

С.И. Носков¹, А.Р. Чекалова¹

¹ Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация

О РАВЕНСТВЕ СУММ ПОЛОЖИТЕЛЬНЫХ И ОТРИЦАТЕЛЬНЫХ ОШИБОК АППРОКСИМАЦИИ В РЕГРЕССИОННЫХ МОДЕЛЯХ

Аннотация. В работе описан элементарный способ учета при вычислении параметров линейной регрессионной модели требования равенства нулю ошибок аппроксимации для методов наименьших модулей и антиробастного оценивания. Он состоит в дополнении систем ограничений соответствующих задач линейного программирования одним равенством. Решен простой численный пример, результаты которого показывают, что учет данного требования нарушает некоторые свойства указанных методов.

Ключевые слова: регрессионная модель, параметры, методы наименьших модулей и антиробастного оценивания, равенство нулю ошибок аппроксимации, задача линейного программирования, адекватность.

S.I. Noskov¹, A.R. Chekalova¹

¹ Irkutsk State Transport University, Irkutsk, Russian Federation

ON THE EQUALITY OF THE SUMS OF POSITIVE AND NEGATIVE APPROXIMATION ERRORS IN REGRESSION MODELS

Abstract. The paper describes an elementary method of taking into account the requirement of zero approximation errors for the methods of least modules and anti-robust estimation when calculating the parameters of a linear regression model. It consists in supplementing the constraint systems of the corresponding linear programming problems with one equality. A simple numerical example has been solved, the results of which show that taking into account this requirement violates some properties of these methods.

Keywords: regression model, parameters, methods of least modules and anti-robust estimation, equality of approximation errors to zero, linear programming problem, adequacy.

Рассмотрим линейное регрессионное уравнение (модель):

$$y_k = \sum_{i=1}^m \alpha_i x_{ki} + \varepsilon_k, \quad k = \overline{1, n}, \quad (1)$$

где y – зависимая (объясняемая), а x_i – i -ая независимая (объясняющая) переменные, α_i – i -ый неизвестный параметр, ε_k – ошибки аппроксимации, k – номер наблюдения, n – их количество (длина выборки).

Часто при построении регрессионных моделей принято требовать выполнения равенства нулю суммы ошибок аппроксимации:

$$\sum_{k=1}^n \varepsilon_k = 0, \quad (2)$$

или, что то же, равенства сумм положительных и отрицательных ошибок. Таково, в частности, одно из условий теоремы Гаусса-Маркова относительно оптимальности оценок параметров, полученных по методу наименьших квадратов (МНК), в классе линейных несмещённых оценок (см., например, [1-9]). Очевидно, что это требование совершенно естественно и из содержательных соображений – положительные и отрицательные ошибки

в идеале должны уравнивать друг друга. Известно, что для линейной регрессии (1) со свободным членом (т.е. при $x_{k1}=1, k = \overline{1, n}$) условие (2) при использовании МНК выполняется автоматически.

Для методов наименьших модулей (МНМ) [10] и антиробастного оценивания параметров (МАО) [10] учет условия (2) также осуществляется элементарным образом. Действительно, введем в рассмотрение неотрицательные переменные $u_k, v_k, k = \overline{1, n}$:

$$u_k = \begin{cases} \varepsilon_k, & \text{если } \varepsilon_k > 0 \\ 0, & \text{в противном случае,} \end{cases}$$

$$v_k = \begin{cases} -\varepsilon_k, & \text{если } \varepsilon_k < 0 \\ 0, & \text{в противном случае.} \end{cases}$$

Тогда реализация МНМ сводится к решению следующей задачи линейного программирования (ЛП):

$$\sum_{i=1}^m \alpha_i x_{ki} + u_k - v_k = y_k, k = \overline{1, n}, \quad (3)$$

$$u_k \geq 0, v_k \geq 0, k = \overline{1, n}, \quad (4)$$

$$\sum_{k=1}^n (u_k + v_k) \rightarrow \min. \quad (5)$$

Учет же условия (2) обеспечивается дополнением задачи (ЛП) (3) – (5) ограничением

$$\sum_{k=1}^n u_k = \sum_{k=1}^n v_k. \quad (6)$$

При этом выполнение следующих из определения переменных u_k, v_k условий $u_k v_k = 0, k = \overline{1, n}$ на ее оптимальном решении достигается, что очевидно, автоматически.

При реализации МАО ограничения (3), (4) дополняются системой неравенств

$$u_k + v_k - r \leq 0, k = \overline{1, n}, \quad (7)$$

а целевая функция (5) в задаче ЛП заменяется на следующую:

$$r + \delta \sum_{k=1}^n (u_k + v_k) \rightarrow \min, \quad (8)$$

где δ – наперед заданное малое положительное число.

Учет условия (2) при использовании МАО также обеспечивается дополнением задачи ЛП (3), (4), (7), (8) ограничением (6).

Аналогичным образом условие (2) может быть учтено при формировании соответствующих задач ЛП для реализации методов смешанного оценивания [11], непрерывной формы метода максимальной согласованности (ММС) [12] и задач линейного булева программирования при использовании дискретной формы ММС [13] или оценивания параметров кусочно-линейных регрессионных моделей [14].

Рассмотрим численный пример. Пусть задана выборка данных:

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

Построим по ней линейную двухфакторную модель без свободного члена:

$$y_k = \alpha_1 x_{k1} + \alpha_2 x_{k2} + \varepsilon_k, k = \overline{1,5}$$

с помощью МНМ и МАО – без учета условия (2) и с ним.

1. МНМ.

$$y_k = 2.11x_{k1} - 1.42x_{k2} + \varepsilon_k, k = \overline{1,5},$$

$$\varepsilon = (2.69, -13.07, 0, 5.5, 0), E = 4.25,$$

где – E – средняя абсолютная ошибка аппроксимации.

С учетом условия (2):

$$y_k = 1.75x_{k1} - 1.14 + \varepsilon_k, k = \overline{1,5},$$

$$\varepsilon = (3.66, -10.73, 1.05, 6.02, 0), E = 4.29.$$

2. МАО.

$$y_k = 0.81x_{k1} + 0.26x_{k2} + \varepsilon_k, k = \overline{1,5},$$

$$\varepsilon = (4.15, -6.04, 1.07, 6.04, -6.04), E = 4.67, r = 6.04.$$

С учетом условия (2):

$$y_k = 0.85x_{k1} + 0.17x_{k2} + \varepsilon_k, k = \overline{1,5},$$

$$\varepsilon = (4.25, -6.11, 1.23, 6.11, -5.48), E = 4.64, r = 6.11.$$

Анализ результатов моделирования позволяет сделать следующие выводы.

1. Учет условия (2) незначительно меняет оценки параметров моделей и значения критериев адекватности, однако обеспечивает требуемое равенство сумм положительных и отрицательных ошибок.

2. Дополнение ограничений задач ЛП (3) – (5) и (3), (4), (7), (8) условием (6) нарушает некоторые свойства МНМ (равенство нулю m ошибок аппроксимации) и МАО (число максимальных по модулю ошибок не меньше $m + 1$) (см., в частности, [15]).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Демиденко Е.З. Линейная и нелинейная регрессии // Финансы и статистика. – Москва, 1981. – С. 302.
2. Дрейпер Н., Смит Г. Прикладной регрессионный анализ // Финансы и статистика. – Москва, 1981. – Т. 1. – С. 366. – Т. 2. – С. 351.
3. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей // Финансы и статистика. – Москва, 1985. – С. 488.
4. Афифи А., Эйзен С. Статистический анализ: поход с использованием ЭВМ // Мир. – Москва, 1982. – С. 486.
5. Джонстон Дж. Эконометрические методы // Статистика. – Москва, 1980. – С. 446.
6. Пирогов Г.Г., Федоровский Ю.П. Проблемы структурного оценивания в эконометрии // Статистика. – Москва, 1979. – С. 327.
7. Поллард Дж. Справочник по вычислительным методам статистики // Финансы и статистика. – Москва, 1982. – С. 344.

8. Прикладная статистика // Финансы и статистика. – Москва, 1980-1990. – Т. 1. – С. 510. – Т. 2. – С. 526.
9. Фишер Ф. Проблемы идентификации в эконометрии // Статистика. – Москва, 1978. – С. 223.
10. Носков С.И. L-множество в многокритериальной задаче оценивания параметров регрессионных уравнений // Информационные технологии и проблемы математического моделирования сложных систем, 2004. – № 1. – С. 164-171.
11. Носков С.И. Метод смешанного оценивания параметров линейной регрессии: особенности применения // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии, 2021. – № 1. – С. 126-132.
12. Носков С.И. Применение непрерывного критерия согласованности поведения при построении регрессионных моделей // Известия Тульского государственного университета. Технические науки, 2021. – № 6. – С. 74-78.
13. Носков С.И. Метод максимальной согласованности в регрессионном анализе // Известия Тульского государственного университета. Технические науки, 2021. – № 10. С. 380-385.
14. Носков С.И., Жукова М.С., Кириллова Т.К., Купитман Ю.О., Хоняков А.А. Уточнение способов идентификации параметров некоторых кусочно-линейных регрессий // Электронный сетевой политематический журнал "Научные труды КубГТУ", 2023. – № 2. – С. 75-81.
15. Носков С.И. О кластеризации данных на основе свойств методов идентификации параметров линейной регрессии // Информационные технологии и математическое моделирование в управлении сложными системами, 2022. – № 4 (16). – С. 82-85.

REFERENCES

1. Demidenko E.Z. Linear and nonlinear regression // Finance and statistics. – Moscow, 1981. – P. 302.
2. Draper N., Smith G. Applied regression analysis // Finance and Statistics. – Moscow, 1981. – V. 1. – P. 366. – V. 2. – P. 351.
3. Ayvazyan S.A., Enyukov I.S., Meshalkyn L.D. Applied statistics. Addiction research // Finance and Statistics. – Moscow, 1985. – P. 488.
4. Afifi A., Eisen S. Statistical analysis: a campaign using a computer // Mir. – Moscow, 1982. – P. 486.
5. Johnston J. Econometric methods // Statistics. – Moscow, 1980. – P. 446.
6. Pirogov G.G., Fedorovsky Yu.P. Problems of structural estimation in econometrics // Statistics. – Moscow, 1979. – P. 327.
7. Pollard J. Handbook of computational methods of statistics // Finance and Statistics. – Moscow, 1982. – P. 344.
8. Applied statistics // Finance and Statistics. – Moscow, 1980-1990. – V. 1. – P. 510. – V. 2. – P. 526.
9. Fischer F. Identification problems in econometrics // Statistics. – Moscow, 1978. – P. 223.
10. Noskov S.I. L-set in the multicriteria problem of estimating the parameters of regression equations // Information technologies and problems of mathematical modeling of complex systems, 2004. – №. 1. – P. 164-171.
11. Noskov S.I. Method of mixed estimation of linear regression parameters: application features // Bulletin of the Voronezh State University. Series: System Analysis and Information Technologies, 2021. – №. 1. – P. 126-132.

12. Noskov S.I. Application of the continuous criterion of consistency of behavior in the construction of regression models // Proceedings of Tula State University. Technical Sciences, 2021. – №. 6. – P. 74-78.

13. Noskov S.I. The method of maximum consistency in regression analysis // Izvestiya Tula State University. Technical Sciences, 2021. – №. 10. – P. 380-385.

14. Noskov S.I., Zhukova M.S., Kirillova T.K., Kupitman Yu.O., Khonyakov A.A. Refinement of methods for identifying parameters of some piecewise linear regressions // Electronic network polythematic journal "Scientific Works of KubSTU", 2023. – №. 2. – P. 75-81.

15. Noskov S.I. On data clustering based on the properties of methods for identifying linear regression parameters // Information technologies and mathematical modeling in the management of complex systems, 2022. – №. 4 (16). – P. 82-85.

Информация об авторах

Сергей Иванович Носков – д.т.н., профессор, профессор кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: sergey.noskov.57@mail.ru

Александра Романовна Чекалова – магистрант кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: chekalova49@gmail.com

Authors

Sergey Ivanovich Noskov, Doctor of Technical Science, Professor, Professor of the Subdepartment of Information Systems and Information Security, Irkutsk State Transport University, Irkutsk, e-mail: sergey.noskov.57@mail.ru

Aleksandra Romanovna Chekalova, Undergraduate student of the Department Information Systems and Information Security Irkutsk State Transport University, Irkutsk, e-mail: chekalova49@gmail.com

Для цитирования

Носков С.И., Чекалова А.Р. О равенстве сумм положительных и отрицательных ошибок аппроксимации в регрессионных моделях// «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2024. – №1. – С. 59-64. – Режим доступа: <http://ismm-irgups.ru/toma/121-2024>, свободный. – Загл. с экрана. – Яз.рус., англ. (дата обращения: 27.02.2024)

For citations

Noskov S.I., Chekalova A.R. On the equality of the sums of positive and negative approximation errors in regression models// Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2024. No. 1. P.59-64. (Accessed 27.02.2024)