}\right\},\ M_2 = \max\left\{\gamma_{\frac{2n-1}{3}}, \gamma_{\frac{2n+2}{3}}\right\}.$$

Маємо

$$M_2 = \delta m_2,\ \delta \geq 1.$$

Нехай $m_2 = \gamma_{\frac{2n-1}{3}}$. Представимо $\gamma_{\frac{2n+3l}{3}}$ у вигляді

$$\gamma_{\frac{2n+2+3l}{3}} = \frac{3l+1}{(l+1)\delta^{-1} + 2l} \cdot m_2.$$

Аналогічно попередньому випадку отримуємо, що $m_2 \ll M_2$, то $m_2 < \gamma_{\gamma_{\frac{2n+3l}{3}}} < 2m_2$, $l = -\frac{2n}{3}, \ldots, -2, -1, 2, \ldots, \frac{n}{3}$.

У випадку $m_2 = \gamma_{\frac{2n+2}{3}}$, маємо

$$\gamma_{\frac{2n+2+3l}{3}} = \frac{3l+1}{(l+1) + 2l\delta^{-1}} \cdot m_2.$$

Якщо $m_2 \ll M_2$, то $m_2 < \gamma_{\gamma_{\frac{2n+3l}{3}}} < 5m_2$, $l = -\frac{2n}{3}, \ldots, -2, -1, 2, \ldots, \frac{n}{3}$.

Крім того,

$$\gamma_i > \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{2n-1}{3}},$$

$$\gamma_i < \gamma_{i+1},\ \text{якщо}\ m = \gamma_{\frac{2n+2}{3}},$$

$$i = 0, 1, \ldots, \frac{2n-4}{3},\ \frac{2n+5}{3}, \ldots, n.$$

Висновки. Доведено, що рівність оцінок МНК та Ейткена параметра a забезпечує родина векторів $(n+1)$-го порядку, i-й елемент вектора з цієї родини — дисперсія відхилення в точці t_i, $i = 0, 1, \ldots, n$. Знайдено два сусідні елементи цього вектору, якіможуть приймати будь-які значення, а значення всіх інших елементів знаходяться між ними. Рівність оцінок МНК та Ейткена параметра b забезпечує інша родина векторів $(n+1)$-го порядку, у яких i-й елемент — також дисперсія відхилення в точці t_i, $i = 0, 1, \ldots, n$.

Список використаних джерел:

1. Демиденко Е. З. Линейная и нелинейная регрессии. М. : Финансы и статистика, 1981. 302 с.

2. Андерсон Т. Статистический анализ временных рядов. М. : Мир, 1976. 756 с.
3. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. 336 с.
4. Савкіна М. Ю. Умови збігу оцінок МНК та Ейткена параметрів моделі лінійної регресії. *Журнал обчислювальної та прикладної математики.* 1980. № 3 (129). С. 34–42.

EQUALITY OF MNC AND AITKEN ESTIMATIONS OF LINEAR REGRESSION MODEL PAPER IN THE CASE OF HETEROSCEDASTIC DEVIATIONS

At the paper in the case of heteroscedastic independent deviations a linear regression model whose function has the form $f(t) = at + b$, where a and b unknown parameters, is studied. Approximate values (observations) of functions $f(t)$ are registered at equidistant points of the segment $[0,1]$. We formulate Theorem 1, which gives conditions on the variances of deviations, in which the Aitken estimation of parameter a coincides with its estimation of MNCs. Under these conditions, the Aitken and MNC estimations of the parameter b will not coincide. We also formulate Theorem 2, which gives the conditions for the coincidence of the Aitken estimation and the MNC estimation of parameter b. Based on Theorems 1 and 2, in this paper the properties of variances of deviations that give equality with these estimations separately for parameter a and for parameter b are investigated. It is shown that for equality estimations of Aitken and MNC of the parameter a the deviations will have the largest and smallest variance in two adjacent observation points located in the middle of the segment [0, 1], for the equality estimations of the parameter b — in the neighborhood of the point 2/3. The asymptotic values of the variances of all deviations are found, if the ratio of the largest to the smallest variance goes to infinity. It is proved that in this case, the variances of all deviations will be no more the smallest variance than 3 times for parameter a and not more than 5 times for parameter b.

Key words: *least square method, regression model, Aitken estimation.*

Одержано 15.02.2019