

Научная статья

УДК 519.237:519.854.33

DOI 10.35266/1999-7604-2024-1-7

ВЫЧИСЛЕНИЕ ПАРАМЕТРОВ ПРОСТОЙ ФОРМЫ ВЛОЖЕННОЙ КУСОЧНО-ЛИНЕЙНОЙ РЕГРЕССИИ МЕТОДОМ СМЕШАННОГО ОЦЕНИВАНИЯ

Сергей Иванович Носков¹✉, Алексей Николаевич Знайдюк²

^{1,2}Иркутский государственный университет путей сообщения, Иркутск, Россия

¹sergey.noskov.57@mail.ru✉, <https://orcid.org/0000-0003-4097-2720>

²znaidyuk00@gmail.com

Аннотация. В работе описан алгоритмический способ определения численных оценок параметров простой формы вложенной кусочно-линейной регрессии методом смешанного оценивания. Его суть состоит в их одновременной идентификации методами наименьших модулей и антиробастного оценивания, каждый из которых «работает» на своей подвыборке данных исходной выборки. Этот способ сводится к решению задачи линейно-булевого программирования. Решен численный пример.

Ключевые слова: регрессионная модель, простая форма вложенной кусочно-линейной регрессии, методы наименьших модулей, антиробастного и смешанного оценивания, задача линейно-булевого программирования

Для цитирования: Носков С. И., Знайдюк А. Н. Вычисление параметров простой формы вложенной кусочно-линейной регрессии методом смешанного оценивания // Вестник кибернетики. 2024. Т. 23, № 1. С. 54–59. DOI 10.35266/1999-7604-2024-1-7.

Original article

CALCULATING PARAMETERS OF A SIMPLE NESTED PIECEWISE LINEAR REGRESSION USING MIXED ESTIMATION

Sergey I. Noskov¹✉, Aleksey N. Znaidyuk²

^{1,2}Irkutsk State Transport University, Irkutsk, Russia

¹sergey.noskov.57@mail.ru✉, <https://orcid.org/0000-0003-4097-2720>

²znaidyuk00@gmail.com

Abstract. The article describes an algorithm for calculating numerical estimates of the parameters of a simple nested piecewise linear regression using mixed estimation. The algorithm aims at simultaneous identification using methods of least modules and antirobust estimation, both of which operate on certain subsamplings from the initial sampling. The approach comes to solving the problem of linear Boolean programming. A numerical problem has been solved.

Keywords: regression model, simple nested piecewise linear regression, methods of the least modules, antirobust and mixed estimation, linear Boolean programming problem

For citation: Noskov S. I., Znaidyuk A. N. Calculating parameters of a simple nested piecewise linear regression using mixed estimation. *Proceedings in Cybernetics*. 2024;23(1):54–59. DOI 10.35266/1999-7604-2024-1-7.

ВВЕДЕНИЕ

При построении математических моделей регрессионного типа исследователи часто эффективно используют различные нелинейные формы связи между независимыми переменными, в том числе кусочно-линейные. Так, в работе [1] рассматривается робастная модель кусочно-линейной регрессии с неизвестным количеством точек переключения. Статья [2] посвящена методу оптимизации невыпуклой кусочно-линейной модели многомерной регрессии, адаптированной для решения проблемы проектирования системы безопасности для крупного автопроизводителя США и решенной с помощью метода ветвей и границ. В [3] представлена процедура автоматической спецификации моделей, основанных на предположениях об аддитивности и кусочно-линейности. Эта процедура позволяет аналитику получить представление о проблеме при исследовании автоматически выбранной модели, что позволяет легко проверить достоверность прогноза. В работе [4] предлагается использовать нечеткую кусочную регрессионную модель как развитие традиционной логистической модели. Предлагаемый метод способен не только обнаруживать точки переключения, но и определять прогнозируемые интервалы, когда на тенденцию роста анализируемого поколения продуктов влияют другие поколения. Для демонстрации эффективности предлагаемой модели используются рыночные доли четырех телевизионных технологий. Результаты показывают, что предложенная модель превосходит логистическую модель, обеспечивая как лучшие, так и худшие возможные доли рынка для соответствующего поколения, а также определяя время воздействия внешних воздействий путем выявления точек переключения.

Целью исследования [5] является применение кусочно-линейной аппроксимации (КЛА) в качестве метода анализа тенденций, который учитывает структурные сдвиги. КЛА использует полные данные для одновременной оценки контрольных точек и постоянно связанных тенденций непосредственно до и после пере-

рыва. Таким образом, КЛА не только обеспечивает простоту интерпретации результатов, но и исключает вероятность неверных заключений за счет однозначного определения текущей тенденции, что делает расчетный результат надежным. Тематическое исследование подтверждает предположение о том, что тенденции продаж некоторых продуктов и необходимость определения подходящего временного интервала для анализа данных претерпевали изменения по крайней мере один раз. Статья [6] посвящена разработке алгоритма решения задач построения многомерной регрессии и классификации с использованием кусочно-линейных предикторов над многогранным разбиением пространства признаков. В работе [7] изучается частично линейная аддитивная регрессия с пространственными данными. Разработана новая процедура оценки неизвестных параметров и аддитивных компонентов регрессии. Предложенный метод подходит для данных большой размерности, нет необходимости решать ограниченную задачу минимизации, не требуются итерационные алгоритмы. В мягких предположениях регулярности установлено асимптотическое распределение оценки вектора неизвестных параметров, а также получены асимптотические распределения оценок неизвестных функций. В работе [8] рассмотрены методы построения цензурированной регрессии, которые обеспечивают мощный и гибкий подход к анализу данных о выживаемости, когда считаются подходящими, в частности стандартные линейные модели. Исследование [9] посвящено способам построения кусочно-экспоненциальных моделей. Широка практическая направленность применения кусочно-линейных регрессий, они, в частности, используются при прогнозировании чрезвычайных ситуаций [10], оценке состояния технического объекта [11], анализе данных испытаний скважин [12].

МАТЕРИАЛЫ И МЕТОДЫ

Пусть при исследовании анализируемого объекта исследователь полагает, что на выходной фактор (зависимую переменную)

у оказывают влияние входные факторы (независимые переменные) $x_1, x_2, \dots, x_m$, т. е. исходит из наличия регрессионной связи:

$$y_k = F(a; x_{k1}, x_{k2}, \dots, x_{km}) + \varepsilon_k, \quad k = \overline{1, n},$$

где k – номер наблюдения,

n – их количество,

F – аппроксимирующая функция,

a – вектор параметров,

ε – ошибки аппроксимации, при этом будем исходить из отсутствия какой-либо информации об их вероятностной природе, как это и принято в рамках логико-алгебраического подхода к анализу данных.

В работе [13] введены в рассмотрение так называемые вложенные кусочно-линейные аппроксимирующие функции первого и второго типов:

– вложенная кусочно-линейная регрессия первого типа:

$$y_k = \min \{ \min_{i \in I^1} \{a_i^1 x_{ki}\}, \dots, \min_{i \in I^G} \{a_i^G x_{ki}\}, \max_{i \in J^1} \{\beta_i^1 x_{ki}\}, \dots, \max_{i \in J^H} \{\beta_i^H x_{ki}\} \} + \varepsilon_k;$$

– вложенная кусочно-линейная регрессия второго типа:

$$y_k = \max \{ \min_{i \in I^1} \{a_i^1 x_{ki}\}, \dots, \min_{i \in I^G} \{a_i^G x_{ki}\}, \max_{i \in J^1} \{\beta_i^1 x_{ki}\}, \dots, \max_{i \in J^H} \{\beta_i^H x_{ki}\} \} + \varepsilon_k,$$

где наперед заданные индексные множества $I^i, i = \overline{1, G}, J^i, i = \overline{1, H}$ представляют собой подмножества множества $\{1, 2, \dots, m\}$ и могут иметь непустые попарные пересечения.

В [14] приводятся, в частности, некоторые возможные конкретизации вложенных кусочно-линейных регрессий, в том числе простая вложенная кусочно-линейная регрессия:

$$y_k = \min \{ \min_{i \in I} \{a_i x_{ki}\}, \max_{i \in J} \{\beta_i x_{ki}\} \} + \varepsilon_k, \quad (1)$$

Наконец, в работе [15] описан алгоритмический способ идентификации параметров модели (1) с помощью метода наименьших модулей (МНМ), состоящего в решении задачи:

$$\sum_{k=1}^n |\varepsilon_k| \rightarrow \min. \quad (2)$$

В [15] показано, что задача (2) может быть сведена к задаче линейно-булевого программирования (ЛБП). Действительно, введем следующие обозначения:

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

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

$$t_k = \min \{h_k, g_k\}, \quad k = \overline{1, n}.$$

Тогда задача оценивания параметров эквивалентна следующей задаче ЛБП:

$$h_k \leq a_i x_{ki}, \quad k = \overline{1, n}, \quad i \in I, \quad (3)$$

$$a_i x_{ki} - h_k \leq (1 - s_{ki})M, \quad k = \overline{1, n}, \quad i \in I, \quad (4)$$

$$\sum_{i \in J} s_{ki} = 1, \quad k = \overline{1, n}, \quad (5)$$

$$g_k \geq \beta_i x_{ki}, \quad k = \overline{1, n}, \quad i \in I, \quad (6)$$

$$\beta_i x_{ki} - g_k \geq (\rho_{ki} - 1)M, \quad k = \overline{1, n}, \quad i \in J, \quad (7)$$

$$\sum_{i \in J} \rho_{ki} = 1, \quad k = \overline{1, n}, \quad (8)$$

$$t_k \leq h_k, \quad k = \overline{1, n}, \quad (9)$$

$$t_k \leq g_k, \quad k = \overline{1, n}, \quad (10)$$

$$h_k - t_k + Mr_k \leq M, \quad k = \overline{1, n}, \quad (11)$$

$$g_k - t_k - Mr_k \leq 0, \quad k = \overline{1, n}, \quad (12)$$

$$t_k + u_k - v_k = y_k, \quad k = \overline{1, n}, \quad (13)$$

$$u_k \geq 0, \quad v_k \geq 0, \quad h_k \geq 0, \quad g_k \geq 0, \quad t_k \geq 0, \quad k = \overline{1, n}, \quad (14)$$

$$s_{ki} \in \{0, 1\} \quad k = \overline{1, n}, \quad i \in I, \quad (15)$$

$$p_{ki} \in \{0, 1\} \quad k = \overline{1, n}, \quad i \in I, \quad (16)$$

$$r_{ki} \in \{0, 1\} \quad k = \overline{1, n}, \quad (17)$$

$$\sum_{k=1}^n (u_k + v_k) \rightarrow \min. \quad (18)$$

Заметим, что для недопустимости получения бессмысленного результата (крайне редкого в реальных ситуациях) полезно несколько скорректировать целевую функцию задачи ЛБП (3)–(18) следующим образом:

$$\sum_{k=1}^n (u_k + v_k) + \delta (\sum_{i \in I} \alpha_i - \sum_{i \in J} \beta_i) \rightarrow \min, \quad (19)$$

где δ – наперед заданная малая положительная константа.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Поставим теперь задачу определения параметров модели (2) с помощью метода смешанного оценивания (МСО) [16]. Его суть состоит в следующем; пусть, исходя из некоторых эмпирических или теоретических принципов, множество номеров всех наблюдений $P = \{1, 2, \dots, n\}$ в простейшем случае может быть раз-

бито на два непересекающихся подмножества $P_1 \subset P$ и $P_2 \subset P$:

$$P_1 \cup P_2 = P, P_1 \cap P_2 = \emptyset.$$

При этом минимизация ошибок аппроксимации на первом из них производится в соответствии с МНМ:

$$L^1(\alpha) = \sum_{k \in P_1} |\varepsilon_k| \rightarrow \min,$$

а на втором – с методом антиробастного оценивания:

$$L^2(\alpha) = \max_{k \in P_2} |\varepsilon_k| \rightarrow \min.$$

В совмещении этих двух задач, т. е. в одновременной минимизации функций $L^1(\alpha)$ и $L^2(\alpha)$, и состоит идея метода смешанного оценивания.

Для реализации МСО при вычислении параметров простой вложенной кусочно-линейной регрессии (1) введем в систему ограничений задачи ЛБП (3)–(17), (19) новые неравенства:

$$u_k + v_k \leq b, k \in P_2, \quad (20)$$

и заменим целевую функцию (19) на следующую:

$$\sum_{k \in P_1} (u_k + v_k) / s + b + \eta \sum_{k \in P_2} (u_k + v_k) + \delta (\sum_{i \in I} \alpha_i - \sum_{i \in J} \beta_i) \rightarrow \min, \quad (21)$$

где s – число элементов в множестве, – заранее заданное малое положительное число.

Поставим задачу построения простой вложенной кусочно-линейной регрессии (1) с помощью МСО на основе данных из работы [15]:

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

Сформируем множества I , J , P_1 и P_2 следующим образом:

$$\begin{aligned} I &= \{1, 2\}, \\ J &= \{2, 3\}, \\ P_1 &= \{2, 3\}, \end{aligned}$$

$$P_2 = \{1, 4\}.$$

Результат решения этой задачи на основе применения МНМ состоит в следующем [15]:

$$\begin{aligned} y_k &= \min\{\min\{2x_{k1}, 4x_{k2}\}, \\ &\max\{0.75x_{k2}, 0.468x_{k3}\}\} + \varepsilon_k, k = \overline{1, 4}, \quad (22) \\ h &= (4, 4, 12, 6), \\ g &= (3, 4.218, 3.75, 6), \\ t &= (3, 4, 3.75, 6), \\ u &= (0, 0, 0, 3), \\ v &= (0, 0, 1.75, 0), \\ L^1 &= 1.75, \\ L^2 &= 3. \end{aligned}$$

В результате применения МСО путем решения задачи ЛБП (3)–(17), (20), (21) получим:

$$\begin{aligned} y_k &= \min\{\min\{2.66x_{k1}, 4x_{k2}\}, \\ &\max\{1.0x_{k2}, 0.624x_{k3}\}\} + \varepsilon_k, k = \overline{1, 4}, \quad (23) \\ h &= (5.33, 4, 16, 8), \\ g &= (4, 5.62, 5, 8), \\ t &= (4, 4, 5, 8), \\ u &= (0, 0, 0, 1), \\ v &= (1, 0, 3, 0), \\ L^1 &= 3, \\ L^2 &= 1. \end{aligned}$$

Таким образом, по критерию L^1 модель (23) на 1.25 уступает модели (22), но по критерию L^2 превосходит ее на 2. При этом все оценки параметров модели (23) не меньше оценок модели (22).

ЗАКЛЮЧЕНИЕ

В работе продолжены исследования, связанные с разработкой различных форм вложенных кусочно-линейных регрессионных моделей. Предложен алгоритмический способ определения численных оценок параметров простой формы такой регрессии методом смешанного оценивания, сводящийся к решению задачи линейно-булевого программирования.

Список источников

1. Shi S., Li Y., Wan C. Robust continuous piecewise linear regression model with multiple change points. *J Supercomput.* 2020;76:3623–3645.
2. Martinez N., Anahideh H., Rosenberger J. M. et al. Global optimization of non-convex piecewise linear regression splines. *J Glob Optim.* 2017;68:563–586.
3. Gascón A., Sánchez-Úbeda E. F. Automatic specification of piecewise linear additive models: Application to forecasting natural gas demand. *Stat Comput.* 2017;28:201–217.
4. Yu J. R., Tseng F.-M. Fuzzy piecewise logistic growth model for innovation diffusion: A case study of the TV Industry. *Int J Fuzzy Syst.* 2015;18:511–522.
5. Moriyama T., Kuwano M., Nakayama M. A statistical method for estimating piecewise linear sales trends. *J Market Anal.* 2023. URL: <https://link.springer.com/article/10.1057/s41270-023-00207-9> (дата обращения: 04.01.2024).
6. Bemporad A. A piecewise linear regression and classification algorithm with application to learning and model predictive control of hybrid systems. *IEEE Transactions on Automatic Control.* 2023;68:3194–3209.
7. Qingguo T., Wenyu C. Estimation for partially linear additive regression with spatial data. *Stat Papers.* 2022;63:2041–2063.
8. Neocleous T., Portnoy S. Partially linear censored quantile regression. *Lifetime Data Anal.* 2009;15:357–378.
9. Demarqui F. N., Loschi R. H., Colosimo E. A. Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Anal.* 2008;14:333–356.
10. Жижин К. С., Благородова Н. В. Использование кусочно-линейной регрессии в прогнозировании чрезвычайных ситуаций // Международный журнал экспериментального образования. 2016. № 5–3. С. 337–338.
11. Клячкин В. Н., Бубыр Д. С. Прогнозирование состояния технического объекта на основе кусочно-линейных регрессий // Радиотехника. 2014. № 7. С. 137–140.
12. Изюмов Б. Д. Кусочно-линейный нечеткий регрессионный анализ данных испытаний скважин // Автоматизация, телемеханизация и связь в нефтяной промышленности. 2013. № 11. С. 22–29.
13. Носков С. И. Подход к формализации вложенной кусочно-линейной регрессии // Международный журнал гуманитарных и естественных наук. 2023. № 1–2. С. 218–220.
14. Носков С. И. Некоторые формы вложенной кусочно-линейной регрессии // Известия Тульского государственного университета. Технические науки. 2023. № 3. С. 467–469.
15. Носков С. И. Идентификация параметров простой формы вложенной кусочно-линейной регрессии // Ученые записки Комсомольского-на-Амуре го-

References

1. Shi S., Li Y., Wan C. Robust continuous piecewise linear regression model with multiple change points. *J Supercomput.* 2020;76:3623–3645.
2. Martinez N., Anahideh H., Rosenberger J. M. et al. Global optimization of non-convex piecewise linear regression splines. *J Glob Optim.* 2017;68:563–586.
3. Gascón A., Sánchez-Úbeda E. F. Automatic specification of piecewise linear additive models: Application to forecasting natural gas demand. *Stat Comput.* 2017;28:201–217.
4. Yu J. R., Tseng F.-M. Fuzzy piecewise logistic growth model for innovation diffusion: A case study of the TV Industry. *Int J Fuzzy Syst.* 2015;18:511–522.
5. Moriyama T., Kuwano M., Nakayama M. A statistical method for estimating piecewise linear sales trends. *J Market Anal.* 2023. URL: <https://link.springer.com/article/10.1057/s41270-023-00207-9> (accessed: 04.01.2024).
6. Bemporad A. A piecewise linear regression and classification algorithm with application to learning and model predictive control of hybrid systems. *IEEE Transactions on Automatic Control.* 2023;68:3194–3209.
7. Qingguo T., Wenyu C. Estimation for partially linear additive regression with spatial data. *Stat Papers.* 2022;63:2041–2063.
8. Neocleous T., Portnoy S. Partially linear censored quantile regression. *Lifetime Data Anal.* 2009;15:357–378.
9. Demarqui F. N., Loschi R. H., Colosimo E. A. Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Anal.* 2008;14:333–356.
10. Zhizhin K. S., Blagorodova N. V. Ispolzovanie kusochno-lineinoi regressii v prognozirovanii chrezvychainykh situatsii. *Mezhdunarodnyi zhurnal eksperimentalnogo obrazovaniia.* 2016;(5–3):337–338. (In Russian).
11. Klyachkin V. N., Buby D. S. Forecasting of technical object state based on piecewise linear regressions. *Radiotekhnika.* 2014;(7):137–140. (In Russian).
12. Izumov B. D. Piecewise-linear fuzzy regression analysis of well testing data. *Automation, Telemechanization and Communication in Oil Industry.* 2013;(11):22–29. (In Russian).
13. Noskov S. I. Approach to formalizing nested piece-linear regression. *International Journal of Humanities and Natural Sciences.* 2023;(1–2):218–220. (In Russian).
14. Noskov S. I. Some forms of nested piece-linear regression. *News of the Tula State University. Technical Sciences.* 2023;(3):467–469. (In Russian).
15. Noskov S. I. Parameter identification of the simple form of a nested piecewise linear regression. *Uchenye zapiski Komsomolskogo-na-Amure gosudarstvennogo tekhnicheskogo universiteta.* 2023;(3):57–61. (In Russian).

- сударственного технического университета. 2023. № 3. С. 57–61.
16. Носков С. И., Перфильева К. С. Эмпирический анализ некоторых свойств метода смешанного оценивания параметров линейного регрессионного уравнения // Наука и бизнес: пути развития. 2020. № 6. С. 62–66.
16. Noskov S. I., Perfilyeva K. S. Empirical analysis of some properties of the mixed estimation method for linear regression equation parameters. *Science and Business: Ways of Development*. 2020;(6):62–66. (In Russian).

Информация об авторах

С. И. Носков – доктор технических наук, профессор.
А. Н. Знайдюк – студент.

Information about the authors

S. I. Noskov – Doctor of Sciences (Engineering), Professor.
A. N. Znaidyuk – Student.