УДК 330.43

ГЕОМЕТРИЯ СИСТЕМ ЛИНЕЙНЫХ РЕГРЕССИОННЫХ УРАВНЕНИЙ

© 2013 А.П. Котенко, М.Б. Букаренко

Самарский государственный аэрокосмический университет имени академика С.П.Королёва
(национальный исследовательский университет)

Поступила в редакцию 02.12.2013

Представлена геометрическая интерпретация метода наименьших квадратов, позволяющая классифицировать возможность идентификации параметров системы линейных регрессионных уравнений. Это даёт возможность в случаях неидентифицируемости или сверхидентифицируемости выдвинуть дополнительные условия, гарантирующие однозначную идентификацию параметров как отдельного уравнения, так и всей системы в целом. Экспоненциальный рост числа вариантов разбиения множества исследуемых факторов на эндогенные и экзогенные усложняет задачу отыскания регрессионных взаимосвязей. В этом случае геометрическая интерпретация системы регрессионных уравнений позволяет целенаправленно подобрать структурную форму модели, обеспечивающую значимость параметров всех уравнений.

Ключевые слова: система линейных регрессионных уравнений, структурная форма модели, приведённая форма модели, идентификация параметров модели.

Системы регрессионных уравнений (например, одномоментные системы эконометрических уравнений [1]) – полезный инструмент моделирования взаимосвязанных явлений. Представим геометрическую интерпретацию метода наименьших квадратов для описания идентифицируемости систем линейных регрессионных уравнений.

Рассмотрим систему линейных регрессионных уравнений

$$\begin{cases} \vec{y}_1 = a_{12}\vec{y}_2 + a_{13}\vec{y}_3 + \ldots + a_{1n}\vec{y}_n + \\ + b_{11}\vec{x}_1 + b_{12}\vec{x}_2 + \ldots + b_{1m}\vec{x}_m + \vec{\varepsilon}_1, \\ \vec{y}_2 = a_{21}\vec{y}_1 + a_{23}\vec{y}_3 + \ldots + a_{2n}\vec{y}_n + \\ + b_{21}\vec{x}_1 + b_{22}\vec{x}_2 + \ldots + b_{2m}\vec{x}_m + \vec{\varepsilon}_2, \\ \ldots \\ \vec{y}_n = a_{n1}\vec{y}_1 + a_{n2}\vec{y}_2 + \ldots + a_{n,n-1}\vec{y}_{n-1} + \\ + b_{n1}\vec{x}_1 + b_{n2}\vec{x}_2 + \ldots + b_{nm}\vec{x}_m + \vec{\varepsilon}_n, \end{cases} \qquad (1)$$

связывающую центрированные эндогенные переменные $\vec{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iN})^T \in \Re^N$, $1 \le i \le n$, экзогенные переменные $\vec{x}_j = (x_{j1}, x_{j2}, \ldots, x_{jN})^T \in \Re^N$, $1 \le j \le m$, и остатки $\vec{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iN})^T \in \Re^N$, $1 \le i \le n$; некоторые структурные коэффициенты a_{ij}, b_{ij} могут быть нулевыми. Ей соответствует приведённая система

$$\begin{cases} \tilde{y}_1 = \beta_{11}\vec{x}_1 + \beta_{12}\vec{x}_2 + \ldots + \beta_{1m}\vec{x}_m, \\ \tilde{y}_2 = \beta_{21}\vec{x}_1 + \beta_{22}\vec{x}_2 + \ldots + \beta_{2m}\vec{x}_m, \\ \ldots \\ \tilde{y}_n = \beta_{n1}\vec{x}_1 + \beta_{n2}\vec{x}_2 + \ldots + \beta_{nm}\vec{x}_m \end{cases} \qquad (2)$$

с регрессионными значениями $\tilde{y}_i = (\tilde{y}_{i1}, \tilde{y}_{i2}, \ldots, \tilde{y}_{iN})^T \in \Re^N$ и приведёнными МНК-коэффициентами $в_{ij}$, среди которых могут встретиться незначимые.

Вектора $\tilde{y}_i$ регрессионных значений эндогенных переменных получаются ортогональным проектированием векторов $\vec{y}_i$ наблюдений эндогенных переменных на линейную оболочку $L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) \subseteq \Re^N$ векторов экзогенных переменных:

$$\tilde{y}_i = \Pr(\vec{y}_i; L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m)) \in L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m),$$

$1 \le i \le n$.

Как правило $\vec{y}_i \ne \tilde{y}_i$, так как вероятность совпадения $\vec{y}_i = \tilde{y}_i$ близка к нулю при достаточно большом числе наблюдений N. Приведённые коэффициенты $в_{ij}$ являются решениями системы N линейных уравнений с $m \le N$ неизвестными

$$\begin{cases} \tilde{y}_{i1} = \beta_{i1}x_{11} + \beta_{i2}x_{21} + \ldots + \beta_{im}x_{m1}, \\ \tilde{y}_{i2} = \beta_{i1}x_{12} + \beta_{i2}x_{22} + \ldots + \beta_{im}x_{m2}, \\ \ldots \\ \tilde{y}_{iN} = \beta_{i1}x_{1N} + \beta_{i2}x_{2N} + \ldots + \beta_{im}x_{mN}. \end{cases} \qquad (3)$$

Котенко Андрей Петрович, кандидат физико-математических наук, доцент кафедры «Математические методы в экономике». E-mail: ako1959@mail.ru
Букаренко Максим Борисович, ассистент кафедры «Математические методы в экономике». E-mail: maxim.bukarenko@gmail.com

Её ранг $r<m$ при строгой мультиколлинеарности векторов регрессоров $\vec{x}_j \in \Re^N$ и система (3) имеет бесчисленное множество решений:

$$\forall i \in \overline{1,n} \Rightarrow \exists \widetilde{y}_i \in L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) \subset \Re^N,$$
$$\dim L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) = r < m.$$

Без строгой мультиколлинеарности $r=m$ и система (3) имеет единственное решение:

$$\forall i \in \overline{1,n} \Rightarrow \exists! \widetilde{y}_i \in L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) \subset \Re^N,$$
$$\dim L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) = r = m.$$

В дальнейшем предполагаем отсутствие строгой мультиколлинеарности, когда вектора экзогенных переменных $\vec{x}_j$ линейно независимы: $\dim L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) = m$. Этого можно добиться переходом к базису размерности $m_1 \in \overline{1,m}$ оболочки $L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m)$, исключив линейно зависимые вектора экзогенных переменных.

Заменяя в системе (1) выборочные значения $\vec{y}_i$ эндогенных переменных регрессионными значениями $\widetilde{y}_i$ из системы (2), получим в линейной оболочке $L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m)$ выборочных значений экзогенных переменных независимые линейные уравнения для идентификации структурных коэффициентов:

$$\begin{aligned}
\widetilde{y}_1 &= a_{12}\widetilde{y}_2 + a_{13}\widetilde{y}_3 + \ldots + a_{1n}\widetilde{y}_n + \\
&\quad + b_{11}\vec{x}_1 + b_{12}\vec{x}_2 + \ldots + b_{1m}\vec{x}_m, \\
\widetilde{y}_2 &= a_{21}\widetilde{y}_1 + a_{23}\widetilde{y}_3 + \ldots + a_{2n}\widetilde{y}_n + \\
&\quad + b_{21}\vec{x}_1 + b_{22}\vec{x}_2 + \ldots + b_{2m}\vec{x}_m, \\
&\ldots \\
\widetilde{y}_n &= a_{n1}\widetilde{y}_1 + a_{n2}\widetilde{y}_2 + \ldots + a_{n,n-1}\widetilde{y}_{n-1} + \\
&\quad + b_{n1}\vec{x}_1 + b_{n2}\vec{x}_2 + \ldots + b_{nm}\vec{x}_m.
\end{aligned} \tag{4}$$

Из (2), (3), (4) получим систему линейных уравнений для идентификации структурных коэффициентов i-го уравнения $\forall i \in \overline{1,n}$:

$$\widetilde{y}_i = \sum_{k=1}^{m} \beta_{ik}\vec{x}_k = \sum_{j=1;\, j\neq i}^{n} a_{ij}\widetilde{y}_j + \sum_{k=1}^{m} b_{ik}\vec{x}_k =$$

$$= \sum_{j=1;\, j\neq i}^{n} a_{ij}\left(\sum_{k=1}^{m} \beta_{jk}\vec{x}_k\right) + \sum_{k=1}^{m} b_{ik}\vec{x}_k \Leftrightarrow$$

$$\Leftrightarrow \beta_{i1} = b_{i1} + \sum_{j=1;\, j\neq i}^{n} \beta_{j1}a_{ij}$$

$$\beta_{i2} = b_{i2} + \sum_{j=1;\, j\neq i}^{n} \beta_{j2}a_{ij}, \ldots, \beta_{im} = b_{im} + \sum_{j=1;\, j\neq i}^{n} \beta_{jm}a_{ij}. \tag{5}$$

Для решения СЛАУ (5) рассмотрим вначале подсистему $s \geq 1$ уравнений

$$\beta_{i1} = b_{i1} + \sum_{j=1;\, j\neq i}^{n} \beta_{j1}a_{ij},$$
$$\beta_{i2} = b_{i2} + \sum_{j=1;\, j\neq i}^{n} \beta_{j2}a_{ij},$$
$$\ldots, \beta_{is} = b_{is} + \sum_{j=1;\, j\neq i}^{n} \beta_{js}a_{ij} \tag{6}$$

с фиксированными структурными коэффициентами b_{ik}: например, если экзогенная переменная $\vec{x}_k$ отсутствует в i-ом уравнении, то $b_{ik}=0$ и не требует идентификации.

Её решение $(a_{i1},\ldots,a_{i,i-1},a_{i,i+1},\ldots,a_{in})$ подставим в оставшиеся $m-s<m$ уравнений СЛАУ (5) и доопределим неизвестные структурные коэффициенты $(b_{i,s+1},b_{i,s+2},\ldots,b_{im})$:

$$b_{i,s+1} = \beta_{i,s+1} - \sum_{j=1;\, j\neq i}^{n} \beta_{j,s+1}a_{ij},$$

$$b_{i,s+2} = \beta_{i,s+2} - \sum_{j=1;\, j\neq i}^{n} \beta_{j,s+2}a_{ij},$$

$$\ldots, b_{im} = \beta_{im} - \sum_{j=1;\, j\neq i}^{n} \beta_{jm}a_{ij}.$$

Обозначим через $L(M_i)$ линейную оболочку объединения $M_i := N_i \cup K_i$ множества векторов регрессионных значений присутствующих в правой части i-го уравнения системы (4) эндогенных переменных

$$N_i := \left\{\widetilde{y}_j \in \Re^N : j \neq i, a_{ij} \neq 0\right\}; \tag{7}$$
$$0 \leq |N_i| \leq |M_i|,\ |N_i| \leq n \leq N;$$

и множества векторов выборочных значений экзогенных переменных

$$K_i := \left\{\vec{x}_k \in \Re^N : b_{ik} \neq 0\right\}; \tag{8}$$
$$0 \leq |K_i| \leq |M_i| = |N_i| + |K_i|,\ |K_i| \leq m \leq N;$$

из правой части этого уравнения: $M_i = N_i \cup K_i \subset L(M_i) \subseteq L(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m) \subseteq \Re^N, 1 \leq i \leq n$.

Идентифицируемость структурных коэффициентов i-го уравнения определяется принадлежностью его левой части $\widetilde{y}_i$ линейной оболочке $L(M_i)$ векторов правой части (7), (8) и их возможной линейной зависимостью.

Возможны следующие случаи:

1) если система M_i векторов (7), (8) линейно независима (при этом $|M_i| = |N_i| + |K_i| \le m \le N$) и вектор из левой части уравнения принадлежит её линейной оболочке (т.е. $\tilde{y}_i \in L(M_i)$), то уравнение точно идентифицируемо и его структурные коэффициенты алгебраически однозначно определяются косвенным МНК (5):

$$\tilde{y}_i = \Pr(\tilde{y}_i, L(\{\tilde{y}_j \in N_i\} \cup \{\vec{x}_k \in K_i\})) \in L(M_i);$$

2) если система M_i линейно независима и $\tilde{y}_i \notin L(M_i)$, то уравнение сверхидентифицируемо и структурные коэффициенты статистически однозначно определяются двухшаговым МНК:

$$\tilde{y}_i \ne \Pr(\tilde{y}_i, L(\{\tilde{y}_j \in N_i\} \cup \{\vec{x}_k \in K_i\})) \in L(M_i);$$

3) если система M_i линейно зависима ($|M_i| = |N_i| + |K_i| < m \le N$) и $\tilde{y}_i \in L(M_i)$, то уравнение неидентифицируемо и его структурные коэффициенты нельзя (ни алгебраически, ни статистически) однозначно выразить через приведённые, однако их можно алгебраически представить как бесконечное множество решений СЛАУ (5);

4) если система M_i линейно зависима и $\tilde{y}_i \notin L(M_i)$, то уравнение неидентифицируемо, однако ортогональную проекцию $\Pr(\tilde{y}_i, L(M_i))$ можно алгебраически представить как непустое бесконечное множество решений неоднородной системы (5), т.е. структурные коэффициенты этого уравнения статистически неоднозначно представимы через приведённые.

Рассмотрим геометрические свойства систем линейных регрессионных уравнений в указанных случаях подробнее.

В случае 1) структурные коэффициенты являются решениями СЛАУ с $|M_i| = |N_i| + |K_i| \ge 1$ неизвестными и $N \ge 2$ уравнениями, из которых $|M_i| \le N$ независимы:

$$\sum_{\tilde{y}_j \in N_i} a_{ij} \tilde{y}_j + \sum_{x_j \in K_i} b_{ij} \vec{x}_j = \tilde{y}_i; \qquad (9)$$

$\tilde{y}_i \in L(M_i)$,
$\dim L(M_i) = |M_i|$, $1 \le i \le n$.

Поскольку вектора $\tilde{y}_i \in \Re^N$ регрессионных значений эндогенных переменных являются линейными комбинациями линейно независимых векторов $\vec{x}_j \in \Re^N$, то ранг системы (9)

$$r = |M_i| = |N_i| + m - |K_i| \Leftrightarrow$$

$$\Leftrightarrow |N_i| + |K_i| = |N_i| + m - |K_i| \Leftrightarrow |K_i| = m - |K_i|.$$

В случае 2) векторное уравнение (9) не разрешимо алгебраически, так как его левая часть $\sum_{y_j \in N_i} a_{ij} \tilde{y}_j + \sum_{x_j \in K_i} b_{ij} \vec{x}_j \in L(M_i)$ в отличие от правой части $\tilde{y}_i \notin L(M_i)$.

Обычным МНК найдём ортогональную проекцию правой части на $L(M_i)$:

$$\sum_{y_j \in N_i} a_{ij} \tilde{y}_j + \sum_{x_j \in K_i} b_{ij} \vec{x}_j = \Pr(\tilde{y}_i, L(M_i)). \qquad (10)$$

Система (10) имеет единственное алгебраическое решение. Следовательно, найдено единственное статистическое решение системы (9).

В случае 3) перейдём от СЛАУ (9) с линейно зависимой системой M_i к СЛАУ с базисной подсистемой $M_i' \subset M_i$, $M_i' := N_i' \cup K_i'$, $N_i' \subseteq N_i$, $K_i' \subseteq K_i$:

$$\sum_{\tilde{y}_j \in N_i'} a_{ij} \tilde{y}_j + \sum_{x_j \in K_i'} b_{ij} \vec{x}_j = \tilde{y}_i; \qquad (11)$$

$\tilde{y}_i \in L(M_i')$,
$1 \le \dim L(M_i') = |M_i'| = |N_i'| + |K_i'| < |M_i|$,
$1 \le i \le n$.

Применяя алгоритм решения СЛАУ (5), получим единственное алгебраическое решение системы (11). Тогда система (9) имеет бесконечное множество решений.

В случае 4) перейдём от СЛАУ (11) с линейно зависимой системой M_i к СЛАУ с базисной подсистемой $M_i' \subset M_i$, $M_i' := N_i' \cup K_i'$, $N_i' \subseteq N_i$, $K_i' \subseteq K_i$:

$$\sum_{\tilde{y}_j \in N_i'} a_{ij} \tilde{y}_j + \sum_{x_j \in K_i'} b_{ij} \vec{x}_j = \Pr(\tilde{y}_i, L(M_i')); \qquad (12)$$

$\tilde{y}_i \in L(M_i')$,
$1 \le \dim L(M_i') = |M_i'| = |N_i'| + |K_i'| < |M_i|$,
$1 \le i \le n$.

Здесь ортогональная проекция $\Pr(\tilde{y}_i, L(M_i'))$ получается обычным МНК. Применяя алгоритм решения СЛАУ (5), получим единственное алгебраическое решение системы (12). Тогда система (9) имеет непустое (и даже бесконечное) множество решений.

Представленная геометрическая иллюстрация действия МНК и его модификаций позволяет исследовать идентифицируемость систем линейных регрессионных уравнений в случае дополнительных условий, связывающих структурные коэффициенты.

СПИСОК ЛИТЕРАТУРЫ

1. *Тихомиров Н.П., Дорохина Е.Ю.* Эконометрика: учебник. М.: Экзамен, 2003. 512 с.

GEOMETRY OF THE SYSTEMS OF THE LINEAR REGRESSION EQUATIONS

Samara State Aerospace University named after Academician S.P.Korolyov
(National Research University)

Is represented the geometric interpretation of the method of least squares, which makes it possible to classify the possibility of the identification of the parameters of the system of linear regression equations. This gives the possibility in the cases of nonidentifiability or superidentifiability to advance the additional conditions, which guarantee the single-valued identification of the parameters of both the separate equation, and the entire system as a whole. An exponential increase in the number of versions of the partition of many factors into the endogenous and predetermined being investigated complicates the task of finding the regression interrelations. In this case the geometric interpretation of the system of regression equations makes it possible to goal-directed select the structural form of model, which ensures the significance of the parameters of all equations.
Key words: system of linear regression equations, the structural form of model, the given form of model, the identification of the parameters of model.

Andrey Kotenko, Candidate of Physics and Mathematical Sciences, Associate Professor at the Mathematical methods in the Economy Department. E-mail: ako1959@mail.ru
Maxim Bukarenko, Assistant Lecturer at the Mathematical methods in the Economy Department.
E-mail: maxim.bukarenko@gmail.com