

УДК 550.388.2

**М.П. Базилевский**

*Иркутский государственный университет путей сообщения,  
г. Иркутск, Российская Федерация*

## **ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ ПОСТРОЕНИЯ КУСОЧНО-ЛИНЕЙНЫХ РЕГРЕССИЙ С НЕЛИНЕЙНЫМИ ГРАНИЦАМИ ПЕРЕКЛЮЧЕНИЯ**

**Аннотация.** Работа посвящена исследованию предложенных автором двух-факторных кусочно-линейных регрессий с нелинейными границами переключения. Для оценивания таких моделей с безусловными и условными переключениями с помощью метода наименьших модулей сформулированы две задачи частично-булевого линейного программирования. В первом случае задача позволяет одновременно определять оценки модели и оптимальные переключения на каждом наблюдении. Но при этом отсутствуют механизм переключения, что делает такие модели непригодными для прогнозирования. Во втором случае одновременно определяются и оценки регрессии, и точки переключения, и условия переключения, заданные в виде нелинейных функций. Проведены эксперименты, показавшие превосходство построенных кусочно-линейных регрессий с нелинейными границами переключения по качеству аппроксимации над соответствующими линейными регрессиями и производственными функциями Леонтьева. Также предложенные модели особо не уступили по качеству индексным регрессиям с неизвестными индексными векторами. Но построенные кусочно-линейные регрессии с нелинейными границами переключения, в отличие от индексных моделей, можно использовать для прогнозирования, поскольку они объясняют механизм переключения функций. И можно добиться того, чтобы этот механизм был объяснен в значительной степени или полностью. Показано, что одной из главных проблем при построении предложенных моделей является выбор малого положительного числа, характеризующего расстояние между областями переключения. Установлено, что выбор этого числа влияет на спецификацию входящей в условия переключения нелинейной функции.

**Ключевые слова.** Кусочно-линейная регрессия, точки переключения, кластеризация, индексная регрессия, метод наименьших модулей, задача частично-булевого линейного программирования.

**Информация о статье.** Дата поступления: 20 мая 2021 г.

**M.P. Bazilevskiy**

*Irkutsk State Transport University,  
Irkutsk, Russian Federation*

## **INVESTIGATION OF THE POSSIBILITY OF CONSTRUCTING PIECEWISE LINEAR REGRESSIONS WITH NONLINEAR SWITCHING BOUNDARIES**

**Abstract.** The research examines two-factor piecewise linear regressions with nonlinear switching boundaries proposed by the author. To estimate such models with unconditional and conditional switchings using the least absolute deviations, two partial Boolean linear programming problems were formulated. In the first case, the problem allows one to simultaneously determine model estimates and optimal switchings on each

observation. But at the same time, there is no switching mechanism, which makes such models unsuitable for forecasting. In the second case, both the regression estimates, and the switching points, and the switching conditions, given in the form of nonlinear functions, are determined simultaneously. Experiments showed the superiority of the constructed piecewise linear regressions with nonlinear switching boundaries in terms of the quality of approximation over the corresponding linear regressions and Leontiev's production functions. Also, the proposed models were not particularly inferior in quality to index regressions with unknown index vectors. But the constructed piecewise linear regressions with nonlinear switching boundaries, in contrast to index models, can be used for forecasting, since they explain the mechanism of switching functions. And it is possible to ensure that this mechanism is explained to a large extent or completely. The study showed that one of the main problems in the construction of the proposed models is the choice of a small positive number characterizing the distance between switching regions. We found that the choice of this number affects the specification of the nonlinear function included in the switching conditions.

**Keywords.** Piecewise linear regression, switching points, clustering, index regression, least absolute deviations, partial Boolean linear programming problem.

**Article info.** Received 20 May 2021.

---

## Введение

При построении регрессионных моделей исследователи чаще всего прибегают к оцениванию по всей заданной выборке одного единственного уравнения, полагая, что его коэффициенты не меняются во времени. В противном случае целесообразно строить кусочно-заданные регрессии с некоторым количеством точек переключения.

Так называемым регрессиям с переключениями посвящено множество научных работ. Например, в работах [1; 2] описаны способы одновременного оценивания параметров линейной регрессии и точек переключения с помощью метода наименьших квадратов (МНК). Работы [3–5] посвящены моделями с марковскими переключениями состояний (режимов).

В работе [6] предложен способ совместного определения точек переключения и параметров кусочно-линейной зависимости с помощью метода наименьших модулей (МНМ). Эта задача сведена к задаче частично-булевого линейного программирования (ЧБЛП). При этом границы переключения являются линейными и зависят только от одной конкретной объясняющей переменной.

Целью данной работы является исследование двухфакторных кусочно-линейных регрессий с нелинейными границами переключения.

### 1. МНМ-оценивание двухфакторных кусочно-линейных регрессий с нелинейными границами переключения

На основе математического аппарата, предложенного в работе [там же], сведем задачу МНМ-оценивания двухфакторной кусочно-линейной регрессии с нелинейными границами переключения к задаче ЧБЛП.

Введем многофакторную модель кусочно-линейной регрессии с нелинейными границами переключения:

$$y_i = \alpha_0 + f(x_{i1}, x_{i2}, \dots, x_{im}) + \varepsilon_i, \quad i = \overline{1, n}, \quad (1)$$

где  $n$  — объем выборки;  $m$  — количество объясняющих переменных;  $y_i, i = \overline{1, n}$  — значения объясняемой переменной  $y$ ;  $x_{ij}, i = \overline{1, n}, j = \overline{1, m}$  — значения объясняющих переменных;  $\alpha_0$  — неизвестный параметр;  $\varepsilon_i, i = \overline{1, n}$  — ошибки аппроксимации;  $f(x_1, x_2, \dots, x_m)$  — функция вида

$$f(x_1, x_2, \dots, x_m) = \begin{cases} \alpha_1 x_1, & \text{при } F(x_1, x_2, \dots, x_m) \leq x_1^*, \\ \alpha_2 x_2, & \text{при } x_1^* < F(x_1, x_2, \dots, x_m) \leq x_2^*, \\ \dots \\ \alpha_m x_m, & \text{при } F(x_1, x_2, \dots, x_m) > x_{m-1}^*, \end{cases}$$

в которой  $\alpha_1, \alpha_2, \dots, \alpha_m$  — неизвестные параметры,  $x_1^*, x_2^*, \dots, x_{m-1}^*$  — неизвестные точки переключения, а функция  $F(x_1, x_2, \dots, x_m)$  имеет, например, вид

$$F(x_1, x_2, \dots, x_m) = \sum_{j=1}^m c_j x_j + \sum_{j=1}^m c_{jj} x_j^2 + \sum_{j=1}^{m-1} \sum_{k=j+1}^m c_{jk} x_j x_k,$$

где  $c_j, c_{jj}, j = \overline{1, m}, c_{jk}, j = \overline{1, m}, k = \overline{j+1, m}$  — неизвестные параметры.

Пусть  $m = 2$ . Тогда модель (1) примет вид

$$y_i = \alpha_0 + f(x_{i1}, x_{i2}) + \varepsilon_i, \quad i = \overline{1, n}, \quad (2)$$

$$\text{где } f(x_1, x_2) = \begin{cases} \alpha_1 x_1, & \text{при } \sum_{j=1}^2 c_j x_j + \sum_{j=1}^2 c_{jj} x_j^2 + \sum_{j=1}^1 \sum_{k=j+1}^2 c_{jk} x_j x_k \leq x_1^*, \\ \alpha_2 x_2, & \text{при } \sum_{j=1}^2 c_j x_j + \sum_{j=1}^2 c_{jj} x_j^2 + \sum_{j=1}^1 \sum_{k=j+1}^2 c_{jk} x_j x_k > x_1^*. \end{cases}$$

С целью учета переключения между функциями  $\alpha_1 x_1$  и  $\alpha_2 x_2$  на  $i$ -м наблюдении введем булевы переменные  $\sigma_i^j, i = \overline{1, n}, j = \overline{1, 2}$  по следующему правилу:

$$\sigma_i^{(1)} = \begin{cases} 0, & \text{при } f(x_1, x_2) = \alpha_1 x_1, \\ 1, & \text{в противном случае,} \end{cases} \quad \sigma_i^{(2)} = \begin{cases} 0, & \text{при } f(x_1, x_2) = \alpha_2 x_2, \\ 1, & \text{в противном случае.} \end{cases}$$

Тогда справедливы ограничения

$$\sigma_i^{(1)} + \sigma_i^{(2)} = 1, \quad i = \overline{1, n}, \quad (3)$$

$$\sigma_i^{(j)} \in \{0,1\}, \quad i = \overline{1,n}, \quad j = \overline{1,2}. \quad (4)$$

Если  $\sigma_i^{(1)} = 0$  для всех наблюдений  $i = \overline{1,n}$ , то модель (2) принимает вид

$$y_i = \alpha_0 + \alpha_1 x_{i1} + \varepsilon_i, \quad i = \overline{1,n},$$

а ее МНМ-оценки [7] находятся на основе решения задачи линейного программирования с целевой функцией

$$\sum_{i=1}^n (u_i^{(1)} + v_i^{(1)}) \rightarrow \min,$$

и с линейными ограничениями

$$y_i = \alpha_0^{(1)} - \alpha_0^{(2)} + \alpha_1^{(1)} x_{i1} - \alpha_1^{(2)} x_{i1} + u_i^{(1)} - v_i^{(1)}, \quad i = \overline{1,n}, \quad (5)$$

$$\alpha_0^{(2)} \geq 0, \quad \alpha_0^{(1)} \geq 0, \quad \alpha_1^{(1)} \geq 0, \quad u_i^{(1)} \geq 0, \quad v_i^{(1)} \geq 0, \quad i = \overline{1,n}, \quad (6)$$

где  $\alpha_0 = \alpha_0^{(1)} - \alpha_0^{(2)}$ ,  $\alpha_1 = \alpha_1^{(1)} - \alpha_1^{(2)}$ ,  $\varepsilon_i = u_i^{(1)} - v_i^{(1)}$ ,  $i = \overline{1,n}$ .

Если  $\sigma_i^{(2)} = 0$  для всех наблюдений  $i = \overline{1,n}$ , то модель (2) принимает вид

$$y_i = \alpha_0 + \alpha_2 x_{i2} + \varepsilon_i^*, \quad i = \overline{1,n},$$

а ее МНМ-оценки находятся в результате решения задачи линейного программирования с целевой функцией

$$\sum_{i=1}^n (u_i^{(2)} + v_i^{(2)}) \rightarrow \min,$$

и с линейными ограничениями

$$y_i = \alpha_0^{(1)} - \alpha_0^{(2)} + \alpha_2^{(1)} x_{i2} - \alpha_2^{(2)} x_{i2} + u_i^{(2)} - v_i^{(2)}, \quad i = \overline{1,n}, \quad (7)$$

$$\alpha_0^{(2)} \geq 0, \quad \alpha_0^{(1)} \geq 0, \quad \alpha_2^{(1)} \geq 0, \quad \alpha_2^{(2)} \geq 0, \quad u_i^{(2)} \geq 0, \quad v_i^{(2)} \geq 0, \quad i = \overline{1,n}, \quad (8)$$

где  $\alpha_0 = \alpha_0^{(1)} - \alpha_0^{(2)}$ ,  $\alpha_2 = \alpha_2^{(1)} - \alpha_2^{(2)}$ ,  $\varepsilon_i^* = u_i^{(2)} - v_i^{(2)}$ ,  $i = \overline{1,n}$ .

При наличии переключений в модели (2) ограничения (5) необходимо заменить на следующие:

$$\alpha_0^{(1)} - \alpha_0^{(2)} + \alpha_1^{(1)} x_{i1} - \alpha_1^{(2)} x_{i1} + u_i^{(1)} - v_i^{(1)} \leq y_i + M \sigma_i^{(1)}, \quad i = \overline{1,n}, \quad (9)$$

$$\alpha_0^{(1)} - \alpha_0^{(2)} + \alpha_1^{(1)} x_{i1} - \alpha_1^{(2)} x_{i1} + u_i^{(1)} - v_i^{(1)} \geq y_i - M \sigma_i^{(1)}, \quad i = \overline{1,n}, \quad (10)$$

где  $M$  — большое положительное число.

Аналогично необходимо заменить ограничения (7) на следующие:

$$\alpha_0^{(1)} - \alpha_0^{(2)} + \alpha_2^{(1)}x_{i2} - \alpha_2^{(2)}x_{i2} + u_i^{(2)} - v_i^{(2)} \leq y_i + M\sigma_i^{(2)}, \quad i = \overline{1, n}, \quad (11)$$

$$\alpha_0^{(1)} - \alpha_0^{(2)} + \alpha_2^{(1)}x_{i2} - \alpha_2^{(2)}x_{i2} + u_i^{(2)} - v_i^{(2)} \geq y_i - M\sigma_i^{(2)}, \quad i = \overline{1, n}. \quad (12)$$

Тогда решение задачи частично-булевого линейного программирования (ЧБЛП) с целевой функцией

$$\sum_{i=1}^n (u_i^{(1)} + v_i^{(1)} + u_i^{(2)} + v_i^{(2)}) \rightarrow \min, \quad (13)$$

и линейными ограничениями (9)–(12), (3), (4), (6), (8) позволяет найти МНМ-оценки двухфакторной модели кусочно-линейной регрессии с безусловными переключениями:

$$y_i = \alpha_0 + f(x_{i1}, x_{i2}) + \varepsilon_i, \quad i = \overline{1, n}, \quad (14)$$

$$\text{где } f(x_{i1}, x_{i2}) = \begin{cases} \alpha_1 x_{i1}, & \text{при } \sigma_i^{(1)} = 0, \\ \alpha_2 x_{i2}, & \text{при } \sigma_i^{(2)} = 0. \end{cases}$$

Фактически решение задачи (13), (9)–(12), (3), (4), (6), (8) позволяет разделить все  $n$  наблюдений на 2 кластера. В первом из них модель регрессии имеет вид  $y_i = \alpha_0 + \alpha_1 x_{i1} + \varepsilon_i$ , во втором —  $y_i = \alpha_0 + \alpha_2 x_{i2} + \varepsilon_i$ . При этом не возникает эффекта мультиколлинеарности. К сожалению, использовать оцененную модель (14) для прогнозирования не представляется возможным, поскольку в ней отсутствует условие принадлежности заданной пары значений объясняющих переменных к одному из кластеров. Однако она может применяться как вспомогательный инструмент, служа эталоном по величине суммы модулей остатков для моделей кусочно-линейной регрессии с условными переключениями (2).

Поставим на условия переключения функции  $f(x_1, x_2)$  в модели (2) следующие линейные ограничения:

$$\begin{aligned} & -M\sigma_i^{(1)} + \sum_{j=1}^2 (c_j^{(1)} - c_j^{(2)})x_j + \sum_{j=1}^2 (c_{jj}^{(1)} - c_{jj}^{(2)})x_j^2 + \\ & + (c_{12}^{(1)} - c_{12}^{(2)})x_1x_2 - (x_1^{(1)*} - x_1^{(2)*}) \leq 0, \quad i = \overline{1, n}, \end{aligned} \quad (15)$$

$$\begin{aligned} & M\sigma_i^{(2)} + \sum_{j=1}^2 (c_j^{(1)} - c_j^{(2)})x_j + \sum_{j=1}^2 (c_{jj}^{(1)} - c_{jj}^{(2)})x_j^2 + \\ & + (c_{12}^{(1)} - c_{12}^{(2)})x_1x_2 - (x_1^{(1)*} - x_1^{(2)*}) \geq \delta, \quad i = \overline{1, n}, \end{aligned} \quad (16)$$

где  $c_{jj} = c_{jj}^{(1)} - c_{jj}^{(2)}$ ,  $c_j = c_j^{(1)} - c_j^{(2)}$ ,  $j = \overline{1, 2}$ ,  $c_{12} = c_{12}^{(1)} - c_{12}^{(2)}$ ,  $x_1^* = x_1^{(1)*} - x_1^{(2)*}$ ,  $\delta$  — малое положительное число, характеризующее расстояние между областями переключения.

Тогда решение задачи ЧБЛП с целевой функцией (13) и с линейными ограничениями (9)–(12), (3), (4), (6), (8), (15), (16) дает МНМ-оценки двухфакторной кусочно-линейной регрессии (2), точку переключения  $x_1^*$  и уравнение границы переключения.

## 2. Вычислительные эксперименты

*Эксперимент № 1.* Для проведения эксперимента были использованы статистические данные (табл. 1) из учебника [8] о сменной добыче угля на одного рабочего  $y$ , мощности пласта  $x_1$  и уровне механизации работ  $x_2$ .

Таблица 1

Статистические данные

Имя	A	B	C	D	E	F	G	H	I	J
$y$	5	10	10	7	5	6	6	5	6	8
$x_1$	8	11	12	9	8	8	9	9	8	12
$x_2$	5	8	8	5	7	8	6	4	5	7

По этим данным с помощью МНМ были построены следующие модели.

### 1. Линейная регрессия

$$\tilde{y} = -4,66667 + x_1 + 0,33333x_2, \quad (17)$$

сумма модулей остатков которой  $I = \sum_{i=1}^n |e_i| = 6,33333$ .

### 2. Функция Леонтьева со свободным членом [9–11]

$$\tilde{y} = -2,5 + \min\{1,04166x_1; 1,875x_2\}, \quad (18)$$

для которой  $I = 6,04166$ .

3. Индексная регрессия [12; 13] с неизвестным индексным вектором

$$\tilde{y} = \text{ind}_G \left\{ \frac{2}{3}x_1; \frac{5}{4}x_2 \right\}, \quad G = (1 \ 2 \ 2 \ 2 \ 1 \ 1 \ 1 \ 2 \ 2 \ 1), \quad (19)$$

для которой  $I = 2,33333$ .

4. Кусочно-линейная регрессия с безусловными переключениями (при  $M = 10\ 000$ )

$$\tilde{y} = \begin{cases} \frac{2}{3}x_1, & \text{если } \sigma_i^{(1)} = 0, \\ \frac{5}{4}x_2, & \text{если } \sigma_i^{(2)} = 0, \end{cases} \quad \sigma^{(1)} = (0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 1 \ 0), \quad (20)$$

для которой  $I = 2,33333$ .

5. Кусочно-линейная регрессия с условными переключениями (при  $M = 10\,000$  и  $\delta = 0,0001$ )

$$\tilde{y} = \begin{cases} \frac{2}{3}x_1, & \text{если } -0,0021176x_2 + 5,88235 \cdot 10^{-6}x_1^2 + \\ & + 0,00015686x_2^2 + 0,00629 \leq 0, \\ \frac{5}{4}x_2, & \text{если } -0,0021176x_2 + 5,88235 \cdot 10^{-6}x_1^2 + \\ & + 0,00015686x_2^2 + 0,00629 \geq 0,0001, \end{cases} \quad (21)$$

$$\sigma^{(1)} = (0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0),$$

для которой  $I = 2,75$ .

Прокомментируем полученные результаты. Во-первых, уравнения индексной (19) и кусочно-линейной регрессий (20), как и их переключения по выборке, оказались полностью идентичными. Во-вторых, модель (21) по величине суммы модулей остатков оказалась несколько хуже регрессии с безусловными переключениями (20). Иными словами, при построении модели (21) осуществлялась попытка «объяснить» механизм переключения функций  $\alpha_1 x_1$  и  $\alpha_2 x_2$  с помощью условий, содержащих заданную нелинейную функцию  $F(x_1, x_2)$ . Но оказалось, что такая функция не смогла на 100 % объяснить этот механизм. Возможно, что для этого требуется использовать другую функцию  $F(x_1, x_2)$ . В-третьих, модели (19)–(21) существенно выиграли по качеству аппроксимации  $y$  (17) и (18). Для регрессий (18) и (19) это естественно, поскольку индексные модели являются обобщением производственных функций Леонтьева [13]. Однако, несмотря на высокое качество, возникает проблема с применением уравнений (19) и (20) для прогнозирования значений переменной  $y$ , связанная с незнанием механизма переключений в будущие моменты времени. В-четвертых, при использовании уравнения (21) при прогнозировании может оказаться так, что не будет выполнено ни первое, ни второе условие переключения. В работе [6] предложены некоторые способы решения этой проблемы. В -пятых, во всех моделях, за исключением линейной регрессии (17), отсутствует эффект мультиколлинеарности.

Таким образом, из всех пяти зависимостей (17)–(21) для прогнозирования лучше всего подходит кусочно-линейная регрессия (21). Значения функции  $F(x_1, x_2) = -0,0021176x_2 + 5,88235 \cdot 10^{-6}x_1^2 + 0,00015686x_2^2 + 0,00629$ , входящей в условия переключения зависимости (21), для каждого наблюдения представлены в табл. 2.

На рис. 1 изображена диаграмма рассеивания и график кривой  $F(x_1, x_2) = -0,0021176x_2 + 5,88235 \cdot 10^{-6}x_1^2 + 0,00015686x_2^2 + 0,00629$ , представляющей собой эллипс. Зеленым цветом выделена область, в которой выполняется первое условие переключения зависимости (21).

Переключения для модели (21)

Имя	$x_1$	$x_2$	$F(x_1, x_2)$	$\sigma_i^{(1)}$
A	8	5	-2,96E - 08	0
B	11	8	0,0001	1
C	12	8	0,0002353	1
D	9	5	9,997E - 05	1
E	8	7	-0,000471	0
F	8	8	-0,000235	0
G	9	6	-0,000292	0
H	9	4	0,0008058	1
I	8	5	-2,96E - 08	0
J	12	7	-1,6E - 09	0

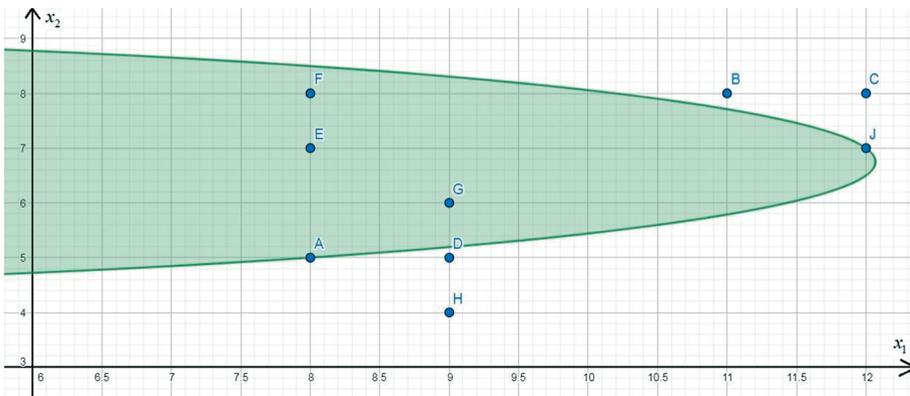


Рис. 1. Диаграмма рассеивания и области переключения для модели (21)

Как видно, точки A, E, F, G, I, J лежат внутри эллипса или на его границе, а точки B, C, D, H — снаружи.

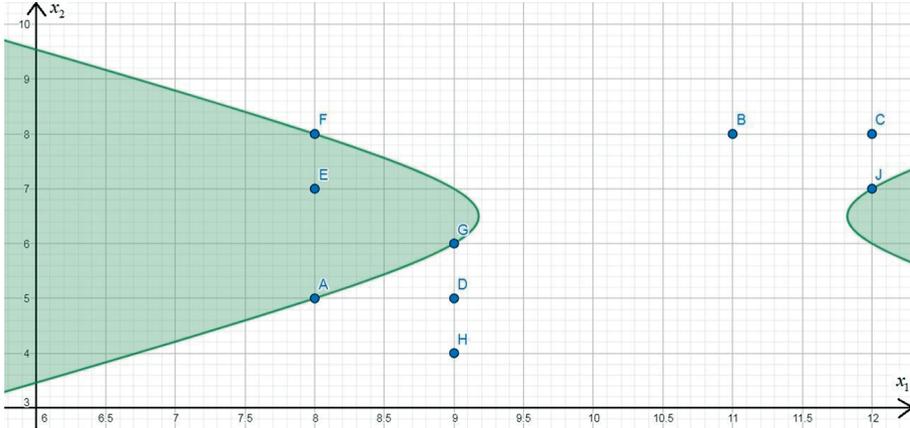
Далее исследовалось влияние заданного малого положительного числа  $\delta$  на оценки кусочно-линейной регрессии (2). Она строилась по тем же данным (табл. 1). При  $M = 10\ 000$ ,  $\delta = 0,001$  было снова получено уравнение (21), но при этом все коэффициенты функции  $F(x_1, x_2)$  оказались больше ровно в 10 раз. При  $M = 10\ 000$ ,  $\delta = 0,01$  и при  $M = 10\ 000$ ,  $\delta = 0,1$  также происходило увеличение ее коэффициентов в 10 раз. А при  $M = 10\ 000$ ,  $\delta = 1$  была получена другая спецификация модели:

$$\tilde{y} = \begin{cases} \frac{2}{3}x_1, & \text{если } 5,25x_1 - 6,5x_2 - 0,25x_1^2 + 0,5x_2^2 - 6 \leq 0, \\ \frac{5}{4}x_2, & \text{если } 5,25x_1 - 6,5x_2 - 0,25x_1^2 + 0,5x_2^2 - 6 > 0, \end{cases} \quad (22)$$

$$\sigma^{(1)} = (0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0),$$

для которой сумма модулей остатков  $I = 2,75$ .

На рис. 2 изображена диаграмма рассеивания и график кривой  $F(x_1, x_2) = 5,25x_1 - 6,5x_2 - 0,25x_1^2 + 0,5x_2^2 - 6$ , представляющей собой гиперболу.



**Рис. 2. Диаграмма рассеивания и области переключения для модели (22)**

Как видно, точки А, Е, F, G, I, J лежат внутри ветвей гиперболы или на их границах, а точки В, С, D, H — снаружи. При этом точки В, С, D и H находятся гораздо дальше от выделенной области переключения, чем на рис. 1.

*Эксперимент № 2.* Для проведения эксперимента были использованы статистические данные (табл. 3) из статьи [14] о потреблении электроэнергии  $y$ , валовом региональном продукте  $x_1$  и строительстве жилых домов  $x_2$  в Иркутской области.

Таблица 3

**Статистические данные**

Имя	A	B	C	D	E	F	G	H	I	J
$y$	52,5	53,6	53,3	55,1	52,4	54,3	56,7	58	56,6	56,3
$x_1$	258,1	330,8	402,7	438,9	458,8	546,1	634,6	738	805,2	907,4
$x_2$	303	331	575	585	602,2	629,5	755,2	871,4	829,2	716,9

По этим данным с помощью МНМ были построены следующие модели.

1. Линейная регрессия

$$\tilde{y} = 50,2886 + 0,002626x_1 + 0,00506x_2, \tag{23}$$

для которой  $I = 7,45736$ .

2. Функция Леонтьева со свободным членом

$$\tilde{y} = 49,71817 + \min\{0,010925x_1; 0,00918x_2\}, \quad (24)$$

для которой  $I = 6,83683$ .

3. Индексная регрессия

$$\tilde{y} = 49,6857 + \text{ind}_G\{0,008587x_1; 0,009288x_2\}, \quad (25)$$

$$G = (2 \ 2 \ 1 \ 2 \ 1 \ 1 \ 2 \ 2 \ 1 \ 2),$$

для которой  $I = 2,58099$ .

4. Кусочно-линейная регрессия с безусловными переключениями (при  $M = 10\ 000$ )

$$\tilde{y} = \begin{cases} 49,68575 + 0,008587x_1, & \text{если } \sigma_i^{(1)} = 0, \\ 49,68575 + 0,009288x_2, & \text{если } \sigma_i^{(2)} = 0, \end{cases} \quad (26)$$

$$\sigma^{(1)} = (1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1),$$

для которой  $I = 2,58099$ .

5. Кусочно-линейная регрессия с условными переключениями (при  $M = 10\ 000$  и  $\delta = 10$ )

$$\tilde{y} = \begin{cases} 49,68575 + 0,0085869x_1, & \text{если } 7,8674x_1 - 20,4321x_2 + \\ & + 0,00202074x_1^2 + 0,0213125x_2^2 - \\ & - 0,0150479x_1x_2 + 4690,497 \leq 0, \\ 49,68575 + 0,0092879x_2, & \text{если } 7,8674x_1 - 20,4321x_2 + \\ & + 0,00202074x_1^2 + 0,0213125x_2^2 - \\ & - 0,0150479x_1x_2 + 4690,497 \geq 10, \end{cases} \quad (27)$$

$$\sigma^{(1)} = (1 \ 1 \ 0 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0 \ 1),$$

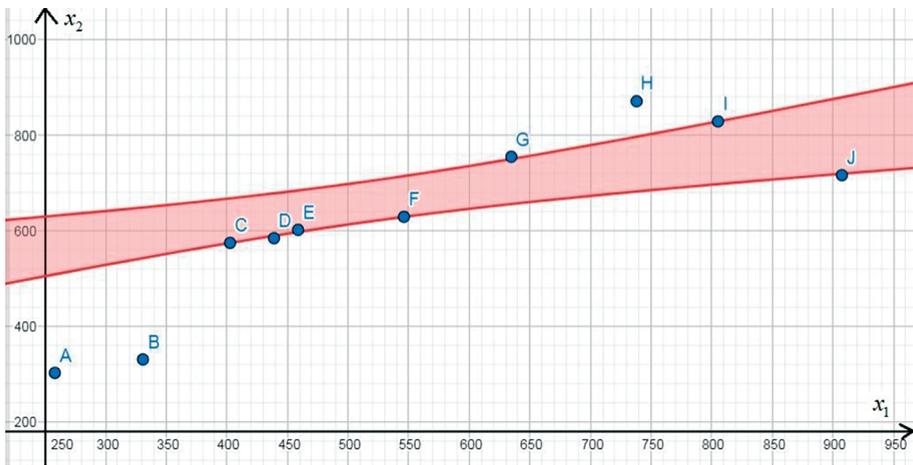
для которой  $I = 2,58099$ .

Полученные результаты во многом схожи с результатами, полученными при построении моделей (17)–(21). Однако теперь кусочно-линейная регрессия (27) оказалась такой же качественной, как и модели (25) и (26), т.е. полученная функция  $F(x_1, x_2)$  смогла на 100 % объяснить механизм переключения функций  $a_1x_1$  и  $a_2x_2$ . Поэтому модель (27) отлично подходит для прогнозирования будущих значений переменной  $y$ .

На рис. 3 изображена диаграмма рассеивания и график кривой

$$F(x_1, x_2) = 7,8674x_1 - 20,4321x_2 + 0,00202074x_1^2 + 0,0213125x_2^2 - 0,0150479x_1x_2 + 4690,497.$$

Красным цветом выделена область, в которой выполняется первое условие переключения зависимости (27).



*Рис. 3. Диаграмма рассеивания и области переключения для модели (27)*

Как видно, точки С, Е, F, I лежат внутри выделенной области или на ее границе, а точки А, В, D, G, H, J — снаружи.

### Заключение

В данной работе предложены многофакторные модели кусочно-линейной регрессии с нелинейными границами переключения. Предложенные модели исследованы для случая двух объясняющих переменных. Задача их оценивания с помощью МНМ сведена к задаче ЧБЛП. Проведенные эксперименты показали превосходство построенных кусочно-линейных регрессий с нелинейными границами переключения по качеству аппроксимации над соответствующими линейными регрессиями и производственными функциями Леонтьева. При этом они особо не уступили по качеству и индексным регрессиям с неизвестными индексными векторами. Предложенные кусочно-линейные регрессии объясняют механизм переключения функций, поэтому их, в отличие от индексных моделей, можно использовать для прогнозирования. Кроме того, они лишены эффекта мультиколлинеарности. Установлено, что одной из главных проблем при построении предложенных моделей является выбор малого положительного числа, характеризующего расстояние между областями переключения.

### СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Корхин А.С. О построении регрессии с переключениями, когда точки переключения неизвестны / А.С. Корхин // Кибернетика и системный анализ. — 2018. — Т. 54, № 3. — С. 116–130.

2. Кнопов П.С. Статистический анализ динамики инфицирования коронавирусом с помощью пошаговой регрессии с переключениями / П.С. Кнопов, А.С. Корхин // *Кибернетика и системный анализ*. — 2020. — Т. 56, № 6. — С. 96–106.

3. Силаева В.А. Оценивание параметров моделей временных рядов с марковскими переключениями режимов / В.А. Силаева, М.В. Силаева, А.М. Силаев // *Компьютерные исследования и моделирование*. — 2018. — Т. 10, № 6. — С. 903–918.

4. Малюгин В.И. Анализ сложных систем на основе многомерных регрессионных моделей с переключениями состояний / В.И. Малюгин, А.Ю. Новопольцев // *Международный конгресс по информатике: информационные системы и технологии : материалы междунар. конгресса*. — Минск, 2016. — С. 458–462.

5. Малюгин В.И. Анализ бизнес-цикла белорусской экономики на основе моделей с переключениями состояний и экзогенными переменными / В.И. Малюгин // *Тенденции экономического развития в XXI веке : материалы II Междунар. науч. конф.* — Минск, 2020. — С. 176–179.

6. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных / С.И. Носков. — Иркутск : Облформпечать, 1996. — 321 с.

7. Носков С.И., Базилевский М.П. Построение регрессионных моделей с использованием аппарата линейно-булевого программирования / С.И. Носков, М.П. Базилевский. — Иркутск : Изд-во ИрГУПС, 2018. — 176 с.

8. Кремер Н.Ш. Теория вероятностей и математическая статистика / Н.Ш. Кремер. — Москва : ЮНИТИ-ДАНА, 2004. — 573 с.

9. Иванова Н.К., Лебедева С.А., Носков С.И. Идентификация параметров некоторых негладких регрессий / Н.К. Иванова, С.А. Лебедева, С.И. Носков // *Информационные технологии и проблемы математического моделирования сложных систем*. — 2016. — № 17. — С. 107–110.

10. Носков С.И. Программный комплекс построения некоторых типов кусочно-линейных регрессий / С.И. Носков, А.А. Хоняков // *Информационные технологии и математическое моделирование в управлении сложными системами*. — 2019. — № 3 (4). — С. 47–55.

11. Носков С.И. Оценивание параметров аппроксимирующей функции с постоянными пропорциями / С.И. Носков // *Современные технологии. Системный анализ. Моделирование*. — 2013. — № 2 (38). — С. 135–136.

12. Носков С.И. Об индексных преобразованиях матриц при построении регрессионных моделей / С.И. Носков, М.П. Базилевский // *Информационные технологии и математическое моделирование в управлении сложными системами*. — 2019. — № 3 (4). — С. 11–16.

13. Базилевский М.П. Оценивание индексных моделей регрессии с помощью метода наименьших модулей / М.П. Базилевский, С.И. Носков // *Вестник Российского нового университета. Серия: Сложные системы: модели, анализ и управление*. — 2020. — № 1. — С. 17–23.

14. Базилевский М.П. МНК-оценивание параметров специфицированных на основе функций Леонтьева двухфакторных моделей регрессии / М.П. Базилевский // *Южно-Сибирский научный вестник*. — 2019. — № 2 (26). — С. 66–70.

## REFERENCES

1. Korkhin A.S. On Constructing a Switching Regression with Unknown Switching Points. *Kibernetika i sistemnyi analiz = Cybernetics and Systems Analysis*, 2018, vol. 54, no. 3, pp. 116–130. (In Russian).

2. Knopov P.S., Korkhin A.S. Statistical Analysis of the Dynamics of Coronavirus Cases using Stepwise Switching Regression. *Kibernetika i sistemnyi analiz = Cybernetics and Systems Analysis*, 2020, vol. 56, no. 6, pp. 96–106. (In Russian).

3. Silaeva V.A., Silaeva M.V., Silaev A.M. Estimation of Models Parameters for Time Series with Markov Switching Regimes. *Komp'yuternye issledovaniya i modelirovanie = Computer Research and Modeling*, 2018, vol. 10, no. 6, pp. 903–918. (In Russian).
4. Malugin V.I., Novopoltsev A.Yu. Analysis of Complex System Based on the Multivariate Regression Models with a Switching States. *Mezhdunarodnyi kongress po informatike: informatsionnye sistemy i tekhnologii. Materialy mezhdunarodnogo kongressa* [International Congress of Informatics: Information Systems and Technologies. Materials of the International Congress]. Minsk, 2016, pp. 458–462. (In Russian).
5. Malugin V.I. Business Cycle Analysis of the Belarusian Economy Based on Regime-Switching Models with Exogenous Variables. *Tendentsii ekonomicheskogo razvitiya v XXI veke. Materialy II Mezhdunarodnoi nauchnoi konferentsii* [Economic Trends in the 21st Century. Materials of the VII International Scientific Conference]. Minsk, 2020, pp. 176–179. (In Russian).
6. Noskov S.I. *Tekhnologiya modelirovaniya ob"ektov s nestabil'nym funktsionirovaniem i neopredelennost'yu v dannykh* [Object Modeling Technology with Unstable Operation and Data Uncertainty]. Irkutsk, Oblinformpechat Publ., 1996. 321 p.
7. Noskov S.I., Bazilevsky M.P. *Postroenie regressionnykh modelei s ispol'zovaniem apparata lineino-bulevogo programmirovaniya* [Construction of Regression Models Using Mixed Integer Linear Programming]. Irkutsk State Railway Transport Engineering University Publ., 2018. 176 p.
8. Kremer N.Sh. *Teoriya veroyatnostei i matematicheskaya statistika* [Theory of Probability and Mathematical Statistics]. Moscow, YuNITI-DANA Publ., 2004. 573 p.
9. Ivanova N.K., Lebedeva S.A., Noskov S.I. Identification of parameters of some nonsmooth regressions. *Informatsionnye tekhnologii i problemy matematicheskogo modelirovaniya slozhnykh sistem = Information technologies and problems of mathematical modeling of complex systems*, 2016, no. 17, pp. 107–110. (In Russian).
10. Noskov S.I., Khonyakov A.A. Software Complex for Building some Types Pieces of Linear Regressions. *Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnyimi sistemami = Information Technology and Mathematical Modeling in the Management of Complex Systems*, 2019, no. 3 (4), pp. 47–55. (In Russian).
11. Noskov S.I. Parameters Estimation of the Fitting Functions with Constant Proportions. *Sovremennye tekhnologii. Sistemnyi analiz. Modelirovanie = Modern Technologies. System Analysis. Modeling*, 2013, no. 2 (38), pp. 135–136. (In Russian).
12. Noskov S.I., Bazilevsky M.P. On Index Transformations of Matrices in Regression Models Construction. *Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnyimi sistemami = Information Technology and Mathematical Modeling in the Management of Complex Systems*, 2019, no. 3 (4), pp. 11–16. (In Russian).
13. Bazilevsky M.P., Noskov S.I. Estimation of Index Regression Models Using the Least Absolute Deviations. *Vestnik Rossiiskogo novogo universiteta. Seriya: Slozhnye sistemy: modeli, analiz i upravlenie = Vestnik of Russian New University. Series: Complex Systems: Models, Analysis, Management*, 2020, no. 1, pp. 17–23. (In Russian).
14. Bazilevsky M.P. Ols-Estimation of Two-Factor Regression Models Specified on Leontiev Functions. *Yuzhno-Sibirskii nauchnyi vestnik = South-Siberian Scientific Bulletin*, 2019, no. 2 (26), pp. 66–70. (In Russian).

### ИНФОРМАЦИЯ ОБ АВТОРЕ

**Базилевский Михаил Павлович** — кандидат технических наук, доцент, кафедра математики, Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация, e-mail: mik2178@yandex.ru.

### INFORMATION ABOUT THE AUTHOR

**Mikhail P. Bazilevskiy** — PhD in Technical Sciences, Associate Professor, Department of Mathematics, Irkutsk State Transport University, Irkutsk, Russian Federation, e-mail: mik2178@yandex.ru.

### ДЛЯ ЦИТИРОВАНИЯ

Базилевский М.П. Исследование возможности построения кусочно-линейных регрессий с нелинейными границами переключения / М.П. Базилевский // *System Analysis & Mathematical Modeling*. — 2021. — Т. 3, № 2. — С. 99–112.

### FOR CITATION

Bazilevskiy M.P. Investigation of the Possibility of Constructing Piecewise Linear Regressions with Nonlinear Switching Boundaries. *System Analysis & Mathematical Modeling*, 2021, vol. 3, no. 2, pp. 99–112. (In Russian).