DOI: 10.25586/RNU.V9187.20.01.P.017 УДК 519.862.6

М.П. Базилевский, С.И. Носков

ОЦЕНИВАНИЕ ИНДЕКСНЫХ МОДЕЛЕЙ РЕГРЕССИИ С ПОМОЩЬЮ МЕТОДА НАИМЕНЬШИХ МОДУЛЕЙ

В схеме регрессионного моделирования ключевым этапом является выбор спецификации модели, т.е. математической формы связи между переменными. Предложена новая спецификация регрессионных моделей – индексная регрессия, являющаяся обобщением производственной функции Леонтьева. Отмечено, что при построении индексных регрессий, наряду со статистической информацией, требуется привлечение еще и экспертной информации о ретроспективном периоде, что относит их к классу экспертно-статистических регрессионных моделей. Задача оценивания неизвестных параметров индексной регрессии по методу наименьших модулей сведена к задаче частично-булевого линейного программирования. С использованием данных Хальда рассмотрен пример построения индексных регрессий.
Ключевые слова: индексная регрессия, производственная функция Леонтьева, метод наименьших модулей, задача частично-булевого линейного программирования.

M.P. Bazilevskiy, S.I. Noskov

ESTIMATION OF INDEX REGRESSION MODELS USING THE LEAST ABSOLUTE DEVIATIONS

In the regression modeling scheme, the key step is the model specification selection, i.e. the mathematical form of the relationship between variables. In this work, a new specification of regression models – index regression, is proposed, which is a generalization of the Leontief production function. It is noted that when constructing index regressions, along with statistical information, it is also necessary to attract expert information about the retrospective period, which classifies them as expert-statistical regression models. The task of estimating the unknown parameters of index regression using the least absolute deviations is reduced to the problem of partial-Boolean linear programming. Using Hald's data, an example of constructing index regressions is considered.
Keywords: index regression, Leontief production function, least absolute deviations, partial-Boolean linear programming problem.

Введение

Одним из основных инструментов интеллектуального анализа данных является регрессионный анализ [8; 9; 10]. Его использование приводит к построению математических моделей влияния одной или нескольких объясняющих переменных на объясняемую переменную. В схеме регрессионного моделирования ключевым этапом является выбор спецификации модели, т.е. математической формы связи между переменными. К настоящему времени разработан значительный арсенал ставших уже классическими способов описания взаимосвязей между моделируемыми показателями [7]. В эконометрике при моделировании социально-экономических явлений особое внимание традиционно уделяется вопросам построения производственных функций [4]. В работах [1; 3; 5] рассмотрены методы оценивания производственных функций Леонтьева. Вместе с тем на сегодняшний день появляются новые формы связи между переменными. Так, в работе [6] с использованием индексных преобразований предложено обобщение производственной функции Леонтьева. Целью данной статьи является разработка математического аппарата для оценивания таких моделей с помощью метода наименьших модулей.

Оценивание индексных моделей регрессии

Рассмотрим способ преобразования произвольной матрицы в вектор – индексное преобразование [Там же]. Пусть дана матрица

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}.$$

Упорядочим элементы каждой строки матрицы A по возрастанию. Тогда она примет вид

$$A^{sort} = \begin{bmatrix} a_{1,c_{11}} & a_{1,c_{12}} & \dots & a_{1,c_{1m}} \\ a_{2,c_{21}} & a_{2,c_{22}} & \dots & a_{2,c_{2m}} \\ \dots & \dots & \dots & \dots \\ a_{n,c_{n1}} & a_{n,c_{n2}} & \dots & a_{n,c_{nm}} \end{bmatrix},$$

где c_{ij}, $i=\overline{1,n}$, $j=\overline{1,m}$ – элементы индексной матрицы $C_{n\times m}$. Элемент c_{ij} – это порядковый номер столбца, который занял j-ю позицию при упорядочивании i-й строки матрицы A по возрастанию.

Пусть задан индексный вектор

$$G = \begin{bmatrix} g_1 & g_2 & \dots & g_n \end{bmatrix},$$

где g_i, $i=\overline{1,n}$ – порядковый номер столбца матрицы A^{sort} для i-го наблюдения.

Тогда с использованием вектора G из матрицы A^{sort} сформируем вектор

$$B = \begin{bmatrix} a_{1,c_{1,g_1}} & a_{2,c_{2,g_2}} & \dots & a_{n,c_{n,g_n}} \end{bmatrix}.$$

Вектор B – индексное преобразование матрицы A по индексному вектору G, которое обозначается

$$B = \operatorname{ind}_G(A).$$

С использованием индексного преобразования в работе [6] сформулирована индексная модель регрессии:

$$y_i = \operatorname{ind}_G\left\{\alpha_1 x_{i1}, \alpha_2 x_{i2}, \dots, \alpha_m x_{im}\right\} + \varepsilon_i,\ i=\overline{1,n}, \tag{1}$$

где y_i, $i=\overline{1,n}$ – наблюдаемые значения объясняемой (выходной) переменной y; x_{ij}, $i=\overline{1,n}$, $j=\overline{1,m}$ – наблюдаемые значения объясняющих (входных) переменных x_1, x_2, …, x_m; ε_i, $i=\overline{1,n}$ – ошибки аппроксимации; α_1, α_2, …, α_m – неизвестные параметры; G – индексный вектор.

Без потери общности будем предполагать, что переменные модели (1) неотрицательны.

Как видно, при построении индексных регрессий (1), наряду со статистической информацией (выборкой), требуется привлечение еще и экспертной информации о ретроспективном периоде, что относит их к классу экспертно-статистических регрессионных моделей. Такая информация формируется посредством задания экспертом – специалистом в данной предметной области вектора G, произвольная компонента g_i которого указывает на независимую переменную из набора $x_1, x_2, \ldots, x_m$, которая должна «срабатывать» в i-м наблюдении при оценивании параметров индексной регрессии (1).

По аналогии с работами [3; 5] приведем способ точного оценивания неизвестных параметров индексной регрессии (1) с использованием метода наименьших модулей, приводящего к задаче

$$J(\alpha_1, \alpha_2, \ldots, \alpha_m) = \sum_{i=1}^{n} |\varepsilon_i| \to \min . \tag{2}$$

Для этого введем в рассмотрение расчетные значения объясняемой переменной z_i:

$$z_i = \mathrm{ind}_G \left\{ \alpha_1 x_{i1}, \alpha_2 x_{i2}, \ldots, \alpha_m x_{im} \right\},\ i = \overline{1,n}, \tag{3}$$

после чего регрессия (1) представима в виде

$$y_i = z_i + \varepsilon_i,\ i = \overline{1,n}.$$

Введем в рассмотрение переменные u_i, v_i, $i = \overline{1,n}$ по правилу

$$u_i = \begin{cases} y_i - z_i, & \text{если } y_i > z_i, \\ 0 & \text{в противном случае,} \end{cases} \quad v_i = \begin{cases} z_i - y_i, & \text{если } z_i > y_i, \\ 0 & \text{в противном случае,} \end{cases}$$

Легко видеть, что имеют место тождества

$$z_i + u_i - v_i = y_i,\ i = \overline{1,n}. \tag{4}$$

Как следует из выражения (3), для любого j расчетное значение объясняемой переменной удовлетворяет либо $z_i \le \alpha_j x_{ij}$, либо $z_i \ge \alpha_j x_{ij}$, $i = \overline{1,n}$. Для учета этого обстоятельства введем mn булевых переменных σ_{ij}, $i = \overline{1,n}$, $j = \overline{1,m}$ по правилу

$$\sigma_{ij} = \begin{cases} 0, \text{ если } z_i \le \alpha_j x_{ij}, \\ 1, \text{ если } z_i \ge \alpha_j x_{ij}. \end{cases}$$

Тогда справедливы следующие ограничения:

$$-M\sigma_{ij} \le \alpha_j x_{ij} - z_i \le M(1 - \sigma_{ij}),\ i = \overline{1,n},\ j = \overline{1,m}, \tag{5}$$

где M – заранее выбранное большое положительное число.

Пусть задан индексный вектор $G = \begin{bmatrix} g_1 & g_2 & \ldots & g_n \end{bmatrix}$, причем $0 \le g_i \le m$, $i = \overline{1,n}$. Тогда сформируем следующие ограничения:

$$\sum_{j=1}^{m} \sigma_{ij} = g_i,\ i = \overline{1,n}. \tag{6}$$

Для каждого i по крайней мере одно из неравенств (5) должно обращаться в строгое равенство $\alpha_j x_{ij} - z_i = 0$. Для достижения этого требования введем еще mn булевых переменных δ_{ij}, $i = \overline{1,n}$, $j = \overline{1,m}$ и сформируем ограничения

$$-M(1 - \delta_{ij}) - M(1 - \sigma_{ij}) \le \alpha_j x_{ij} - z_i \le M(1 - \delta_{ij}) + M\sigma_{ij},\ i = \overline{1,n},\ j = \overline{1,m}, \tag{7}$$

$$\sum_{j=1}^{m} \delta_{ij} = 1,\ i = \overline{1,n}. \tag{8}$$

Из задания переменных u_i и v_i, $i=\overline{1,n}$ следуют равенства

$$|\varepsilon_i|=u_i+v_i,\ u_i v_i=0,$$

позволяющие представить функционал (2) в виде

$$J\left(\alpha_1,\alpha_2,...,\alpha_m\right)=\sum_{i=1}^{n}\left(u_i+v_i\right)\to\min. \qquad (9)$$

Таким образом, задача (2) поиска значений неизвестных параметров α_j, $j=\overline{1,m}$ индексной регрессии (1) по методу наименьших модулей свелась к задаче частично-булевого линейного программирования (4)–(9) с $2mn+3n+m$ переменными и $4mn+3n$ ограничениями.

Пример построения индексной регрессии

Для построения индексной регрессии были использованы данные Хальда из монографии [8] по пяти переменным, русскоязычное описание которых приведено в работе [2]:

y – тепло, выделяющееся при производстве цемента, кал/г;

x_j, $j=\overline{1,4}$ – переменные, характеризующие содержание четырех веществ в клинкере, %.

Объем выборки составляет 13 наблюдений.

Оценивание индексных регрессий осуществлялось с использованием пакета решения задач математического программирования LPSolve. Большое положительное число M = 1 000 000.

Оцененная для индексного вектора $G_0=\left(0,0,0,0,0,0,0,0,0,0,0,0,0\right)$ индексная регрессия

$$\tilde{y}=\mathrm{ind}_{G_0}\left\{72{,}5x_1;\ 2{,}8258x_2;\ 13{,}0375x_3;\ 17{,}1167x_4\right\}. \qquad (10)$$

Для регрессии (10) сумма моделей остатков $J_0=\sum_{i=1}^{13}|e_i|=168{,}729$.

Стоит заметить, что модель (10) также можно записать в виде

$$\tilde{y}=\min\left\{72{,}5x_1;\ 2{,}8258x_2;\ 13{,}0375x_3;\ 17{,}1167x_4\right\},$$

т.е. она представляет собой производственную функцию Леонтьева.

Оцененная для $G_1=\left(1,1,1,1,1,1,1,1,1,1,1,1,1\right)$ индексная регрессия:

$$\tilde{y}=\mathrm{ind}_{G_1}\left\{46{,}55x_1;\ 2{,}4659x_2;\ 13{,}0375x_3;\ 17{,}1167x_4\right\}; \qquad (11)$$

для $G_2=\left(2,2,2,2,2,2,2,2,2,2,2,2,2\right)$:

$$\tilde{y}=\mathrm{ind}_{G_2}\left\{74{,}3x_1;\ 1{,}6088x_2;\ 12{,}5888x_3;\ 4{,}4577x_4\right\}; \qquad (12)$$

для $G_3=\left(3,3,3,3,3,3,3,3,3,3,3,3,3\right)$:

$$\tilde{y}=\mathrm{ind}_{G_3}\left\{5{,}519x_1;\ 1{,}7241x_2;\ 13{,}0375x_3;\ 1{,}6477x_4\right\}. \qquad (13)$$

Суммы модулей ошибок для моделей (11)–(13) $J_1=67{,}951$, $J_2=63{,}804$, $J_3=76{,}334$.

Оцененная для $G_4=\left(4,4,4,4,4,4,4,4,4,4,4,4,4\right)$ индексная регрессия:

$$\tilde{y}=\mathrm{ind}_{G_4}\left\{5{,}519x_1;\ 1{,}7241x_2;\ 3{,}6435x_3;\ 1{,}4288x_4\right\}. \qquad (14)$$

Для регрессии (14) сумма модулей остатков $J_4=91{,}742$.

Снова заметим, что модель (14) можно записать в виде

$$\tilde{y} = \max\left\{5{,}519x_1;\ 1{,}7241x_2;\ 3{,}6435x_3;\ 1{,}4288x_4\right\}.$$

Таким образом, лучшей индексной регрессией по величине суммы модулей остатков оказалась модель (12), для которой индексный вектор $G_2 = (2,2,2,2,2,2,2,2,2,2,2,2,2)$.

В таблице 1 приведена подробная информация об остатках индексных регрессий (10)–(14).

Таблица 1

Остатки индексных регрессий

Остаток	Индексный вектор				
	G_0	G_1	G_2	G_3	G_4
e_1	5,029	0,275	2,9667	0,275	–7,2308
e_2	1,8	2,7872	0	–11,3818	0
e_3	0	0	3,5889	0	7,7519
e_4	0	11,1553	–13,1111	10,1568	20,4442
e_5	17,675	17,675	12,2412	6,2481	6,2481
e_6	–8,1375	–8,1375	–4,1	–8,1375	14,3759
e_7	0	0	–11,5265	–19,7093	–19,7093
e_8	0	–3,9447	–1,8	0	–7,6565
e_9	–51,9	0	–4,9692	0	0
e_{10}	63,75	0	0	0	0
e_{11}	11,3	–14,8383	9,5	14,8370	0
e_{12}	–4,0375	–4,0375	0	–0,4889	–0,4889
e_{13}	5,1	5,1	0	5,1	–7,8370

Как видно по таблице 1, для каждой из индексных регрессий (10)–(14) ровно 4 остатка оказались нулевыми.

В таблице 2 приведена подробная информация о том, какая переменная из набора x_1, x_2, x_3, x_4 «сработала» в i-м наблюдении при оценивании индексных регрессий (10)–(14).

Таблица 2

Переменные, «сработавшие» в *i*-м наблюдении

Номер наблюдения	Индексный вектор				
	G_0	G_1	G_2	G_3	G_4
1	2	3	3	3	4
2	1	2	1	4	4
3	3	3	3	3	2
4	2	2	3	4	4
5	3	3	2	2	2
6	3	3	3	3	2
7	4	4	2	2	2
8	1	2	1	4	3
9	1	1	4	2	2

Окончание табл. 2

Номер наблюдения	**Индексный вектор**				
	G_0	G_1	G_2	G_3	G_4
10	3	2	4	1	1
11	1	2	1	2	3
12	3	3	3	2	2
13	3	3	2	3	2

Как следует из таблицы 2, для каждой из индексных регрессий (10)–(14) были задействованы и «срабатывали» абсолютно все переменные из набора x_1, x_2, x_3, x_4.

Заключение

В работе рассмотрено индексное преобразование матриц, на основе которого предложено обобщение производственной функции Леонтьева – индексная регрессия. Задача оценивания неизвестных параметров индексной регрессии по методу наименьших модулей сведена к задаче частично-булевого линейного программирования. С использованием данных Хальда рассмотрен пример построения индексных регрессий.

Литература

1. *Базилевский М.П.* МНК-оценивание параметров специфицированных на основе функций Леонтьева двухфакторных моделей регрессии // Южно-Сибирский научный вестник. 2019. № 2 (26). С. 66–70.
2. *Ершов Э.Б.* Выбор регрессии, максимизирующий несмещенную оценку коэффициента детерминации // Прикладная эконометрика. 2008. № 4 (12). С. 71–83.
3. *Иванова Н.К., Лебедева С.А., Носков С.И.* Идентификация параметров некоторых негладких регрессий // Информационные технологии и проблемы математического моделирования сложных систем. 2016. Вып. 17. С. 111–114.
4. *Клейнер Г.Б.* Производственные функции: Теория, методы, применение. М.: Финансы и статистика, 1986. 239 с.
5. *Носков С.И.* Оценивание параметров аппроксимирующей функции с постоянными пропорциями // Современные технологии. Системный анализ. Моделирование. 2013. № 2 (38). С. 135–136.
6. *Носков С.И., Базилевский М.П.* Об индексных преобразованиях матриц при построении регрессионных моделей // Информационные технологии и математическое моделирование в управлении сложными системами. 2019. № 3. С. 11–16.
7. *Носков С.И., Базилевский М.П.* Построение регрессионных моделей с использованием аппарата линейно-булевого программирования. Иркутск: ИрГУПС, 2018. 176 с.
8. *Draper N.R., Smith H.* Applied Regression Analysis. Hoboken: John Wiley & Sons, 1998. 736 p.
9. *Harrell Jr., Frank E.* Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. [S. l.]: Springer Series in Statistics, 2015. 582 p.
10. *Kuhn M., Johnson K.* Applied Predictive Modeling. Berlin: Springer, 2018. 600 p.

Literatura

1. *Bazilevskij M.P.* MNK-otsenivanie parametrov spetsifitsirovannykh na osnove funktsij Leont'eva dvukhfaktornykh modelej regressii // Yuzhno-Sibirskij nauchnyj vestnik. 2019. № 2 (26). S. 66–70.

2. *Ershov E.B.* Vybor regressii, maksimiziruyushchij nesmeshchennuyu otsenku koeffitsienta determinatsii // Prikladnaya ekonometrika. 2008. № 4 (12). S. 71–83.
3. *Ivanova N.K., Lebedeva S.A., Noskov S.I.* Identifikatsiya parametrov nekotorykh negladkikh regressij // Informatsionnye tekhnologii i problemy matematicheskogo modelirovaniya slozhnykh sistem. 2016. Vyp. 17. S. 111–114.
4. *Klejner G.B.* Proizvodstvennye funktsii: Teoriya, metody, primenenie. M.: Finansy i statistika, 1986. 239 s.
5. *Noskov S.I.* Otsenivanie parametrov approksimiruyushchej funktsii s postoyannymi proportsiyami // Sovremennye tekhnologii. Sistemnyj analiz. Modelirovanie. 2013. № 2 (38). S. 135–136.
6. *Noskov S.I., Bazilevskij M.P.* Ob indeksnykh preobrazovaniyakh matrits pri postroenii regressionnykh modelej // Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami. 2019. № 3. S. 11–16.
7. *Noskov S.I., Bazilevskij M.P.* Postroenie regressionnykh modelej s ispol'zovaniem apparata linejno-bulevogo programmirovaniya. Irkutsk: IrGUPS, 2018. 176 s.
8. *Draper N.R., Smith H.* Applied Regression Analysis. Hoboken: John Wiley & Sons, 1998. 736 p.
9. *Harrell Jr., Frank E.* Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. [S. l.]: Springer Series in Statistics, 2015. 582 p.
10. *Kuhn M., Johnson K.* Applied Predictive Modeling. Berlin: Springer, 2018. 600 p.

DOI: 10.25586/RNU.V9187.20.01.P.023 УДК 123.2+004.58

М.Г. Груба, Ю.С. Фоменко, С.В. Калиниченко

АНАЛИЗ ВОЗДЕЙСТВИЯ ЗВУКОВОЙ ИНФОРМАЦИИ НА ОПЕРАТОРА СРЕДСТВ АВТОМАТИЗАЦИИ

Дается краткая характеристика основных показателей звука. Рассматриваются способы воздействия звука на психофизическое состояние человека. Обсуждается необходимость создания программного или аппаратно-программного средства обнаружения возможного информационно-психологического воздействия на состояние человека.
Ключевые слова: воздействие звука, речь, частоты, диапазон.

M.G. Gruba, Yu.S. Fomenko, S.V. Kalinichenko

ANALYSIS OF THE INFLUENCE OF SOUND INFORMATION ON THE OPERATOR OF AUTOMATION MEANS

A brief description of the main indicators of sound is given. The methods of influence of sound on the psychophysical state of a person are considered. The necessity of creating a software or hardware-software tool for detecting a possible information-psychological effect on a person's condition is discussed.
Keywords: sound exposure, speech, frequencies, range.