

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ТЕЛЕКОММУНИКАЦИИ
INFORMATION TECHNOLOGY AND TELECOMMUNICATIONS

УДК 519. 852

DOI: 10.21822/2073-6185-2023-50-4-115-120



Оригинальная статья/ Original article

Вычисление оценок параметров однородной вложенной кусочно-линейной регрессии

С.И. Носков, С.И. Белинская

Иркутский государственный университет путей сообщения,
664074, г. Иркутск, ул. Чернышевского, 15, Россия

Резюме. Цель. Целью исследования является разработка алгоритма идентификации параметров однородной вложенной кусочно-линейной регрессионной модели первого типа методом наименьших модулей. **Метод.** Оценивание ее неизвестных параметров осуществляется с помощью сведения к задаче линейно-булевого программирования. Решение не должно вызывать вычислительных трудностей ввиду значительного количества эффективных программных средств – это, например, хорошо себя зарекомендовавшая и размещенная в сети Internet в свободном доступе программа LPsolve. **Результат.** Сформированная задача линейно-булевого программирования имеет приемлемую размерность для решения практических проблем моделирования. **Вывод.** Результаты решения численного примера указывают на эффективность предложенного в работе способа вычисления оценок параметров однородной вложенной кусочно-линейной регрессионной модели первого типа методом наименьших модулей.

Ключевые слова: однородная вложенная кусочно-линейная регрессия, идентификация параметров, метод наименьших модулей, задача линейно-булевого программирования, размерность

Для цитирования: С.И. Носков, С.И. Белинская. Вычисление оценок параметров однородной вложенной кусочно-линейной регрессии. Вестник Дагестанского государственного технического университета. Технические науки. 2023; 50(4): 115-120. DOI:10.21822/2073-6185-2023-50-4-115-120

Computing Parameter Estimates of a Homogeneous Nested Piecewise Linear Regression

S.I. Noskov, S.I. Belinskaya

Irkutsk State Transport University,
15 Chernyshevskogo St., Irkutsk 664074, Russia

Abstract. Objective. The aim of the study is to develop an algorithm for identifying the parameters of a homogeneous nested piecewise linear regression model of the first type by the method of least modules. **Method.** Estimation of its unknown parameters is carried out with the help of reduction to the problem of linear Boolean programming. Its solution should not cause computational difficulties due to a significant number of effective software tools - for example, the well-established and freely available program LPsolve on the Internet. **Result.** The generated linear programming problem has an acceptable dimension for solving practical modeling problems. **Conclusion.** The results of solving a numerical example indicate the effectiveness of the method proposed in the work for calculating parameter estimates for a homogeneous nested piecewise linear regression model of the first type by the method of least modules.

Keywords: homogeneous nested piecewise linear regression, parameter identification, least moduli method, linear boolean programming problem, dimension.

For citation: S.I. Noskov, S.I. Belinskaya. Computing Parameter Estimates of a Homogeneous Nested Piecewise Linear Regression. Herald of Daghestan State Technical University. Technical Sciences. 2023; 50(4): 115-120. DOI:10.21822/2073-6185-2023-50-4-115-120

Введение. При моделировании различных сложных систем весьма часто используются конструкции, связанные с применением операций вложенности. Так, в работе [1] предлагается новый подход к решению непараметрических задач последовательного обнаружения разладки. Вложенные последовательные планы выборочного контроля позволяют использовать тест Колмогорова-Смирнова и другие непараметрические тесты для последовательного обнаружения точек разладки. Точные формулы для математического ожидания и дисперсии момента ложного обнаружения основаны на точных распределениях непараметрических статистик для малых объемов выборки.

В [2] представлены результаты исследования, расширяющего известную модель Брока и Дурлауфа, позволяя учитывать ненаблюдаемую неоднородность предпочтений между альтернативами выбора путем изучения вложенной логит-модели. При этом ослабляется предположение о глобальных взаимодействиях, и вместо этого рассматриваются локальные взаимодействия в рамках нескольких гипотетических социальных и пространственных сетевых структур.

Статья [3] посвящена использованию двухуровневой вложенной логит-модели для оценки влияния региональных стимулов на выбор местоположения новых фирм в Индии. В работе [4] представлена вложенная гибридная эволюционная модель для сокращения времени ожидания транспортных средств на светофорах и улучшения мобильности в дорожной сети. В [5] предложен подход, основанный на вложенной оптимизации, для решения задач расширения пропускной способности систем разнокачественного водоснабжения. Он должен учитывать, как будет выполняться условие удовлетворения спроса и требования к качеству воды за счет использования нескольких источников с разным качеством воды, а также временное и пространственное распределение доступных водных ресурсов и воздействие на окружающую среду.

Работа [6] посвящена планированию компьютерных экспериментов методом «черного ящика». При работе с его несколькими функциями часто возникает необходимость создавать проекты для всех черных ящиков вместе, а не по отдельности. Эти так называемые вложенные планы особенно полезны в качестве обучающих и тестовых наборов для подбора и проверки метамоделей. Кроме того, вложенные схемы могут использоваться для связывания параметров и последовательных оценок. В статье вводятся одномерные вложенные максиминные планы. Доказано, что потеря заполняемости по отношению к традиционным максиминным схемам составляет не более 14.64% и 19.21 % при вложении двух и трех конструкций соответственно.

В [7] представлена методология проксимальной вложенной выборки для объективного сравнения альтернативных байесовских моделей визуализации для приложений, которые используют изображения для принятия решений в условиях неопределенности. Методология основана на вложенной выборке и методе Монте-Карло, специально предназначенном для сравнения моделей, и использует проксимальные методы цепи Маркова для эффективного масштабирования до больших задач и работы с логарифмически вогнутыми и не обязательно гладкими моделями.

В статье [8] представлена удобная в вычислительном отношении процедура расчета критерия Вальда для вложенных гипотез. Подобно обобщенному критерию Шротера (1983), предложенная процедура не требует явного вывода ограничений, подразумеваемых нулевой гипотезой, и, следовательно, ее использование может исключить сложный этап проверки линейных и нелинейных гипотез. Показано, что традиционный критерий Вальда, обобщенный критерий Шротера и предлагаемая в статье процедура асимптотически эквивалентны. Иллюстрируется ее использование в модели динамической регрессии. В работе [9] исследуются вложенные ковариационные модели, определяемые как линейные комбинации основных ковариационных функций, которые очень популярны во многих областях прикладной статистики, в частности в геостатистике. Известным ограничением вложенных моделей является то, что константы в линейной комбинации обязаны быть

неотрицательными, чтобы сохранить положительную определенность (допустимость). В данной статье исследуются вложенные модели на d -мерных сферах. Указан точный интервал допустимости констант, входящих в линейные комбинации. В частности, показывается, что хотя бы одна константа может быть отрицательной. Одним из следствий данного факта является то, что можно получить вложенную модель, достигающую отрицательных корреляций. Приводятся теоремы характеризации для произвольных линейных комбинаций, а также для невыпуклых комбинаций, включающих две функции ковариации.

Постановка задачи. В работе [10] путем обобщения некоторых кусочно-линейных регрессионных моделей [11-13] введены два типа вложенных кусочно-линейных регрессий:

- вложенная кусочно-линейная регрессия первого типа:

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

- вложенная кусочно-линейная регрессия второго типа:

$$y_k = \max\{\min_{i \in I^1}\{a_i^1 x_{ki}\}, \dots, \min_{i \in I^G}\{a_i^G x_{ki}\}, \max_{i \in J^1}\{\beta_i^1 x_{ki}\}, \dots, \max_{i \in J^H}\{\beta_i^H x_{ki}\}\} + \varepsilon_k, \quad k = \overline{1, n} \quad (2)$$

Здесь k – номер наблюдения, y – зависимая переменная, $x_i, i = \overline{1, m}$ – независимые переменные, a_i^j и β_i^j – подлежащие оцениванию параметры, $\varepsilon_k, k = \overline{1, n}$ – ошибки аппроксимации, n – количество наблюдений (длина выборки). Индексные множества $I^i, i = \overline{1, G}, J^j, j = \overline{1, H}$ являются подмножествами множества $\{1, 2, \dots, m\}$ и могут иметь непустые пересечения.

В [10] указано, что в формулах (1) и (2) использован первый порядок вложенности. Он может быть также вторым, третьим и т.д.

В работе [14] описаны некоторые частные формы вложенных кусочно-линейных регрессий:

- простая вложенная кусочно-линейная регрессия первого типа

$$y_k = \min\{\min_{i \in I}\{a_i x_{ki}\}, \max_{i \in J}\{\beta_i x_{ki}\}\} + \varepsilon_k; \quad (3)$$

- простая вложенная кусочно-линейная регрессия второго типа

$$y_k = \max\{\min_{i \in I}\{a_i x_{ki}\}, \max_{i \in J}\{\beta_i x_{ki}\}\} + \varepsilon_k; \quad (4)$$

- однородная вложенная кусочно-линейная регрессия первого типа

$$y_k = \min\{\min_{i \in I^1}\{a_i^1 x_{ki}\}, \dots, \min_{i \in I^G}\{a_i^G x_{ki}\}\} + \tilde{\varepsilon}_k; \quad (5)$$

- однородная вложенная кусочно-линейная регрессия второго типа

$$y_k = \max\{\max_{i \in J^1}\{\beta_i^1 x_{ki}\}, \dots, \max_{i \in J^H}\{\beta_i^H x_{ki}\}\} + \varepsilon_k. \quad (6)$$

Разумеется, представления (3) – (6) далеко не исчерпывают всех возможных способов конкретизации вложенных кусочно-линейных регрессий. Некоторые из таких способов отражены в [14]. Поставим задачу вычисления оценок параметров однородной вложенной кусочно-линейной регрессии первого типа (5) методом наименьших модулей (МНМ) (его описание см., например, в работах [11-13, 15]):

$$I_1(\alpha) = \sum_{k=1}^n |\tilde{\varepsilon}_k| \rightarrow \min. \quad (7)$$

Методы исследования. В работе [16] задача идентификации параметров простой вложенной кусочно-линейной регрессии первого типа (3) с помощью МНМ сведена к задаче линейно-булевого программирования (ЛБП) следующим образом.

Введем следующие обозначения.

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

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

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

Тогда указанная задача ЛБП примет следующий вид:

$$h_k \leq a_i x_{ki}, k = \overline{1, n}, i \in I, \quad (8)$$

$$a_i x_{ki} - h_k \leq (1 - s_{ki})M, k = \overline{1, n}, i \in I, \quad (9)$$

$$\sum_{i \in I} s_{ki} = 1, k = \overline{1, n}, \quad (10)$$

$$g_k \geq \beta_i x_{ki}, k = \overline{1, n}, i \in J, \quad (11)$$

$$\beta_i x_{ki} - g_k \geq (p_{ki} - 1)M, k = \overline{1, n}, i \in J, \quad (12)$$

$$\sum_{i \in J} p_{ki} = 1, k = \overline{1, n}, \quad (13)$$

$$t_k \leq h_k, k = \overline{1, n}, \quad (14)$$

$$t_k \leq g_k, k = \overline{1, n}, \quad (15)$$

$$h_k - t_k + Mr_k \leq M, k = \overline{1, n}, \quad (16)$$

$$g_k - t_k - Mr_k \leq 0, k = \overline{1, n}, \quad (17)$$

$$t_k + u_k - v_k = y_k, k = \overline{1, n}, \quad (18)$$

$$u_k \geq 0, v_k \geq 0, h_k \geq 0, g_k \geq 0, t_k \geq 0, k = \overline{1, n}, \quad (19)$$

$$s_{ki} \in \{0, 1\}, k = \overline{1, n}, i \in I, \quad (20)$$

$$p_{ki} \in \{0, 1\}, k = \overline{1, n}, i \in J, \quad (21)$$

$$r_k \in \{0, 1\}, k = \overline{1, n}, \quad (22)$$

$$\sum_{k=1}^n (u_k + v_k) \rightarrow \min, \quad (23)$$

где M – заранее заданное большое положительное число.

Обсуждение результатов. Преобразуем задачу (8) – (23) так, чтобы новая сформированная задача ЛБП была эквивалентна задаче (7).

Введем обозначения:

$$v_{kj} = \min_{i \in I^j} \{a_i^j x_{ki}\}, k = \overline{1, n}, j = \overline{1, G},$$

$$w_k = \min_{j=1, G} v_{kj}, k = \overline{1, n}.$$

Тогда по аналогии с (8) – (23) эквивалентная (7) задача ЛБП примет вид:

$$v_{kj} \leq a_i^j x_{ki}, k = \overline{1, n}, i \in I^j, j = \overline{1, G}, \quad (24)$$

$$a_i^j x_{ki} - v_{kj} \leq (1 - s_{kij})M, k = \overline{1, n}, i \in I^j, j = \overline{1, G}, \quad (25)$$

$$\sum_{i \in I^j} s_{kij} = 1, k = \overline{1, n}, j = \overline{1, G}, \quad (26)$$

$$w_k \leq v_{kj}, k = \overline{1, n}, j = \overline{1, G}, \quad (27)$$

$$v_{kj} - w_k \leq (1 - r_{kj})M, k = \overline{1, n}, j = \overline{1, G}, \quad (28)$$

$$\sum_{j=1}^G r_{kj} = 1, k = \overline{1, n}, \quad (29)$$

$$w_k + u_k - v_k = y_k, k = \overline{1, n}, \quad (30)$$

$$s_{kij} \in \{0, 1\}, k = \overline{1, n}, i \in I^j, j = \overline{1, G}, \quad (31)$$

$$r_{kj} \in \{0, 1\}, k = \overline{1, n}, j = \overline{1, G}, \quad (32)$$

$$\sum_{k=1}^n (u_k + v_k) \rightarrow \min. \quad (33)$$

Решение задачи ЛБП (24) – (33) не должно вызывать вычислительных трудностей ввиду значительного количества эффективных программных средств – это, например, хорошо себя зарекомендовавшая и размещенная в сети Internet в свободном доступе программа LPsolve.

Проиллюстрируем на простом примере предложенный способ построения однородной вложенной кусочно-линейной регрессии первого типа. В качестве исходных данных возьмем выборку из [16], несколько изменив в ней вектор наблюдений для зависимой переменной:

$$X = \begin{pmatrix} 2 & 4 & 3 \\ 7 & 1 & 9 \\ 6 & 5 & 8 \\ 3 & 8 & 4 \end{pmatrix}, \quad y = \begin{pmatrix} 3 \\ 4 \\ 12 \\ 9 \end{pmatrix}.$$

Будем строить однородную кусочно-линейную регрессию первого типа с индексными множествами I^1 и I^2 , равными:

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

В результате решения задачи ЛБП (24) – (33) получим следующую регрессионную модель:

$$y_k = \min\{\min\{2x_{k1}, 4x_{k2}\}, \min\{4x_{k2}, 1.5x_{k3}\} + \varepsilon_k, k = \overline{1, 4}.$$

При этом некоторые вещественные переменные задачи примут значения:

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

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

$$v = (1, 0, 0, 0).$$

Таким образом, в задаче (7) $\sum_{k=1}^n |\tilde{\varepsilon}_k| = 4$.

Тот факт, что $a_2^1 = a_1^2$ объясняется тем обстоятельством, что вторая компонента индексного множества I^1 совпадает с первой компонентой множества I^2 .

Вывод. В работе поставлена задача идентификации параметров однородной вложенной кусочно-линейной регрессии первого типа методом наименьших модулей. Получены следующие результаты:

1. Поставленная задача сведена к задаче линейно-булевого программирования.
2. Сформированная задача ЛБП имеет приемлемую размерность для решения практических проблем моделирования.
3. Представленный подход может быть применен к расчету параметров и других форм вложенной кусочно-линейной регрессии.
4. Решен численный пример.

Библиографический список:

1. Lumelskii Ya. P., Feigin P. D. Nonparametric Tests and Nested Sequential Sampling Plans for Change-Point Detection // Journal of Mathematical Sciences. – 2017. – V. 221. – P. 566–579.
2. Dugundji E.R., Gulyás L. Structure and emergence in a nested logit model with social and spatial interactions // Computational and Mathematical Organization Theory. – 2013. – V. 19. – P. 151–203.
3. Ramaul N.K., Ramaul P. Regional Incentives and Location Choice of New Firms in India: A Nested Logit Model // Journal of Quantitative Economics. – 2018. – V. 16. – P. 501–525.
4. Srivastava S., Sahana K.S. Nested hybrid evolutionary model for traffic signal optimization // Applied Intelligence. – 2017. – V. 46. – P. 113–123.
5. Vieira J., Cunha M.C. Nested Optimization Approach for the Capacity Expansion of Multiquality Water Supply Systems under Uncertainty // Water Resources Management. – 2017. – V. 31. P. 1381–1395.
6. Van Dam E.R., den Hertog B.H. One-dimensional nested maximin designs // Journal of Global Optimization. – 2010. – V. 46. – P. 287–306.
7. Cai X., McEwen J.D., Pereyra M. Proximal nested sampling for high-dimensional Bayesian model selection // Statistics and Computing. – 2022. – V. 32. – Article number: 87.
8. Palm F.C., Kodde D.A. Computing Wald criteria for nested hypotheses // Statistical Papers. – 1988. – V. 29. – P. 169–190.
9. Peron A., Porcu E., Emery X. Admissible nested covariance models over spheres cross time // Stochastic Environmental Research and Risk Assessment. – 2018. – V. 32. – P. 3053–3066.
10. Носков С.И. Подход к формализации вложенной кусочно-линейной регрессии // Международный журнал гуманитарных и естественных наук. – 2023. – № 1-2 (76). – С. 218-220.
11. Носков С.И., Хоняков А.А. Программный комплекс построения некоторых типов кусочно-линейных регрессий // Информационные технологии и математическое моделирование в управлении сложными системами. – 2019. – № 3 (4). – С. 47-55.
12. Носков С.И. Идентификация параметров кусочно-линейной функции риска // Транспортная инфраструктура Сибирского региона. – 2017. – Т. 1. – С. 417-421.
13. Носков С.И., Лоншаков Р. В. Идентификация параметров кусочно-линейной регрессии // Информационные технологии и проблемы математического моделирования сложных систем. – 2008. – № 6. – С. 63-64.
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., Попов Е.С., Середкин С.П., Тирских В.В., Торопов В.Д. Вариантное регрессионное моделирование производства электроэнергии в Российской Федерации // Вестник Дагестанского государственного технического университета. Технические науки. - 2023. - Т. 50. - № 1. - С. 123-129.
16. Носков С.И. Идентификация параметров простой формы вложенной кусочно-линейной регрессии // Ученые записки Комсомольского-на-Амуре государственного технического университета. - 2023. - № 3 (67). - С. 57-61.

References:

1. Lumelskii Ya. P., Feigin P. D. Nonparametric Tests and Nested Sequential Sampling Plans for Change-Point Detection. *Journal of Mathematical Sciences*. 2017; 221: 566–579.
2. Dugundji E.R., Gulyás L. Structure and emergence in a nested logit model with social and spatial interactions. *Computational and Mathematical Organization Theory*. 2013;19:151–203.
3. Ramaul N.K., Ramaul P. Regional Incentives and Location Choice of New Firms in India: A Nested Logit Model. *Journal of Quantitative Economics*. 2018;16: 501–525.
4. Srivastava S., Sahana K.S. Nested hybrid evolutionary model for traffic signal optimization. *Applied Intelligence*. 2017; 46: 113–123.
5. Vieira J., Cunha M.C. Nested Optimization Approach for the Capacity Expansion of Multiquality Water Supply Systems under Uncertainty. *Water Resources Management*. 2017;31:1381–1395.
6. Van Dam E.R., den Hertog B.H. One-dimensional nested maximin designs. *Journal of Global Optimization*. 2010;46: 287–306.
7. Cai X., McEwen J.D., Pereyra M. Proximal nested sampling for high-dimensional Bayesian model selection. *Statistics and Computing*. 2022; 32- Article number: 87.
8. Palm F.C., Kodde D.A. Computing Wald criteria for nested hypotheses. *Statistical Papers*. 1988; 2: 169–190.
9. Peron A., Porcu E., Emery X. Admissible nested covariance models over spheres cross time. *Stochastic Environmental Research and Risk Assessment*. 2018; 32: 3053–3066.
10. Noskov S.I. An approach to the formalization of nested piecewise linear regression. *International Journal of the Humanities and Natural Sciences*. 2023; 1-2 (76): 218-220. (In Russ)
11. Noskov S.I., Khonyakov A.A. Software complex for constructing some types of piecewise linear regressions. *Information technologies and mathematical modeling in the management of complex systems*. 2019; 3 (4); 47-55. (In Russ)
12. Noskov S.I. Identification of the parameters of the piecewise linear risk function. *Transport infrastructure of the Siberian region*. 2017;1: 417-421. (In Russ)
13. Noskov S. I., Lonshakov R. V. Identification of parameters of piecewise linear regression. *Information technologies and problems of mathematical modeling of complex systems*. 2008; 6:63-64. (In Russ)
14. Noskov S.I. Some forms of nested piecewise linear regression. *Izvestia of the Tula State University. Technical Sciences*. 2023; 3:467-469. (In Russ)
15. Noskov S.I., Popov E.S., Seredkin S.P., Tirkikh V.V., Toropov V.D. Variant regression modeling of electricity production in the Russian Federation. *Herald of the Daghestan State Technical University. Technical Science*. 2023; 50(1):123-129. DOI: 10.21822/2073-6185-2023-50-1-123-129 (In Russ)
16. Noskov S.I. Identification of parameters of a simple form of nested piecewise linear regression. *Scientific notes of the Komsomolsk-on-Amur State Technical University*. 2023; 3 (67):57-61. (In Russ)

Сведения об авторах:

Носков Сергей Иванович, доктор технических наук, профессор, профессор кафедры информационных технологий и защиты информации; sergey.noskov.57@mail.ru

Белинская Софья Иосифовна, кандидат физико-математических наук, доцент, доцент кафедры информационных технологий и защиты информации; sofiabel@mail.ru

Information about authors:

Sergey I. Noskov, Dr. Sci. (Eng), Prof., Prof. of the Department of Information Technologies and Information Security; sergey.noskov.57@mail.ru

Sofia I. Belinskaya, Cand. Sci. (Physical and Mathematical), Assoc. Prof., Assoc. Prof. of the Department of Information Technologies and Information Security; sofiabel@mail.ru

Конфликт интересов/Conflict of interest.

Авторы заявляют об отсутствии конфликта интересов/The authors declare no conflict of interest.

Поступила в редакцию/Received 18.09.2023.

Одобрена после рецензирования/ Revised 20.10.2023.

Принята в печать/Accepted for publication 20.10.2023.