

**ПОСТРОЕНИЕ КУСОЧНО-ЛИНЕЙНОЙ АВТОРЕГРЕССИОННОЙ МОДЕЛИ
ПРОИЗВОЛЬНОГО ПОРЯДКА**

Носков Сергей Иванович

*доктор технических наук,
профессор кафедры*

*«Информационные системы и защита информации»,
ФГБОУ ВО «Иркутский государственный университет путей сообщения»,
Иркутск, Россия*

E-mail: sergey.noskov.57@mail.ru

Актуальность исследования обусловлена необходимостью расширения арсенала форм связи между переменными в регрессионных моделях.

Объект: кусочно-линейная авторегрессионная модель произвольного порядка.

Предмет: вычислительный аппарат решения задач линейно-булевого программирования.

Цель: разработка алгоритма оценивания параметров кусочно-линейной регрессии.

Методы: регрессионный анализ, математическое программирование.

Результаты: в работе сформулирована задача построения кусочно-линейной авторегрессионной модели произвольного порядка на основе метода наименьших модулей. Предложен алгоритм ее решения, сводящийся к задаче линейно-булевого программирования приемлемой размерности для реальных прикладных проблем. Разработана кусочно-линейная авторегрессионная модель обеспеченности жильем на статистической информации Иркутской области, обладающая высокой адекватностью. Модель может быть успешно использована при решении различных проблем прогнозного характера.

Ключевые слова: регрессионная модель, авторегрессия, метод наименьших модулей, задача линейно-булевого программирования, обеспеченность жильем.

**CONSTRUCTION OF A PIECE-LINEAR AUTOREGRESSION MODEL
OF AN ARBITRARY ORDER**

Sergey I. Noskov

*Doctor of Technical Sciences
Professor of the Department*

*of Information Systems and Information Security,
Irkutsk State Railway University
Irkutsk, Russia*

E-mail: sergey.noskov.57@mail.ru

The relevance of the study is due to the need to expand the arsenal of forms of communication between variables in regression models.

Object: piecewise linear autoregressive model of arbitrary order.

Subject: computing apparatus for solving problems of linear-Boolean programming.

Purpose: development of an algorithm for estimating the parameters of piecewise linear regression.

Methods: regression analysis, mathematical programming.

Results: the paper formulated the problem of constructing a piecewise linear autoregressive model of an arbitrary order based on the method of least modules. An algorithm for solving it is proposed, which reduces to a linear Boolean programming problem of acceptable dimension for

real applied problems. A piecewise linear autoregressive model of housing provision based on the statistical information of the Irkutsk region has been developed, which has a high adequacy. The model can be successfully used in solving various predictive problems. Keywords: regression model, autoregression, least modules method, linear Boolean programming problem, housing supply.

Keywords: regression model, autoregression, least modules method, linear Boolean programming problem, housing supply.

Введение

Весьма часто при построении математических моделей динамических процессов самой различной природы успешно применяются авторегрессионные соотношения, в правые части которых включаются значения зависимых переменных в прошлые по отношению к текущему моменты времени (или за прошлые периоды времени). Так, в работе [1] с помощью авторегрессионной нейронной сети изучается влияние 66 климатических факторов на спрос на электроэнергию. В [2] с использованием векторной авторегрессионной модели, а также реальной и номинальной кривых Нельсона – Зигеля изучается доходность активов. Получены важные результаты, показывающие, что отсроченные аннуитеты с защитой от инфляции важны при наличии риска реального трудового дохода, однако так называемые номинальные отсроченные аннуитеты покупаются как более дешевая альтернатива, если реальная доходность низка или отрицательна. В статье [3] на основе модели авторегрессионного интегрированного скользящего среднего исследуются будущие цены на интеллектуальные системы, позволяя предлагать более доступные товары и услуги. Работа [4] посвящена созданию авторегрессионной модели первого порядка для нерегулярных дискретных временных рядов при изучении разных типов сверхновых, обнаружении и характеристике внесолнечных планет и классифицировании звезд. Весьма интересным представляется исследование [5], в котором изучается вопрос снижения зависимости Саудовской Аравии от нефти и диверсификации ее экономики. Рассматриваются данные о ценах на нефть и инвестициях в акции исламских государств, оценивается их взаимодействие и взаимовлияние с помощью векторного авторегрессионного моделирования и оценки функции импульсного отклика.

Результаты и обсуждение

Пусть поведение некоторого динамического процесса описывается зависимой переменной y_t , $t = \overline{1, n}$. Тогда линейная авторегрессионная модель порядка p представима в виде:

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \varepsilon_t, t = \overline{p+1, n}, \quad (1)$$

где α_i , $i = \overline{1, p}$ – подлежащие оцениванию параметры, ε_t , $t = \overline{p+1, n}$ – ошибки аппроксимации, не имеющие вероятностной природы. Таким образом, на значения переменной y в момент времени t (или за период времени t) влияют ее прошлые значения до момента $t-p$ включительно. Отметим, что проблемам построения авторегрессионных моделей посвящена обширная литература (см., например, [6–8]).

В настоящей работе мы займемся решением проблемы построения кусочно-линейной авторегрессии произвольного порядка p , имеющей вид

$$y_t = \min \{ \alpha_1 y_{t-1}, \alpha_2 y_{t-2}, \dots, \alpha_p y_{t-p} \} + \varepsilon_t, t = \overline{p+1, n}. \quad (2)$$

При этом потребуем выполнения условия $n-p > p$ (длина выборки данных должна быть больше числа оцениваемых параметров).

Введем следующие обозначения:

$$z_t = \min \{ \alpha_1 y_{t-1}, \alpha_2 y_{t-2}, \dots, \alpha_p y_{t-p} \}, t = \overline{p+1, n}. \quad (3)$$

Тогда модель (2) представима в виде:

$$y_t = z_t + \varepsilon_t.$$

Неизвестные параметры $\alpha_i, i = \overline{1, p}$ будем оценивать посредством реализации метода наименьших модулей (МНМ), минимизируя функцию $J(\alpha)$:

$$J(\alpha) = \sum_{t=p+1}^n |y_t - z_t| \rightarrow \min. \quad (4)$$

Следует указать, что построению кусочно-линейной регрессии методом наименьших квадратов с произвольными независимыми переменными посвящены, в частности, работы [9, 10].

Для решения задачи (4) воспользуемся подходом, развитым в работах [11– 14] при использовании МНМ для построения кусочно-линейных многофакторных регрессий.

Введем в рассмотрение переменные u_i и v_i следующим образом:

$$u_t = \begin{cases} y_t - z_t, & y_t > z_t \\ 0, & \text{в противном случае,} \end{cases}$$

$$v_t = \begin{cases} z_t - y_t, & z_t > y_t \\ 0, & \text{в противном случае.} \end{cases}$$

Тогда будут справедливы тождества

$$z_t + u_t - v_t = y_t, t = \overline{p+1, n}. \quad (5)$$

Из соотношения (3) следует справедливость неравенств

$$z_t \leq \alpha_i y_{t-i}, t = \overline{p+1, n}, i = \overline{1, p}, \quad (6)$$

причем для каждого t по крайней мере одно из них должно обращаться в строгое равенство.

Для достижения последнего условия введем в рассмотрение $(n-p)p$ булевых переменных $\sigma_{ii}, t = \overline{p+1, n}, i = \overline{1, p}$ и сформируем ограничения:

$$\alpha_i y_{t-i} - z_t \leq (1 - \sigma_{ii}) M, t = \overline{p+1, n}, i = \overline{1, p}, \quad (7)$$

$$\sum_{i=1}^p \sigma_{ii} = 1, t = \overline{p+1, n}, \quad (8)$$

$$\sigma_{ii} \in \{0, 1\}, t = \overline{p+1, n}, i = \overline{1, p}, \quad (9)$$

где M – заранее выбранное большое положительное число.

Из определения переменных u_i и v_i следуют следующие равенства

$$|\varepsilon_t| = u_t + v_t, u_t v_t = 0,$$

что позволяет представить задачу (4) в виде

$$J(\alpha) = \sum_{t=p+1}^n (u_t + v_t) \rightarrow \min. \quad (10)$$

Таким образом, задача (4) поиска значений неизвестных параметров $\alpha_i, i = \overline{1, p}$ кусочно-линейной авторегрессии (2) с помощью МНМ свелась к задаче линейно-булевого програм-

мирования (5) – (10) с $(n - p)p + 3(n - p) + p$ переменными (из которых $(n - p)p$ – булевы) и $2((n - p)p + n - p)$ ограничениями.

Важнейшим социальным индикатором эффективности функционирования любого региона является общая площадь жилых помещений, приходящаяся в среднем на одного жителя, выраженная в квадратных метрах. Построим кусочно-линейную авторегрессионную модель (2) динамики этого показателя на основе статистической информации по Иркутской области [15] за 1995–2020 годы (рисунок).



Рисунок – Динамика переменной y за 1995–2020 годы

Решение задачи (5) – (10) с этими исходными данными позволяет представить модель (2) в виде кусочно-линейной авторегрессии третьего порядка:

$$y_t = \min \{1.016y_{t-1}, 1.054y_{t-2}, 1.083y_{t-3}\}, t = \overline{4, 26}, \quad (11)$$

$$E=0.77, \text{ КСП}=18.$$

Здесь в качестве критериев оценки адекватности модели использованы средняя относительная ошибка аппроксимации E и критерий согласованности поведения расчетных и фактических значений зависимой переменной КСП (см., например, [16]):

$$E = \frac{100\%}{n - p} \sum_{t=p+1}^n \frac{|y_t - z_t|}{y_t},$$

$$КСП = \sum_{t=p+2}^n \text{sign}[(y_t - y_{t-1})(z_t - z_{t-1})],$$

где

$$\text{sign}(a) = \begin{cases} 1, & a \geq 0 \\ 0, & \text{в противном случае.} \end{cases}$$

Как следует из значений E и КСП, модель (11) обладает весьма высокими аппроксимационными характеристиками. Ее анализ позволяет сделать вывод об увеличении значения параметра с ростом порядка авторегрессионной составляющей, что вполне согласуется с динамикой переменной y , представленной на рис. 1.

Кусочно-линейная авторегрессионная модель (11) обеспеченности жильем в Иркутской области может быть успешно использована при решении различных проблем прогнозного характера.

Заключение и выводы

В работе описан способ построения кусочно-линейной авторегрессионной модели произвольного порядка методом наименьших модулей. Показано, что оценивание ее параметров может быть сведено к задаче линейно-булевого программирования. Построена кусочно-линейная авторегрессионная модель обеспеченности жильем на статистическом материале Иркутской области, обладающая высокими аппроксимационными характеристиками. Модель может быть успешно использована при решении различных прогнозных проблем.

Литература

1. Chreng, K. Electricity demand prediction for sustainable development in Cambodia using recurrent neural networks with ERA5 reanalysis climate variables / K. Chreng, H. S. Lee, S. Tuy // *Energy Reports*. – 2022. – № 8. – P. 76–81.
2. Owadally, I. Optimal investment for a retirement plan with deferred annuities allowing for inflation and labour income risk / I. Owadally, C. Jang, A. Clare // *European Journal of Operational Research*. – 2021. – № 3 (295). – P. 1132–1146.
3. Forecasting E-commerce products prices by combining an autoregressive integrated moving average (ARIMA) model and Google Trends data / S. Carta, A. Medda, A. Pili, [et al.] // *Future Internet*. – 2018. – № 1 (11). – P. 5–11.
4. Eyheramendy, S. An irregular discrete time series model to identify residuals with autocorrelation in astronomical light curves / S. Eyheramendy, F. Elorrieta, W. Palma // *Monthly Notices of the Royal Astronomical Society*. – 2018. – № 4 (481). – P. 4311–4322.
5. Jawadi, F. Toward a new deal for Saudi Arabia: oil or Islamic stock market investment? / F. Jawadi, N. Jawadi, A. I. Cheffou // *Applied Economics*. – 2018. – № 50 (59). – P. 6355–6363.
6. Драница, Ю. П. Быстрый алгоритм построения нестационарной векторной линейной авторегрессии / Ю. П. Драница, А. Ю. Драница, О. В. Алексеевская. – Текст : непосредственный // *Дифференциальные уравнения и процессы управления*. – 2011. – № 4. – С. 138–177.
7. Тырсин, А. Н. Построение моделей авторегрессии временных рядов при наличии помех / А. Н. Тырсин. – Текст : непосредственный // *Математическое моделирование*. – 2005. – Т. 17, № 5. – С. 10–16.
8. Гельру, Я. Д. Векторная модель авторегрессии показателей производственной деятельности строительного предприятия / Я. Д. Гельруд, Е. А. Угрюмов, В. Л. Рыбак. – Текст : непосредственный // *Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика*. – 2018. – Т. 7, № 3. – С. 19–30.
9. Базилевский, М. П. Исследование возможности построения кусочно-линейных регрессий с нелинейными границами переключения / М. П. Базилевский. – Текст : непосредственный // *System Analysis and Mathematical Modeling*. – 2021. – Т. 3, № 2. – С. 99–112.
10. Базилевский, М. П. МНК-оценивание параметров специфицированных на основе функций Леонтьева двухфакторных моделей регрессии / М. П. Базилевский. – Текст : непосредственный // *Южно-Сибирский научный вестник*. – 2019. – № 2 (26). – С. 66–70.
11. Носков, С. И. Оценивание параметров аппроксимирующей функции с постоянными пропорциями / С. И. Носков. – Текст : непосредственный // *Современные технологии. Системный анализ. Моделирование*. – 2013. – № 2 (38). – С. 135–136.
12. Иванова, Н. К. Идентификация параметров некоторых негладких регрессий / Н. К. Иванова, С. А. Лебедева, С. И. Носков. – Текст : непосредственный // *Информационные технологии и проблемы математического моделирования сложных систем*. – 2016. – № 17. – С. 107–110.
13. 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.

14. Базилевский, М. П. Алгоритм построения линейно-мультипликативной регрессии / М. П. Базилевский, С. И. Носков. – Текст : непосредственный // Современные технологии. Системный анализ. Моделирование. – 2011. – № 1 (29). – С. 88–92.

15. Федеральная служба государственной статистики. – URL: <https://rosstat.gov.ru/folder/210/document/13205> (дата обращения: 09.06.2022). – Текст : электронный.

16. Носков, С. И. Обобщенный критерий согласованности поведения в регрессионном анализе / С. И. Носков. – Текст : непосредственный // Информационные технологии и математическое моделирование в управлении сложными системами. – 2018. – № 1 (1). – С. 14–20.