Моделирование и анализ данных
2025, *15*(4), 38—57.
https://doi.org/10.17759/mda.2025150403
ISSN: 2219-3758 (печатный)
ISSN: 2311-9454 (online)

Modelling and Data Analysis
2025, *15*(4), 38—57.
https://doi.org/10.17759/mda.2025150403
ISSN: 2219-3758 (print)
ISSN: 2311-9454 (online)

МЕТОДЫ ОПТИМИЗАЦИИ | OPTIMIZATION METHOD

Научная статья | Original paper

УДК 519.854

Отбор назначенного числа наиболее сочетающихся при переключении методов оценивания параметров линейной регрессии

М.П. Базилевский
Иркутский государственный университет путей сообщения
Иркутск, Российская Федерация
✉ mik2178@yandex.ru

Резюме

Контекст и актуальность. Существует множество разных методов оценивания параметров линейной регрессионной модели. Актуальной проблемой является выбор в некотором смысле наилучшего метода. **Цель.** Исследовать задачи отбора назначенного числа наиболее сочетающихся методов оценивания параметров линейной регрессии и идентификации правила переключения этих методов. **Гипотеза.** Сочетание методов даёт более адекватную линейную регрессию, чем каждый из методов по-отдельности. **Методы и материалы.** Для формирования множества методов использовано L_v-оценивание. **Результаты.** Сформулирована задача отбора назначенного числа наиболее сочетающихся в смысле минимума остаточной суммы модулей методов и идентификации правила их переключения. Правило формируется благодаря присвоению каждому используемому методу маркера, переключение между которыми осуществляется с помощью целочисленной функции «пол». Задача отбора сочетающихся методов при известной и неизвестной маркировке сведена к задаче частично булевого линейного программирования. На основе реальных данных о смертности от ишемической болезни сердца построена списочная модель, демонстрирующая, что сочетание трех методов L_1, L_2 и $L_{1,2}$ даёт лучшую точность, чем метод L_1 сам по себе. **Выводы.** Получаемые предложенным способом списочные модели, состоящие из списка оцененных разными методами линейных регрессий и правила их переключения, могут быть использованы для прогнозирования. К тому же предложенный способ может быть адаптирован для отбора наиболее сочетающихся спецификаций моделей для конкретного метода и для решения задачи отбора наиболее сочетающихся нейронных сетей.

Ключевые слова: регрессионный анализ, отбор сочетающихся методов, списочная регрессия, целочисленная функция «пол», метод наименьших модулей, Lν-оценивание, задача частично булевого линейного программирования

Для цитирования: Базилевский, М.П. (2025). Отбор назначенного числа наиболее сочетающихся при переключении методов оценивания параметров линейной регрессии. *Моделирование и анализ данных, 15*(4), 38—57. https://doi.org/10.17759/mda.2025150403

Selection a fixed number of the most compatible methods for estimating linear regression parameters when switching

M.P. Bazilevskiy
Irkutsk State Transport University, Irkutsk, Russian Federation
✉ mik2178@yandex.ru

Abstract

Context and relevance. There are many different methods for estimating the parameters of a linear regression model. In some sense, the pressing issue is to choose the best method. **Objective.** To study the problems of selecting the designated number of most compatible methods for estimating linear regression parameters and identifying a switching rule for these methods. **Hypothesis.** The combination of methods produces a more adequate linear regression than any of the methods individually. **Methods and materials.** L_v-estimation was used to form a set of methods. **Results.** The problem of selecting a designated number of methods that are most compatible in terms of minimizing the residual sum of their modules and identifying a switching rule is formulated. A rule is formed by assigning a marker to each method used, which is switched between using the integer function «floor». The problem of choosing compatible methods with known and unknown markers reduces to a mixed 0–1 integer linear programming problem. Based on real data on mortality from coronary heart disease, a list model is constructed to demonstrate that a combination of three methods (L_1, L_2, and $L_{1.2}$) yields better accuracy than method L_1 alone. **Conclusions.** The list model obtained by the proposed way consists of a list of linear regressions estimated using different methods and a switching rule. This list can be used for forecasting. Moreover, the proposed way can be adapted to select the most suitable model specifications for a specific method and to solve the problem of choosing the best neural network.

Keywords: regression analysis, selection of compatible methods, list regression, integer floor function, least absolute deviations method, Lv-estimation, mixed 0–1 integer linear programming problem

For citation: Bazilevskiy, M.P. (2025). Selection a fixed number of the most compatible methods for estimating linear regression parameters when switching. *Modelling and Data Analysis, 15*(4)*,* 38—57. (In Russ.). https://doi.org/10.17759/mda.2025150403

Введение

Регрессионный анализ, несмотря на появление более сложных методов машинного обучения в настоящее время, по-прежнему находит широкое применение при решении самых разных прикладных задач анализа данных. В отличие от нейронных сетей в виде «черных ящиков», регрессионная модель даёт понятные коэффициенты, которые можно легко интерпретировать. Такая прозрачность позволяет руководителям и аналитикам довольно просто принимать эффективные управленческие решения. Анализ научной литературы показывает, что регрессионный анализ традиционно часто используется при построении моделей управления экономикой. Так, в статье (Акбердина и др., 2024) построено 125 регрессионных моделей, которые могут быть использованы для управления процессами развития промышленности Свердловской области. В работе (Курлов, Косухина, Курлов, 2022) идентифицирована регрессионная модель, с помощью которой можно управлять цифровой зрелостью предприятия. В статье (Карпенко, 2021) построены регрессионные модели управления развитием региональных социально-экономических систем Российской Федерации. Однако модели управления с помощью регрессионного анализа строятся и в других сферах. Например, в работе (Холодилин и др., 2024) оценены регрессионные модели для управления тележкой, оснащенной манипулятором робота, а в статье (Чигиринский, Ингеманссон, 2021) — для оперативного управления технологическими режимами лезвийной обработки.

Существует целое множество различных методов оценивания неизвестных параметров регрессионных моделей. Наиболее простым из них является метод наименьших квадратов (МНК), суть которого состоит в минимизации суммы квадратов остатков регрессии. Одной из модификаций МНК является метод Лассо (Wang et al., 2024), при котором функция потерь дополняется компонентой регуляризации для обнуления незначимых коэффициентов. На сегодняшний день актуальны ещё две модификации МНК — полностью модифицированный МНК (FM-OLS) и динамический МНК (DOLS) (Khan, Hassan, 2024; Idroes et al., 2024). В первой из них оценки МНК корректируются на эндогенность и автокорреляцию во временных рядах, во второй добавляются лагированные и опережающие разности регрессоров.

Как известно, при наличии выбросов в данных МНК-оценивание может давать неудовлетворительные результаты. Поэтому в рамках регрессионного анализа существует довольно много робастных (устойчивых к выбросам) методов оценивания.

Условно они разделяются на *M*-оценки и *L*-оценки. *M*-оценки обобщают метод максимального правдоподобия (ММП) (Liu, Liu, 2024; Sengupta et al., 2024) и МНК. В них квадратичная функция потерь заменяется на более робастную. В работе (De Menezes et al., 2021) представлен обзор 48-ми робастных методов *M*-оценивания, среди которых смешанная нормальная оценка (Contaminated Normal estimator), оценка Хубера (Wang et al., 2022), оценка коррентропии (Correntropy estimator), оценка Смита и др. *L*-оценки — это линейные комбинации порядковых статистик. В монографии (Демиденко, 1981) рассмотрены так называемые L_v-оценки. При $v = 2$ имеем L_2-оценки, соответствующие МНК. При $v = 1$ имеем L_1-оценки или оценки метода наименьших модулей (МНМ) (Голованов, Тырсин, 2024; Liu, Yang, 2020), суть которого состоит в минимизации суммы модулей остатков. При $v = \infty$ имеем L_∞-оценки, означающие минимизацию максимального остатка регрессии. Известно, что при $v = 1$ и $v = \infty$ задача минимизации сводится к задаче линейного программирования (см., например, (Wagner, 1959)). Для нахождения L_v-оценок при $1 \le v < 2$ применяется итеративный МНК (Демиденко, 1981), также известный как метод вариационно-взвешенных квадратичных приближений (Мудров, Кушко, 1976).

В статье (Носков, Баенхаева, 2016) рассматривается метод множественного оценивания регрессионных моделей, состоящий в минимизации одновременно двух критериев — L_1 и L_∞. А в работе (Базилевский, 2024а) с некоторыми допущениями предложена двухкритериальная задача оценивания регрессий по критериям L_1 и L_2.

Возникает проблема, связанная с тем, каким методом оценивания регрессии из существующего многообразия лучше всего пользоваться на практике. Существуют определенные рекомендации на этом счёт. Например, при выполнении условий теоремы Гаусса — Маркова (Демиденко, 1981) рекомендуется использовать МНК, при наличии выбросов в данных — МНМ, и т.д. Безусловно, выбор метода зависит от целей исследования. Допустим, что цель состоит в обеспечении максимальной точности аппроксимации во всех наблюдениях выборки в смысле некоторого критерия качества, например, суммы модулей остатков. Тогда при оценивании параметров модели несколькими разными методами будет получен следующий результат: в одних наблюдениях точнее окажется первый метод, в других — второй, и т.д. В этой связи возникает необходимость в разработке технологии, которая позволяла бы вбирать из заранее сформированного множества методов оценивания их наилучшие качества так, чтобы точность аппроксимации по всей выборке была выше, чем при использовании любого из этих методов в отдельности.

В данной статье предлагается организовывать переключение методов в каждом наблюдении выборки с помощью специального правила, включающего в себя целочисленную функцию «пол» (Грэхем, Кнут, Паташник, 1998). На основе этой функции в статье (Базилевский, 2024б) были разработаны регрессионные модели, предназначенные для моделирования в ситуации, когда выходная переменная принимает целые значения. А в работе (Базилевский, 2025а) были предложены списочные регрессии, в которых функция «пол» позволяет прогнозировать номер аналитического выражения из некоторого сформированного списка. Причем, задача МНМ-оценивания списочных регрессий была сведена к задаче частично булевого линейного программирования (ЧБЛП), для решения которой можно применять развитое за последние десятилетия алгоритмическое и программное обеспечение (Koch et al., 2022).

Целью данной работы является исследование задач отбора назначенного числа наиболее сочетающихся методов оценивания параметров линейной регрессии и идентификации правила переключения этих методов.

Формулировка списочной регрессионной модели

Предположим, что имеется выборка данных объема n, содержащая значения y_i, $i=\overline{1,n}$, объясняемой (выходной) переменной y и значения x_{ij}, $i=\overline{1,n}$, $j=\overline{1,l}$, l объясняющих (входных) переменных x_1, x_2, …, x_l. Пусть по этой выборке оцениваются неизвестные параметры α_j, $j=\overline{0,l}$, модели множественной линейной регрессии:

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

где ε_i, $i=\overline{1,n}$ — ошибки модели.

Для оценки параметров линейной регрессии (1) будем использовать L_ν-метод (Демиденко, 1981), состоящий в решении задачи

$$\sum_{i=1}^{n}\left|\varepsilon_i\right|^{\nu}\to\min\,. \tag{2}$$

Как отмечено в монографии (Демиденко, 1981), для функции в (2) при $\nu\ge 1$ существует единственный локальный минимум, который совпадает с глобальным. Для решения задачи (2) будем использовать итеративный МНК, который, как доказано в книге (Мудров, Кушко, 1976), сходится при $1\le\nu<2$. Для реализации итеративного МНК можно воспользоваться специальной программой, реализованной на языке программирования hansl пакета Gretl и подробно описанной в статье (Базилевский, 2025в). Для начала работы с ней необходимо указать следующие параметры: точность *Eps*, большое и малое числа M и d. На каждой итерации требуемые весовые коэффициенты $w_i^{(k)}$ вычисляются по следующей формуле:

$$w_i^{(k)}=\begin{cases}1/\left|e_i^{(k-1)}\right|, \text{ если } \left|e_i^{(k-1)}\right|\ge d,\\ M, \text{ если } \left|e_i^{(k-1)}\right|<d,\end{cases}\quad k=1,2,\ldots\,, \tag{3}$$

где $e_i^{(k-1)}$ — остатки регрессии в i-м наблюдении на (k-1)-й итерации, т.е. разности между фактическим и расчетными значениями переменной y.

Программа работает по следующему алгоритму.

Итерация 0. Находится вектор МНК-оценок $\tilde{\alpha}^{(0)}$ регрессии (1).

Итерация 1. Вычисляются остатки $e_i^{(0)}$, $i=\overline{1,n}$. По формуле (3) находятся веса $w_i^{(1)}$, по которым с помощью взвешенного МНК находится вектор оценок $\tilde{\alpha}^{(1)}$. Если $\sum\left|\tilde{\alpha}^{(1)}-\tilde{\alpha}^{(0)}\right|\le Eps$, то оценки найдены. Иначе переход к следующей итерации.

Итерация 2. Вычисляются остатки $e_i^{(1)}$, $i=\overline{1,n}$. По формуле (3) находятся веса $w_i^{(2)}$, по которым с помощью взвешенного МНК находится вектор оценок $\tilde{\alpha}^{(2)}$. Если $\sum\left|\tilde{\alpha}^{(2)}-\tilde{\alpha}^{(1)}\right|\le Eps$, то оценки найдены. Иначе переход к следующей итерации. И так далее.

Для формирования множества оцененных разными методами регрессионных моделей выберем равномерным образом несколько точек внутри промежутка $1\le\nu<2$. Пусть в каждой из этих точек и на концах промежутка $1\le\nu<2$ построены L_ν-методом p линейных регрессионных моделей:

$$\tilde{y}_{ik}=\tilde{\alpha}_0^{(k)}+\sum_{j=1}^{l}\tilde{\alpha}_j^{(k)}x_{ij}\ ,\ i=\overline{1,n}\ ,\ k=\overline{1,p}\ , \tag{4}$$

где $\tilde{\alpha}_j^{(k)}$ — L_ν-оценка при j-й входной переменной, полученная k-м методом; $\tilde{y}_{ik}$, $i=\overline{1,n}$, $k=\overline{1,p}$ — расчетные значения выходной переменной y, полученные всеми p методами. Для L_ν-оценивания регрессий в точках внутри промежутка $1\le\nu<2$ и при $\nu=1$ можно использовать итеративный МНК, а при $\nu=2$ — обычный МНК.

Проведем маркировку методов, т.е. установим соответствие между номерами регрессий (4) и целыми числами (маркерами) 1, 2, …, p. Будем считать, что первому методу (при $\nu=1$) соответствует маркер 1, второму методу — 2, …, последнему методу (при $\nu=2$) — p. Тогда введём следующую списочную конструкцию (Базилевский, 2025a):

$$y_i=\varepsilon_i^*+\begin{cases}\tilde{y}_{i1},\ \text{если}\ N_i=1,\\ \tilde{y}_{i2},\ \text{если}\ N_i=2,\\ \ldots\\ \tilde{y}_{ip},\ \text{если}\ N_i=p,\end{cases}\quad i=\overline{1,n}\ , \tag{5}$$

где ε_i^*, $i=\overline{1,n}$ — ошибки модели; N_i — неизвестное значение маркера, сработавшего в i-м наблюдении.

Для удобства представим списочную регрессию (5) в виде

$$y_i=\begin{bmatrix}\tilde{y}_{i1}\\ \tilde{y}_{i2}\\ \ldots\\ \tilde{y}_{ip}\end{bmatrix}_{N_i}+\varepsilon_i^*\ ,\ i=\overline{1,n}\ . \tag{6}$$

Для организации правила переключения маркеров в модели (6) используем целочисленную функцию «пол» (Грэхем, Кнут, Паташник, 1998), обозначаемую $\lfloor x\rfloor$ и введенную К.Э. Айверсоном в начале 60-х годов. Запись $\lfloor x\rfloor$ означает округление числа x до ближайшего целого в меньшую сторону. Например, $\lfloor 5{,}9\rfloor=5$.

Предположим, что значения маркеров зависят от наблюдаемых значений входных переменных x_1, x_2, …, x_l. Тогда переключения маркеров в наблюдениях будем осуществлять с помощью следующего правила:

$$N_i = \lfloor \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_l x_{il} \rfloor,\ i = \overline{1,n}, \qquad (7)$$

где β_j, $j = \overline{0,l}$ — неизвестные параметры.

Таким образом, сформулирована списочная регрессионная модель (6), (7), оценивание которой позволяет получить правило переключения p методов для любого наблюдения выборки.

Задача отбора назначенного числа наиболее сочетающихся методов при известной маркировке

Для оценки неизвестных параметров списочной регрессионной модели (6), (7) будем использовать МНМ (расстояние городских кварталов), что приводит к необходимости решения задачи

$$\sum_{i=1}^{n} \left| \varepsilon_i^* \right| \to \min. \qquad (8)$$

Сведём задачу (8) к задаче математического программирования. Для этого введем булевы переменные σ_{ij}, $i = \overline{1,n}$, $j = \overline{1,p}$ по правилу:

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

Тогда выражение (6) можно переписать в виде

$$y_i = \sum_{j=1}^{p} \tilde{y}_{ij} \cdot \sigma_{ij} + \varepsilon_i^*,\ i = \overline{1,n}. \qquad (9)$$

В каждом наблюдении должен срабатывать только 1 метод, поэтому справедливыми будут ограничения

$$\sum_{j=1}^{p} \sigma_{ij} = 1,\ i = \overline{1,n}. \qquad (10)$$

Учитывая (9), представим функцию потерь (8) в виде

$$\sum_{i=1}^{n} \left| y_i - \sum_{j=1}^{p} \tilde{y}_{ij} \cdot \sigma_{ij} \right| \to \min. \qquad (11)$$

В соответствии с приёмом, предложенным в работе (Wagner, 1959), введём неотрицательные вещественные переменные u_i, v_i, $i=\overline{1,n}$:

$$u_i = \begin{cases} y_i - \sum_{j=1}^{p} \tilde{y}_{ij} \cdot \sigma_{ij}, \text{ если } y_i - \sum_{j=1}^{p} \tilde{y}_{ij} \cdot \sigma_{ij} > 0, \\ 0, \text{ в противном случае,} \end{cases}$$

$$v_i = \begin{cases} -y_i + \sum_{j=1}^{p} \tilde{y}_{ij} \cdot \sigma_{ij}, \text{ если } y_i - \sum_{j=1}^{p} \tilde{y}_{ij} \cdot \sigma_{ij} < 0, \\ 0, \text{ в противном случае.} \end{cases}$$

Тогда выражение (9) можно справедливо заменить ограничениями

$$y_i = \sum_{j=1}^{p} \tilde{y}_{ij} \cdot \sigma_{ij} + u_i - v_i, \ i=\overline{1,n}, \tag{12}$$

а функционал (11) заменить целевой функцией

$$\sum_{i=1}^{n} (u_i + v_i) \to \min. \tag{13}$$

Отметим, что использованный приём довольно часто применяется в научных исследованиях (см., например, (Konno, Yamamoto, 2009; Носков, 2022)).

Таким образом, решение задачи ЧБЛП с целевой функцией (13), линейными ограничениями (10), (12), а также с ограничениями на булевость переменных σ_{ij}, $i=\overline{1,n}$, $j=\overline{1,p}$, вида

$$\sigma_{ij} \in \{0,1\}, \ i=\overline{1,n}, \ j=\overline{1,p}, \tag{14}$$

и условиями неотрицательности переменных u_i, v_i, $i=\overline{1,n}$, вида

$$u_i \ge 0, \ v_i \ge 0, \ i=\overline{1,n}, \tag{15}$$

позволяет найти оптимальное с точки зрения критерия (8) переключение L_v-методов в каждом наблюдении выборки. Такая задача сама по себе никакой научной ценности не представляет, поскольку не идентифицируется правило переключения методов.

Интегрируем далее в задачу (10), (12) — (15) ограничения для переключения в наблюдениях методов по правилу (7). По-прежнему считаем, что методы маркированы числами 1, 2, …, p. Введем линейные ограничения так, как это сделано в статье (Базилевский, 2025a):

$$\sum_{j=1}^{p} j \cdot \sigma_{ij} \le \beta_0 + \sum_{j=1}^{l} \beta_j x_{ij} \le \sum_{j=1}^{p} j \cdot \sigma_{ij} + 1 - \Delta, \ i=\overline{1,n}, \tag{16}$$

где Δ — близкое к нулю положительное число.

Например, если в *i*-м наблюдении $\beta_0 + \sum_{j=1}^{l} \beta_j x_{ij} = 2,7$, то двойное неравенство (16) будет выполняться только при $\sigma_{i2} = 1$, т.е. значение 2,7 с помощью функции «пол» преобразуется в маркер 2, поэтому сработает второй метод L_v-оценивания.

Исходя из вышеизложенного, решение задачи ЧБЛП (10), (12) — (16) позволяет найти МНМ-оценки списочной регрессии (6), (7), т.е. идентифицировать правило оптимального по критерию (8) переключения *p* методов в наблюдениях выборки. Зная это правило, можно в зависимости от значений входных переменных x_1, x_2, …, x_l прогнозировать значения выходной переменной *y*.

Как указано во введении, методов оценки параметров линейной регрессии существует весьма внушительное количество, поэтому и число *p* может быть довольно большим. В этой связи возникает задача выделения из используемых *p* методов ровно *m* штук, гарантирующих минимум критерия (8). Назовем её задачей отбора наиболее сочетающихся при переключении методов (ОСМ).

Введем булевы переменные δ_j, $j = \overline{1,p}$, по правилу:

$$\delta_j = \begin{cases} 1, & \text{если j-й метод срабатывает хотя бы в одном наблюдении,} \\ 0, & \text{в противном случае.} \end{cases}$$

Тогда булевы переменные δ_j, $j = \overline{1,p}$, необходимо связать с переменными σ_{ij}, $i = \overline{1,n}$, $j = \overline{1,p}$, линейными ограничениями

$$\delta_j \le \sum_{i=1}^{n} \sigma_{ij} \le n \cdot \delta_j, \quad j = \overline{1,p}. \tag{17}$$

Например, если $\delta_j = 0$, то j-й метод не будет срабатывать ни в одном наблюдении выборки, поскольку (17) имеет вид $\sum_{i=1}^{n} \sigma_{ij} = 0$. А если $\delta_j = 1$, то двойное неравенство (17) принимает вид $1 \le \sum_{i=1}^{n} \sigma_{ij} \le n$, т.е. j-й метод должен срабатывать хотя бы в одном наблюдении выборки.

Число отбираемых методов должно быть равно *m* ($2 \le m \le p$), что фиксируется линейным ограничением

$$\sum_{j=1}^{p} \delta_j = m. \tag{18}$$

Таким образом, решение задачи ЧБЛП (10), (12) — (18) с ограничениями булевости переменных δ_j, $j = \overline{1,p}$, вида

$$\delta_j \in \{0,1\}, \quad j = \overline{1,p}, \tag{19}$$

идентифицирует m наиболее сочетающихся при переключении по критерию (8) методов и правило их переключения в наблюдениях выборки. Тем самым, задача ОСМ сведена к задаче ЧБЛП.

Отметим, что в результате решения задачи ЧБЛП (10), (12) — (19) при $m<p$ происходит сокращение списка методов, а, следовательно, и исключаются некоторые маркеры. Поэтому после решения при необходимости нужно уменьшить все маркеры и идентифицированный свободный член β_0 в правиле (7) на одинаковую величину так, чтобы минимальный маркер был равен 1. От этого качество списочной регрессии не изменится.

Важно, что задача (10), (12) — (19) сформулирована для случая, когда каждый L_v-метод маркирован соответствующим его порядку целым числом от 1 до p. Если провести маркировку методов иначе, то нет гарантии, что будет получено точно такое же решение задачи ЧБЛП. Возникает проблема выбора ещё и оптимальной маркировки методов по критерию (8).

Задача отбора назначенного числа наиболее сочетающихся методов при неизвестной маркировке

Пусть каждому из p L_v-методов нужно поставить в соответствие один маркер из имеющегося набора 1, 2, …, p. Причём, разные методы не должны иметь одинаковый маркер. Для этого можно воспользоваться математическим приёмом, предложенным в работе (Базилевский, 2025б) для маркировки категориальной переменной.

Сформируем переменные z_j, $j=\overline{1,p}$, где z_j — маркер, соответствующий j-му L_v-методу. Понятно, что z_j, $j=\overline{1,p}$, — целые и разные числа в диапазоне от 1 до p.

Введём бинарные переменные ρ_{jk}, $j=\overline{1,p}$, $k=\overline{1,p}$, по правилу:

$$\rho_{jk}=\begin{cases}1, & \text{если в списке j-й метод имеет маркер k,}\\ 0, & \text{в противном случае.}\end{cases}$$

Тогда переменные z_j, $j=\overline{1,p}$, и ρ_{jk}, $j=\overline{1,p}$, $k=\overline{1,p}$, должны быть связаны соотношениями

$$\sum_{k=1}^{p} k\cdot\rho_{jk}=z_j\,,\ \ j=\overline{1,p}\,. \tag{20}$$

Ограничения на булевость переменных ρ_{jk}, $j=\overline{1,p}$, $k=\overline{1,p}$, имеют вид

$$\rho_{jk}\in\{0,1\}\,,\ \ j=\overline{1,p}\,,\ \ k=\overline{1,p}\,. \tag{21}$$

Поскольку каждому методу должен быть назначен только 1 маркер, то справедливы следующие линейные ограничения:

$$\sum_{k=1}^{p}\rho_{jk}=1\,,\ \ j=\overline{1,p}\,, \tag{22}$$

$$\sum_{j=1}^{p} \rho_{jk} = 1\,,\ k = \overline{1,p}\,. \tag{23}$$

В итоге с помощью ограничений (20) — (23) в задаче математического программирования появляется контроль над всеми возможными комбинациями перестановок маркеров из набора 1, 2, …, p.

Поскольку теперь маркировка методов неизвестна, то будут несправедливы линейные ограничения (16). Для решения этой проблемы сформируем переменные r_{ij}, $i = \overline{1,n}$, $j = \overline{1,p}$, по правилу:

$$r_{ij} = \begin{cases} z_j, & \text{если в i-м наблюдении срабатывает j-й метод,} \\ 0, & \text{в противном случае.} \end{cases}$$

Используя подход, рассмотренный в статье (Базилевский, 2020), для переменных r_{ij}, $i = \overline{1,n}$, $j = \overline{1,p}$, введём следующие ограничения:

$$-p \cdot \left(1 - \sigma_{ij}\right) \le r_{ij} - z_j \le p \cdot \left(1 - \sigma_{ij}\right),\ i = \overline{1,n},\ j = \overline{1,p}\,, \tag{24}$$

$$r_{ij} \le p \cdot \sigma_{ij}\,,\ i = \overline{1,n},\ j = \overline{1,p}\,, \tag{25}$$

$$r_{ij} \ge 0\,,\ i = \overline{1,n},\ j = \overline{1,p}\,. \tag{26}$$

Например, если $\sigma_{ij} = 1$, т.е. в i-м наблюдении срабатывает j-й метод, то из ограничений (24) — (26) следуют соотношения $r_{ij} = z_j$, $r_{ij} \in [0, p]$. А если $\sigma_{ij} = 0$, то $r_{ij} = 0, -z_j \in [-p, p]$.

Тогда вместо (16) необходимо использовать следующие ограничения:

$$\sum_{j=1}^{p} r_{ij} \le \beta_0 + \sum_{j=1}^{l} \beta_j x_{ij} \le \sum_{j=1}^{p} r_{ij} + 1 - \Delta\,,\ i = \overline{1,n}\,. \tag{27}$$

Таким образом, решение задачи ЧБЛП (10), (12) — (15), (17) — (27) позволяет идентифицировать m наиболее сочетающихся при переключении по критерию (8) методов, правило их переключения в наблюдениях выборки, а также оптимальную маркировку методов.

Хотелось бы подчеркнуть следующие важные обстоятельства.

1. Достоинство сформулированной задачи ЧБЛП (10), (12) — (15), (17) — (27) в том, что для неё не нужно решать проблему, связанную с выбором для ограничений достаточно большого положительного числа M. Эта проблема встречается во многих научных работах (см., например, (Базилевский, 2025a; Konno, Yamamoto, 2009; Носков, 2022; Bertsimas, King, Mazumder, 2016)).
2. В статье (Носков, Попов, 2024) описан метод смешанного оценивания параметров линейной регрессии. Его суть состоит в том, что сначала исходная выборка разбивается на 2 части, а затем одновременно применяется L_1-метод на первой выборке

и L_∞-метод на второй, для чего решается единая задача линейного программирования, приводящая к единой регрессионной модели. Рассмотренная в данной статье технология не имеет ничего общего с методом смешанного оценивания, который, к сожалению, только искажает оценки линейной регрессии, ухудшая её качество как по L_1-критерию, так и по L_∞-критерию. Предложенная в данной статье технология позволяет идентифицировать правило переключения L_ν-методов так, что в сочетании они дают лучшие результаты, чем по-отдельности. При этом L_ν-оценки моделей, полученные разными методами, сохраняются в первоначальном виде.

3. В работе (Носков, 2022) рассматривается задача выбора в каждом наблюдении наилучшего варианта регрессионной модели. Нетрудно переформулировать эту задачу в задачу выбора наилучшего метода оценивания регрессионной модели. Однако такая задача будет отличаться от предложенной в данной статье технологии по следующим причинам. Во-первых, в статье (Носков, 2022) для переключения моделей использовано абсолютно другое правило, выраженное в виде мультиарной операции max, не зависящей от входных переменных. Во-вторых, в работе (Носков, 2022) отсутствует контроль количества используемых для переключения моделей. В-третьих, не упоминается, как выбирать в задаче ЧБЛП большое число M.

Вычислительные эксперименты

Главная цель вычислительных экспериментов заключалась в том, чтобы показать работоспособность предложенного математического аппарата, продемонстрировав, что при ОСМ точность аппроксимации выше, чем при использовании только МНМ. Эксперимент № 1 проводился для подтверждения того, что при изменении маркировки методов меняется результат решения задачи ЧБЛП (10), (12) — (19). Эксперимент № 2 проводился для поиска оптимального с точки зрения точности и сложности списочной модели числа сочетающихся методов, для чего решалась задача ЧБЛП (10), (12) — (15), (17) — (27) при m=2, 3, …, p.

Для проведения экспериментов использовались встроенные в эконометрический пакет Gretl данные data4–7.gdt о влиянии на переменную *chd* — коэффициент смертности от ишемической болезни сердца на 100000 человек населения в США, следующих восьми (l = 8) факторов:

cal — потребление кальция на душу населения (в граммах);

unemp — процент безработных среди гражданского населения на 1000 человек в возрасте от 16 лет и старше;

cig — потребление сигарет на душу населения в фунтах табака лицами в возрасте от 18 лет и старше;

edfat — потребление пищевых жиров и масла на душу населения в фунтах, включая сало, маргарин и сливочное масло;

meat — потребление мяса на душу населения в фунтах, включая говядину, телятину, свинину, баранину;

spirits — потребление крепких спиртных напитков на душу населения в галлонах для лиц от 18 лет и старше;

beer — потребление солодовых напитков на душу населения в галлонах для лиц от 18 лет и старше;

wine — потребление вина на душу населения в галлонах для лиц от 18 лет и старше.

Объем выборки $n = 34$.

Для удобства обозначим переменную *chd* как y, *cal* — x_1, *unemp* — x_2, *cig* — x_3, *edfat* — x_4, *meat* — x_5, *spirits* — x_6, *beer* — x_7, *wine* — x_8.

Предварительно были найдены L_v-оценки линейной регрессии при $v = 1$, 1,2, 1,4, 1,6, 1,8, для чего была использована описанная в статье (Базилевский, 2025в) и выше программа, реализующая итеративный МНК. В ней были указаны следующие параметры: $Eps = 0{,}0001$, $M = 100000$ и $d = 0{,}000001$. L_v-оценки при $v = 2$ были найдены с помощью обычного МНК. Для удобства будем обозначать далее используемые методы как L_1 (МНМ), $L_{1,2}$, $L_{1,4}$, $L_{1,6}$, $L_{1,8}$ и L_2 (МНК). Все вычисленные разными методами L_v-оценки представлены в таблице.

Таблица / Table

Lν-оценки линейной регрессии

Lν-estimates of linear regression

Метод	**Оценки линейной регрессии**								
	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
L_1	344,386	-99,702	0,117	7,677	1,204	0,0997	29,068	-3,057	-12,885
$L_{1,2}$	297,791	-72,793	-0,295	7,779	1,889	0,0307	28,642	-2,937	-14,741
$L_{1,4}$	274,067	-71,950	-0,529	8,336	2,325	0,0395	26,757	-3,066	-13,157
$L_{1,6}$	248,743	-71,396	-0,605	9,284	2,589	0,0759	24,239	-3,218	-9,444
$L_{1,8}$	233,659	-70,375	-0,616	9,838	2,730	0,0988	22,676	-3,348	-6,645
L_2	226,002	-69,983	-0,613	10,116	2,810	0,1116	21,716	-3,467	-4,562

Точность полученных в экспериментах моделей было решено оценивать по величине суммы модулей остатков *SAE*. Для линейной регрессии, оцененной с помощью МНМ, величина $SAE = 190{,}3049$.

Для решения задач ЧБЛП был использован решатель Cardinal Optimizer (COPT) (Ge et al., 2023), для управления которым был использован язык программирования Python.

Эксперимент № 1. Сначала методы были промаркированы следующим образом: L_1 присвоен маркер 1, $L_{1,2}$ — 2, $L_{1,4}$ — 3, $L_{1,6}$ — 4, $L_{1,8}$ — 5, L_2 — 6. В результате решения задачи ЧБЛП (10), (12) — (19) при $m = 2$, $\Delta = 0{,}01$ была получена следующая списочная модель:

$$\tilde{y} = \begin{bmatrix} L_1 \\ L_{1,2} \end{bmatrix}_{N^{(1)}}, \tag{28}$$

где

$$N^{(1)} = -14{,}782 + 3{,}497x_1 + 0{,}0221x_2 + 0{,}415x_3 +$$

$$+0,228x_4-0,00346x_5+0,00668x_6-0,0557x_7+0,177x_8\,. \tag{29}$$

Для модели (28), (29) *SAE* = 180,5825. Можно сделать следующий вывод: для выбранной маркировки двумя наилучшими методами из шести оказались L_1 и $L_{1,2}$. В сочетании они обеспечивают более высокую точность аппроксимации по критерию *SAE*, чем метод L_1. Причём, метод L_1 сработал в наблюдениях при переключении по правилу (29) 16 раз, а $L_{1,2}$ — 18 раз.

Затем методы были промаркированы по схеме: L_1 — 6, $L_{1,2}$ — 2, $L_{1,4}$ — 3, $L_{1,6}$ — 4, $L_{1,8}$ — 5, L_2 — 1. Решение задачи ЧБЛП (10), (12) — (19) при $m = 2$, $\Delta = 0{,}01$ даёт следующую списочную модель:

$$\tilde{y}=\begin{bmatrix} L_{1,8} \\ L_1 \end{bmatrix}_{N^{(2)}}, \tag{30}$$

где

$$N^{(2)}=1,880-0,207x_1+0,00297x_2+0,0142x_3+$$
$$+0,0049x_4+0,000745x_5-0,00283x_6-0,00774x_7+0,00828x_8\,. \tag{31}$$

Заметим, что в результате решения был выбран метод L_1, имеющий маркер 6, и $L_{1,8}$, имеющий маркер 5. Поэтому, как описано выше, свободный член в правиле (31) был уменьшен на 4. Тогда в списке (30) методу $L_{1,8}$ соответствует маркер 1, а методу L_1 — 2.

Для модели (30), (31) *SAE* = 168,025. Таким образом, для выбранной маркировки двумя наилучшими методами из шести оказались L_1 и $L_{1,8}$. В сочетании они также обеспечивают более высокую точность аппроксимации, чем метод L_1. Но при этом по критерию *SAE* качество модели (30), (31) выше, чем у (28), (29), что подтверждает зависимость результата оценивания списочной регрессии от маркировки методов. В модели (30), (31) метод $L_{1,8}$ сработал 9 раз, метод L_1 — 25 раз.

Эксперимент № 2. Сначала была решена задача ЧБЛП (10), (12) — (15), (17) — (27) при $m = 2$, $\Delta = 0{,}01$. В результате была получена оптимальная маркировка методов: L_1 — 2, $L_{1,2}$ — 1, $L_{1,4}$ — 4, $L_{1,6}$ — 5, $L_{1,8}$ — 6, L_2 — 3. Ей соответствует следующая списочная модель:

$$\tilde{y}=\begin{bmatrix} L_1 \\ L_2 \end{bmatrix}_{N^{(3)}}, \tag{32}$$

где

$$N^{(3)}=0,685+1,767x_1-0,00357x_2-0,0152x_3-$$
$$-0,019x_4-0,0000376x_5+0,0544x_6+0,0232x_7-0,0152x_8\,. \tag{33}$$

Для модели (32), (33) *SAE* = 165,1142. Естественным образом её качество оказалось выше, чем у регрессии (28), (29) и (30), (31), построенных в результате организации маркировки методов вручную. В модели (32), (33) метод L_1 сработал 26 раз, метод L_2 — 8 раз.

При $m = 3$ и $\Delta = 0{,}01$ оптимальная маркировка методов имеет вид: L_1 — 3, $L_{1,2}$ — 2, $L_{1,4}$ — 5, $L_{1,6}$ — 6, $L_{1,8}$ — 1, L_2 — 4. Ей соответствует списочная модель

$$\tilde{y} = \begin{bmatrix} L_{1,2} \\ L_1 \\ L_2 \end{bmatrix}_{N^{(4)}}, \tag{34}$$

где

$$N^{(4)} = 30{,}986 - 11{,}518x_1 - 0{,}0485x_2 - 0{,}217x_3 - \\ -0{,}479x_4 + 0{,}0241x_5 - 0{,}0936x_6 + 0{,}11x_7 + 1{,}183x_8\,. \tag{35}$$

Для модели (34), (35) *SAE* = 157,6717. Её качество заметно улучшилось по сравнению с моделью (32), (33). В модели (34), (35) метод $L_{1,2}$ сработал 13 раз, метод L_1 — 16 раз, метод L_2 — 5 раз.

При $m = 4$ и $\Delta = 0{,}01$ оптимальная маркировка методов имеет вид: L_1—3, $L_{1,2}$—5, $L_{1,4}$ — 6, $L_{1,6}$ — 1, $L_{1,8}$ — 2, L_2 — 4. Ей соответствует списочная модель

$$\tilde{y} = \begin{bmatrix} L_{1,8} \\ L_1 \\ L_2 \\ L_{1,2} \end{bmatrix}_{N^{(5)}}, \tag{36}$$

$$N^{(5)} = -47{,}277 + 16{,}483x_1 - 0{,}386x_2 + 0{,}233x_3 + \\ +0{,}794x_4 - 0{,}00832x_5 - 2{,}161x_6 + 0{,}393x_7 - 5{,}483x_8\,. \tag{37}$$

Для модели (36), (37) *SAE* = 156,2552. Её качество улучшилось незначительно по сравнению с моделью (34), (35). В модели (36), (37) метод $L_{1,8}$ сработал 3 раза, метод L_1 — 12 раз, метод L_2 — 8 раз, метод $L_{1,2}$ — 11 раз.

Заметим, что в правилах (33), (35) и (37) свободный член был уменьшен на 1 для обеспечения в соответствующих списках минимального маркера 1.

Задача ЧБЛП (10), (12) — (15), (17) — (27) также была решена при $m = 5$ и $m = 6$. При этом величины *SAE* составили 156,2226 и 156,1886 соответственно. Поэтому оптимальной списочной моделью с точки зрения точности и сложности следует признать регрессию (34), (35), осуществляющей переключение трех методов: L_1, L_2 и $L_{1,2}$.

Заключение

В статье сформулирована задача ОСМ, состоящая в выделении из имеющегося числа p методов оценивания параметров линейной регрессии m наиболее сочетающихся методов по критерию суммы модулей остатков. Переключение методов осуществляется благодаря их маркировке и применению целочисленной функции «пол». Сформулированная задача сведена к задаче ЧБЛП для случая известной и неизвестной маркировки методов. В случае неизвестной маркировки решение задачи ЧБЛП позволяет определить m наиболее сочетающихся методов, их оптимальную маркировку и правило переключения.

Резюмируя, хотелось бы обратить внимание на следующие моменты.

Предложенный в статье способ отбора методов при их неизвестной маркировке подразумевает решение задачи ЧБЛП, содержащей $p \cdot (n + p + 1)$ булевых переменных. Из этого следует, что чем больше объем выборки n и количество используемых методов p, тем выше вычислительная сложность задачи ЧБЛП. В этом на данный момент прослеживается определенная ограниченность предложенного способа при обработке больших данных. Тем не менее, рассмотренная в статье задача обработки выборки объема $n = 34$ с числом входных переменных $l = 8$ и с количеством методов $p = 6$ с использованием пакета COPT была решена довольно эффективно.

Достоинство сформулированных задач ЧБЛП в том, что для них не требуется решать проблему выбора достаточно большого положительного числа M.

Полученную в результате решения задач ЧБЛП списочную модель, состоящую из списка оцененных разными методами линейных регрессий и правила их переключения, можно использовать для прогнозирования значений выходной переменной y. Например, для получения прогноза по модели (34), (35) нужно сначала подставить значения входных переменных в правило (35) и найти маркер метода. Затем для соответствующего маркеру метода из списка (34) вновь с помощью значений входных переменных вычислить прогнозное значение выходной переменной y.

Предложенный в статье способ, предназначенный для обработки данных из любых предметных областей, описан для отбора только из класса L_ν-методов оценивания. Однако для этого могут быть использованы абсолютно любые методы оценки линейных регрессий, многие из которых упомянуты во введении. К тому же предложенный способ ОСМ только для линейной регрессии может быть адаптирован в способ отбора наиболее сочетающихся спецификаций моделей для конкретного метода. Более того, разработанный способ можно легко адаптировать для решения задачи отбора наиболее сочетающихся нейронных сетей.

Список источников / References

1. Акбердина, В.В., Шориков, А.Ф., Коровин, Г.Б., Сиротин, Д.В. (2024). Идентификация параметров агент-ориентированной модели управления промышленным комплексом региона. *Экономика региона*, 20(1), 48—62. https://doi.org/10.17059/ekon.reg.2024-1-4
Akberdina, V.V., Shorikov, A.F., Korovin, G.B., Sirotin, D.V. (2024). Parameter Identification

of the Agent-Based Model for Managing a Regional Industrial Complex. *Economy of Regions*, 20(1), 48—62. (In Russ.). https://doi.org/10.17059/ekon.reg.2024-1-4

2. Базилевский, М.П. (2020). Отбор оптимального числа информативных регрессоров по скорректированному коэффициенту детерминации в регрессионных моделях как задача частично целочисленного линейного программирования. *Прикладная математика и вопросы управления*, (2), 41—54. https://doi.org/10.15593/2499-9873/2020.2.03
Bazilevskii, M.P. (2020). Selection an Optimal Number of Variables in Regression Models using Adjusted Coefficient of Determination as a Mixed Integer Linear Programming Problem. *Applied Mathematics and Control Sciences*, (2), 41—54. (In Russ.). https://doi.org/10.15593/2499-9873/2020.2.03
3. Базилевский, М.П. (2024). Двухкритериальное оценивание линейных регрессионных моделей методами наименьших модулей и квадратов. *International Journal of Open Information Technologies*, 12(6), 76—81.
Bazilevskii, M.P. (2024). Two-criteria Estimation of Linear Regression Models Using Least Absolute Deviations and Squares. *International Journal of Open Information Technologies*, 12(6), 76—81. (In Russ.).
4. Базилевский, М.П. (2024). Оценивание с помощью метода наименьших модулей регрессионных моделей с целочисленными функциями пол и потолок. *International Journal of Open Information Technologies*, 12(10), 56—61.
Bazilevskii, M.P. (2024). Estimation using Least Absolute Deviations Method of Regression Models with Integer Floor and Ceiling Functions. *International Journal of Open Information Technologies*, 12(10), 56—61. (In Russ.).
5. Базилевский, М.П. (2025). Идентификация неизвестных параметров списочных регрессионных моделей методом наименьших модулей. *System Analysis and Mathematical Modeling*, 7(2), 248—262.
Bazilevskii, M.P. (2025). Identification of Unknown Parameters in List Regression Models Using the Least Absolute Deviations. *System Analysis and Mathematical Modeling*, 7(2), 248—262. (In Russ.).
6. Базилевский, М.П. (2025). Оптимальная маркировка значений категориальной зависимой переменной при построении регрессионных моделей с целочисленной функцией пол. *Современные наукоёмкие технологии*, (5), 22—26. https://doi.org/10.17513/snt.40385
Bazilevskii, M.P. (2025). Optimal Marking of Categorical Dependent Variable Values when Constructing Regression Models using Integer Floor Function. *Modern High Technologies*, (5), 22—26. (In Russ.). https://doi.org/10.17513/snt.40385
7. Базилевский, М.П. (2025). Сравнительный анализ разных подходов к оценке параметров регрессионных моделей с помощью метода наименьших модулей на примере моделирования стоимости домов по выборке большого объема. *Инженерный вестник Дона*, (6). http://www.ivdon.ru/ru/magazine/archive/n6y2025/10130.
Bazilevskii, M.P. (2025). Comparative Analysis of Different Approaches to Estimating the Parameters of Regression Models using the Least Absolute Deviations Method using the Example of Modeling House Prices Based on a Large Sample. *Engineering Journal of Don*, (6), http://www.ivdon.ru/ru/magazine/archive/n6y2025/10130. (In Russ.).
8. Голованов, О.А., Тырсин, А.Н. (2024). Спуск по узловым прямым и симплекс-алгоритм — два варианта регрессионного анализа на основе метода наименьших

модулей. *Заводская лаборатория. Диагностика материалов*, 90(5), 79—87. https://doi.org/10.26896/1028-6861-2024-90-5-79-87
Golovanov, O.A., Tyrsin, A.N. (2024). Descent along Nodal Straight Lines and Simplex Algorithm: Two Variants of Regression Analysis Based on the Least Absolute Deviation Method. *Industrial Laboratory. Diagnostics of Materials*, 90(5), 79—87. (In Russ.). https://doi.org/10.26896/1028-6861-2024-90-5-79-87

9. Грэхем, Р., Кнут, Д., Паташник, О. (1998). *Конкретная математика. Основание информатики*: Пер. с англ. М.: Мир.
Grehkhem, R., Knut, D., Patashnik, O. (1998). *Konkretnaya matematika. Osnovanie informatiki*: Per. s angl. Moscow: World. (In Russ.).

10. Демиденко, Е.З. (1981). *Линейная и нелинейная регрессии*. М.: Финансы и статистика.
Demidenko, E.Z. (1981). *Linear and Nonlinear Regressions*. Moscow: Finance and Statistics. (In Russ.).

11. Карпенко, П.А. (2021). Математическое описание концептуальной модели управления развитием региональных социально-экономических систем Российской Федерации. *Вестник Алтайской академии экономики и права*, (9—1), 69—74.
Karpenko, P.A. (2021). Mathematical Description of the Conceptual Model of Managing the Development of Regional Socio-Economic Systems of the Russian Federation. *Vestnik Altaiskoi akademii ehkonomiki i prava*, (9—1), 69—74. (In Russ.).

12. Курлов, В.В., Косухина, М.А., Курлов, А.В. (2022). Модель оценки цифровой зрелости промышленного предприятия. *Экономика и управление*, 28(5), 439—451. http://doi.org/10.35854/1998-1627-2022-5-439-451
Kurlov, V.V., Kosukhina, M.A., Kurlov, A.V. (2022). Model for Assessing the Digital Maturity of an Industrial Enterprise. *Economics and Management*, 28(5), 439—451. (In Russ.). http://doi.org/10.35854/1998-1627-2022-5-439-451

13. Мудров, В.И., Кушко, В.Л. (1976). *Методы обработки измерений*. М.: Сов. радио.
Mudrov, V.I., Kushko, V.L. (1976). *Metody obrabotki izmerenii*. Moscow: Sov. Radio. (In Russ.).

14. Носков, С.И. (2022). Кусочно-линейная свёртка вариантов регрессионной модели объекта. *Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии*, (3), 15—21.
Noskov, S.I. (2022). Piece-Linear Convolution of Options of the Object Regression Model. *Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies*, (3), 15—21. (In Russ.).

15. Носков, С.И., Баенхаева, А.В. (2016). Множественное оценивание параметров линейного регрессионного уравнения. *Современные технологии. Системный анализ. Моделирование*, 3(51), 133—138.
Noskov, S.I., Baenkhaeva, A.V. (2016). Multiple Estimation of Parameters for the Linear Regression Equation. *Modern Technologies. System Analysis. Modeling*, 3(51), 133—138. (In Russ.).

16. Носков, С.И., Попов, Е.С. (2024). Метод смешанного оценивания параметров регрессионной модели с произвольным составом групп наблюдений. *Модели, системы, сети в экономике, технике, природе и обществе*, 3(51), 98—104.
Noskov, S.I., Popov, E.S. (2024). The Method of Mixed Estimation of Regression Model Parameters with an Arbitrary Composition of Observation Groups. *Models, Systems, Networks in Economics, Technology, Nature and Society*, 3(51), 98—104. (In Russ.).

17. Холодилин, И.Ю., Григорьев, М.А., Кушнарёв, В.А., Савостеенко, Н.В., Спицин, Д.В., Осипов, О.И. (2024). Набор регрессионных моделей для управления тележкой, оснащенной манипулятором робота, интегрированным в технологический процесс. *Вестник Южно-Уральского государственного Университета. Серия «Энергетика»*, 24(4), 47—55. https://doi.org/10.14529/power240406
Kholodilin, I.YU., Grigor'ev, M.A., Kushnarev, V.A., Savosteenko, N.V., Spitsin, D.V., Osipov, O.I. (2024). A Set of Regression Models for Controlling a Trolley Equipped with a Robot Manipulator Integrated into the Technological Process. *Bulletin of the South Ural State University. Ser. Power Engineering*, 24(4), 47—55. (In Russ.). https://doi.org/10.14529/power240406
18. Чигиринский, Ю.Л., Ингеманссон, А.Р. (2021). Математические модели оперативного управления технологическими режимами лезвийной обработки. *Наукоемкие технологии в машиностроении*, (7), 20—28. https://doi.org/10.30987/2223-4608-2021-7-20-28
Chigirinskii, YU.L., Ingemansson, A.R. (2021). Mathematical Model of Technological Mode On-line Control at Blade Processing. *Science Intensive Technologies in Mechanical Engineering*, (7), 20—28. (In Russ.). https://doi.org/10.30987/2223-4608-2021-7-20-28
19. Bertsimas, D., King, A., Mazumder, R. (2016). Best Subset Selection via a Modern Optimization lens. *The annals of statistics*, 44(2), 813—852.
20. De Menezes, D.Q.F., Prata, D.M., Secchi, A.R., Pinto, J.C. (2021). A Review on Robust M-estimators for Regression Analysis. *Computers & Chemical Engineering*, 147, 107254. https://doi.org/10.1016/j.compchemeng.2021.107254
21. Ge, D., Huangfu, Q., Wang, Z., Wu, J., Ye, Y. (2023). Cardinal Optimizer (COPT) user guide. https://guide.coap.online/copt/en-doc.
22. Idroes, G.M., Hardi, I., Hilal, I.S., Utami R.T., Noviandy T.R., Idroes R. (2024). Economic Growth and Environmental Impact: Assessing the Role of Geothermal Energy in Developing and Developed Countries. *Innovation and Green Development*, 3(3), 100144. https://doi.org/10.1016/j.igd.2024.100144
23. Khan, Y., Hassan, T. (2024). Promoting Sustainable Development: Evaluating the Influence of Natural Resources, High-tech Export and Corruption on CO2 Emissions in Developing Economies. *Resources Policy*, 88, 104511. https://doi.org/10.1016/j.resourpol.2023.104511
24. Koch, T., Berthold, T., Pedersen, J., Vanaret, C. (2022). Progress in Mathematical Programming Solvers from 2001 to 2020. *EURO Journal on Computational Optimization*, 10, 100031. https://doi.org/10.1016/j.ejco.2022.100031
25. Konno, H., Yamamoto, R. (2009). Choosing the Best Set of Variables in Regression Analysis using Integer Programming. *Journal of Global Optimization*, 44(2), 273—282. https://doi.org/10.1007/s10898-008-9323-9
26. Liu, Y., Liu, B. (2024). A Modified Uncertain Maximum Likelihood Estimation with Applications in Uncertain Statistics. *Communications in Statistics-Theory and Methods*, 53(18), 6649—6670. https://doi.org/10.1080/03610926.2023.2248534
27. Liu, Z., Yang, Y. (2020). Least Absolute Deviations Estimation for Uncertain Regression with Imprecise Observations. *Fuzzy Optimization and Decision Making*, 19(1), 33—52. https://doi.org/10.1007/s10700-019-09312-w
28. Sengupta, D., Klein, S., Raine, J.A., Golling, T. (2024). CURTAINs Flows for Flows: Constructing Unobserved Regions with Maximum Likelihood Estimation. *SciPost Physics*, 17(2), 046. https://doi.org/10.21468/SciPostPhys.17.2.046

29. Wagner, H.M. (1959). Linear Programming Techniques for Regression Analysis. *Journal of the American Statistical Association*, 54(285), 206—212.
30. Wang, S., Chen, Y., Cui, Z., Lin, L., Zong, Y. (2024). Diabetes Risk Analysis based on Machine Learning LASSO Regression Model. *Journal of Theory and Practice of Engineering Science*, 4(01), 58—64. https://doi.org/10.53469/jtpes.2024.04(01).08
31. Wang, Y., Wang, B., Peng, C., Li, X., Yin, H. (2022). Huber Regression Analysis with a Semi-Supervised Method. *Mathematics*, 10(20), 3734. https://doi.org/10.3390/math10203734

Информация об авторе

Михаил Павлович Базилевский, кандидат технических наук, доцент кафедры математики, факультет «Управление на транспорте и информационные технологии», Иркутский государственный университет путей сообщения (ФГБОУ ВО ИрГУПС), Иркутск, Российская Федерация, ORCID: https://orcid.org/0000-0002-3253-5697, e-mail: mik2178@yandex.ru

Information about the author

Mikhail P. Bazilevskiy, Candidate of Science (Technical), Associate Professor, Department of Mathematics, Irkutsk State Transport University, Irkutsk, Russian Federation, ORCID: https://orcid.org/0000-0002-3253-5697, e-mail: mik2178@yandex.ru

Вклад автора

Базилевский М.П. — идеи исследования; аннотирование, написание и оформление рукописи; планирование исследования; контроль за проведением исследования; применение статистических, математических или других методов для анализа данных; проведение эксперимента; сбор и анализ данных; визуализация результатов исследования.

Contribution of the author

Mikhail P. Bazilevskiy — ideas; annotation, writing and design of the manuscript; planning of the research; control over the research; application of statistical, mathematical or other methods for data analysis; conducting the experiment; data collection and analysis; visualization of research results.

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.

Conflict of interest

The author declare no conflict of interest.

Поступила в редакцию 13.10.2025 Received 2025.10.13
Поступила после рецензирования 23.10.2025 Revised 2025.10.23
Принята к публикации 29.10.2025 Accepted 2025.10.29
Опубликована 28.12.2025 Published 2025.12.28