

ПРИБЛИЖЕННОЕ ОЦЕНИВАНИЕ С ПОМОЩЬЮ МЕТОДА НАИМЕНЬШИХ КВАДРАТОВ ОДНОЙ ФОРМЫ НЕЭЛЕМЕНТАРНЫХ МОДУЛЬНЫХ ЛИНЕЙНЫХ РЕГРЕССИЙ

М. П. Базилевский

Иркутский государственный университет путей сообщения, Иркутск, Россия
mik2178@yandex.ru

Аннотация. *Актуальность и цели.* Проблема поиска новых структурных спецификаций регрессионных моделей с интересными интерпретационными свойствами является актуальной в настоящее время. Цель исследования состоит в формализации на основе симбиоза предложенных ранее неэлементарных и модульных линейных регрессий новой структурной спецификации, разработке алгоритма ее приближенного оценивания с помощью метода наименьших квадратов и демонстрации ее эффективности на примере моделирования индекса потребительских цен в Республике Алтай. *Материалы и методы.* Для оценивания регрессионных моделей применялся метод наименьших квадратов в сочетании с методом «всех возможных регрессий». *Результаты.* Сформулирована новая структурная спецификация регрессионных моделей – неэлементарная модульная линейная регрессия, обобщающая многие известные модели. Предложен алгоритм ее приближенного оценивания. Построенная с помощью него неэлементарная модульная линейная регрессия оказалась на 39,3 % лучше по качеству аппроксимации неэлементарной регрессии без модулей и на 63,7 % лучше линейной регрессии. *Выводы.* С помощью предложенных моделей неэлементарной модульной линейной регрессии можно успешно решать задачи прогнозирования, а также задачи выявления новых знаний об объекте исследования.

Ключевые слова: линейная регрессия, неэлементарная регрессия, модульная регрессия, функция Леонтьева, метод наименьших квадратов, мультиколлинеарность, индекс потребительских цен

Для цитирования: Базилевский М. П. Приближенное оценивание с помощью метода наименьших квадратов одной формы неэлементарных модульных линейных регрессий // Модели, системы, сети в экономике, технике, природе и обществе. 2024. № 2. С. 119–129. doi: 10.21685/2227-8486-2024-2-8

APPROXIMATE LEAST SQUARES ESTIMATION OF ONE FORM OF NON-ELEMENTARY MODULAR LINEAR REGRESSIONS

M.P. Bazilevskiy

Irkutsk State Transport University, Irkutsk, Russia
mik2178@yandex.ru

Abstract. *Background.* The problem of finding new structural specifications of regression models with interesting interpretive properties is currently relevant. The purpose of the study is to formalize a new structural specification, based on a symbiosis of previously pro-

posed non-elementary and modular linear regressions, to develop an algorithm for its approximate estimation using the ordinary least squares method and to demonstrate its effectiveness using the example of modeling the consumer price index in the Altai Republic. *Materials and methods.* To estimate the regression models, the ordinary least squares method was used in combination with the «all possible regressions» method. *Results.* A new structural specification of regression models is formulated – non-elementary modular linear regression, which generalizes many well-known models. An algorithm for its approximate estimation is proposed. The non-elementary modular linear regression constructed using it turned out to be 39.3 % better in terms of the approximation quality of non-elementary regression without modules, and 63.7 % better than linear regression. *Conclusions.* Using the proposed models of non-elementary modular linear regression, one can successfully solve forecasting problems, as well as the problem of identifying new knowledge about the object of study.

Keywords: linear regression, non-elementary regression, modular regression, Leontief function, ordinary least squares method, multicollinearity, consumer price index

For citation: Bazilevskiy M.P. Approximate least squares estimation of one form of non-elementary modular linear regressions. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve* = *Models, systems, networks in economics, technology, nature and society*. 2024;(2):119–129. (In Russ.). doi: 10.21685/2227-8486-2024-2-8

Введение

Весьма актуальной научной проблемой является поиск скрытых и практически полезных знаний в статистических данных. Для решения этой проблемы все чаще применяются методы Data mining [1, 2]. Эффективным инструментом методологии Data mining считается регрессионный анализ [3, 4], конечным продуктом которого выступает регрессионная модель, которую можно использовать либо для прогнозирования, либо для интерпретации исследуемого процесса или явления. Ежегодно с помощью регрессионного анализа во всем мире решается множество прикладных задач различного характера. Например, в работе [5] построены регрессионные модели для прогнозирования масс фюзеляжа, крыла и оперения самолета, в работе [6] проведен регрессионный анализ влияния климатических факторов на электропотребление объектов нефтедобычи, в работе [7] регрессионный анализ применен для оптимизации содержания пищевых волокон в рецептуре хлебобулочного изделия, а в работе [8] – для обоснования рациональных параметров двухлетнего сошника.

Как отмечено в работе [9], спецификация модели, т.е. выбор ее общего вида, это первый и, быть может, важнейший шаг регрессионного анализа. Структурных спецификаций регрессионных моделей существует великое множество. Описание многих из них можно найти в работах [10, 11], а процесс выявления, изучения и внедрения в существующую практику регрессионного анализа новых хорошо интерпретируемых моделей продолжается и по сей день. Так, например, в статье [12] введена в рассмотрение кусочно-линейная регрессия на основе симбиоза функции Леонтьева и регрессии с максимальным вкладом переменных. Чуть позже были разработаны хорошо зарекомендовавшие себя неэлементарные линейные регрессии (НЛР) [13], включающие в свое уравнение как объясняющие переменные, так и преобразованные с помощью бинарных операций min или max все возможные комбинации их пар. В работе [14] для снижения временных издержек при моде-

лировании были предложены некоторые стратегии построения НЛР известным методом «всех возможных регрессий» [9]. В работе [15] на основе аппарата математического программирования разработана эффективная технология построения НЛР, включающих в себя регрессоры с бинарными операциями $\min$ и $\max$. В работе [16] НЛР обобщены с использованием не только бинарных, но и тернарных, кватернарных, ..., l -арных операций $\min$ и $\max$. Параллельно были введены модели модульной линейной регрессии [17, 18].

Цель данной работы заключается в следующем: формализовать на основе симбиоза НЛР и модульной регрессии новую регрессионную спецификацию, разработать алгоритм нахождения ее приближенных оценок с помощью метода наименьших квадратов (МНК) и продемонстрировать ее эффективность на примере моделирования индекса потребительских цен (ИПЦ) в Республике Алтай.

Материалы и методы

Рассмотрим ранее предложенную и исследованную в статье [16] НЛР с бинарными, ..., l -арными неэлементарными операциями $\min$ ($\max$):

$$\begin{aligned} y_i = & \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} + \sum_{j=1}^{p_1} \alpha_j^{\min,2} \min \left\{ x_{i,\mu_{j1}^{(1)}}, k_j^{\min,2} x_{i,\mu_{j2}^{(1)}} \right\} + \\ & + \sum_{j=1}^{p_1} \alpha_j^{\max,2} \max \left\{ x_{i,\mu_{j1}^{(1)}}, k_j^{\max,2} x_{i,\mu_{j2}^{(1)}} \right\} + \sum_{j=1}^{p_2} \alpha_j^{\min,3} \min \left\{ x_{i,\mu_{j1}^{(2)}}, k_{j1}^{\min,3} x_{i,\mu_{j2}^{(2)}}, k_{j2}^{\min,3} x_{i,\mu_{j3}^{(2)}} \right\} + \\ & + \sum_{j=1}^{p_2} \alpha_j^{\max,3} \max \left\{ x_{i,\mu_{j1}^{(2)}}, k_{j1}^{\max,3} x_{i,\mu_{j2}^{(2)}}, k_{j2}^{\max,3} x_{i,\mu_{j3}^{(2)}} \right\} + K + \\ & + \alpha_1^{\min,l} \min \left\{ x_{i1}, k_1^{\min,l} x_{i2}, \dots, k_{l-1}^{\min,l} x_{il} \right\} + \\ & + \alpha_1^{\max,l} \max \left\{ x_{i1}, k_1^{\max,l} x_{i2}, \dots, k_{l-1}^{\max,l} x_{il} \right\} + \varepsilon_i, \quad i = \overline{1, n}, \end{aligned} \quad (1)$$

где n – объем исходной выборки данных; l – общее число имеющихся объясняющих переменных; y_i – значение с индексом i выбранной объясняемой переменной; x_{ij} – значение под номером i объясняющей переменной с индексом j ; $\min$ ($\max$) – неэлементарные операции, определяющие минимум (максимум) двух, ..., l чисел; $\mu_{jh}^{(s-1)}$ ($s = \overline{2, l}$, $j = \overline{1, p_{s-1}}$, $h = \overline{1, s}$) – элемент строки с индексом j и столбца с номером h массива M_{s-1} размерности $p_{s-1} \times s$ ($p_{s-1} = C_l^s$), содержащего по строкам все комбинации сочетаний без повторений из l переменных по s ; α_j , $j = \overline{0, l}$ – неизвестные коэффициенты линейной комбинации переменных; $\alpha_j^{\min,s}$, $\alpha_j^{\max,s}$ ($s = \overline{2, l}$, $j = \overline{1, p_{s-1}}$) – неизвестные коэффициенты при s -арных неэлементарных операциях $\min$ и $\max$, преобразующих j -ю комбинацию сочетаний переменных; $k_{jh}^{\min,s}$, $k_{jh}^{\max,s}$ ($s = \overline{2, l}$, $j = \overline{1, p_{s-1}}$, $h = \overline{1, s-1}$) – неизвестные коэффициенты с номером h в операциях $\min$ и $\max$ s -й арности, преобразующих комбинацию сочетаний входных переменных под номером j ; ε_i – ошибка аппроксимации под номером i .

Используя математическую операцию модуль, впервые введем следующее обобщение НЛР (1):

$$\begin{aligned}
 y_i = & \alpha_0 + \sum_{j=1}^l \alpha_j \cdot |x_{ij} - \lambda_j| + \sum_{j=1}^{p_1} \alpha_j^{\min,2} \cdot \min \left\{ \left| x_{i,\mu_{j1}^{(1)}} - \lambda_{j1}^{\min,2} \right|, k_j^{\min,2} \cdot \left| x_{i,\mu_{j2}^{(1)}} - \lambda_{j2}^{\min,2} \right| \right\} + \\
 & + \sum_{j=1}^{p_1} \alpha_j^{\max,2} \cdot \max \left\{ \left| x_{i,\mu_{j1}^{(1)}} - \lambda_{j1}^{\max,2} \right|, k_j^{\max,2} \cdot \left| x_{i,\mu_{j2}^{(1)}} - \lambda_{j2}^{\max,2} \right| \right\} + \\
 & + \sum_{j=1}^{p_2} \alpha_j^{\min,3} \cdot \min \left\{ \left| x_{i,\mu_{j1}^{(2)}} - \lambda_{j1}^{\min,3} \right|, k_{j1}^{\min,3} \cdot \left| x_{i,\mu_{j2}^{(2)}} - \lambda_{j2}^{\min,3} \right|, k_{j2}^{\min,3} \cdot \left| x_{i,\mu_{j3}^{(2)}} - \lambda_{j3}^{\min,3} \right| \right\} + \\
 & + \sum_{j=1}^{p_2} \alpha_j^{\max,3} \cdot \max \left\{ \left| x_{i,\mu_{j1}^{(2)}} - \lambda_{j1}^{\max,3} \right|, k_{j1}^{\max,3} \cdot \left| x_{i,\mu_{j2}^{(2)}} - \lambda_{j2}^{\max,3} \right|, k_{j2}^{\max,3} \cdot \left| x_{i,\mu_{j3}^{(2)}} - \lambda_{j3}^{\max,3} \right| \right\} + K + \\
 & + \alpha_1^{\min,l} \cdot \min \left\{ \left| x_{i1} - \lambda_1^{\min,l} \right|, k_1^{\min,l} \cdot \left| x_{i2} - \lambda_2^{\min,l} \right|, \dots, k_{l-1}^{\min,l} \cdot \left| x_{il} - \lambda_{l-1}^{\min,l} \right| \right\} + \\
 & + \alpha_1^{\max,l} \cdot \max \left\{ \left| x_{i1} - \lambda_1^{\max,l} \right|, k_1^{\max,l} \cdot \left| x_{i2} - \lambda_2^{\max,l} \right|, \dots, k_{l-1}^{\max,l} \cdot \left| x_{il} - \lambda_{l-1}^{\max,l} \right| \right\} + \varepsilon_i, \quad i = \overline{1, n}, \quad (2)
 \end{aligned}$$

где λ_j , $j = \overline{1, l}$, $\lambda_{jh}^{\min,s}$, $\lambda_{jh}^{\max,s}$, $s = \overline{2, l}$, $j = \overline{1, p_{s-1}}$, $h = \overline{1, s}$ – неизвестные параметры.

Будем называть спецификацию (2) моделью неэлементарной модульной линейной регрессии (НМЛР). Очевидно, что если значения всех объясняющих переменных неотрицательны, т.е. $x_{ij} \geq 0$, то НЛР (1) является частным случаем НМЛР (2) при $\lambda_j = 0$, $j = \overline{1, l}$, $\lambda_{jh}^{\min,s} = 0$, $\lambda_{jh}^{\max,s} = 0$, $s = \overline{2, l}$, $j = \overline{1, p_{s-1}}$, $h = \overline{1, s}$. Скорее всего, аналогичный вывод можно получить для ситуации, когда переменные принимают любые значения.

Рассмотрим еще частные случаи НМЛР (2):

1) если $\alpha_j^{\min,s} = 0$, $\alpha_j^{\max,s} = 0$ ($s = \overline{2, l}$, $j = \overline{1, p_{s-1}}$), то имеем модульную линейную регрессию [18] вида

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j \cdot |x_{ij} - \lambda_j| + \varepsilon_i, \quad i = \overline{1, n}, \quad (3)$$

которая при $\lambda_j = 0$, $j = \overline{1, l}$ вырождается в линейную регрессию

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} + \varepsilon_i, \quad i = \overline{1, n};$$

2) если $\alpha_j^{\min,s} = 0$, $\alpha_j^{\max,s} = 0$ ($s = \overline{3, l}$, $j = \overline{1, p_{s-1}}$), то имеем НМЛР только с бинарными операциями $\min$ и $\max$ вида

$$\begin{aligned}
 y_i = & \alpha_0 + \sum_{j=1}^l \alpha_j \cdot |x_{ij} - \lambda_j| + \sum_{j=1}^{p_1} \alpha_j^{\min,2} \cdot \min \left\{ \left| x_{i,\mu_{j1}^{(1)}} - \lambda_{j1}^{\min,2} \right|, k_j^{\min,2} \cdot \left| x_{i,\mu_{j2}^{(1)}} - \lambda_{j2}^{\min,2} \right| \right\} + \\
 & + \sum_{j=1}^{p_1} \alpha_j^{\max,2} \cdot \max \left\{ \left| x_{i,\mu_{j1}^{(1)}} - \lambda_{j1}^{\max,2} \right|, k_j^{\max,2} \cdot \left| x_{i,\mu_{j2}^{(1)}} - \lambda_{j2}^{\max,2} \right| \right\} + \varepsilon_i, \quad i = \overline{1, n}, \quad (4)
 \end{aligned}$$

которая при $\lambda_j = 0$, $j = \overline{1, l}$, $\lambda_{jh}^{\min,2} = 0$, $\lambda_{jh}^{\max,2} = 0$, $j = \overline{1, p_1}$, $h = \overline{1, 2}$ вырождается в обычную НЛР с бинарными операциями $\min$ и $\max$ [15] вида

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j \cdot x_{ij} + \sum_{j=1}^{p_l} \alpha_j^{\min,2} \cdot \min \left\{ x_{i,\mu_{j1}^{(1)}}, k_j^{\min,2} \cdot x_{i,\mu_{j2}^{(1)}} \right\} + \sum_{j=1}^{p_l} \alpha_j^{\max,2} \cdot \max \left\{ x_{i,\mu_{j1}^{(1)}}, k_j^{\max,2} \cdot x_{i,\mu_{j2}^{(1)}} \right\} + \varepsilon_i, \quad i = \overline{1, n}; \quad (5)$$

3) если $\alpha_j = 0$ ($j = \overline{0, l}$), $\alpha_j^{\min,s} = 0$, $\alpha_j^{\max,s} = 0$ ($s = \overline{2, l-1}$, $j = \overline{1, p_{s-1}}$), $\alpha_1^{\max,l} = 0$, то имеем модульную функцию Леонтьева вида

$$y_i = \alpha_1^{\min,l} \cdot \min \left\{ |x_{i1} - \lambda_1^{\min,l}|, k_1^{\min,l} \cdot |x_{i2} - \lambda_2^{\min,l}|, \dots, k_{l-1}^{\min,l} \cdot |x_{il} - \lambda_{l-1}^{\min,l}| \right\} + \varepsilon_i, \quad i = \overline{1, n},$$

4) которая при $\lambda_j^{\min,l} = 0$, $j = \overline{1, l-1}$ вырождается в обычную функцию Леонтьева [10] $y_i = \alpha_1^{\min,l} \cdot \min \left\{ x_{i1}, k_1^{\min,l} \cdot x_{i2}, \dots, k_{l-1}^{\min,l} \cdot x_{il} \right\} + \varepsilon_i$, $i = \overline{1, n}$;

5) если $\alpha_j = 0$ ($j = \overline{0, l}$), $\alpha_j^{\min,s} = 0$, $\alpha_j^{\max,s} = 0$ ($s = \overline{2, l-1}$, $j = \overline{1, p_{s-1}}$), то имеем модульную НМЛР вида

$$y_i = \alpha_1^{\min,l} \cdot \min \left\{ |x_{i1} - \lambda_1^{\min,l}|, k_1^{\min,l} \cdot |x_{i2} - \lambda_2^{\min,l}|, \dots, k_{l-1}^{\min,l} \cdot |x_{il} - \lambda_{l-1}^{\min,l}| \right\} + \alpha_1^{\max,l} \cdot \max \left\{ |x_{i1} - \lambda_1^{\max,l}|, k_1^{\max,l} \cdot |x_{i2} - \lambda_2^{\max,l}|, \dots, k_{l-1}^{\max,l} \cdot |x_{il} - \lambda_{l-1}^{\max,l}| \right\} + \varepsilon_i, \quad i = \overline{1, n},$$

которая при $\lambda_j^{\min,l} = 0$, $\lambda_j^{\max,l} = 0$, $j = \overline{1, l-1}$ вырождается в кусочно-линейную регрессию [12].

Для разработки нового алгоритма нахождения оценок НМЛР рассмотрим сначала уже известный и исследованный ранее алгоритм приближенного МНК-оценивания линейной регрессии с модулями (3) [18].

Шаг 1. Вычисляется минимальное $x_{\min}^j$ и максимальное $x_{\max}^j$ значения каждой входной переменной под номером j .

Шаг 2. На каждом отрезке $[x_{\min}^j, x_{\max}^j]$ равномерно выбирается p точек.

Шаг 3. Методом «всех возможных регрессий», используя $(p+2)^l$ комбинаций точек с отрезков $[x_{\min}^j, x_{\max}^j]$ вместо параметров λ_j , с помощью МНК находятся оценки параметров α_j , $j = \overline{1, l}$.

Шаг 4. Выбирается модель с минимальной величиной суммы квадратов остатков.

Чем больше число p разбиений отрезков, тем ближе МНК-оценки модульной регрессии к оптимальным.

Аналогичный алгоритм приближенного оценивания известен для НЛР (5) [13]. Разница в том, что для перебора параметров $k_j^{\min,2}$ и $k_j^{\max,2}$ необходимо формировать интервалы $(k_j^{\text{нижн}}, k_j^{\text{верхн}})$, где

$$k_j^{\text{нижн}} = \min \left\{ \frac{x_{1,\mu_{j1}}}{x_{1,\mu_{j2}}}, \frac{x_{2,\mu_{j1}}}{x_{2,\mu_{j2}}}, \dots, \frac{x_{n,\mu_{j1}}}{x_{n,\mu_{j2}}} \right\}, \quad k_j^{\text{верхн}} = \max \left\{ \frac{x_{1,\mu_{j1}}}{x_{1,\mu_{j2}}}, \frac{x_{2,\mu_{j1}}}{x_{2,\mu_{j2}}}, \dots, \frac{x_{n,\mu_{j1}}}{x_{n,\mu_{j2}}} \right\}.$$

Точки на концах этих интервалов нельзя использовать из-за возникновения совершенной мультиколлинеарности факторов.

Разработать алгоритм оценивания сразу для всей конструкции (2), обеспечивающий близость решения к оптимальному, пока затруднительно, поэтому была сформулирована следующая упрощенная форма НМЛР (4) при $\lambda_{jh}^{\min,2} = 0$, $\lambda_{jh}^{\max,2} = 0$, $j = \overline{1, p_1}$, $h = \overline{1, 2}$:

$$y_i = \alpha_0 + \sum_{j=1}^l \alpha_j \cdot |x_{ij} - \lambda_j| + \sum_{j=1}^{p_1} \alpha_j^{\min,2} \cdot \min \left\{ x_{i, \mu_{j1}^{(1)}}, k_j^{\min,2} \cdot x_{i, \mu_{j2}^{(1)}} \right\} + \sum_{j=1}^{p_1} \alpha_j^{\max,2} \cdot \max \left\{ x_{i, \mu_{j1}^{(1)}}, k_j^{\max,2} \cdot x_{i, \mu_{j2}^{(1)}} \right\} + \varepsilon_i, \quad i = \overline{1, n}. \quad (6)$$

Тогда алгоритм приближенного оценивания НМЛР (6) аналогичен двум предыдущим лишь с разницей в том, что необходимо формировать и отрезки $[x_{\min}^j, x_{\max}^j]$ для перебора параметров λ_j , и интервалы $(k_j^{\text{нижн}}, k_j^{\text{верхн}})$ для перебора параметров $k_j^{\min,2}$ и $k_j^{\max,2}$. Причем между модульными и бинарными регрессорами в (6) возможно наличие мультиколлинеарности, что необходимо учитывать при реализации алгоритма.

Результаты и обсуждение

Актуальной в настоящее время задачей является совершенствование механизмов прогнозирования инфляции [19, 20] в Российской Федерации и в ее субъектах. Было принято решение апробировать предложенный в данной статье математический аппарат на примере моделирования ИПЦ в Республике Алтай. Для этого использовались ежегодные статистические данные (<https://rosstat.gov.ru/>) за период 2000–2022 гг. по следующим переменным: y – индекс потребительских цен (%); x_1 – индекс цен производителей на строительную продукцию (%); x_2 – индекс цен производителей сельскохозяйственной продукции (%); x_3 – индекс цен приобретения промышленных товаров и услуг (%); x_4 – индекс тарифов на грузовые перевозки (%). Эти данные для воспроизводимости экспериментов представлены в табл. 1.

Таблица 1

Статистические данные по Республике Алтай

y	x_1	x_2	x_3	x_4
1	2	3	4	5
123,6	133,9	116,3	130,8	125,1
113,7	97	128,6	116,7	132,9
116,4	113,8	114,3	111,3	109
113,5	108,1	110,9	112,8	103,8
112,1	122,8	106,9	125	100
111,1	96,9	127	127,3	104,4
107,9	118,2	127,8	117,6	108,1
111,3	112,6	111,1	109	107,5
111,8	125	112,4	112,5	115,2
111,2	96	102,5	106,1	92,5

Окончание табл. 1

1	2	3	4	5
108,6	115,8	107,1	106,6	100
106,4	100,6	125,5	109,3	100
107	120,1	110,1	108,7	110,4
106,4	91,8	107,5	112,9	100,2
110,1	102,5	107,8	106,8	100
112,5	105,4	118	104,3	107,1
103,8	88,3	100,8	107,9	100
101,6	103,3	104,2	104,5	100,3
102,9	109,3	105,9	117,8	103,9
102	104,5	108,7	95,9	100,6
104,2	102,4	99,7	103,4	100
108,3	113,4	102,3	114,4	95,9
113,9	111,1	107,7	110,1	100

Уравнение оцененной по этим статистическим данным с помощью МНК линейной регрессии имеет вид

$$\hat{y} = 54,622 + 0,118x_1 + 0,023x_2 + 0,2x_3 + 0,163x_4, \quad (7)$$

(1,303) (0,175) (1,526) (1,247)

где в круглых скобках под коэффициентами при переменных указаны наблюдаемые статистики t -критерия Стьюдента. Коэффициент детерминации регрессии (7) $R^2 = 0,471361$, т.е. модель объясняет примерно 47 % вариации переменной y , поэтому качество ее аппроксимации следует признать ниже среднего уровня. К тому же абсолютно все МНК-оценки построенной модели незначимы по t -критерию Стьюдента для выбранного уровня значимости $\alpha = 0,05$ (критическое значение t -критерия составляет 2,101).

Затем с помощью МНК оценивались следующие разновидности обычной НЛР (5):

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_1^{\min,2} \min\{x_3, k_1^{\min,2} x_4\} + \alpha_1^{\max,2} \max\{x_3, k_1^{\max,2} x_4\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_3 + \alpha_1^{\min,2} \min\{x_2, k_1^{\min,2} x_4\} + \alpha_1^{\max,2} \max\{x_2, k_1^{\max,2} x_4\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_4 + \alpha_1^{\min,2} \min\{x_2, k_1^{\min,2} x_3\} + \alpha_1^{\max,2} \max\{x_2, k_1^{\max,2} x_3\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_3 + \alpha_1^{\min,2} \min\{x_1, k_1^{\min,2} x_4\} + \alpha_1^{\max,2} \max\{x_1, k_1^{\max,2} x_4\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1 x_2 + \alpha_2 x_4 + \alpha_1^{\min,2} \min\{x_1, k_1^{\min,2} x_3\} + \alpha_1^{\max,2} \max\{x_1, k_1^{\max,2} x_3\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1 x_3 + \alpha_2 x_4 + \alpha_1^{\min,2} \min\{x_1, k_1^{\min,2} x_2\} + \alpha_1^{\max,2} \max\{x_1, k_1^{\max,2} x_2\} + \varepsilon.$$

Их оценивание проводилось с использованием специально созданного скрипта на языке эконометрического пакета Gretl. Поскольку при оценивании любой линейной регрессии Gretl автоматически обнаруживает мультиколлинеарность и исключает факторы ее возникновения, то было решено для перебора

параметров $k_j^{\min,2}$ и $k_j^{\max,2}$ формировать отрезки $[k_j^{\text{нижн}}, k_j^{\text{верхн}}]$. Каждый из этих отрезков равномерно разбивался десятью точками. Это значит, что для приближенного оценивания каждой из шести НЛР с учетом точек на концах отрезков приходилось оценивать 144 линейных регрессий.

Лучшей из оцененных моделей по величине R^2 оказалась следующая НЛР:

$$\hat{y} = 38,304 + \underset{(1,159)}{0,096}x_1 + \underset{(2,050)}{0,247}x_3 + \underset{(2,272)}{0,637}\min\{x_2, 1.107x_4\} - \underset{(-1,498)}{0,331}x_2, \quad (8)$$

для которой $R^2 = 0,553682$. Как видно, в этой модели появился один значимый по t -критерию регрессор $\min\{x_2, 1.107x_4\}$. Заметим, что последний регрессор x_2 равносильно преобразованию $\max\{x_2, 0.9296x_4\}$.

После чего с помощью МНК оценивались следующие разновидности НМЛР (6):

$$y = \alpha_0 + \alpha_1|x_1 - \lambda_1| + \alpha_2|x_2 - \lambda_2| + \alpha_1^{\min,2}\min\{x_3, k_1^{\min,2}x_4\} + \alpha_1^{\max,2}\max\{x_3, k_1^{\max,2}x_4\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1|x_1 - \lambda_1| + \alpha_2|x_3 - \lambda_2| + \alpha_1^{\min,2}\min\{x_2, k_1^{\min,2}x_4\} + \alpha_1^{\max,2}\max\{x_2, k_1^{\max,2}x_4\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1|x_1 - \lambda_1| + \alpha_2|x_4 - \lambda_2| + \alpha_1^{\min,2}\min\{x_2, k_1^{\min,2}x_3\} + \alpha_1^{\max,2}\max\{x_2, k_1^{\max,2}x_3\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1|x_2 - \lambda_1| + \alpha_2|x_3 - \lambda_2| + \alpha_1^{\min,2}\min\{x_1, k_1^{\min,2}x_4\} + \alpha_1^{\max,2}\max\{x_1, k_1^{\max,2}x_4\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1|x_2 - \lambda_1| + \alpha_2|x_4 - \lambda_2| + \alpha_1^{\min,2}\min\{x_1, k_1^{\min,2}x_3\} + \alpha_1^{\max,2}\max\{x_1, k_1^{\max,2}x_3\} + \varepsilon,$$

$$y = \alpha_0 + \alpha_1|x_3 - \lambda_1| + \alpha_2|x_4 - \lambda_2| + \alpha_1^{\min,2}\min\{x_1, k_1^{\min,2}x_2\} + \alpha_1^{\max,2}\max\{x_1, k_1^{\max,2}x_2\} + \varepsilon.$$

Их оценивание также проводилось в Gretl. Для перебора параметров $k_j^{\min,2}$ и $k_j^{\max,2}$ формировались отрезки $[k_j^{\text{нижн}}, k_j^{\text{верхн}}]$, а для $\lambda_j - [x_{\min}^j, x_{\max}^j]$. Каждый из отрезков также равномерно разбивался десятью точками. Но теперь задача существенно усложнилась – для приближенного оценивания каждой из шести НМЛР приходилось оценивать уже 20 736 линейных регрессий.

Лучшей из оцененных моделей по величине R^2 оказалась следующая НМЛР:

$$\hat{y} = 93,568 - \underset{(-5,074)}{0,728}|x_2 - 115.464| + \underset{(3,485)}{0,317}|x_4 - 103.518| + \underset{(2,998)}{1,166}x_3 - \underset{(-2,237)}{0,67} \cdot \max\{x_1, 1.047x_3\}, \quad (9)$$

для которой $R^2 = 0,77156$. В этом уравнении абсолютно все коэффициенты значимы по t -критерию. Регрессор x_3 равносильно преобразованию $\min\{x_1, 0.761x_3\}$.

Заключение

Впервые сформулирована относящаяся к нелинейным по оцениваемым параметрам спецификациям модель неэлементарной модульной линейной регрессии. Она обобщает разработанные и исследованные ранее модели линейной, неэлементарной, модульной, кусочно-линейной регрессии, функции Леонтьева. Она служит стойким фундаментом для введения в рассмотрение в будущем новых форм регрессионных моделей. Предложен алгоритм приближенного МНК-оценивания одной из упрощенных возможных форм неэлементарной модульной регрессии. Построенная с помощью него модель оказалась на 39,3 % лучше по качеству аппроксимации неэлементарной регрессии без модулей и на 63,7 % лучше линейной регрессии. При этом построенная модель одержала абсолютную победу по количеству значимых по t -критерию регрессоров. Дальнейшие исследования планируется посвятить поиску алгоритмов эффективного оценивания и интерпретации [21] предложенных регрессий.

Список литературы

1. Tsui K. L., Chen V., Jiang W. [et al.]. Data mining methods and applications // Springer handbook of engineering statistics. London : Springer London, 2023. P. 797–816.
2. Dogan A., Birant D. Machine learning and data mining in manufacturing // Expert Systems with Applications. 2021. Vol. 166. P. 114060.
3. Montgomery D. C., Peck E. A., Vining G. G. Introduction to linear regression analysis. John Wiley & Sons, 2021.
4. Fox J. Regression diagnostics: An introduction. Sage publications, 2019.
5. Ресулкулыева Г., Серебрянский С. А. Весовая модель конструкции фюзеляжа, крыла и оперения самолета на основе регрессионного анализа // Управление развитием крупномасштабных систем (MLSD'2022). 2022. С. 918–924.
6. Гилев Д. В., Елтышев Д. К. Регрессионный анализ влияния климатических факторов на электропотребление объектов нефтедобычи // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. 2020. № 35. С. 152–168.
7. Садыгова М. К., Кириллова Т. В., Каневская И. Ю. Оптимизация содержания пищевых волокон в рецептуре хлебобулочного изделия методом регрессионного анализа // XXI век: итоги прошлого и проблемы настоящего плюс. 2020. Т. 9, № 1. С. 135–140.
8. Демчук Е. В., Мяло В. В., Браулик Р. А., Бардола А. С. Обоснование рациональных параметров двухленточного сошника методом регрессионного анализа // Вестник Омского государственного аграрного университета. 2020. № 4 (40). С. 125–131.
9. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. М. : ЮНИТИ, 1998. 1005 с.
10. Клейнер Г. Б. Производственные функции: теория, методы, применение. М. : Финансы и статистика, 1986. 239 с.
11. Хацкевич Г. А., Проневич А. Ф., Чайковский М. В. Двухфакторные производственные функции с заданной предельной нормой замещения // Экономическая наука сегодня. 2019. № 10. С. 169–181.
12. 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.

13. Базилевский М. П. Оценка линейно-неэлементарных регрессионных моделей с помощью метода наименьших квадратов // Моделирование, оптимизация и информационные технологии. 2020. Т. 8, № 4 (31).
14. Базилевский М. П. Отбор информативных операций при построении линейно-неэлементарных регрессионных моделей // International Journal of Open Information Technologies. 2021. Т. 9, № 5. С. 30–35.
15. Базилевский М. П. Метод построения неэлементарных линейных регрессий на основе аппарата математического программирования // Проблемы управления. 2022. № 4. С. 3–14.
16. Базилевский М. П. Обобщение неэлементарных линейных регрессий // Моделирование и анализ данных. 2023. Т. 13, № 2. С. 85–98.
17. Базилевский М. П., Ойдопова А. Б. Оценка модульных линейных регрессионных моделей с помощью метода наименьших модулей // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. 2023. № 45. С. 130–146.
18. Базилевский М. П. Программное обеспечение для оценивания модульных линейных регрессий // Информационные и математические технологии в науке и управлении. 2023. № 3 (31). С. 136–146.
19. Юревич М. А. Инфляционные ожидания и инфляция: наука и прогноз // Journal of Economic Regulation (Вопросы регулирования экономики). 2021. Т. 12, № 2. С. 22–35.
20. Карабутов Н. Н. Взаимосвязь индексов, определяющих уровень инфляции // Экономика. Налоги. Право. 2020. Т. 13, № 1. С. 42–48.
21. Molnar C. Interpretable machine learning // Lulu.com. 2020.

References

1. Tsui K.L., Chen V., Jiang W. et al. Data mining methods and applications. *Springer handbook of engineering statistics*. London: Springer London, 2023:797–816.
2. Dogan A., Birant D. Machine learning and data mining in manufacturing. *Expert Systems with Applications*. 2021;166:114060.
3. Montgomery D.C., Peck E.A., Vining G.G. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.
4. Fox J. *Regression diagnostics: an introduction*. Sage publications, 2019.
5. Resulkulyeva G., Serebryanskiy S.A. A weight model of the fuselage, wing and tail structure of an aircraft based on regression analysis. *Upravlenie razvitiem krupnomasshtabnykh sistem (MLSD'2022) = Management of the development of large-scale systems (MLSD'2022)*. 2022:918–924. (In Russ.)
6. Gilev D.V., Eltyshv D.K. Regression analysis of the influence of climatic factors on the power consumption of oil production facilities. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotehnika, informatsionnye tekhnologii, sistemy upravleniya = Bulletin of the Perm National Research Polytechnic University. Electrical engineering, information technology, control systems*. 2020;(35):152–168. (In Russ.)
7. Sadygova M.K., Kirillova T.V., Kanevskaya I.Yu. Optimization of the content of dietary fibers in the formulation of a bakery product by regression analysis. *XXI vek: itogi proshlogo i problemy nastoyashchego plyus = XXI century: results of the past and problems of the present plus*. 2020;9(1):135–140. (In Russ.)
8. Demchuk E.V., Myalo V.V., Braulik R.A., Bardola A.S. Substantiation of rational parameters of a two-tape coulter by regression analysis. *Vestnik Omskogo gosudarstvennogo agrarnogo universiteta = Bulletin of Omsk State Agrarian University*. 2020;(4):125–131. (In Russ.)
9. Ayvazyan S.A., Mkhitarian V.S. *Prikladnaya statistika i osnovy ekonometriki = Applied statistics and fundamentals of econometrics*. Moscow: YuNITI, 1998:1005. (In Russ.)

10. Kleyner G.B. *Proizvodstvennye funktsii: teoriya, metody, primeneniye* = *Production functions: theory, methods, application*. Moscow: Finansy i statistika, 1986:239. (In Russ.)
11. Khatskevich G.A., Pronevich A.F., Chaykovskiy M.V. Two-factor production functions with a given marginal rate of substitution. *Ekonomicheskaya nauka segodnya* = *Economics today*. 2019;(10):169–181. (In Russ.)
12. Noskov S.I., Khonyakov A.A. A software package for constructing some types of piecewise linear regressions. *Informatsionnye tekhnologii i matematicheskoe modelirovaniye v upravlenii slozhnymi sistemami* = *Information technologies and mathematical modeling in the management of complex systems*. 2019;(3):47–55. (In Russ.)
13. Bazilevskiy M.P. Estimation of linear non-elementary regression models using the least squares method. *Modelirovaniye, optimizatsiya i informatsionnye tekhnologii* = *Modeling, optimization and information technologies*. 2020;8(4).
14. Bazilevskiy M.P. Selection of informative operations in the construction of linear non-elementary regression models. *International Journal of Open Information Technologies*. 2021;9(5):30–35. (In Russ.)
15. Bazilevskiy M.P. Method of constructing non-elementary linear regressions based on mathematical programming apparatus. *Problemy upravleniya* = *Management problems*. 2022;(4):3–14. (In Russ.)
16. Bazilevskiy M.P. Generalization of non-elementary linear regressions. *Modelirovaniye i analiz dannykh* = *Modeling and data analysis*. 2023;13(2):85–98. (In Russ.)
17. Bazilevskiy M.P., Oydopova A.B. Estimation of modular linear regression models using the method of least modules. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniya* = *Bulletin of the Perm National Research Polytechnic University. Electrical engineering, information technology, control systems*. 2023;(45):130–146. (In Russ.)
18. Bazilevskiy M.P. Software for estimating modular linear regressions. *Informatsionnye i matematicheskie tekhnologii v nauke i upravlenii* = *Information and mathematical technologies in science and management*. 2023;(3):136–146. (In Russ.)
19. Yurevich M.A. Inflationary expectations and inflation: science and forecasting. *Journal of Economic Regulation (Voprosy regulirovaniya ekonomiki)* = *Journal of Economic Regulation (Issues of economic regulation)*. 2021;12(2):22–35. (In Russ.)
20. Karabutov N.N. The relationship of indices determining the level of inflation. *Ekonomika. Nalogi. Pravo* = *Economy. Taxes. Right*. 2020;13(1):42–48. (In Russ.)
21. Molnar C. Interpretable machine learning. *Lulu.com*. 2020.

Информация об авторах / Information about the authors

Михаил Павлович Базилевский

кандидат технических наук, доцент,
доцент кафедры математики,
Иркутский государственный университет
путей сообщения
(Россия, г. Иркутск, ул. Чернышевского, 15)
E-mail: mik2178@yandex.ru

Mikhail P. Bazilevskiy

Candidate of technical sciences,
associate professor, associate professor
of the sub-department of mathematics,
Irkutsk State Transport University
(15 Chernyshevskogo street, Irkutsk, Russia)

**Автор заявляет об отсутствии конфликта интересов /
The author declares no conflicts of interests.**

Поступила в редакцию/Received 12.02.2024

Поступила после рецензирования/Revised 02.05.2024

Принята к публикации/Accepted 17.06.2024