ЭКОНОМИКА
И МАТЕМАТИЧЕСКИЕ МЕТОДЫ 1971 том VII, вып. 2

ОЦЕНКА ПАРАМЕТРОВ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ ПРИ СЕРИЙНО-КОРРЕЛИРОВАННЫХ ИСПЫТАНИЯХ

А. И. БЕЛЯЕВСКИЙ

(Ленинград)

1. В классической регрессионной модели рассматриваются наблюдения $(y_1, \ldots, y_N)$, удовлетворяющие условиям

$$M\{y_\xi\} = bx_\xi,$$

$$Cov\{y_\xi y_\eta\} = \sigma^2\delta_{\eta\xi}, \quad \xi, \eta = 1, \ldots, N,$$

где b, σ^2 — неизвестные параметры, подлежащие оценке, x_ξ, $\xi = 1, \ldots, N$ — известные коэффициенты, $\delta_{\xi\eta}$ = символ Кронекера, $M\{\ \}$, $Cov\{\ \}$ — операторы математического ожидания и ковариации соответственно.

Во многих практических случаях эта модель не оправдывается, так как условие некоррелированности наблюдений не выполняется.

Рассмотрим следующую задачу.

Имеется некоторая сельскохозяйственная зона, в которую входит ряд сельскохозяйственных предприятий. Необходимо выявить производственную функцию, описывающую в среднем для данной зоны линейную зависимость какого-либо производственного показателя, например, урожайности той или иной сельскохозяйственной культуры, от метеорологического фактора. Каждому испытанию соответствует пара наблюдений $(y_{\alpha i}, x_{\alpha i})$, где $y_{\alpha i}$ — наблюдаемое значение урожайности в α-м году (или в α-й серии наблюдений) на i-м предприятии; $x_{\alpha i}$ — соответствующее значение метеофактора.

Очевидно, $Cov\{y_{\alpha i} y_{\alpha j}\} \neq 0$, т. е. корреляция между наблюдениями, относящимися к одной серии испытаний не равна нулю, хотя бы в силу пространственной зависимости метеорологических факторов, не учитываемых в уравнении регрессии, но тем не менее влияющих на урожайность. Предположение, что корреляция наблюдений, относящихся к разным годам, нулевая, представляется интуитивно менее ограничительным. Поэтому для данной задачи, более близкой к практике, является модель, в которой допускается возможность корреляций внутри серий испытаний, а именно

$$y_\alpha = bx_\alpha + u_\alpha, \quad \alpha = 1, \ldots, n, \tag{1}$$

где $y_\alpha^* = (y_{\alpha_1}, \ldots, y_{\alpha p})$ — вектор-строка наблюдений, относящихся к α-й серии испытаний; $x_\alpha^* = (x_{\alpha_1}, \ldots, x_{\alpha p})$, $\alpha = 1, \ldots, n$,— векторы известных коэффициентов; b — неизвестная скалярная величина.

Для векторов-возмущений u_α, $\alpha = 1, \ldots, n$, выполняются условия

$$\begin{gathered} M\{u_\alpha\} = 0, \\ Cov\{u_\alpha u_\beta^*\} = W\delta_{\alpha\beta}, \end{gathered} \tag{2}$$

где W — положительно определенная $p \times p$ матрица. В общем случае матрица W также не известна.

Если $W = \sigma^2 E$ (где E — единичная матрица, σ^2 — неизвестный скалярный параметр), то модель, определяемая соотношениями (1), (2), сводится к классической регрессионной схеме. К ней же можно привести рассматриваемую схему, если матрица W пропорциональна заданной матрице B, т. е. $W = \sigma^2 B$, где σ^2 — параметр, подлежащий оценке (см., например, [1]).

В [2] рассмотрена относительная эффективность оценки методом наименьших квадратов параметра относительно эффективной оценки, получаемой при известной матрице W, и приведены необходимые и достаточные условия, при которых эта относительная эффективность равна единице.

В данной статье предлагается процедура нахождения оценок b и W, являющихся более эффективными, чем оценки, получаемые по методу наименьших квадратов; при этом для нахождения оценок используется метод максимального правдоподобия.

2. Предположим, что случайный вектор y имеет нормальное распределение с математическим ожиданием bx и матрицей ковариаций W, т. е. $y \in N\{bx, W\}$. Тогда плотность распределения для выборки $\{y_\alpha\}$, $\alpha = 1, \ldots, n$,— функция правдоподобия -- имеет вид

$$p(y_1, \ldots y_n) = \frac{1}{(2\pi)^{pn/2}} |W|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{\alpha}^{n} (y_\alpha - bx_\alpha)^* \times \right.$$
$$\left. \times W^{-1}(y_\alpha - bx_\alpha)\right\}. \tag{3}$$

Допустим, что матрица W известна. Тогда, дифференцируя логарифм функции правдоподобия по параметру b, получим уравнение для оценки максимального правдоподобия b

$$\sum_{\alpha}^{n} x_\alpha^* W^{-1}(y_\alpha - bx_\alpha) = 0,$$

откуда

$$\hat{b} = \left(\sum_{\alpha}^{n} x_\alpha^* W^{-1} x_\alpha\right)^{-1} \sum_{\alpha}^{n} x_\alpha^* W^{-1} y_\alpha.$$

Легко показать достаточность оценки $\hat{b}$. В самом деле, функция правдоподобия (3) может быть представлена в виде

$$p(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{pn/2}} |W|^{-n/2} \exp\left\{-\frac{1}{2}\sum_{\alpha}^{n} (y_\alpha - \hat{b}x_\alpha)^* \times \right.$$
$$\left. \times W^{-1}(y_\alpha - \hat{b}x_\alpha)\right\} \times$$
$$\times \exp\left\{-\frac{1}{2}(\hat{b} - b)^2 \sum_{\alpha}^{n} x_\alpha^* W^{-1} x_\alpha\right\}.$$

Таким образом, оценка b удовлетворяет критерию факторизации [3] и является, тем самым, достаточной оценкой для параметра b. Помимо того, для частной производной логарифма функции правдоподобия по параметру b выполняется равенство

$$\frac{\partial \ln p(y_1, \ldots, y_n)}{\partial b} = (\hat{b} - b) \sum_{\alpha}^{n} x_\alpha{}^* W^{-1} x_\alpha. \tag{4}$$

Тогда, в соответствии с (4), выполняются условия эффективности оценки $\hat{b}$ [3] в том смысле, что для дисперсии оценки достигается нижняя грань в неравенстве Крамера — Рао. Итак, среди всех оценок параметра b обладает наименьшей дисперсией оценка $\hat{b}$.

Пусть теперь b — заданный параметр. Тогда оценка максимального правдоподобия для неизвестной матрицы W имеет вид

$$n\hat{W} = \sum_{\alpha}^{n} (y_\alpha - bx_\alpha)(y_\alpha - bx_\alpha)^*. \tag{5}$$

Выражение (5) нетрудно получить, дифференцируя функцию правдоподобия по параметрам ω^{ij}, $i, j = 1, \ldots, p$,— элементам обратной матрицы W^{-1}.

При подстановке (5) в (3) получим, что функция правдоподобия зависит явным образом лишь от параметра W и оценки $\hat{W}$. Отсюда следует, что $\hat{W}$ — достаточная оценка для W. Далее

$$\frac{\partial \ln p(y_1, \ldots, y_n)}{\partial \omega^{ij}} = \omega_{ij} - \frac{1}{n} \sum_{\alpha}^{n} (y_{\alpha i} - bx_{\alpha i})(y_{\alpha i} - bx_{\alpha j}) = \omega_{ij} - \hat{\omega}_{ij}. \tag{6}$$

Тогда ввиду равенства (6) и достаточности оценки (5) выполняется условие эффективности и для $\hat{W}$.

Пусть теперь b и W — неизвестные параметры. Дифференцируя (3) по b и ω^{ij}, получим уравнения для оценки максимального правдоподобия

$$\begin{gathered} b = \left(\sum_{\alpha}^{n} x_\alpha{}^* W^{-1} x_\alpha \right)^{-1} \sum_{\alpha}^{n} x_\alpha{}^* W^{-1} y_\alpha, \\ W = \frac{1}{n} \sum_{\alpha}^{n} (y_\alpha - bx_\alpha)(y_\alpha - bx_\alpha)^*. \end{gathered} \tag{7}$$

Приведенные выше доказательства эффективности оценок $\hat{b}$ и W могут служить интуитивным подтверждением разумности оценок, удовлетворяющих системе (7) при неизвестных b и W.

Будем решать систему (7) методом последовательных приближений. Пусть в качестве первого приближения выбрана оценка наименьших квадратов $b_1 = \left(\sum_{\alpha}^{n} x_\alpha{}^* x_\alpha \right)^{-1} \sum_{\alpha}^{n} x_\alpha{}^* y_\alpha$; последовательно получаемые оценки удовлетворяют соотношениям

$$\begin{gathered} b_k = \left(\sum_{\alpha}^{n} x_\alpha{}^* W_{k-1}^{-1} x_\alpha \right)^{-1} \left(\sum_{\alpha}^{n} x_\alpha{}^* W_{k-1}^{-1} y_\alpha \right), \\ nW_{k-1} = \sum_{\alpha}^{n} (y_\alpha - b_{k-1} x_\alpha)(y_\alpha - b_{k-1} x_\alpha)^*. \end{gathered} \tag{8}$$

3. Докажем несмещенность оценок b_k. Рассмотрим предварительно оценку b_1. Ее несмещенность следует непосредственно из (3). Обозначим $u_\alpha = y_\alpha - bx_\alpha$, $u_\alpha^{(k)} = y_\alpha - b_k x_\alpha$. Можно получить

$$u_\alpha^{(1)} = y_\alpha - b_1 x_\alpha = y_\alpha - \left(\sum_\beta^n x_\beta{}^* x_\beta\right)^{-1}\left(\sum_\beta^n x_\beta{}^* y_\beta\right)x_\alpha =$$

$$= u_\alpha - \left[\left(\sum_\beta^n x_\beta{}^* x_\beta\right)^{-1}\sum_\beta^n x_\beta{}^* u_\beta\right]x_\alpha.$$

Таким образом, одновременное изменение знака у всех u_α (у всех компонент) влечет за собой соответствующее изменение знака у всех $u_\alpha^{(k)}$, $\alpha = 1, \ldots, n$.

Отсюда $W_1 = \sum\limits_\alpha u_\alpha^{(1)} u_\alpha^{(1)*}$ сохраняет свое значение при одновременном изменении знака у всех u_α.

Рассмотрим теперь величину

$$b_2 - b = \left(\sum_\alpha x_\alpha{}^* W_1^{-1} x_\alpha\right)^{-1}\sum_\alpha x_\alpha{}^* W u_\alpha. \tag{9}$$

Из (9) видно, что одновременное изменение знака у всех u_α, $\alpha = 1, \ldots, n$, приводит лишь к изменению знака у разности $b_2 - b$.

Так как из (3) следует, что $p(u_1, \ldots, u_n) = p(-u_1, \ldots, -u_n)$, получим, что $b_2 - b$ имеет ту же самую плотность вероятности, что и $b - b_2$.

Следовательно, оценка b_2 симметрично распределена относительно b и b_2 — несмещенная оценка параметра b. Далее можно представить

$$u_\alpha^{(2)} = u_\alpha - \left[\left(\sum_\beta x_\beta{}^* W_1^{-1} x_\beta\right)^{-1}\sum_\beta x_\beta{}^* W_1^{-1} u_\beta\right]x_\alpha.$$

Так как W_1^{-1} — четная функция относительно $u_1, \ldots, u_n$, то $u_\alpha^{(2)}$ при одновременном изменении знака у всех u_α, $\alpha = 1, \ldots, n$, также меняет знак на обратный. Отсюда W_2^{-1} — четная функция относительно $u_1, \ldots, u_n$ и, следовательно, b_3 — несмещенная оценка параметра b.

По индукции следует доказательство несмещенности оценок b_k, $k = 1, 2, \ldots$.

Покажем теперь состоятельность оценок b_k, W_{k-1}. С этой целью найдем дисперсию оценки b_1. Как нетрудно видеть

$$D(b_1) = \left(\sum_\alpha^n x_\alpha{}^* x_\alpha\right)^{-2}\left(\sum_\alpha^n x_\alpha{}^* W x_\alpha\right).$$

Пусть λ_1 — максимальное собственное число матрицы W. Тогда для всех α имеем $x_\alpha{}^* W x_\alpha \leqslant \lambda_1(x_\alpha{}^* x_\alpha)$ и, следовательно $\sum\limits_\alpha^n x_\alpha{}^* W x_\alpha \leqslant \lambda_1 \sum\limits_\alpha^n x_\alpha{}^* x_\alpha$.

Отсюда ясно, что при $n \to \infty$ дисперсия оценки b_1 стремится к нулю и, следовательно, b_1 — состоятельная оценка параметра b.

Рассмотрим теперь матрицу $W_1 = n^{-1} \sum_{\alpha} (y_\alpha - b_1 x_\alpha)(y_\alpha - b_1 x_\alpha)^*$. Перепишем ее в виде

$$W_1 = n^{-1} \sum_{\alpha}^{n} (y_\alpha - bx_\alpha)(y_\alpha - bx_\alpha)^* + n^{-1} \sum_{\alpha}^{n} (y_\alpha - bx_\alpha) x_\alpha^* (b - b_1) +$$

$$+ n^{-1} \sum_{\alpha}^{n} x_\alpha (y_\alpha - bx_\alpha)^* (b - b_1) + n^{-1} \sum_{\alpha}^{n} x_\alpha x_\alpha^* (b - b_1)^2.$$

Первый член разложения совпадает с эффективной и состоятельной оценкой $\hat{W}$ матрицы W. Остальные члены стремятся к нулю по вероятности. Отсюда W_1 — состоятельная оценка матрицы W. Но если W_1 состоятельна, то по теореме Слуцкого [4] W_1^{-1} — состоятельная оценка матрицы W^{-1}. Значит, оценка b_2 сходится по вероятности к оценке $\hat{b}$ и, тем самым, также является состоятельной.

Повторяя вышеприведенные рассуждения, можно прийти к утверждению о состоятельности следующих итераций b_k, W_{k-1}.

Ввиду состоятельности $W_1, W_2, \ldots,$ получаем, что оценки b_k, начиная с $k = 2$, асимптотически эффективны и асимптотически имеют то же самое распределение, что и оценка $\hat{b}$. Поэтому при достаточно больших n с вероятностью, сколь угодно близкой к 1,

$$\frac{D(b_1)}{D(b_2)} \geqslant 1. \tag{10}$$

Можно получить оценку для сравнительной эффективности сверху, имеющую место при достаточно больших n.

Выражение для дисперсии эффективной оценки $\hat{b}$

$$D(\hat{b}) = \left(\sum_{\alpha} x_\alpha^* W^{-1} x_\alpha \right)^{-1}.$$

Отсюда для достаточно больших n

$$\frac{D(b_1)}{D(b_2)} = \frac{\left(\sum_{\alpha}^{n} x_\alpha^* W x_\alpha \right) \left(\sum_{\alpha}^{n} x_\alpha^* W x_\alpha \right)}{\left(\sum_{\alpha}^{n} x_\alpha^* x_\alpha \right)^2}.$$

Используя неравенство Канторовича [1], получим

$$1 \leqslant \frac{D(b_1)}{D(b_2)} \leqslant \frac{1}{4} \left[\left(\frac{\lambda_1}{\lambda_n} \right)^{1/2} + \left(\frac{\lambda_n}{\lambda_1} \right)^{1/2} \right]^2, \tag{11}$$

где λ_1 и λ_n — соответственно максимальное и минимальное собственные числа матрицы W. При $\lambda_1 = \ldots = \lambda_n$ сравнительная эффективность равна единице и, таким образом, оценка b_2 не будет эффективнее оценки b_1.

Введем теперь следующие обозначения: $\tilde{x}_\alpha = W^{-1/2} x_\alpha$, $\tilde{y}_\alpha = W^{-1/2} y_\alpha$.

В преобразованных координатах оценка наименьших квадратов $\tilde{b}_1$ имеет вид

$$\tilde{b}_1 = \left(\sum_{\alpha} \tilde{x}_\alpha{}^* \tilde{x}_\alpha\right)^{-1} \sum_{\alpha} \tilde{x}_\alpha{}^* \tilde{y}_\alpha = \left(\sum_{\alpha}^{n} x_\alpha{}^* W_1{}^{-1} x_\alpha\right)^{-1} \sum_{\alpha}^{n} x_\alpha{}^* W_1{}^{-1} y_\alpha = b_2. \tag{12}$$

Аналогично (8) имеем для модели в преобразованных координатах

$$\tilde{b}_2 = \left(\sum_{\alpha}^{n} \tilde{x}_\alpha{}^* \tilde{W}_1{}^{-1} x_\alpha\right)^{-1} \left(\sum_{\alpha}^{n} \tilde{x}_\alpha{}^* \tilde{W}_1{}^{-1} y_\alpha\right), \tag{13}$$

где

$$\tilde{W}_1 = n^{-1} \sum_{\alpha}^{n} (\tilde{y}_\alpha - \tilde{b}_1 \tilde{x}_\alpha)(\tilde{y}_\alpha - \tilde{b}_1 x_\alpha)^*.$$

С учетом (12) получим: $n\tilde{W}_1 = W_1{}^{1/2} W_2 W_1{}^{-1/2}$. Подставим это выражение в (13)

$$\tilde{b}_2 = \left(\sum_{\alpha}^{n} x_\alpha{}^* W_1{}^{+1/2} W_2{}^{-1} W_1{}^{+1/2} x_\alpha\right)^{-1} \sum_{\alpha}^{n} (x_\alpha{}^* W_1{}^{+1/2} W_2 W_1{}^{1/2} y_\alpha) =$$

$$= \left(\sum_{\alpha}^{n} x_\alpha{}^* W_2{}^{-1} x_\alpha\right)^{-1} \sum_{\alpha}^{n} x_\alpha{}^* W_2{}^{-1} y_\alpha = b_3.$$

Для достаточно больших n имеем для сравнительной эффективности оценок $\tilde{b}_2$ и $\tilde{b}_1$ неравенства, аналогичные (10), (11). Отсюда следует выполнение этих неравенств для сравнительной эффективности оценок b_3 и b_2.

Точно также можно получить подобные утверждения для любой пары оценок b_{k+1} и b_k.

Суммируем теперь полученные результаты.

1) Рассмотренные оценки b_k при достаточно больших n с вероятностью, сколь угодно близкой к 1, удовлетворяют неравенству

$$\frac{D(b_k)}{D(b_{k+1})} \geqslant 1.$$

Тем самым, в указанном смысле происходит улучшение оценок при увеличении k.

2) Оценка b_k для любого k является несмещенной оценкой параметра b.

3) Оценки b_k, W_k для любого k состоятельны.

4) Оценки b_k, начиная с $k = 2$, асимптотически эффективны.

Вышеприведенные результаты дополняют [5], где для матрицы W специального вида рассмотрена оценка b_2 и доказана ее состоятельность и асимптотическая эффективность.

4. В заключение рассмотрим вопрос о возможных способах проверки статистических гипотез для параметров W и b.

Возьмем вектора $y_i = (y_{1i}, \ldots, y_{ni})^*$, $i = 1, \ldots, p$, и соответственно вектора $x_i = (x_{1i}, \ldots, x_{ni})^*$. Для введенных векторов имеем $M\{y_i\} = bx_i$.

Применим к каждому из векторов y_i ортогональное преобразование C такое, что первые p строк матрицы C принадлежат пространству, натянутому на вектора $x_1, \ldots, x_p$.

Пусть $\tilde{y}_i = Cy_i$; тогда для каждого $\tilde{y}_{\alpha i}$, $\alpha > p$, имеем по построению $M\{\tilde{y}_{\alpha i}\} = 0$. Разобьем матрицу Y, составленную из элементов $\tilde{y}_{\alpha i}$, $\alpha = 1, \ldots, n$; $i = 1, \ldots, p$, так, что $\tilde{y}_{\alpha i}$ принадлежит α-й строке и i-му столбцу матрицы $\tilde{Y}$ следующим образом

$$\tilde{Y} = \begin{pmatrix} \tilde{Y}_1 \\ \tilde{Y}_2 \end{pmatrix},$$

где $\tilde{Y}_1$ составлена из первых строк матрицы $\tilde{Y}$; $\tilde{Y}_2$ — из оставшихся $n - p$ строк.

Для любых α, β имеем $Cov(\tilde{y}_{\alpha i}\tilde{y}_{\beta j}) = Cov(c_\alpha^* y_i y_j^* c_\beta) = \omega_{ij} c_\alpha^* c_\beta$, где c_α^*, c_β^* — соответственно α-я и β-я строки матрицы C.

Ввиду ортогональности c_α^* и c_β^*

$$Cov(\tilde{y}_{\alpha i}\tilde{y}_{\beta j}) = \omega_{ij}\delta_{\alpha\beta}. \tag{14}$$

Учитывая, что $M\{\tilde{Y}_2\} = 0$, получим

$$M\{\tilde{Y}_2^*\tilde{Y}_2\} = W. \tag{15}$$

Статистика $\tilde{Y}_2^*\tilde{Y}_2 = \sum\limits_{\alpha=p+1}^{n} \tilde{y}_\alpha \tilde{y}_\alpha^*$ где $\tilde{y}_\alpha$ — α-я строка матрицы Y дает несмещенную оценку матрицы W, не зависящую от значений параметра b, и $\tilde{Y}_2^*\tilde{Y}_2$ принадлежит распределению Уитшарта с $n - p$ степенями свободы. Поэтому она может быть использована для проверки гипотезы о сферичности W [6], что эквивалентно проверке гипотезы о том, что наблюдения являются независимыми и равноточными, т. е. соответствуют классической схеме.

Приведем более явное выражение для $\tilde{Y}_2^*\tilde{Y}_2$. Как нетрудно показать [6],

$$\tilde{Y}_2^*\tilde{Y}_2 = \sum_{\alpha=1}^{n} (y_\alpha - \hat{B}x_\alpha)(y_\alpha - \hat{B}x_\alpha)^*,$$

где $B = \left(\sum\limits_{\alpha=1}^{n} x_\alpha x_\alpha^*\right)^{-1} \left(\sum\limits_{\alpha}^{n} x_\alpha y_\alpha^*\right)$ — матрица оценок коэффициентов многомерной регрессии B вектора y на вектор x по совместным наблюдениям $\{y_\alpha, x_\alpha\}$, $\alpha = 1, \ldots, n$, а $(n - p)^{-1}\tilde{Y}_2^*\tilde{Y}_2$ — матрица оценок остаточных ковариаций.

Аналогично проверка гипотезы $b = 0$ может быть заменена проверкой гипотезы $B = 0$ в модели многомерной регрессии [6], что приводит к критерию вида

$$\left|\sum_{\alpha}^{n} (y_\alpha - \hat{b}x_\alpha)(y_\alpha - Bx_\alpha)\right| \Big/ \left|\sum_{\alpha}^{n} y^\alpha y_\alpha^*\right| = \frac{|Y_2^*Y_2|}{|Y_1^*Y_1 + Y_2^*Y_2|}. \tag{16}$$

Изложенные выше критерии не совпадают с критериями отношения правдоподобия, так как не учитывают ограничений, вводимых моделью (1), на коэффициенты многомерной регрессии. Однако они удовлетворяют требованиям инвариантности, т. е. не зависят от выбора частной системы координат, служащей для выражения результатов наблюдения. В самом деле, критерии, зависящие только от $\tilde{Y}_2^*\tilde{Y}_2$, являются максимальными инвариантами [6] относительно сдвига $\tilde{Y}_1$ и ортогональных преобразований.

Критерий же вида (15) является максимальным инвариантом относительно вообще линейных преобразований. Эти свойства построенных критериев — естественное основание их применения.

ЛИТЕРАТУРА

1. С. Р. Рао. Линейные статистические методы и их применения. М., «Наука», 1968.
2. G. S. Watson. Linear Least Squares Regression. Annals of Math. Statist., 1967, v. 38, N 2.
3. С. Уилкс. Математическая статистика. М., «Наука», 1967.
4. Г. Крамер. Математические методы статистики. М., Изд-во иностр. лит., 1948.
5. R. W. Rarks. Efficient Estimation of a System of Regression Equations when Disturbances are Both Serially and Contemparencously Correlated. Journal of the American Statist. Assoc. 1967, v. 63, N 318.
6. Т. Андерсон. Введение в многомерный статистический анализ. М., Физматгиз, 1963.

Поступила в редакцию
3 IX 1969