УДК 519.86

Состоятельность МНК-оценок коэффициентов множественной линейной регрессионной модели

Л. М. Гафарова, И. Г. Завьялова, Н. Н. Мустафин

Национальный исследовательский университет «МИЭТ»

Авторы рассматривают регрессионные модели — основной инструмент исследования и прогнозирования экономических процессов — в целях исследования оценок параметров регрессии, полученных методом наименьших квадратов. Оценки являются несмещенными, состоятельными и эффективными. Поскольку в доступной литературе доказательства состоятельности приводятся исключительно для простой линейной регрессии (однофакторной), авторы делают попытку доказать состоятельность оценок коэффициентов линейной регрессии для многомерного случая.

Ключевые слова: множественная линейная регрессия; прогнозирование экономических процессов; метод наименьших квадратов; состоятельность оценок коэффициентов линейной регрессии.

Рассмотрим модель множественной линейной регрессии

$$M(X/\overline{T}=\bar{t}\,)=\beta_1 t_1+\beta_2 t_2+\ldots+\beta_k t_k,$$

где $\overline{T}=(T_1, T_2, \ldots, T_k)$ — вектор факторов регрессии; $\overline{\beta}=(\beta_1, \beta_2, \ldots, \beta_k)$ — вектор коэффициентов регрессии; k — число факторов регрессии, фиксированное [1, с. 197]; $(X_1, \ldots, X_n)$ — значения случайной величины X, наблюдаемые в n экспериментах, $n > k$; $(t_i^{(1)}, \ldots, t_i^{(n)})$ — значения i-го фактора регрессии в n экспериментах, $i = 1, \ldots, k$.

$$\begin{cases} X_1=\beta_1 t_1^{(1)}+\beta_2 t_2^{(1)}+\ldots+\beta_k t_k^{(1)}+\varepsilon_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_n=\beta_1 t_1^{(n)}+\beta_2 t_2^{(n)}+\ldots+\beta_k t_k^{(n)}+\varepsilon_n. \end{cases}$$

Здесь $\bar{\varepsilon}=(\varepsilon_1, \ldots, \varepsilon_n)$ — вектор остатков (ошибок наблюдений). Запишем эту систему в матричном виде

$$\overline{X}=T\cdot\overline{\beta}+\bar{\varepsilon},$$

где матрица $T=\begin{pmatrix} t_1^{(1)} & \ldots & t_k^{(1)} \\ \ldots & \ldots & \ldots \\ t_1^{(n)} & \ldots & t_k^{(n)} \end{pmatrix}.$

Вектор остатков $\bar{\varepsilon}$ удовлетворяет следующим предположениям:

1) $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, n$, т. е. все остатки имеют нормальное распределение с нулевым математическим ожиданием и дисперсией σ^2;

2) $\operatorname{cov}(\varepsilon_i, \varepsilon_j)=\sigma^2\delta_{ij}$, где

$$\delta_{ij}=\begin{bmatrix} 1, \; i=j, \\ 0, \; i\neq j, \; i=1, \ldots, n; \; j=1, \ldots, k, \end{bmatrix}$$

т. е. остатки некоррелированы. Из этих условий следует также независимость остатков.

Оптимальные оценки $\overline{\beta}^*=(\beta_1^*, \ldots, \beta_k^*)$ коэффициентов регрессии получаются методом наименьших квадратов (МНК). Это оценки, которые доставляют минимум сумме квадратов остатков в n экспериментах.

$$\sum_{i=1}^{n}\varepsilon_i^2=(\bar{\varepsilon},\bar{\varepsilon})=\|\varepsilon\|^2=(\overline{X}-T\overline{\beta})(\overline{X}-T\overline{\beta})'\to\min.$$

Оценки $\overline{\beta}^*$ являются решением *нормальной системы* $T'\cdot T\cdot\overline{\beta}^*=T'\cdot\overline{X}$ или $A\overline{\beta}^*=T'\cdot\overline{X}$, где $A=T'\cdot T$ — информационная матрица.

Если A — невырожденная матрица, т. е. $\mathrm{rank}(A) = k \leqslant n$, то в этом и только в этом случае нормальная система имеет единственное решение $\bar{\beta}^* = A^{-1} \cdot T' \cdot \bar{X}$ [2, с. 168]. Из этого равенства следует несмещенность $\bar{\beta}^*$: $M(\bar{\beta}^*) = \bar{\beta}$ [2, с. 168].

Сложнее доказать состоятельность оценок $\bar{\beta}^*$, т. е. доказать, что $\beta_i^* \xrightarrow[n\to\infty]{P} \beta_i$ ($\bar{\beta}_i^*$ сходятся по вероятности к β_i), $i = 1, 2, \ldots, k$ (для любого $\varepsilon > 0$ $P\left\{\left|\beta_i^* - \beta_i\right| > \varepsilon\right\} \xrightarrow[n\to\infty]{} 0$).

Согласно неравенству Чебышева,

$$P\{|\beta_i^* - \beta_i| > \varepsilon\} = P\{|\beta_i^* - M\beta_i^*| > \varepsilon\} \leqslant \frac{D\beta_i^*}{\varepsilon^2}.$$

Докажем, что для $i = 1, \ldots, k$, дисперсии оценок коэффициентов регрессии $D\beta_i^* \xrightarrow[n\to\infty]{} 0$.

Дисперсии $D\beta_i^*$, $i = 1, \ldots, k$, стоят на главной диагонали ковариационной матрицы вектора $\bar{\beta}^* - K_{\bar{\beta}^*}$.

$K_{\bar{\beta}^*} = \sigma^2 \cdot A^{-1}$, где σ^2 — дисперсия ошибок наблюдений; A — информационная матрица [1, с. 201]. Перечислим свойства информационной матрицы $A = T' \cdot T$.

1. Матрица A симметрична, положительно полуопределена, и $\mathrm{rank}(A) = \mathrm{rank}(T)$ (см. [3]).

2. Поскольку матрица A невырожденная, ее собственные значения не равны нулю, и матрица A положительно определена [4, стр. 130].

3. Любая симметричная положительно определенная матрица A может быть приведена к диагональному виду D переходом к новому базису $D = P^{-1} \cdot A \cdot P$; на главной диагонали матрицы D стоят собственные значения матрицы A (действительные и положительные) — $\lambda_1, \lambda_2, \ldots, \lambda_k$; P — матрица перехода к новому базису [5, с. 127].

В новом базисе матрица обратного оператора D^{-1} тоже диагональная с собственными значениями на главной диагонали $1/\lambda_1, 1/\lambda_2, \ldots, 1/\lambda_k$ [4, с. 136].

4. Характеристические многочлены матриц A и D одинаковые:

$$\begin{gathered}\det(D - \lambda E) = \det(P^{-1} \cdot A \cdot P - \lambda E) = \\ = \det(P^{-1} \cdot A \cdot P - P^{-1} \cdot \lambda E P) = \\ = \det(P^{-1}(A - \lambda E)P) = \det(P^{-1}) \times \\ \times \det(A - \lambda E) \cdot \det(P) = \det(A - \lambda E).\end{gathered}$$

Значит, совпадают коэффициенты характеристических многочленов при одинаковых степенях λ. Характеристический многочлен диагональной матрицы D $\det(D - \lambda E) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots (\lambda_k - \lambda)$. Приравняем коэффициенты при одинаковых степенях λ:

– при $\lambda^k : (-1)^k = (-1)^k$;

– при $\lambda^{k-1} : a_{11} + a_{22} + \ldots + a_{kk} = \lambda_1 + \lambda_2 + \ldots \lambda_k$ (след матрицы инвариантен относительно преобразования $P^{-1} \cdot A \cdot P$);

– при λ^{k-2} : $(a_{11} \cdot a_{22} - a_{12}^2) + (a_{11} \cdot a_{33} - a_{13}^2) + \ldots + (a_{k-1k-1} \cdot a_{kk} - a_{k-1k}^2) = \lambda_1\lambda_2 + \lambda_1\lambda_3 + \ldots \lambda_{k-1}\lambda_k$ и т. д.;

– при λ^0 : $\det A = \lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_k$.

Таким образом, суммы миноров порядка r, симметричных относительно главной диагонали, у матриц A и D совпадают, $r = 1, 2, \ldots, k$. В частности, сумма таких миноров порядка $(k - 1)$ матрицы A равна $\lambda_2 \cdot \ldots \cdot \lambda_k + \lambda_1 \cdot \lambda_3 \cdot \ldots \cdot \lambda_k + \ldots + \lambda_1 \cdot \ldots \cdot \lambda_{k-1}$.

Выпишем матрицу $A = T' \cdot T$

$$A = \begin{pmatrix} \sum\limits_{j=1}^{n} (t_1^{(j)})^2 & \ldots & \sum\limits_{j=1}^{n} t_1^{(j)} \cdot t_k^{(j)} \\ \sum\limits_{j=1}^{n} t_2^{(j)} t_1^{(j)} & \sum\limits_{j=1}^{n} (t_j^{(j)})^2 & \ldots \\ \ldots & \ldots & \ldots \\ \sum\limits_{j=1}^{n} t_k^{(j)} t_1^{(j)} & \ldots & \sum\limits_{j=1}^{n} (t_k^{(j)})^2 \end{pmatrix}.$$

Предположим, что существует такое $\delta > 0$, что для всех элементов матрицы T $|t_i^{(j)}| \geqslant \delta$, $i = 1, \ldots, k$, $j = 1, \ldots, n$. В противном случае, если есть подпоследовательность из $\{t_i^{(j)}\}$, сходящаяся к нулю, то для выполнения этих условий достаточно начало отсчета перенести в другую точку, отдаленную от всех предельных точек не менее чем на некоторое $\delta > 0$ (при этом все значения факторов регрессии сместятся на $\delta > 0$). Следовательно, не ограничивая общности, можно считать, что $|t_i^{(j)}| \geqslant \delta$. Это означает, что все диагональные элементы матрицы A стремятся к бесконечности $a_{ii} \xrightarrow[n\to\infty]{} \infty$, $i = 1, 2, ..., k$. Поскольку след матрицы A и след матрицы D совпадают, $\lambda_1 + \lambda_2 + ... + \lambda_k \xrightarrow[n\to\infty]{} \infty$ и $\det A \xrightarrow[n\to\infty]{} \infty$.

Аналогично, определитель любой подматрицы матрицы A, симметричной относительно главной диагонали (подматрица соответствует некоторому подмножеству факторов), стремится к бесконечности при $n \to \infty$.

Если обозначить вектор-столбцы матрицы T как $\bar{t}_1, ..., \bar{t}_k$, то матрицу A можно переписать в виде

$$A = \begin{pmatrix} \|\bar{t}_1\|^2 & (\bar{t}_1,\bar{t}_2) & ... & (\bar{t}_1,\bar{t}_k) \\ (\bar{t}_2,\bar{t}_1) & \|\bar{t}_2\|^2 & ... & (\bar{t}_2,\bar{t}_k) \\ ... & ... & ... & ... \\ (\bar{t}_k,\bar{t}_1) & (\bar{t}_k,\bar{t}_2) & ... & \|\bar{t}_k\|^2 \end{pmatrix},$$

где $\|\bar{t}_j\|$ — норма вектора $\bar{t}_j$; $(\bar{t}_i,\bar{t}_j)$ — скалярное произведение векторов $\bar{t}_i$ и $\bar{t}_j$ (матрица Грама). $\det A$ — определитель Грама векторов $\bar{t}_1, \bar{t}_2, ..., \bar{t}_k$. Он равен квадрату k-мерного объема параллелепипеда, построенного на этих векторах [6, § 5.1.5—5.1.7].

Вернемся к доказательству того, что $D\beta_i^* \xrightarrow[n\to\infty]{} 0$, $i = 1, 2, ..., k$.

$$D\beta_i^* \leqslant \sigma^2 trA^{-1} =$$

$$= \sigma^2(1/\lambda_1 + 1/\lambda_2 + ... + 1/\lambda_k) =$$

$$= \sigma^2 \cdot \frac{\lambda_2 \cdot ... \cdot \lambda_k + \lambda_1\lambda_3 \cdot ... \cdot \lambda_k + \lambda_1\lambda_2 \cdot ... \cdot \lambda_{k-1}}{\lambda_1 \cdot \lambda_2 \cdot ... \cdot \lambda_k} \leqslant$$

$$\leqslant \sigma^2 \cdot \frac{k \cdot \Delta_{k-1,\max}}{\det A},$$

где $\Delta_{k-1,\max}$ — максимальный минор порядка $(k-1)$, симметричный относительно главной диагонали (см. свойство 4 информационной матрицы) и равный квадрату объема $(k-1)$-мерного параллелепипеда, построенного на $(k-1)$ векторе из множества $\{\bar{t}_1, \bar{t}_2, ..., \bar{t}_k\}$ (кроме вектора $\bar{t}_e$). Отсюда

$$\det A = \Delta_{k-1,\max} \cdot \|\bar{t}_e\|^2 \cdot \sin\alpha,$$

где α — угол между вектором $\bar{t}_e$ и линейной оболочкой векторов, на которых построен $(k-1)$-мерный параллелепипед [6, § 5.1.7]; $\|\bar{t}_e\|^2 = a_{ee} \xrightarrow[n\to\infty]{} \infty$.

Поскольку $\det A \xrightarrow[n\to\infty]{} \infty$, существует $0 < \gamma < 1$ такое, что при любом n $|\sin\alpha| \geqslant 1 - \gamma$.

Итак, $D\beta_i^* \leqslant k\sigma^2 \times$

$$\times \frac{\Delta_{k-1,\max}}{\Delta_{k-1,\max} \cdot \|\bar{t}_e\|^2 \cdot \sin\alpha} \leqslant \frac{\sigma^2 k}{a_{ee} \cdot (1-\gamma)} \xrightarrow[n\to\infty]{} 0.$$

Состоятельность МНК-оценок коэффициентов множественной линейной регрессии доказана.

Литература

1. ***Вуколов Э. А.*** Основы статистического анализа. М.: Форум: ИНФРА-М, 2004. 462 с.: ил. (Профессиональное образование).
2. ***Чернова Н. И.*** Лекции по математической статистике. Новосибирск: НГУ, 2003. 179 с.
3. ***Bhatia R.*** Positive Definite Matrices. Princeton: Princeton University Press, 2015. 240 p. (Princeton Series in Applied Mathematics).
4. Сборник задач по математике для втузов [в 4 ч.] / Под ред. А. В. Ефимова, В. С. Поспелова. Ч. 1. М.: Изд-во Физико-математической литературы, 2003. 288 с.: ил.
5. ***Бугров Я. С., Никольский С. М.*** Элементы линейной алгебры и аналитической геометрии. 2-е изд., перераб. и доп. М.: Наука, 1984. 190 с.: ил. (Высшая математика).
6. ***Овсянников А. Я.*** Линейная алгебра. Екатеринбург: Изд-во Гуманит. ун-та, 2004. 293 с.

Гафарова Любовь Михайловна — старший преподаватель кафедры высшей математики № 2 (ВМ-2) МИЭТ. **E-mail: hm2@miet.ru**

Завьялова Ирина Геннадьевна — кандидат физико-математических наук, доцент кафедры ВМ-2 МИЭТ.
E-mail: irinazavialova@gmail.com

Мустафин Наиль Нухович — старший преподаватель кафедры ВМ-2 МИЭТ.
E-mail: hm2@miet.ru