

**М.П. Базилевский***Иркутский государственный университет путей сообщения,
г. Иркутск, Российская Федерация*

Идентификация неизвестных параметров списочных регрессионных моделей методом наименьших модулей

Аннотация. Регрессионный анализ, как ёмкая составная часть машинного обучения, находит широкое применение при решении прикладных задач выявления зависимостей между выходной переменной и одной или несколькими входными переменными. Многие процессы в природе нелинейны, поэтому применение линейных регрессий для идентификации зависимостей между переменными часто приводит к плохим результатам. Оценивание же нелинейных регрессий всегда характеризовалось сложностью вычислительных процедур. Однако современное развитие вычислительной техники позволяет вновь вернуться к созданию новых эффективных спецификаций нелинейных регрессионных моделей и методов их параметризации. Одна из возможных таких спецификаций разработана в данной статье. В её основе лежит предложенная ранее регрессия с целочисленной функцией «пол», которую целесообразно применять при наличии целочисленной выходной переменной. Возникла идея, что целочисленную функцию в модели регрессии можно применить для прогнозирования номера функции из некоторого наперёд заданного списка уравнений. Введенная спецификация названа списочной регрессией. Она состоит из назначенного исследователем списка функций нескольких переменных и правила, по которому осуществляется переключение между функциями. Задача идентификации неизвестных параметров списочной регрессии с помощью метода наименьших модулей сведена к задаче частично-булевого линейного программирования. В результате оценки находятся и параметры функций из списка, и коэффициенты для правила их переключения. На основе выборки данных небольшого объема проведены вычислительные эксперименты. Каждая из шести оцененных списочных регрессий оказались по качеству лучше линейной регрессии и производственной функции Леонтьева.

Ключевые слова. Регрессионный анализ, списочная регрессия, метод наименьших модулей, задача частично-булевого линейного программирования, целочисленная функция «пол», функция Леонтьева.

Информация о статье. Дата поступления: 12 марта 2025 г.; дата принятия к публикации: 11 июня 2025 г.; дата онлайн-размещения: 8 июля 2025 г.

Original article

M.P. Bazilevskiy*Irkutsk State Transport University,
Irkutsk, Russian Federation*

Identification of Unknown Parameters in List Regression Models Using the Least Absolute Deviations

Abstract. Regression analysis, as a capacious component of machine learning, is widely used in solving applied problems of identifying dependencies between an output

variable and one or more input variables. Many processes in nature are nonlinear, so the use of linear regressions to identify dependencies between variables often leads to poor results. The estimation of nonlinear regressions has always been characterized by the complexity of computational procedures. However, the modern development of computing technology allows us to return to the creation of new effective specifications for nonlinear regression models and methods for their parameterization. One of the possible such specifications is developed in this article. It is based on the previously proposed regression with an integer function «floor», which is advisable to use in the presence of an integer output variable. The idea arose that an integer function in a regression model could be used to predict a function number from a predefined list of equations. This type of regression is called list regression, and it consists of a set of functions with several variables that are assigned by the researcher, as well as a rule for switching between these functions. The problem of estimating the unknown parameters of these list regressions using the least absolute deviation method is reduced to mixed 0-1 integer linear programming. After evaluating the results, both the parameters for the functions in the list and the coefficients for the switching rule were found. Computational experiments were conducted using a small sample of data, and each of the six list regressions evaluated was found to be superior in quality to linear regression and the Leontief production function.

Keywords. Regression analysis, list regression, least absolute deviations, mixed 0-1 integer linear programming, integer function «floor», Leontief function.

Article info. Received 12 March, 2025; Accepted 11 June, 2025; Available online 8 July, 2025.

Введение

Выявление математической зависимости между выходной переменной и одной или несколькими входными переменными по имеющимся статистическим данным, по-прежнему, остаётся актуальной научной задачей. Для её решения может быть применён инструментарий регрессионного анализа [1]. При этом часто исследователи ограничиваются построением лишь самой простой модели множественной линейной регрессии с помощью метода наименьших квадратов. Так, например, в [2] оцениваются параметры линейной регрессии для прогнозирования стоимости квартиры, в [3] — расхода топлива автомобиля, в [4] — стоимости Bitcoin, в [5] — урожайности яровой пшеницы. Линейные модели просто интерпретируются, однако многие зависимости в природе носят нелинейный характер, поэтому при попытке их описания с помощью линейных регрессий может быть получен неадекватный результат.

Среди нелинейных регрессий особенно распространены модели, в которых объясняющие переменные преобразованы с помощью элементарных функций. Например, полиномиальные регрессии [6; 7], относящиеся к нелинейным по факторам, но линейным по параметрам. Допускают линеаризацию с помощью логарифмирования степенные регрессии, более известные в экономике как производственные функции Кобба-Дугласа [8; 9]. Исследований, посвященных построению нелинейных по параметрам регрессионных моделей, гораздо меньше ввиду трудности их оценки. Сре-

ди них хотелось бы выделить работы, в которых для оценки нелинейных спецификаций регрессий использован метод наименьших модулей (МНМ) [10]. Так, в [11] описан алгоритм оценки параметров производственной функции Леонтьева [12], в [13] — псевдобулевой регрессии, в [14] — модульной регрессии, в [15] — вполне интерпретируемой неэлементарной линейной регрессии, в [16] — «широкой» модульной регрессии, в [17] — «глубокой» модульной регрессии. В основе алгоритмов, реализующих оценивание с помощью МНМ перечисленных моделей, лежит, как отмечено в [18], хорошо развитый за последние два десятка лет аппарат частично-булевого линейного программирования (ЧБЛП) [19].

Настоящая статья посвящена разработке и исследованию новой нелинейной спецификации регрессионных моделей. Её появление предопределила работа [20], в которой предложены регрессии с целочисленными функциями «пол» и «потолок». Возникла идея использовать эти целочисленные функции в модели для прогнозирования в зависимости от значений объясняющих переменных номера «срабатывающей» функции из некоторого наперёд заданного списка. Цель статьи состоит в сведении задачи оценки неизвестных параметров новой модельной спецификации к задаче ЧБЛП.

1. Списочная регрессия и идентификация её параметров

Обозначим выходную (объясняемую) переменную как y , а входные (объясняющие) переменные как x_1, x_2, \dots, x_l . Пусть для этих переменных сформирована выборка объема n . Рассмотрим предложенную в [20] регрессию с целочисленной функцией «пол»:

$$y_i = \left\lfloor \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right\rfloor + \varepsilon_i, \quad i = \overline{1, n}, \quad (1)$$

где $y_i, x_{i1}, x_{i2}, \dots, x_{il}, i = \overline{1, n}$ — значения переменных; $\alpha_0, \alpha_1, \dots, \alpha_l$ — неизвестные параметры; $\varepsilon_i, i = \overline{1, n}$ — ошибки регрессии; скобками $\lfloor \rfloor$ обозначена целочисленная функция «пол» [21], округляющая содержащееся в них значение до ближайшего целого числа в меньшую сторону. Например, $\lfloor 10,99 \rfloor = 10$. В этой связи модель (1) целесообразно использовать тогда, когда значения переменной y целые.

В [20] сформулирована следующая задача оценки с помощью МНМ неизвестных параметров регрессии (1):

$$\sum_{i=1}^n \left| y_i - \left\lfloor \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right\rfloor \right| \rightarrow \min. \quad (2)$$

И в той же статье [20] задача (2) была сведена к следующей задаче частично-целочисленного линейного программирования (ЧЦЛП):

$$\sum_{i=1}^n (u_i + v_i) \rightarrow \min, \quad (3)$$

$$y_i = \theta_i + u_i - v_i, \quad i = \overline{1, n}, \quad (4)$$

$$\theta_i \in \mathbb{Z}, \quad (5)$$

$$\theta_i \leq \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \leq \theta_i + 1 - \Delta, \quad i = \overline{1, n}, \quad (6)$$

$$u_i \geq 0, \quad v_i \geq 0, \quad i = \overline{1, n}, \quad (7)$$

где \mathbb{Z} — множество целых чисел; $\theta_i, i = \overline{1, n}$ — целочисленные переменные, равные прогнозным по модели значениям выходной переменной y ; Δ — близкое к нулю положительное число; $u_i, v_i, i = \overline{1, n}$ — неотрицательные переменные, удовлетворяющие условиям:

$$u_i = \begin{cases} y_i - \left[\alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right], & \text{если } y_i - \left[\alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right] > 0, \\ 0, & \text{в противном случае,} \end{cases}$$

$$v_i = \begin{cases} -y_i + \left[\alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right], & \text{если } y_i - \left[\alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right] < 0, \\ 0, & \text{в противном случае.} \end{cases}$$

Для решения оптимизационной задачи ЧЦЛП (3)–(7) можно воспользоваться любой современной программой-решателем. Аналогичным образом в [20] введена регрессия с целочисленной функцией «потолок».

Рассмотрим далее новую спецификацию регрессионных моделей, фундаментом которой будет регрессия (1).

Пусть имеется следующий список из p различных функций нескольких переменных:

$$L = [F_1(\beta^{(1)}, x_1, x_2, \dots, x_l), \\ F_2(\beta^{(2)}, x_1, x_2, \dots, x_l), \\ \dots \\ F_p(\beta^{(p)}, x_1, x_2, \dots, x_l)], \quad (8)$$

где $\beta^{(1)}, \dots, \beta^{(p)}$ — векторы неизвестных параметров, которые могут иметь различное число компонент.

Например, для трех переменных x_1 , x_2 и x_3 список из трех функций может иметь вид:

$$L = [\beta_0^{(1)} + \beta_1^{(1)}x_1 + \beta_2^{(1)}x_2x_3), \\ \beta_0^{(2)} + \beta_1^{(2)}x_1^2 + \beta_2^{(2)}\sqrt{x_2} + \beta_3^{(2)}\sin x_3, \\ \beta_0^{(3)} + \beta_1^{(3)}x_1\sqrt{x_2^2 + 7\ln x_3}].$$

Список функций (8), обозначенный скобками[], не следует отождествлять с множеством функций, обозначаемым скобками {}. Разница в том, что в последнем случае порядок следования элементов не важен, а в списке — важен, т.е. за каждой функцией в списке закреплён конкретный номер. При перестановке местами двух функций в списке получаем новый список. Обозначим номер функции в списке (8) как N . Очевидно, что $N \in \{1, 2, \dots, p\}$.

Введём регрессионную модель вида

$$y_i = F_N(\beta^{(N)}, x_1, x_2, \dots, x_l) + \varepsilon_i, \quad i = \overline{1, n}. \quad (9)$$

Если $N = 1$, то функцией регрессии (9) будет функция под номером 1 из списка (8), если $N = 2$ — функция под номером 2 и т.д. Тем самым, при постоянном номере N имеем традиционную задачу оценки параметров регрессионной модели (9). Интерес вызывает ситуация, когда N в (9) меняется в соответствии с некоторым правилом. Поскольку N принимает только целые значения, то для организации в регрессии (9) переключений функций из списка L введено следующее правило:

$$N_i = \left\lfloor \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right\rfloor, \quad i = \overline{1, n}, \quad (10)$$

где $N_i \in \{1, 2, \dots, p\}$ — номер функции, выбранной из списка L в i -м наблюдении. Таким образом, по правилу (10) осуществляется трансформация значений объясняющих переменных в номер функции из списка L .

Учитывая (10), представим регрессию (9) в виде

$$y_i = F_{\left\lfloor \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right\rfloor} \left(\beta^{\left\lfloor \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \right\rfloor}, x_{i1}, x_{i2}, \dots, x_{il} \right) + \varepsilon_i, \quad i = \overline{1, n}. \quad (11)$$

Сформулированную регрессионную модель (11) будем называть списочной регрессией.

Пусть список (8) состоит из l линейных функций одной переменной и имеет вид

$$L = [\beta_0^{(1)} + \beta_1^{(1)}x_1, \\ \beta_0^{(2)} + \beta_1^{(2)}x_2, \\ \beta_0^{(l)} + \beta_1^{(l)}x_l]. \quad (12)$$

С учётом (12) списочная регрессия (11) принимает вид

$$y_i = \beta_0^{(N_i)} + \beta_1^{(N_i)}x_{i, N_i} + \varepsilon_i, \quad i = \overline{1, n}, \quad (13)$$

где N_i — номера сработавших в наблюдениях функций, которые находятся по правилу (10), причём, $N_i \in \{1, 2, \dots, l\}$.

Перейдем к решению вопроса оценки неизвестных параметров регрессии (13) с помощью МНМ.

Пусть d_{ij} — значение j -й функции из списка (12), вычисленное в i -м наблюдении, т.е.

$$d_{ij} = \beta_0^{(j)} + \beta_1^{(j)}x_{ij}, \quad i = \overline{1, n}, \quad j = \overline{1, l}. \quad (14)$$

Как следует из (13), в каждом наблюдении должна «срабатывать» ровно одна функция из списка (12). Для этого введем вспомогательные переменные z_{ij} , $i = \overline{1, n}$, $j = \overline{1, l}$ по правилу:

$$z_{ij} = \begin{cases} d_{ij}, & \text{при срабатывании в } i\text{-м наблюдении } j\text{-й функции,} \\ 0, & \text{в противном случае.} \end{cases}$$

Для переключения функций используем следующие ограничения:

$$-M\sigma_{ij} \leq z_{ij} \leq M\sigma_{ij}, \quad i = \overline{1, n}, \quad j = \overline{1, l}, \quad (15)$$

$$-M(1 - \sigma_{ij}) \leq z_{ij} - d_{ij} \leq M(1 - \sigma_{ij}), \quad i = \overline{1, n}, \quad j = \overline{1, l}, \quad (16)$$

где M — достаточно большое положительное число; σ_{ij} , $i = \overline{1, n}$, $j = \overline{1, l}$ — булевы переменные, удовлетворяющие условиям

$$\sigma_{ij} = \begin{cases} 1, & \text{если в } i\text{-м наблюдении включается } j\text{-я функция,} \\ 0, & \text{в противном случае.} \end{cases}$$

Если $\sigma_{ij} = 0$, то двойные неравенства (15) и (16) принимают вид

$$0 \leq z_{ij} \leq 0, \\ -M \leq z_{ij} - d_{ij} \leq M,$$

откуда $z_{ij} = 0$, $d_{ij} \in [-M, M]$; если $\sigma_{ij} = 1$, то (15), (16) имеют вид

$$-M \leq z_{ij} \leq M, \\ 0 \leq z_{ij} - d_{ij} \leq 0,$$

откуда $z_{ij} = d_{ij}$, $z_{ij} \in [-M, M]$.

Поскольку в каждом наблюдении должна срабатывать ровно одна функция из списка, то следует ввести соответствующие ограничения на значения булевых переменных:

$$\sum_{j=1}^l \sigma_{ij} = 1, \quad i = \overline{1, n}. \quad (17)$$

Для трансформации значений объясняющих переменных в номер функции из списка (12) по правилу (10) использованы следующие ограничения:

$$\sum_{j=1}^l j \cdot \sigma_{ij} \leq \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij} \leq \sum_{j=1}^l j \cdot \sigma_{ij} + 1 - \Delta, \quad i = \overline{1, n}. \quad (18)$$

Ограничение (18) получено из (6) заменой целочисленных переменных θ_i на $\sum_{j=1}^l j \cdot \sigma_{ij}$. Поскольку должны выполняться равенства (17), то сумма $\sum_{j=1}^l j \cdot \sigma_{ij}$ численно равна номеру сработавшей в i -м наблюдении функции из списка (10).

По аналогии с (4) введём следующие ограничения:

$$y_i = \sum_{j=1}^l z_{ij} + u_i - v_i, \quad i = \overline{1, n}, \quad (19)$$

где $u_i + v_i$, $i = \overline{1, n}$ — модули ошибок списочной регрессии (13).

Избавимся от переменных α_{ij} , подставив (14) в (16):

$$-M(1 - \sigma_{ij}) \leq z_{ij} - \beta_0^{(j)} - \beta_1^{(j)} x_{ij} \leq M(1 - \sigma_{ij}), \quad i = \overline{1, n}, \quad j = \overline{1, l}. \quad (20)$$

Тогда решение задачи ЧБЛП с целевой функцией (3), с линейными ограничениями (7), (15), (17) — (20) и условиями булевости переменных $\sigma_{ij} \in \{0, 1\}$, $i = \overline{1, n}$, $j = \overline{1, l}$, приводит к идентификации неизвестных параметров списочной регрессии (13) с помощью МНМ. Причём, одновременно идентифицируются как параметры из списка функций (12), так и параметры из правила (10).

Из двойных неравенств (18) следует, что для любого наблюдения номер N_i функции из списка всегда попадает в диапазон от 1 до l . Однако при прогнозировании по новым данным номера функций могут принимать выходящие за рамки этого диапазона значения. Во избежание такой ситуации представим оцененную с помощью МНМ списочную регрессию (13) в виде следующей кусочно-заданной функции:

$$\tilde{y} = \begin{cases} \tilde{\beta}_0^{(1)} + \tilde{\beta}_1^{(1)}x_1, & \text{если } \left[\tilde{\alpha}_0 + \sum_{j=1}^l \tilde{\alpha}_j x_j \right] \leq 1, \\ \tilde{\beta}_0^{(2)} + \tilde{\beta}_1^{(2)}x_2, & \text{если } \left[\tilde{\alpha}_0 + \sum_{j=1}^l \tilde{\alpha}_j x_j \right] = 2, \\ \dots \\ \tilde{\beta}_0^{(l)} + \tilde{\beta}_1^{(l)}x_l, & \text{если } \left[\tilde{\alpha}_0 + \sum_{j=1}^l \tilde{\alpha}_j x_j \right] \geq l. \end{cases}$$

Из этого математического представления следует, что выбор номера функции из списка будет осуществлен для любых значений объясняющих переменных.

Стоит отметить, что механизмы функционирования оцененной списочной регрессии (13) и производственной функции Леонтьева, исследованной, например, в [11] и имеющей вид

$$y_i = \min\{\alpha_1 x_{i1}, \alpha_2 x_{i2}, \dots, \alpha_r x_{ir}\} + \varepsilon_i, \quad i = \overline{1, n}, \quad (21)$$

подобны.

В оцененной производственной функции Леонтьева в каждом наблюдении также выбирается функция из исходного набора, но делается это в соответствии с одним и тем же правилом — берётся функция с минимальным значением. В списочной же регрессии выбор функции осуществляется по её номеру из списка, причём, правило выбора (10) можно регулировать.

2. Вычислительные эксперименты

Проводилось сравнение качества аппроксимации оцененных с помощью МНМ моделей линейной и списочной регрессии, а также производственной функции Леонтьева. Для этого были использованы встроенные в пакет Gretl статистические данные data4-1 (табл. 1) по следующим переменным:

- y — цена дома (в тысячах долларов);
- x_1 — площадь помещений (в кв. футах);
- x_2 — количество спален;
- x_3 — количество ванных комнат.

Таблица 1

Статистические данные

№	y	x_1	x_2	x_3
1	199,9	1065	3	1,75
2	228	1254	3	2
3	235	1300	3	2
4	285	1577	4	2,5

Окончание табл. 1

№	y	x_1	x_2	x_3
5	239	1600	3	2
6	293	1750	4	2
7	285	1800	4	2,75
8	365	1870	4	2
9	295	1935	4	2,5
10	290	1948	4	2
11	385	2254	4	3
12	505	2600	3	2,5
13	425	2800	4	3
14	415	3000	4	3

Сначала с помощью МНМ было найдено уравнение линейной регрессии:

$$\tilde{y} = 107,628 + 0,145x_1 - 6,962x_2 - 20,339x_3. \quad (22)$$

Для (22) сумма модулей остатков $J = \sum_{i=1}^n |y_i - \tilde{y}_i|$ равна 343,665.

Затем в результате решения задачи ЧБЛП в пакете LPSolve с помощью МНМ было найдено уравнение производственной функции Леонтьева [11] (21):

$$\tilde{y} = \min\{0,1708x_1, 120,8333x_2, 145x_3\}, \quad (23)$$

для которой $J = 403,142$.

Для повышения точности функции Леонтьева оценивалась её расширенная спецификация вида

$$y_i = \min\{\beta_0^{(1)} + \beta_1^{(1)}x_{i1}, \beta_0^{(2)} + \beta_1^{(2)}x_{i2}, \dots, \beta_0^{(l)} + \beta_1^{(l)}x_{il}\} + \varepsilon_i, \quad i = \overline{1, n}. \quad (24)$$

Оцененная с помощью МНМ регрессия (24) имеет вид:

$$\tilde{y} = \min\{31,122 + 0,157x_1, 512,288 - 24,322x_2, -292,288 + 292,644x_3\}, \quad (25)$$

для которой $J = 270,723$.

Как видно, модель (23) хуже по качеству, чем линейная регрессия (22). Включение в (22) дополнительных свободных членов привело к построению модели (25), превосходящей по качеству и (22), и (23).

После чего с помощью МНМ в зависимости от выбранного списка (12) функций оценивались параметры списочных регрес-

сий (13). В каждом случае в LPSolve решалась задача ЧБЛП (3), (7), (15), (17) — (20) при $M = 10\,000$, $\Delta = 0,0001$. Ниже представлены полученные результаты.

1. Для списка функций $[\beta_0^{(1)} + \beta_1^{(1)}x_1, \beta_0^{(2)} + \beta_1^{(2)}x_2, \beta_0^{(3)} + \beta_1^{(3)}x_3]$ получена модель

$$\tilde{y} = \begin{cases} -757 + 0,6x_1, & \text{при } N \leq 1, \\ 70 + 55x_2, & \text{при } N = 2, \\ 955 - 180x_3, & \text{при } N \geq 3, \end{cases} \quad (26)$$

где $N = \lfloor 2,8051 + 0,000956x_1 - 1,0217x_2 + 0,7095x_3 \rfloor$.

2. Для списка $[\beta_0^{(1)} + \beta_1^{(1)}x_1, \beta_0^{(2)} + \beta_1^{(2)}x_3, \beta_0^{(3)} + \beta_1^{(3)}x_2]$ модель

$$\tilde{y} = \begin{cases} 240,9497 + 0,0279x_1, & \text{при } N \leq 1, \\ -59,24 + 148,08x_3, & \text{при } N = 2, \\ 745 - 80x_2, & \text{при } N \geq 3, \end{cases} \quad (27)$$

где $N = \lfloor 4,064 + 0,0013x_1 - 1,1505x_2 \rfloor$.

3. Для списка $[\beta_0^{(1)} + \beta_1^{(1)}x_2, \beta_0^{(2)} + \beta_1^{(2)}x_1, \beta_0^{(3)} + \beta_1^{(3)}x_3]$ модель

$$\tilde{y} = \begin{cases} 33 + 65x_2, & \text{при } N \leq 1, \\ -118,186 + 0,2232x_1, & \text{при } N = 2, \\ 905 - 160x_3, & \text{при } N \geq 3, \end{cases} \quad (28)$$

где $N = \lfloor 1,0003 - 1,0005x_2 + 1,99965x_3 \rfloor$.

4. Для списка $[\beta_0^{(1)} + \beta_1^{(1)}x_2, \beta_0^{(2)} + \beta_1^{(2)}x_3, \beta_0^{(3)} + \beta_1^{(3)}x_1]$ модель

$$\tilde{y} = \begin{cases} 775 - 90x_2, & \text{при } N \leq 1, \\ -59,24 + 148,08x_3, & \text{при } N = 2, \\ 240,9497 + 0,0279x_1, & \text{при } N \geq 3, \end{cases} \quad (29)$$

где $N = \lfloor 0,3479 + 1,2347x_2 - 0,001174x_1 \rfloor$.

5. Для списка $[\beta_0^{(1)} + \beta_1^{(1)}x_3, \beta_0^{(2)} + \beta_1^{(2)}x_1, \beta_0^{(3)} + \beta_1^{(3)}x_2]$ модель

$$\tilde{y} = \begin{cases} 905 - 160x_3, & \text{при } N \leq 1, \\ -118,186 + 0,2232x_1, & \text{при } N = 2, \\ 33 + 65x_2, & \text{при } N \geq 3, \end{cases} \quad (30)$$

где $N = \lfloor 3,3558 - 1,2881x_3 - 0,000678x_1 + 1,1017x_2 \rfloor$.

6. Для списка $[\beta_0^{(1)} + \beta_1^{(1)}x_3, \beta_0^{(2)} + \beta_1^{(2)}x_2, \beta_0^{(3)} + \beta_1^{(3)}x_1]$ модель

$$\tilde{y} = \begin{cases} 955 - 180x_3, & \text{при } N \leq 1, \\ 70 + 55x_2, & \text{при } N = 2, \\ -757 + 0,6x_1, & \text{при } N \geq 3, \end{cases} \quad (31)$$

где $N = \lfloor 1,9552 - 1,393x_3 + 1,1055x_2 - 0,000303x_1 \rfloor$.

Для модели (26) $J = 101,1$, для (27) — 109,495, для (28) — 131,452, для (29) — 109,495, для (30) — 131,452, для (31) — 101,1. Как видно, любая из списочных регрессий (26)–(31) оказалась точнее как линейной регрессии (22), так и производственных функций Леонтьева (23), (25). Также видно, что среди всех списочных регрессий есть пары моделей с одинаковой величиной J : (26) и (31), (27) и (29), (28) и (30). Лучшими двумя регрессиями по величине J признаются модели (26) и (31), не отличающиеся по качеству, но отличающиеся правилом переключения функций из списка.

В табл. 2 для каждой списочной регрессии в соответствии со списками приведены номера функций, сработавших в каждом наблюдении, а также величины остатков $e_i = y_i - \tilde{y}_i$, $i = \overline{1,14}$.

Таблица 2

Характеристики моделей (26)–(31)

№	Модель (26)		Модель (27)		Модель (28)		Модель (29)		Модель (30)		Модель (31)	
	N	e	N	e	N	e	N	e	N	e	N	e
1	2	-35,1	2	0	1	-28,1	2	0	3	28,1	2	-35,1
2	2	-7	2	-8,92	1	0	2	-8,92	3	0	2	-7
3	2	0	2	-1,92	1	7	2	-1,92	3	7	2	0
4	2	-5	1	0,052	1	-8	3	0,052	3	8	2	-5
5	2	4	2	2,08	2	0,066	2	2,08	2	0,066	2	4
6	1	0	1	3,225	1	0	3	3,225	3	0	3	0
7	2	-5	1	-6,17	2	1,426	3	-6,17	2	1,426	2	-5
8	1	0	1	71,877	1	72	3	71,877	3	72	3	0
9	2	5	1	0,064	1	2	3	0,064	3	2	2	5
10	2	0	1	-5,29	1	-3	3	-5,29	3	3	2	0
11	3	-30	2	0	2	0,093	2	0	2	0,093	1	-30
12	3	0	3	0	3	0	1	0	1	0	1	0
13	3	10	3	0	3	0	1	10	1	0	1	10
14	3	0	3	-10	3	-10	1	0	1	10	1	0

По табл. 2 видно, что среди моделей (26)–(31) нет ни одной такой, в которой хотя бы одна функция из соответствующего ей списка совсем не срабатывала ни в одном наблюдении.

Заключение

В статье предложены списочные регрессионные модели, относящиеся к классу нелинейных по параметрам регрессий. Задача идентификации их неизвестных параметров с помощью МНМ сведена к задаче ЧБЛП. При идентификации определяются как параметры входящих в список функций, так и параметры правила их переключения. Причём, перестановка функций в списке сказывается на результатах параметризации. При прогнозировании по списочным регрессиям сначала по значениям объясняющих переменных устанавливается номер функции из списка, затем по выбранной функции вычисляется прогнозное значение выходной переменной. В списочных регрессиях, в отличие от производственных функций Леонтьева, можно регулировать математическую форму правила переключения функций в списке. Проведенные на конкретных данных эксперименты показали превосходство по качеству аппроксимации всех построенных списочных регрессий над линейной регрессией и функциями Леонтьева.

Список использованной литературы

1. Montgomery D.C. Introduction to Linear Regression Analysis / D.C. Montgomery, E.A. Peck, G.G. Vining. — New York : Wiley, 2021. — 641 p.
2. Рязанцева Е.А. Применение корреляционно-регрессионного инструментария для анализа социально-значимых факторов / Е.А. Рязанцева. — DOI 10.24412/2782-4845-2023-5-65-74. — EDN MLIMKZ // ЭФО: Экономика. Финансы. Общество. — 2023. — № 1 (5). — С. 65–74.
3. Зуев М.А. Применение метода линейной регрессии для прогнозирования расхода топлива автомобиля / М.А. Зуев, В.М. Шибаев, К.С. Баланев. — DOI 10.47813/2782-2818-2024-4-2-0298-0305. — EDN QTRIU // Современные инновации, системы и технологии. — 2024. — Т. 4, № 2. — С. 298–305.
4. Мошев И.А. Анализ влияния независимых факторов на колебания стоимости Bitcoin с использованием линейной регрессии / И.А. Мошев, Ю.Р. Габдрахманова, Д.Б. Владимирова // Актуальные вопросы научных исследований : материалы X Междунар. науч.-практ. конф., Саратов, 21 авг. 2023 г. — Саратов, 2023. — С. 66–75.
5. Валиев А.А. Прогнозирование урожайности яровой пшеницы с применением регрессионного анализа / А.А. Валиев. — EDN HDQRFU // Научное сопровождение технологий агропромышленного комплекса: теория, практика, инновации : материалы Междунар. науч.-практ. конф., Казань, 24–25 марта 2022 г. — Казань, 2022. — С. 64–70.
6. Зенин Р.С. Методики прогнозирования параметров электроэнергии с использованием различных моделей регрессионного анализа / Р.С. Зенин, Ю.П. Кузьменко // Лучшая исследовательская статья : сб. статей. — Петрозаводск, 2024. — С. 64–72.
7. Мирончук В.А. Прогнозирование экономических циклов с использованием машинного обучения / В.А. Мирончук, К.А. Иванцов, Е.С. Гордеев. — DOI 10.54861/27131211_2024_5_67. — EDN VVVQHT // Прогрессивная экономика. — 2024. — № 5. — С. 67–84.
8. Абрамов Т.Е. Моделирование деятельности высокотехнологичного инновационного предприятия при помощи производственной функции типа Кобба-Дугласа / Т.Е. Абрамов, М.В. Баранов, В.В. Соколянский. — DOI 10.18334/cvr.2.2.112051. — EDN HQIQNJ // Экономика высокотехнологичных производств. — 2021. — Т. 2, № 2. — С. 93–106.

9. Афанасьев А.А. Использование производственной функции Кобба-Дугласа, построенной по панельным данным, при анализе обрабатывающих производств России / А.А. Афанасьев. — DOI 10.18334/ce.16.6.114851. — EDN QYZEYS // Креативная экономика. — 2022. — Т. 16, № 6. — С. 2363–2380.

10. Тырсин А.Н. Оценивание линейных регрессионных уравнений с помощью метода наименьших модулей / А.Н. Тырсин, К.Е. Максимов. — EDN PBOXDD // Заводская лаборатория. Диагностика материалов. — 2012. — Т. 78, № 7. — С. 65–71.

11. Носков С.И. Программный комплекс построения некоторых типов кусочно-линейных регрессий / С.И. Носков, А.А. Хоняков. — EDN UTFPOD // Информационные технологии и математическое моделирование в управлении сложными системами. — 2019. — № 3 (4). — С. 47–55.

12. Клейнер Г.Б. Производственные функции: Теория, методы, применение / Г.Б. Клейнер. — Москва : Финансы и статистика, 1986. — 239 с.

13. Bonates T.O. Optimization in Logical Analysis of Data. Doct. Diss. / T.O. Bonates. — Rutgers University, 2007. — 105 p.

14. Базилевский М.П. Оценивание модульных линейных регрессионных моделей с помощью метода наименьших модулей / М.П. Базилевский, А.Б. Ойдопова. — DOI 10.15593/2224-9397/2023.1.06. — EDN MEKQHE // Вестник Пермского национального исследовательского политехнического университета. Электротехника, информационные технологии, системы управления. — 2023. — № 45. — С. 130–146.

15. Базилевский М.П. Интеграция ограничений на коэффициенты интеркорреляций в оптимизационную задачу и условия построения вполне интерпретируемых неэлементарных линейных регрессий / М.П. Базилевский. — DOI 10.17223/19988605/69/4. — EDN FCXAEY // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. — 2024. — № 69. — С. 31–41.

16. Базилевский М.П. Оценивание регрессионных моделей с регрессорами в виде модулей линейных комбинаций объясняющих переменных / М.П. Базилевский. — DOI 10.17150/2713-1734.2024.6(3).269-281. — EDN MLCNDR // System Analysis and Mathematical Modeling. — 2024. — Т. 6, № 3. — С. 269–281.

17. Базилевский М.П. Оценивание неизвестных параметров многослойной модульной регрессии методом наименьших модулей / М.П. Базилевский. — DOI 10.26102/2310-6018/2024.45.2.039. — EDN INJWYZ // Моделирование, оптимизация и информационные технологии. — 2024. — Т. 12, № 2 (45). — URL: <https://moitvvt.ru/ru/journal/article?id=1581>.

18. Progress in Mathematical Programming Solvers From 2001 to 2020 / T. Koch, T. Berthold, J. Pedersen, C. Vanaret // EURO Journal on Computational Optimization. — 2022. — Vol. 10. — P. 100031.

19. Wolsey L.A. Integer Programming / L.A. Wolsey. — 2nd ed. — John Wiley, Sons, 2020. — 336 p.

20. Базилевский М.П. Оценивание с помощью метода наименьших модулей регрессионных моделей с целочисленными функциями пол и потолок / М.П. Базилевский. — EDN ZKNDPZ // International Journal of Open Information Technologies. — 2024. — Т. 12, № 10. — С. 56–61.

21. Грэхем Р. Конкретная математика. Основание информатики / Р. Грэхем, Д. Кнут, О. Паташник. — Москва : Мир, 1998. — 703 с.

References

1. Montgomery D.C., Peck E.A., Vining G.G. *Introduction to Linear Regression Analysis*. New York, Wiley, 2021. 641 p.

2. Ryazantseva E.A. The Use of Correlation-Regression Tools for the Analysis of Socially Significant Factors. *EFO: Ekonomika. Finansy. Obshchestvo = EFO: Economics. Finance. Society*, 2023, no. 1, pp. 65–74. (In Russian). EDN: MLMKZ. DOI: 10.24412/2782-4845-2023-5-65-74.
3. Zuev M.A., Shibaev V.M., Balanov K.S. Application of the Linear Regression Method to Predict Vehicle Fuel Consumption. *Sovremennye innovatsii, sistemy i tekhnologii = Modern Innovations, Systems and Technologies*, 2024, vol. 4, no. 2, pp. 298–305. (In Russian). EDN: QTRIU. DOI: 10.47813/2782-2818-2024-4-2-0298-0305.
4. Moshev I.A., Gabdrakhmanova Yu.R., Vladimirova D.B. Analysis of the influence of independent factors on Bitcoin price fluctuations using linear regression. Current issues in scientific research. *Materials of the X International Scientific Conference, Saratov, August 21, 2023*. Saratov, 2023, pp. 66–75. (In Russian).
5. Valiev A.A. Forecasting the Yield of Spring Wheat Using Regression Analysis. Scientific support of agro-industrial complex technologies: theory, practice, innovation. *Materials of International Scientific Conference, Kazan', March 24–25, 2022*. Kazan', 2022, pp. 64–70. (In Russian). EDN: HDQRFU.
6. Zenin R.S., Kuz'menko Yu.P. *Methods of forecasting electric power parameters using different regression analysis models. Best research paper. Collected Papers*. Petrozavodsk, 2024, pp. 64–72. (In Russian).
7. Mironchuk V.A., Ivantsov K.A., Gordeev E.S. Forecasting Economic Cycles Using Machine Learning. *Progressivnaya ekonomika = Progressive economy*, 2024, no. 5, pp. 67–84. (In Russian). EDN: VVVQHT. DOI: 10.54861/27131211_2024_5_67.
8. Abramov T.E., Baranov M.V., Sokolyanskiy V.V. Modeling the Activity of a High-Tech Innovative Enterprise Using A Cobb-Douglas Type Production Function. *Ekonomika vysokotekhnologichnykh proizvodstv = High-Tech Enterprises Economy*, 2021, vol. 2, no. 2, pp. 93–106. (In Russian). EDN: HQIQNJ. DOI: 10.18334/evp.2.2.112051.
9. Afanasev A.A. Using the Cobb-Douglas Production Function Based on Panel Data in the Analysis of Manufacturing Industries in Russia. *Kreativnaya ekonomika = Creative Economy*, 2022, vol. 16, no. 6, pp. 2363–2380. (In Russian). EDN: QYZEYS. DOI: 10.18334/ce.16.6.114851.
10. Tyrsin A.N., Maksimov K.E. Estimation of the Linear Regression Equations Using the Least-Modules Method. *Zavodskaya laboratoriya. Diagnostika materialov = Industrial laboratory. Materials diagnostics*, 2012, vol. 78, no. 7, pp. 65–71. (In Russian). EDN: PBOXDD.
11. Noskov S.I., Khonyakov A.A. Software Complex for Building Some Types Pieces of Linear Regressions. *Informatsionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnyimi sistemami = Information Technology and Mathematical Modeling in the Management of Complex Systems*, 2019, no. 3, pp. 47–55. (In Russian). EDN: UTFPOD.
12. Kleiner G.B. *Production Functions: Theory, Methods, Application*. Moscow, Finansy i statistika Publ., 1986. 239 p.
13. Bonates T.O. *Optimization in Logical Analysis of Data. Doct. Diss.* Rutgers University, 2007. 105 p.
14. Bazilevskiy M.P., Oydopova A.B. Estimation of Modular Linear Regression Models Using the Least Absolute Deviations. *Vestnik Permskogo natsional'nogo issledovatel'skogo politekhnicheskogo universiteta. Elektrotekhnika, informatsionnye tekhnologii, sistemy upravleniya = Bulletin of Perm National Research Polytechnic University. Electrical engineering, information technology, control systems*, 2023, no. 45, pp. 130–146. (In Russian). EDN: MEKQHE. DOI: 10.15593/2224-9397/2023.1.06.

15. Bazilevskiy M.P. Integration of Constraints on Intercorrelation Coefficients Into the Optimization Problem and Conditions for Constructing Quite Interpretable Non-Elementary Linear Regressions. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika = Tomsk State University Journal. Management, computer engineering and computer science*, 2024, no. 69, pp. 31–41. (In Russian). EDN: FCXAEY. DOI: 10.17223/19988605/69/4.

16. Bazilevskiy M.P. Estimation of Regression Models with Regressors in the Explanatory Variables Linear Combinations Modules Form. *System Analysis and Mathematical Modeling*, 2024, vol. 6, no. 3, pp. 269–281. (In Russian). EDN: ML-CDNR. DOI: 10.17150/2713-1734.2024.6(3).269-281.

17. Bazilevskiy M.P. Unknown Parameters Estimation for Multilayer Modular Regression Using the Least Absolute Deviations Method. *Modelirovanie, optimizatsiya i informacionnye tehnologii = Modeling, Optimization and Information Technology*, 2024, vol. 12, no. 2. (In Russian). EDN: INJWYZ. DOI: 10.26102/2310-6018/2024.45.2.039.

18. Koch T., Berthold T., Pedersen J., Vanaret C. Progress in Mathematical Programming Solvers From 2001 to 2020. *EURO Journal on Computational Optimization*, 2022, vol. 10, pp. 100031.

19. Wolsey L.A. *Integer Programming*. 2nd ed. John Wiley, Sons, 2020. 336 p.

20. Bazilevskii M.P. Estimation using Least Absolute Deviations Method of Regression Models with Integer Floor and Ceiling Functions. *International Journal of Open Information Technologies*, 2024, vol. 12, no. 10, pp. 56–61. (In Russian). EDN: ZKNDPZ.

21. Graham R., Knuth D., Patashnik O. *Concrete Mathematics: A Foundation for Computer Science*. 1994. 657 p. (Russ. ed.: Graham R., Knuth D., Patashnik O. *Concrete Mathematics: A Foundation for Computer Science*. Moscow, Mir Publ., 1998. 703 p.).

Информация об авторе

Базилевский Михаил Павлович — кандидат технических наук, доцент, кафедра математики, Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация, e-mail: mik2178@yandex.ru, <https://orcid.org/0000-0002-3253-5697>, SPIN-код: 4347-5028, AuthorID РИНЦ: 679277.

Information about the Author

Mikhail P. Bazilevskiy — PhD in Technical Sciences, Associate Professor, Department of Mathematics, Irkutsk State Transport University, Irkutsk, Russian Federation, e-mail: mik2178@yandex.ru, <https://orcid.org/0000-0002-3253-5697>, SPIN-Code: 4347-5028, AuthorID RSCI: 679277.

Для цитирования

Базилевский М.П. Идентификация неизвестных параметров списочных регрессионных моделей методом наименьших модулей / М.П. Базилевский. — DOI 10.17150/2713-1734.2025.7(2).248-262. — EDN WWEDJC // *System Analysis & Mathematical Modeling*. — 2025. — Т. 7, № 2. — С. 248–262.

For Citation

Bazilevskiy M.P. Identification of Unknown Parameters in List Regression Models Using the Least Absolute Deviations. *System Analysis & Mathematical Modeling*, 2025, vol. 7, no. 2, pp. 248–262. (In Russian). EDN: WWEDJC. DOI: 10.17150/2713-1734.2025.7(2).248-262.