Носков С. И.
S. I. Noskov

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ ПРОСТОЙ ФОРМЫ ВЛОЖЕННОЙ КУСОЧНО-ЛИНЕЙНОЙ РЕГРЕССИИ

PARAMETER IDENTIFICATION OF THE SIMPLE FORM OF A NESTED PIECEWISE LINEAR REGRESSION

Носков Сергей Иванович – доктор технических наук, профессор кафедры «Информационные системы и защита информации» Иркутского государственного университета путей сообщения (Россия, Иркутск); Россия, 664074, ул. Чернышевского, д. 15; тел. 8(914)902-24-94. E-mail: sergey.noskov.57@mail.ru.
Sergey I. Noskov – Doctor of Engineering, Professor, Information Systems and Information Protection Department, Irkutsk State University of Railway Transport (Russia, Irkutsk); Russia, 664074, st. Chernyshevsky, 15; tel. 8(914)902-24-94. E-mail: sergey.noskov.57@mail.ru.

Аннотация. В работе на основе прошлых публикаций автора показано, что задача идентификации параметров простой формы вложенной кусочно-линейной регрессионной модели с использованием метода наименьших модулей может быть сведена к задаче линейно-булевого программирования приемлемой для практических проблем размерности. Решён простой численный пример.

Summary. Based on the author's previous publications, the paper shows that the problem of identifying the parameters of a simple form of a nested piecewise linear regression model using the method of least modules can be reduced to a linear Boolean programming problem of dimension acceptable for practical problems. A simple numerical example has been solved.

Ключевые слова: кусочно-линейная регрессия, функция риска, оценивание параметров, вложенная кусочно-линейная модель, метод наименьших модулей.

Key words: piecewise linear regression, risk function, parameter estimation, nested piecewise linear model, method of least modules.

УДК 330.4

Введение. Методы регрессионного анализа широко применяются для исследования различных объектов. При этом используются как обычные, гладкие модельные связи, так и различные негладкие конструкции. Так, в работе [1] разработан новый метод оценки адекватности функции гладкой регрессии на основе непараметрической регрессии и бутстреп-подхода. Он позволяет пользователям выявлять систематические несоответствия в расчётных и модельных траекториях и проверять различные гипотезы. Рассчитываются доверительные интервалы для оценок непараметрической регрессии, если описываемая функция адекватна, что позволяет выявлять контуры возможного несоответствия. В работе [2] с помощью модели регрессии плавного перехода исследуется нелинейная зависимость между размером внешнего долга и экономическим ростом в Иране и Малайзии за период 1973-2017 гг. В работе [3] предлагается использовать кусочно-линейную модель прогрессивной шкалы налогообложения. Продемонстрированы преимущества предложенной модели налогообложения по сравнению с «плоской» шкалой. Кусочно-линейная шкала позволит налогоплательщикам легче перейти на прогрессивную налоговую ставку. Для оценки действенности предлагаемого варианта налогообложения разработана экономико-математическая модель, предполагающая освобождение от налога лиц с доходами ниже прожиточного минимума и увеличение ставки налога по сравнению с действующим для лиц с завышенными доходами. В работе [4] разрабатывается достаточно гладкая функция регрессии с использованием метода ядра при выборе переменных для регрессионной модели. При этом вычислительная процедура не содержит понятия регуляризации. Работа [5] посвящена оцениванию параметров гладкой регрессии в классе условно параметрических моделей ковариантных откликов. В работе [6] рассмотрена задача идентификации неизвестных нестационарных кусочно-линейных параметров для линейной регрессионной модели. В статье [7] рассматривается метод кусочно-линейного нечёткого регрессионного анализа.

Оценивание параметров некоторых кусочно-линейных регрессий. В работах [8–11] приводятся алгоритмы решения задач оценивания неизвестных параметров двух негладких регрессионных зависимостей.

1. Кусочно-линейная регрессия вида

$$y_k = \min\{\alpha_1 x_{k1}, \alpha_2 x_{k2}, ..., \alpha_m x_{km}\} + \varepsilon_k,\ k = \overline{1,n}, \tag{1}$$

где x_i, y – соответственно входные (независимые) и выходная (зависимая) переменные, значения которых известны; α_i, $i = \overline{1,m}$ – подлежащие оцениванию параметры; ε_k – ошибки аппроксимации; n – длина выборки данных. Без потери общности будем предполагать неотрицательность всех переменных модели (1).

В работе [8] показано, что задача идентификации параметров α_i, $i = \overline{1,m}$ зависимости (1) с использованием метода наименьших модулей (МНМ) [12]

$$\sum_{k=1}^{n} |\varepsilon_k| \to \min \tag{2}$$

сводится к следующей задаче линейно-булевого программирования (ЛБП):

$$z_k + u_k - v_k = y_k,\ k = \overline{1,n}, \tag{3}$$

$$z_k \le \alpha_i x_{ki},\ \ k = \overline{1,n},\ \ i = \overline{1,m}, \tag{4}$$

$$\alpha_i x_{ki} - z_k \le (1 - \sigma_{ki})M,\ k = \overline{1,n},\ \ i = \overline{1,m}, \tag{5}$$

$$\sum_{i=1}^{m} \sigma_{ki} = 1,\ k = \overline{1,n}, \tag{6}$$

$$u_k \ge 0,\ v_k \ge 0,\ z_k \ge 0,\ k = \overline{1,n}, \tag{7}$$

$$\sigma_{ki} \in \{0,1\},\ \ k = \overline{1,n},\ \ i = \overline{1,m}, \tag{8}$$

$$\sum_{k=1}^{n} (u_k + v_k) \to \min, \tag{9}$$

где M – заданное большое положительное число.

Таким образом, задача (2) поиска значений неизвестных параметров α_i, $i = \overline{1,m}$ кусочно-линейной регрессии (1) с помощью МНМ сводится к задаче ЛБП (3) – (9) с $mn + 3n + m$ переменными (из которых mn – булевы) и $2(mn + n)$ ограничениями.

Кусочно-линейная модель (1) применяется при исследовании технических, социально-экономических и других систем, когда необходимо отразить то обстоятельство, что увеличение значения выходной переменной невозможно без пропорционального увеличения значений входных факторов (см., например [13]).

2. Функция риска вида

$$y_k = \max\{\alpha_1 x_{k1}, \alpha_2 x_{k2}, ..., \alpha_m x_{km}\} + \varepsilon_k,\ k = \overline{1,n}. \tag{10}$$

Она применяется тогда, когда необходимо формализовать возможность уменьшения значения выходной переменной только при пропорциональном уменьшении значений входных показателей (см., например [10; 14]).

Задача оценивания параметров α_i, $i=\overline{1,m}$ зависимости (10) с использованием метода наименьших модулей также сводится к задаче ЛБП (3) – (5), (6) – (8), (9) после замены ограничений (4), (5) на следующие [11]:

$$z_k \ge \alpha_i x_{ki}, \; k=\overline{1,n}, \; i=\overline{1,m}, \tag{11}$$

$$\alpha_i x_{ki} - z_k \ge (\sigma_{ki}-1)M, \; k=\overline{1,n}, \; i=\overline{1,m}. \tag{12}$$

Вычисление оценок параметров простой формы вложенной кусочно-линейной модели. В работе [15] предложены вложенные кусочно-линейные регрессионные зависимости первого и второго типов:

$$\begin{gathered} y_k = \min\{\min_{i\in I^1}\{\alpha_i^1 x_{ki}\},\ldots,\min_{i\in I^G}\{\alpha_i^G x_{ki}\}, \\ \max_{i\in J^1}\{\beta_i^1 x_{ki}\},\ldots,\max_{i\in J^H}\{\beta_i^H x_{ki}\}\}+\varepsilon_k, k=\overline{1,n} \end{gathered} \tag{13}$$

и

$$\begin{gathered} y_k = \max\{\min_{i\in I^1}\{\alpha_i^1 x_{ki}\},\ldots,\min_{i\in I^G}\{\alpha_i^G x_{ki}\}, \\ \max_{i\in J^1}\{\beta_i^1 x_{ki}\},\ldots,\max_{i\in J^H}\{\beta_i^H x_{ki}\}\}+\varepsilon_k, k=\overline{1,n}, \end{gathered} \tag{14}$$

где индексные множества $I^i, i=\overline{1,G}$, J^i, $i=\overline{1,H}$ являются подмножествами множества $\{1,2,\ldots,m\}$ и могут (однако не обязательно должны) иметь непустые всевозможные попарные пересечения.

Поставим задачу оценивания неизвестных параметров простой формы вложенной кусочно-линейной модели первого типа:

$$y_k = \min\{\min_{i\in I}\{\alpha_i x_{ki}\}, \max_{i\in J}\{\beta_i x_{ki}\}\}+\varepsilon_k, k=\overline{1,n} \tag{15}$$

с использованием метода наименьших модулей. Воспользуемся для этого описанным выше подходом, применённым для идентификации параметров кусочно-линейных моделей (1) и (10).

Отметим, что в настоящей работе мы не рассматриваем вопросы интерпретации вложенных кусочно-линейных моделей (13) – (15).

Введём следующие обозначения:

$$h_k = \min_{i\in I}\{\alpha_i x_{ki}\}, \; k=\overline{1,n},$$

$$g_k = \max_{i\in J}\{\beta_i x_{ki}\}, \; k=\overline{1,n},$$

$$t_k = \min(h_k, g_k), \; k=\overline{1,n}.$$

Тогда задача оценивания параметров $\alpha_i, i\in I$, $\beta_i, i\in J$, по аналогии с задачами (3) – (9) и (3), (6) – (8), (11), (12), (9), сводится к следующей задаче ЛБП:

$$h_k \le \alpha_i x_{ki}, \; k=\overline{1,n}, \; i\in I, \tag{16}$$

$$\alpha_i x_{ki} - h_k \le (1-s_{ki})M, \; k=\overline{1,n}, \; i\in I, \tag{17}$$

$$\sum_{i\in I} s_{ki} = 1, \quad k=\overline{1,n}, \tag{18}$$

$$g_k \ge \beta_i x_{ki}, \; k=\overline{1,n}, \; i\in J, \tag{19}$$

$$\beta_i x_{ki} - g_k \ge (p_{ki} - 1)M\ ,\ k = \overline{1,n},\ \ i \in J\ , \tag{20}$$

$$\sum_{i \in J} p_{ki} = 1,\ \ k = \overline{1,n}, \tag{21}$$

$$t_k \le h_k\ ,\ \ k = \overline{1,n}, \tag{22}$$

$$t_k \le g_k\ ,\ \ k = \overline{1,n}, \tag{23}$$

$$h_k - t_k + Mr_k \le M\ ,\ \ k = \overline{1,n}, \tag{24}$$

$$g_k - t_k - Mr_k \le 0\ ,\ \ k = \overline{1,n}, \tag{25}$$

$$t_k + u_k - v_k = y_k\ ,\ k = \overline{1,n}, \tag{26}$$

$$u_k \ge 0, v_k \ge 0, h_k \ge 0, g_k \ge 0, t_k \ge 0, k = \overline{1,n}\ , \tag{27}$$

$$s_{ki} \in \{0,1\},\ k = \overline{1,n},\ \ i \in I\ , \tag{28}$$

$$p_{ki} \in \{0,1\},\ k = \overline{1,n},\ \ i \in J\ , \tag{29}$$

$$r_k \in \{0,1\},\ k = \overline{1,n}, \tag{30}$$

$$\sum_{k=1}^{n} (u_k + v_k) \to \min\ . \tag{31}$$

Пример. Пусть задана следующая выборка данных:

$$X = \begin{pmatrix} 2 & 4 & 3 \\ 7 & 1 & 9 \\ 6 & 5 & 8 \\ 3 & 8 & 4 \end{pmatrix},\ y = \begin{pmatrix} 3 \\ 4 \\ 2 \\ 9 \end{pmatrix}.$$

Поставим задачу построения простой формы вложенной кусочно-линейной регрессии первого типа с индексными множествами I и J, равными

$$I = \{1,2\}, J = \{2,3\}.$$

В результате решения задачи ЛБП (16) – (31) получим следующую модель:

$$y_k = \min\{\min\{2x_{k1}, 4x_{k2}\}, \max\{0.75x_{k2}, 0.468x_{k3}\} + \varepsilon_k\ ,\ k = \overline{1,4}\ ,$$

при этом вещественные переменные задачи примут значения

$$h = (4,4,12,6),$$

$$g = (3,4.218,3.75,6),$$

$$t = (3,4,3.75,6),$$

$$u = (0,0,0,3),$$

$$v = (0,0,1.75,0).$$

Вызывает интерес исследование сформулированной в работе задачи для интервально заданной выборки с применением соответствующих методов (см., например, [16–18]), а также попытка применить предложенные автором вложенные кусочно-линейные регрессии для анализа минимаксных задач управления риском в сложных производственных системах [19].

Выводы. В работе задача идентификации параметров простой формы вложенной кусочно-линейной регрессионной модели с использованием метода наименьших модулей сведена к задаче линейно-булевого программирования. Решён численный пример.

ЛИТЕРАТУРА

1. Brown S., Heathcote A. On the Use of Nonparametric Regression in Assessing Parametric Regression Models December // Journal of Mathematical Psychology. – 2002. – № 46 (6). – P. 716-730.
2. Shoorekchal M., Eltejaei E. External debt and economic growth in Iran and Malaysia: A smooth transition regression model // Iranian Economic Review. – 2021. – № 25 (2). – P. 323-335.
3. Mkhitaryan V. S., Shishov V. F., Iskorkin D. V. On the Analysis of the Efficiency of a Piecewise-Linear Model for Progressive // Voprosy Statistiki. – 2020. – № 27 (6). – P. 79-85.
4. Ikeda S., Sato Y. Kernel methods for regression model based on variable selection // International Journal of Knowledge Engineering and Soft Data Paradigms. – 2009. – № 1 (1). – P. 49-62.
5. Pal J., Banerjee M. Estimation of smooth regression functions in monotone response models // Journal of Statistical Planning and Inference. – 2008. – № 138 (10). – P. 3125-3143.
6. Идентификация кусочно-линейных параметров регрессионных моделей нестационарных детерминированных систем / В. Цзянь, А. А. Бобцов, С. А. Колюбин, А. А. Пыркин, Л. В. Туан // Автоматика и телемеханика. – 2018. – № 12. – С. 71-82.
7. Изюмов, Б. Д. Кусочно-линейный нечёткий регрессионный анализ данных испытаний скважин / Б. Д. Изюмов // Автоматизация, телемеханизация и связь в нефтяной промышленности. – 2013. – № 11. – С. 22-29.
8. Носков, С. И. Идентификация параметров кусочно-линейной регрессии / С. И. Носков, Р. В. Лоншаков // Информационные технологии и проблемы математического моделирования сложных систем. – 2008. – № 6. – С. 63-64.
9. 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.
10. Носков, С. И. Применение функции риска для моделирования экономических систем / С. И. Носков, А. А. Хоняков // Южно-Сибирский научный вестник. – 2020. – № 5 (33). – С. 85-92.
11. Носков, С. И. Идентификация параметров кусочно-линейной функции риска / С. И. Носков // Транспортная инфраструктура Сибирского региона. – 2017. – Т. 1. – С. 417-421.
12. Носков, С. И. О методе смешанного оценивания параметров линейной регрессии / С. И. Носков // Информационные технологии и математическое моделирование в управлении сложными системами. – 2019. – № 1 (2). – С. 41-45.
13. Носков, С. И. Анализ динамики отдачи от показателей в производственной функции с постоянными пропорциями / С. И. Носков // Вестник СамГУПС. – 2022. – № 4 (58). – С. 109-113.
14. Носков, С. И. Применение функции риска для модельного описания колебания цен на рынке недвижимости / С. И. Носков, А. А. Хоняков // Инженерно-строительный вестник Прикаспия. – 2021. – № 3 (37). – С. 77-82.
15. Носков, С. И. Подход к формализации вложенной кусочно-линейной регрессии / С. И. Носков // Международный журнал гуманитарных и естественных наук. – 2023. – № 1-2 (76). – С. 218-220.
16. Kreinovich V., Lakeyev A. V., Noskov S. I. Approximate linear algebra is intractable // Linear Algebra and its Applications. – 1996. – Т. 232. – № 1-3. – С. 45-54.
17. Lakeyev A. V., Noskov S. I. A description of the set of solutions of a linear equation with intervally defined operator and right-hand side // Doklady Mathematics. – 1993. – Т. 47. № 3. – P. 518-523.
18. Носков, С. И. Точечная характеризация множеств решений интервальных систем линейных алгебраических уравнений / С. И. Носков // Информационные технологии и математическое моделирование в управлении сложными системами. – 2018. – № 1. – С. 8-13.
19. Золотова, Т. В. Линейные минимаксные задачи управления риском в сложных производственных системах / Т. В. Золотова // Учёные записки Комсомольского-на-Амуре государственного технического университета. Науки о природе и технике. – 2010. – № II-1 (2). – С. 8-13.