

УДК 519.237.7

ИДЕНТИФИЦИРУЕМОСТЬ МОДЕЛЕЙ СТРУКТУРНЫХ УРАВНЕНИЙ С ЛАТЕНТНЫМИ ПЕРЕМЕННЫМИ

С. В. Стафеев

*Институт прикладных математических исследований КарНЦ РАН,
ФИЦ «Карельский научный центр РАН», Петрозаводск, Россия*

В статье получены условия почти всюду локальной идентифицируемости для одного класса моделей структурных уравнений с латентными переменными.

Ключевые слова: структурные уравнения; латентные переменные; идентифицируемость.

S. V. Stafeev. IDENTIFIABILITY OF STRUCTURAL EQUATION MODELS WITH LATENT VARIABLES

In this paper the problem of identifiability of structural equation models with latent variables is considered. The conditions of the almost everywhere local identifiability for such models are obtained.

Key words: structural equations; latent variables; identifiability.

МОДЕЛЬ

Пусть $G = (V, E)$ – смешанный граф с множеством вершин $V = \{1, \dots, n\}$ и множеством ребер E . Множество ребер E графа G состоит из ориентированных ($i \rightarrow j$) и двуориентированных ($i \leftrightarrow j$) ребер.

Рассмотрим следующую, связанную с графом G , систему структурных уравнений с латентными переменными [2]:

$$X = B^t X + \Lambda^t H + \varepsilon, \quad (1)$$

где $X = (X_1, \dots, X_n)^t$ – вектор наблюдаемых случайных величин с матрицей ковариаций $\Sigma = (\sigma_{ij})$; $H = (H_1, \dots, H_k)^t$ – вектор независимых нормально распределенных латентных (скрытых) случайных величин с $\mathbf{M}(H_j) = 0$, и $\mathbf{M}(H_j)^2 = 1$; $\Lambda = (\lambda_{ij}) \in \mathbb{R}^{k \times n}$; $B = (b_{ij}) \in \mathbb{R}^{n \times n}$, причем $b_{ij} = 0$, если $i \rightarrow j \notin E$. Вектор остатков $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^t$ имеет нормальное

распределение с нулевым вектором математических ожиданий и неизвестной положительно определенной матрицей ковариаций $\Omega = (\omega_{ij})$, причем $\omega_{ij} = 0$, если $i \leftrightarrow j \notin E$. Мы будем предполагать, что векторы H и ε являются независимыми.

Таким образом, с помощью графа G постулируются некоторые нули в матрицах B и Ω . Вектор параметров $\theta = (\Lambda, B, \Omega)$ модели (1) состоит из матрицы Λ и ненулевых элементов матриц B и Ω . Пусть \mathbf{PD}_V – множество $n \times n$ симметричных положительно определенных матриц, \mathbf{B}_G – множество матриц, для которых матрица $I - B$ обратима. Определим параметрическое множество модели (1): $\Omega_G =$

$$\{\theta : \Lambda = (\lambda_{ij}) \in \mathbb{R}^{k \times n}, B \in \mathbf{B}_G, \Omega \in \mathbf{PD}_V\}.$$

Пусть $\theta \in \Omega_G$. Тогда матрица ковариаций наблюдаемых случайных величин имеет вид:

$$\Sigma = \Sigma(\theta) = (I - B)^{-t}(\Omega + \Lambda\Lambda^t)(I - B)^{-1}. \quad (2)$$

Пусть \mathbf{M}_G – множество всех матриц, допускающих разложение (2). Таким образом, граф G задает следующее семейство нормальных распределений:

$$\mathbb{N}_G = \{N(0, \Sigma) : \Sigma \in \mathbf{M}_G\}. \quad (3)$$

УСЛОВИЯ ИДЕНТИФИЦИРУЕМОСТИ

Одной из наиболее важных проблем, связанных с моделями, содержащими латентные переменные (в частности, с моделью (1)), является проблема параметрической идентифицируемости [4, 6]. Данная проблема заключается в ответе на вопрос о возможности однозначного определения неизвестных параметров модели по совместному распределению наблюдаемых случайных величин.

Рассмотрим вероятностно-статистическую модель $\mathcal{M}(\Theta)$, заданную семейством распределений $\{\mathbb{P}_\theta, \theta \in \Theta\}$.

Определение 1. [6] Модель $\mathcal{M}(\Theta)$ называется идентифицируемой, если любым различным $\theta \in \Theta$ и $\theta' \in \Theta$ соответствуют различные распределения \mathbb{P}_θ и $\mathbb{P}_{\theta'}$.

Модель $\mathcal{M}(\Theta)$ называется локально идентифицируемой, для любого $\theta \in \Theta$ найдется конечное число таких $\theta' \in \Theta$, $\theta' \notin \theta$, что $\mathbb{P}_{\theta'} = \mathbb{P}_\theta$.

Определение 2. [6] Модель $\mathcal{M}(\Theta)$ называется почти всюду (п.в.) идентифицируемой (п.в. локально идентифицируемой), если она является идентифицируемой (локально идентифицируемой) при $\theta \in \Theta' \subseteq \Theta$, а множество $\Theta \setminus \Theta'$ имеет меру нуль.

В нашем случае модель задается семейством распределений (3). Легко видеть, что матрица Λ модели (1) может быть определена только с точностью до ортогонального преобразования. Поэтому целесообразно считать, что модель (1) идентифицируема, если по матрице Σ матрица Λ определяется с точностью до ортогонального преобразования, а элементы матриц Ω и B определяются однозначно. Для того чтобы оставаться в рамках определений 1 и 2, будем считать матрицы Λ и Λ' различимыми, если не существует такой ортогональной матрицы Q , что $\Lambda' = Q\Lambda$. Таким образом, модель (1) называется идентифицируемой, если любым различным $\theta \in \Theta_G$ и $\theta' \in \Theta_G$ соответствуют различные матрицы ковариаций $\Sigma(\theta)$ и $\Sigma(\theta')$.

В работе [4] получены достаточные условия п.в. локальной параметрической идентифицируемости модели (1) в случае, когда G является простым ациклическим графом (т. е. гра-

фом, в котором любые две различные вершины могут быть соединены только одним ребром и он не содержит ориентированных циклов) и $k = 1$. В данной работе получены достаточные условия п.в. локальной идентифицируемости для произвольного k .

Пусть $\bar{G} = (V, \bar{E})$ – дополнительный к G граф. Образует граф $\mathcal{G} = (\mathbf{V}, \mathbf{E})$, где $\mathbf{V} = \{\mathbf{i} = (i_1, \dots, i_k), 1 \leq i_1 < i_2 < \dots < i_k \leq k\}$, а $\mathbf{E} = \{(\mathbf{i}, \mathbf{j}) | \mathbf{i} = (i_1, \dots, i_k), \mathbf{j} = (j_1, \dots, j_k), (i_s, j_l) \in \bar{E}, s, l = 1, \dots, k\}$. (Заметим, что граф \mathcal{G} использовался в работе [1] для решения вопроса об идентифицируемости модели факторного анализа с зависимыми остатками.)

Теорема 1. Пусть G – простой ациклический граф. Модель (1) будет п.в. локально идентифицируемой при $\theta \in \Theta_G$, если граф \mathcal{G} содержит подграф $\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$, каждая компонента связности которого содержит нечетный простой цикл и $\cup_{\mathbf{i} \in \mathbf{V}'} \mathbf{i} = V$.

Доказательство. Соотношение (2) задает отображение

$$\varphi : \Theta_G \rightarrow \mathbb{R}^{\frac{n(n+1)}{2}}. \quad (4)$$

Пусть s – число ребер графа G . Элементы матрицы Λ по матрице $\Lambda \Lambda^t$, максимальный ранг которой равен k , восстанавливаются с точностью до ортогонального преобразования, поэтому матрица Якоби $J(\theta)$ отображения φ имеет ранг, меньший или равный $R = s + n + nk - \frac{1}{2}k(k-1)$. Нетрудно показать, что модель будет локально идентифицируемой в некоторой окрестности $\theta^* \in \Theta$, если ранг матрицы $J(\theta^*)$ равен R .

Так как элементы матрицы $J(\theta)$ являются полиномами относительно параметров модели, то ее ранг будет почти всюду при $\theta \in \Theta$ максимален [4, 6]. (Максимальный ранг матрицы $J(\theta)$ называют размерностью модели.) Поэтому, если мы покажем, что существует такое $\theta^* \in \Theta$, что ранг матрицы Якоби $J(\theta^*)$ равен R , теорема будет доказана.

Произведем невырожденную замену параметров:

$$A = (I - B)^{-t} \Lambda. \quad (5)$$

Соотношение (2) принимает следующий вид:

$$\Sigma = (I - B)^{-t} \Omega (I - B)^{-1} + AA^t. \quad (6)$$

Очевидно, что ранг матриц Якоби отображений, заданных соотношениями (2) и (6), совпадает. Далее находим матрицу Якоби отображения (6).

Используя свойства матричных производных, мы получаем:

- для $i = 1, \dots, k, j = 1, \dots, n$

$$(\sigma_{sl})'_{a_{ij}} = \begin{cases} 2a_{ij}, & \text{если } s = l = i; \\ a_{sj}, & \text{если } l = i, s \neq i; \\ a_{lj}, & \text{если } l \neq i, s = i; \\ 0, & \text{в противном случае;} \end{cases}$$

- для $i \leftrightarrow j \in E$

$$\Sigma'_{\omega_{ij}} = (I - B)^{-1}(\Omega'_{\omega_{ij}})(I - B)^{-t};$$

- для $i \leftarrow j \in E$

$$\begin{aligned} \Sigma'_{b_{ij}} &= [(I - B)^{-1}]'_{b_{ij}}(\Omega)(I - B)^{-t} \\ &+ (I - B)^{-1}(\Omega)[(I - B)^{-t}]'_{b_{ij}} \\ &- (I - B)^{-1}(I - B)'_{b_{ij}}(I - B)^{-1}(\Omega)(I - B)^{-t} \\ &- (I - B)^{-1}(\Omega)(I - B)^{-t}(I - B^t)'_{b_{ij}}(I - B)^{-t} \\ &= (I - B)^{-1}B'_{b_{ij}}\Sigma + \Sigma(B^t)'_{b_{ij}}(I - B)^{-t}. \end{aligned}$$

Выберем $\theta^* \in \Theta$, при котором: $b_{ij} = 0$, $i \leftarrow j \in E$, $\omega_{ij} = 0$, $i \leftrightarrow j \in E$, $\omega_{jj} = 1$, $j = 1, \dots, n$. Для $r, l = 1, \dots, n$, мы получаем:

- для $i \leftrightarrow j \in E$

$$(\sigma_{rl})'_{\omega_{ij}} = \begin{cases} 1, & \text{если } i = r, j = l \text{ или} \\ & i = l, j = r, \\ 0, & \text{в противном случае;} \end{cases}$$

- для $i \leftarrow j \in E$

$$(\sigma_{rl})'_{b_{ij}} = \begin{cases} 1, & \text{если } i = r, j = l \text{ или} \\ & i = l, j = k, \\ 0, & \text{в противном случае.} \end{cases}$$

Пусть $C = AA^t$ и $E_U = \{i - j : (i, j) \in E, (i, i), i = 1, \dots, n\}$. Обозначим:

$$\mathbf{c}'_{ij} = \frac{\partial c_{ij}}{\partial a}, \sigma'_{ij} = \left\{ \frac{\partial \sigma_{ij}}{\partial \omega}, \frac{\partial \sigma_{ij}}{\partial b} \right\}.$$

Образуем матрицы:

$$M_{11} = (\mathbf{c}'_{ij})_{i-j \in E_U}, M_{21} = (\mathbf{c}'_{ij})_{i-j \in \bar{E}},$$

$$M_{12} = (\sigma'_{ij})_{i-j \in E_U}, M_{22} = (\sigma'_{ij})_{i-j \in \bar{E}}.$$

Легко видеть, что матрица $J(\theta^*)$ с помощью перестановки строк и столбцов может быть приведена к виду

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

где матрица M_{12} – единичная матрица, а M_{22} – матрица, все элементы которой равны нулю.

Очевидно матрица $J(\theta^*)$ имеет ранг R , если ранг матрицы M_{21} равен $nk - \frac{1}{2}k(k-1)$. Предположим, что любые k строк матрицы A линейно независимы. (Заметим, что это условие выполняется п.в. при $A \in \mathbb{R}^{n \times k}$.) Применяя результат работы [1], получаем, что если граф \mathcal{G} содержит компоненту связности

$\mathcal{G}' = (\mathbf{V}', \mathbf{E}')$ с нечетным простым циклом и $\cup_{i \in \mathbf{V}'} i = V$, то все элементы матрицы C однозначно восстанавливаются по множеству элементов $\{c_{ij}, i - j \in \bar{E}\}$. Отсюда следует, что ранг матрицы M_{21} п.в. при $A \in \mathbb{R}^{n \times k}$ равен $nk - \frac{1}{2}k(k-1)$. Теорема доказана.

Предположим, что \bar{G} – полный граф, содержащий не менее $2k+1$ вершин. Тогда очевидно, что граф \mathcal{G} является связным. Граф \bar{G} также содержит нечетный цикл длины $2k+1$:

$$\mathbf{i}_1 - \dots - \mathbf{i}_{2k+1} - \mathbf{i}_1, \text{ где} \\ \mathbf{i}_l = \{(2(k-m) + l + 3) \bmod (2k+1), \\ l = 1, \dots, 2k+1\}.$$

Таким образом, в этом случае условия теоремы выполнены. Например, для случая $k=3$ цикл длины 7 будет:

$$(1, 6, 4) - (2, 7, 5) - (3, 1, 6) - (4, 2, 7) - (5, 3, 1) - (6, 4, 2) - (7, 5, 3) - (1, 6, 4).$$

Верна следующая теорема.

Теорема 2. Пусть $n \geq 4k+1$. Если G является лесом, в котором нет вершины с не менее чем с $n-k$ смежными вершинами, то модель (1) п.в. локально идентифицируема при $\theta \in \Theta$.

Доказательство. Ввиду того что лес является двудольным графом, множество вершин графа G можно представить в виде объединения двух непересекающихся подмножеств V^1 и V^2 несоединенных между собой вершин. Пусть (для определенности) множество V^1 содержит не менее $2k+1$ вершин. Индуцированный подграф графа \bar{G} с множеством вершин V^1 является связным и содержит нечетный цикл.

Если любая вершина множества V^2 дополнительного графа \bar{G} соединена ребрами с более чем k вершинами множества V^1 , то нетрудно показать, что условия теоремы 1 выполнены. Предположим, что некоторая вершина $i_1 \in V^2$ соединена ребрами с менее чем k вершинами множества V^1 . Тогда, как легко видеть, любая вершина множества $V^2 \setminus i_1$ соединена с не менее чем k вершинами множества V^1 , так как в противном случае граф G имел бы неориентированные циклы. Таким образом, и в этом случае условия теоремы 1 выполнены. Теорема доказана.

СТРУКТУРНЫЕ УРАВНЕНИЯ И УСЛОВНЫЕ НЕЗАВИСИМОСТИ

С помощью смешанного графа G удобно представлять условные независимости среди случайных величин множества $\{X_1, \dots, X_n\}$. Введем необходимые определения [6].

Определение 3. Простая цепь, принадлежащая графу G , содержит тупиковую вершину k , если она содержит один из подграфов:
 $i_1 \rightarrow k \leftarrow i_2$, $i_1 \rightarrow k \leftrightarrow i_2$, $i_1 \leftrightarrow k \leftarrow i_2$,
 $i_1 \leftrightarrow k \leftrightarrow i_2$.

Определение 4. Вершины i и j называются d -связанными множеством вершин $S \subseteq V$, если граф G содержит такую простую цепь, соединяющую i и j , у которой все тупиковые вершины принадлежат S , а не тупиковые не принадлежат. Вершины i и j называются d -отделимыми множеством вершин S , если они не являются d -связанными.

Из d -отделимости вершин i и j множеством вершин S следует условная независимость соответствующих случайных величин X_i и X_j при данном $X_S = \{X_s, s \in S\}$ [6].

Определим граф $\hat{G} = (V, \hat{E})$, где множество ребер \hat{E} состоит из всех ребер множества \bar{E} , кроме таких ребер $i - j \in \bar{E}$, для которых найдется простая цепь графа G , которая соединяет i и j и содержит только тупиковые вершины. Очевидно, что i и j будут d -отделимы множеством $V \setminus \{i, j\}$, при $i - j \in \hat{E}$. В случае нормального распределения с матрицей ковариаций Σ это будет равносильно равенству нулю соответствующих элементов матрицы Σ^{-1} .

Определим граф $G = (\mathbf{V}, E)$, где $\mathbf{V} = \{\mathbf{i} = (i_1, \dots, i_k), 1 \leq i_1 < i_2 < \dots < i_k \leq k\}$, а $E = \{(\mathbf{i}, \mathbf{j}) | \mathbf{i} = (i_1, \dots, i_k), \mathbf{j} = (j_1, \dots, j_k), (i_s, j_l) \in \hat{E}, s, l = 1, \dots, k\}$.

Очевидно, что граф G является подграфом графа \mathcal{G} .

В следующей теореме сформулированы достаточные условия п.в. идентифицируемости модели (1).

Теорема 3. Пусть G – простой ациклический граф. Модель (1) будет п.в. идентифицируемой при $\theta \in \Theta$, если граф G содержит компоненту связности $G' = ((\mathbf{V}', E'))$ с нечетным простым циклом и $\cup_{\mathbf{i} \in \mathbf{V}'} \mathbf{i} = V$.

Доказательство. Пусть $\Sigma_G = (I - B)^{-t} \Omega (I - B)^{-1}$. Из определения графа $\hat{G} = (V, \hat{E})$ следует, что для элементов матрицы $\Sigma_G^{-1} = (\sigma_G^{ij})$ выполнено следующее условие.

$$\sigma_G^{ij} = 0, \quad \text{если } i - j \in \hat{E}. \quad (7)$$

Легко видеть, что из (6) следует

$$\Sigma^{-1} = \Sigma_G^{-1} - \Sigma_G^{-1} A (I_k + A^t \Sigma_G^{-1} A) A^t \Sigma_G^{-1}. \quad (8)$$

Нетрудно показать, что каждый $(k+1) \times (k+1)$ минор матрицы $D = \Sigma_G^{-1} A (I_k + A^t \Sigma_G^{-1} A) A^t \Sigma_G^{-1}$ равен нулю, а при предположении, что любые k строк матрицы A линейно независимы, каждый $k \times k$ минор матрицы D не равен нулю. Теперь, используя результат работы [1] и соотношение (8), можно показать, что множество элементов матрицы D п.в. однозначно выражается через множество элементов $\{d_{ij}, i - j \in \hat{E}\}$. Таким образом, по матрице Σ мы можем однозначно восстановить элементы матрицы Σ_G . В [3] показано, что по Σ_G параметры $\{B, \Omega\}$ п.в. определяются однозначно. Используя (6), мы однозначно определяем AA^t , а значит, (с помощью (5)) и матрицу $\Lambda \Lambda^t$. Таким образом, матрица Λ определена с точностью до ортогонального преобразования. Теорема доказана.

Финансовое обеспечение исследований осуществлялось из средств федерального бюджета на выполнение государственного задания КарНЦ РАН (Институт прикладных математических исследований КарНЦ РАН).

ЛИТЕРАТУРА

1. Стафеев С. В. Об условиях глобальной идентифицируемости для моделей факторного анализа // Труды КарНЦ РАН. 2011. № 2. С. 111–114.
2. Boollen K. A. Structural equations with latent variables. New York: John Wiley and Sons, 1990.
3. Carlos B., Pearl J. A new identification condition for recursive models with correlated errors // Struct. Equ. Model. 2002. Vol. 9, no. 4. P. 459–474.
4. Drton M. Algebraic problems in structural equation modeling // The 50th Anniv. of Grobner Bases, Mathematical Society of Japan. Tokyo, Japan, 2018. P. 35–86. doi: 10.2969/aspm/07710035.
5. Leung D., Drton M., Hara H. Identifiability of directed Gaussian graphical models with one latent source // Electron. J. Statist. 2016. Vol. 10, no. 1. P. 394–422. doi: 10.1214/16-EJS1111.
6. Maathuis M., Drton M., Lauritzen S., Wainwright M. (Eds) Handbook of Graphical Models. Chapman & Hall/CRC., 2018. 536 p.

Поступила в редакцию 17.05.2019

REFERENCES

1. *Stafeev S. V.* Ob usloviyakh global'noi identifikatsionnosti dlya modeli faktornogo analiza [On global identifiability conditions of factor analysis models]. *Trudy KarNTs RAN* [Trans. KarRC RAS]. 2011. No. 2. P. 111–114.
2. *Boollen K. A.* Structural equations with latent variables. New York: John Wiley and Sons, 1990.
3. *Carlos B., Pearl J.* A new identification condition for recursive models with correlated errors. *Struct. Equ. Model.* 2002. Vol. 9, no. 4. P. 459–474.
4. *Drton M.* Algebraic problems in structural equation modeling. *The 50th Anniv. of Grobner Bases, Mathematical Society of Japan.* Tokyo, Japan, 2018. P. 35–86. doi: 10.2969/aspm/07710035.
5. *Leung D., Drton M., Hara H.* Identifiability of directed Gaussian graphical models with one latent source. *Electron. J. Statist.* 2016. Vol. 10, no. 1. P. 394–422. doi: 10.1214/16-EJS1111.
6. *Maathuis M., Drton M., Lauritzen S., Wainwright M. (Eds)* Handbook of Graphical Models. Chapman & Hall/CRC., 2018. 536 p.

Received May 17, 2019

СВЕДЕНИЯ ОБ АВТОРЕ:

Стафеев Сергей Вячеславович
младший научный сотрудник, к. ф.-м. н.
Институт прикладных математических исследований
КарНЦ РАН, Федеральный исследовательский центр
«Карельский научный центр РАН»
ул. Пушкинская, 11, Петрозаводск,
Республика Карелия, Россия, 185910
эл. почта: stafeev@krc.karelia.ru
тел.: (8142) 763370

CONTRIBUTOR:

Stafeev, Sergei
Institute of Applied Mathematical Research,
Karelian Research Centre,
Russian Academy of Sciences
11 Pushkinskaya St., 185910 Petrozavodsk,
Karelia, Russia
e-mail: stafeev@krc.karelia.ru
tel.: (8142) 763370