

ИДЕНТИФИКАЦИЯ ПАРАМЕТРОВ КОМБИНИРОВАННОЙ КУСОЧНО-ЛИНЕЙНОЙ РЕГРЕССИОННОЙ МОДЕЛИ

Носков Сергей Иванович

*доктор технических наук, профессор,
профессор кафедры «Информационные системы и защита информации»,
Иркутский государственный университет путей сообщения
Иркутск, Россия
E-mail: sergey.noskov.57@mail.ru*

Предмет исследования: задача оценивания параметров комбинированной кусочно-линейной регрессионной модели.

Цель исследования: применить аппарат линейно-булевого программирования для идентификации ее параметров.

Методы и объекты исследования: объектом исследования является совокупность показателей изобретательской активности в Российской Федерации, методами – регрессионный анализ и математическое программирование.

Основные результаты исследования: описан подход к оцениванию параметров комбинированной кусочно-линейной регрессионной модели посредством применения метода наименьших модулей, что позволяет свести эту задачу к задаче линейно-булевого программирования. Построена комбинированная кусочно-линейная модель изобретательской активности в Российской Федерации, которая может использоваться при проведении различных аналитических и прогнозных расчетов. Выходной переменной модели является количество заявок на выдачу патента, а входными – объем валового внутреннего продукта, численность аспирантов и профессорско-преподавательского состава.

Ключевые слова: регрессионная модель, кусочно-линейная регрессия, метод наименьших модулей, задача линейного-булевого программирования, изобретательская активность, патенты.

IDENTIFICATION OF PARAMETERS OF A COMBINED PIECE-LINEAR REGRESSION MODEL

Sergey I. Noskov

*Doctor of Technical Sciences, Professor,
Professor of the Department of Information Systems and Information Security,
Irkutsk State Transport University
Irkutsk, Russia
E-mail: sergey.noskov.57@mail.ru*

Subject of research: the problem of estimating the parameters of a combined piecewise linear regression model.

Purpose of research: to apply the apparatus of linear Boolean programming to identify its parameters.

Methods and objects of research: the object of research is a set of indicators of inventive activity in the Russian Federation, the methods are regression analysis and mathematical programming.

Main results of research: an approach to estimating the parameters of a combined piecewise linear regression model by using the method of least modules is described, which makes it possible to reduce this problem to a linear Boolean programming problem. A combined piecewise linear model of inventive activity in the Russian Federation has been constructed, which can be used in various analytical and predic-

tive calculations. The output variable of the model is the number of patent applications, and the input variable is the volume of gross domestic product, the number of graduate students and teaching staff.

Keywords: regression model, piecewise linear regression, least modules method, linear Boolean programming problem, inventive activity, patents.

Введение

Весьма часто при исследовании методами математического моделирования сложных технических и социально-экономических объектов исследователи вынуждены применять наряду с линейными различные нелинейные конструкции, в частности, кусочного типа. Так, в работе [1] рассматриваются тендерные предложения китайских компаний, зарегистрированных на бирже. При этом посредством применения непараметрического метода и модели кусочно-линейной регрессии доказывается, что ценообразование предложений имеет эффект привязки. В [2] рассмотрено рассеяние гармонической во времени акустической волны на ограниченном анизотропном неоднородном препятствии, погруженном в неограниченную анизотропную однородную среду, в предположении, что граница препятствия является липшицевой поверхностью. Математическая модель сформулирована в виде краевой задачи переноса для эллиптического уравнения в частных производных типа Гельмгольца второго порядка с кусочно-липшиц-непрерывными переменными коэффициентами. Статья [3] посвящена применению метода кусочной линеаризации для решения проблемы ограничения потока природного газа в трубопроводе, что достигается преобразованием исходной модели в задачу смешанного целочисленного программирования. В [4] изучаются максимально допустимые скорости судов на маршруте из-за таких географических особенностей, как каналы, протоки и океанские течения. При этом сначала формулируется задача нелинейного программирования с последующим ее преобразованием в задачу линейного программирования путем использования кусочно-линейных аппроксимирующих функций. Кроме того, представлена эвристическая процедура для сокращения времени вычислений для задач большой размерности. В работе [5] для модельного описания затрат на топливо для генерирующих установок используется кусочно-квадратичная функция. В [6] осуществляется детальный анализ экономической модели объема производства с дефицитом при использовании общей функции нормы стоимости запасов и кусочно-линейных вогнутых производственных затратах.

Результаты и обсуждение

Оценивание параметров комбинированной кусочно-линейной регрессии как задача линейно-булевого программирования

Пусть при исследовании некоторого сложного объекта произвольной природы методами регрессионного анализа поведение некоторого выходного фактора (зависимой переменной) y определяется значениями входных показателей (независимых переменных) $x_i, i = \overline{1, m}$. Наиболее часто при этом (см., например, [7-9]) используется линейная модель (регрессионное уравнение):

$$y_k = \sum_{i=1}^m \alpha_i x_{ki} + \varepsilon_k, \quad k = \overline{1, n}, \quad (1)$$

где k – номер наблюдения, n – длина выборки, ε_k – ошибки аппроксимации, не имеющие вероятностной природы, а представляющие собой погрешности модельного описания реального процесса, α – вектор оцениваемых параметров. Значительно менее употребительными являются кусочно-линейные модели (см., например, [10-14]), обладающие некоторыми замечательными содержательными свойствами:

$$y_k = \min \{ \alpha_1 x_{k1}, \alpha_2 x_{k2}, \dots, \alpha_m x_{km} \} + \varepsilon_k, \quad k = \overline{1, n}. \quad (2)$$

и

$$y_k = \max \{ \alpha_1 x_{k1}, \alpha_2 x_{k2}, \dots, \alpha_m x_{km} \} + \varepsilon_k, \quad k = \overline{1, n}. \quad (3)$$

Одним из таких свойств является невозможность наращивания (снижения) значений зависимой переменной без зафиксированного в моделях (2) или (3) пропорционального роста (уменьшения) значений всех независимых факторов. В работах [10-14] подробно описаны способы оценивания параметров регрессионных уравнений (2) и (3).

В определенной мере обобщающей модельные конструкции (1) и (2) формой является комбинированная кусочно-линейная модель (в работе [15] несколько схожая модель названа линейно-неэлементарной) вида:

$$y_k = \sum_{i \in I_1} \alpha_i x_{ki} + \min_{j \in I_2} \{ \beta_j x_{kj} \} + \varepsilon_k. \quad (4)$$

Здесь I_1 и I_2 – индексные множества, такие, что:

$$I_1 \cup I_2 = \{1, 2, \dots, m\}.$$

При этом выполнения обычно накладываемого в подобных случаях условия

$$I_1 \cap I_2 = \emptyset$$

не требуется, поскольку допускается вхождение некоторых независимых переменных одновременно и в линейную, и в кусочно-линейную составляющие модели (4). Следовательно, в общем случае $|I_1| + |I_2| \geq m$, где $|A|$ – число элементов в множестве A .

Идентификацию параметров $\alpha_i, i \in I_1, \beta_j, j \in I_2$ будем проводить на основе применения метода наименьших модулей (МНМ) посредством минимизации функции $J(\alpha, \beta)$:

$$J(\alpha, \beta) = \sum_{k=1}^n |y_k - \sum_{i \in I_1} \alpha_i x_{ki} - \min_{j \in I_2} \{ \beta_j x_{kj} \}| \rightarrow \min. \quad (5)$$

Используем применяемый при оценивании параметров кусочно-линейных регрессий способ сведения МНМ для них к задаче линейно-булевого программирования (ЛБП). Она для задачи (5) примет вид:

$$\sum_{i \in I_1} \alpha_i x_{ki} + z_k + u_k - v_k = y_k, \quad k = \overline{1, n}, \quad (6)$$

$$z_k \leq \beta_j x_{kj}, \quad k = \overline{1, n}, \quad j \in I_2, \quad (7)$$

$$\beta_j x_{kj} - z_k \leq (1 - \sigma_{kj}) M, \quad k = \overline{1, n}, \quad j \in I_2, \quad (8)$$

$$\sum_{j \in I_2} \sigma_{kj} = 1, \quad k = \overline{1, n}, \quad (9)$$

$$\sigma_{ij} \in \{0, 1\}, \quad k = \overline{1, n}, \quad j \in I_2, \quad (10)$$

$$J(\alpha, \beta) = \sum_{k=1}^n (u_k + v_k) \rightarrow \min. \quad (11)$$

Здесь M – заранее выбранное большое положительное число.

Задача ЛБП (6) – (11) содержит $|I_1| + |I_2| + n(2 + |I_2|)$ переменных, из которых $n|I_2|$ – булевы, и $2n(1 + |I_2|)$ ограничений.

Комбинированная кусочно-линейная модель изобретательской активности

Используем описанный выше алгоритм оценивания неизвестных параметров комбинированной кусочно-линейной регрессии (4) для построения модели изобретательской актив-

ности в Российской Федерации. Отметим при этом, что подобные проблемы рассматривались в ряде публикаций (см., например, [16–19]).

Введем в рассмотрение переменные:

y – количество заявок на выдачу патента в Российской Федерации (шт.);

x_1 – объем валового внутреннего продукта (млрд. руб.);

x_2 – численность аспирантов на конец года (чел.);

x_3 – численность профессорско-преподавательского состава (тыс. чел.).

В качестве информационной базы исследованием будем использовать статистическую информацию по выделенным факторам за 2000–2020 гг. [19]). Сформируем индексные множества I_1 и I_2 следующим образом: $I_1 = \{1\}$, $\{2, 3\}$. При оценивании параметров регрессии (5) воспользуемся программой [20]. В результате получим следующую комбинированную кусочно-линейную модель:

$$y = -2421.9 + 0.177x_1 + \min\{0.242x_2, 101.12x_3\}, \quad (12)$$

$$E = 5.68\%, s = (2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2).$$

Здесь E – средняя относительная ошибка аппроксимации, s – вектор срабатываний [21], произвольная компонента которого s_k рассчитывается по правилу:

$$s_k = \begin{cases} 2, & \text{если } 0.242x_{k2} \leq 101.12x_{k3} \\ 3, & \text{в противном случае.} \end{cases}$$

Отметим, что модель (12) обладает весьма приемлемой точностью, что позволяет эффективно ее использовать при проведении различных аналитических и прогнозных расчетов. Анализ вектора срабатываний s позволяет сделать вывод о том, что рост количества заявок на выдачу патента в Российской Федерации в 2010–2012 годах сдерживался численностью профессорско-преподавательского состава, в остальные годы ретроспективного периода – численностью аспирантов. Это обстоятельство указывает на высокую значимость деятельности последних в сфере изобретательской деятельности.

Заключение и выводы

В работе рассмотрена комбинированная кусочно-линейная регрессионная модель, слагаемыми которой являются линейная и кусочная составляющие. Описан алгоритмический способ оценивания ее параметров посредством применения метода наименьших модулей, что позволяет свести эту задачу к задаче линейно-булевого программирования. Разработана комбинированная кусочно-линейная модель изобретательской активности в Российской Федерации, которая может использоваться при проведении различных аналитических и прогнозных расчетов.

Литература

1. Wang, Y. Market reaction to tender offers: Insights from China / Y. Wang, Y. Qian // Journal of University of Science and Technology of China. – 2021. – 51(12). – P. 894–911.
2. Gorgisheli, S. Boundary-transmission problems of the theory of acoustic waves for piecewise inhomogeneous anisotropic multi-component lipschitz domains / S. Gorgisheli, M. Mrevlishvili, D. Natroshvili // Transactions of A. Razmadze Mathematical Institute. – 2020. – 174(3). – P. 303–324.
3. Research on double-layer optimized configuration of multi-energy storage in regional integrated energy system with connected distributed wind / Q. Zhang, Z. Ren, R. Ma [et al] // Energies. – 2019. – 12(20). – P. 3964–3976.
4. Planning of vessel speed and fuel bunkering over a route with speed limits / I. E. Nielsen, N. A. Dung Do, J. Jang, G. Bocewicz // Maritime Economics and Logistics. – 2016. – 18(4). – P. 414–435.
5. Solving economic dispatch problem using particle swarm optimization by an evolutionary technique for initializing particles / R. Rahmani, M. F. Othman, R. Yusof, M. Khalid // Journal of Theoretical and Applied Information Technology. – 2012. – 46(2). – P. 526–536.

6. Deterministic inventory/production model with general inventory cost rate function and piecewise linear concave production costs / Z. P. Bayindir, S. I. Birbil, J. B. G. A Frenk // *European Journal of Operational Research*. – 2007. – 179(1). – P. 114–123.
7. Агеев, С. П. Моделирование энергопотребления поточной линии производства пилопродукции / С. П. Агеев, В. И. Жабин, В. И. Мелехов. – Текст : непосредственный // *Актуальные проблемы лесного комплекса*. – 2012. – № 34. – С. 21–24.
8. Карпенко, С. М. Прогнозирование электропотребления на горнопромышленных предприятиях с использованием статистических методов / С. М. Карпенко, Н. В. Карпенко, Г. Ю. Безгинов. – Текст : непосредственный // *Горная промышленность*. – 2022. – № 1. – С. 82–88.
9. Радковская, Е. В. Экономико-статистический анализ промышленного развития регионов России / Е. В. Радковская, Г. В. Радковский. – Текст : непосредственный // *Фундаментальные исследования*. – 2019. – № 10. – С. 69–75.
10. Носков, С. И. Построение кусочно-линейной авторегрессионной модели произвольного порядка / С. И. Носков. – Текст : непосредственный // *Вестник Югорского государственного университета*. – 2022. – № 2 (65). – С. 89–94.
11. Носков, С. И. Идентификация параметров кусочно-линейной регрессии / С. И. Носков, Р. В. Лоншаков. – Текст : непосредственный // *Информационные технологии и проблемы математического моделирования сложных систем*. – 2008. – № 6. – С. 63–64.
12. Носков, С. И. Программный комплекс построения некоторых типов кусочно-линейных регрессий / С. И. Носков, А. А. Хоняков. – Текст : непосредственный // *Информационные технологии и математическое моделирование в управлении сложными системами*. – 2019. – № 3 (4). – С. 47–55.
13. Носков, С. И. Идентификация параметров кусочно-линейной функции риска / С. И. Носков. – Текст : непосредственный // *Транспортная инфраструктура Сибирского региона*. – 2017. – Т. 1. – С. 417–421.
14. Носков, С. И. Применение функции риска для моделирования экономических систем / С. И. Носков, А. А. Хоняков. – Текст : непосредственный // *Южно-Сибирский научный вестник*. – 2020. – № 5 (33). – С. 85–92.
15. Базилевский, М. П. Оценивание линейно-неэлементарных регрессионных моделей с помощью метода наименьших квадратов / М. П. Базилевский // *Моделирование, оптимизация и информационные технологии*. – 2020. – Т. 8. – № 4(31).
16. Миролюбова, Т. В. Методические подходы к анализу факторов, влияющих на развитие инновационной системы университета / Т. В. Миролюбова, Л. О. Соломатова. – Текст : непосредственный // *Фундаментальные исследования*. – 2012. – № 11. – С. 1004–1010.
17. Гуриева, Л. К. Анализ инновационной ситуации в регионах Южного федерального округа / Л. К. Гуриева, И. С. Шапка. – Текст : непосредственный // *Известия высших учебных заведений. Северо-Кавказский регион. Общественные науки*. – 2005. – № 4(132). – С. 51–57.
18. Дмитриева, О. А. Регрессионная модель зависимости количества патентов от объема затрат на финансирование научных исследований и разработок / О. А. Дмитриева, Т. Н. Батова. – Текст : непосредственный // *Альманах научных работ молодых ученых Университета ИТМО*. – 2016. – С. 40–42.
19. Носков, С. И. Модификация непрерывной формы метода максимальной согласованности при построении линейной регрессии / С. И. Носков, Ю. А. Бычков. – Текст : непосредственный // *Известия Тульского государственного университета. Технические науки*. – 2022. – № 5. – С. 88–94.
20. Свидетельство о государственной регистрации программы для ЭВМ № 2022617773 Российская Федерация. Программа построения комбинированной кусочно-линейной регрессионной модели: № 2022616675 : заявл. 14.04.2022 : опубл. 25.04.2022 / С. И. Носков, А. А. Хоняков ; заявитель ФГБОУВО «Иркутский государственный университет путей сообщения». – Текст : непосредственный.
21. Носков, С. И. Кусочно-линейные регрессионные модели объемов перевозки пассажиров железнодорожным транспортом / С. И. Носков, А. А. Хоняков. – Текст : непосредственный // *Модели, системы, сети в экономике, технике, природе и обществе*. – 2021. – № 4 (40). – С. 80–89.