

СПОСОБ РАЗРЕШЕНИЯ АЛЬТЕРНАТИВНОСТИ В ОЦЕНКАХ ПАРАМЕТРОВ РЕГРЕССИОННЫХ МОДЕЛЕЙ

С. И. Носков

Иркутский государственный университет путей сообщения, Иркутск, Россия
sergey.noskov.57@mail.ru

Аннотация. *Актуальность и цели.* Регрессионный анализ является весьма эффективным средством математического моделирования сложных систем различного характера и масштаба, позволяет выявлять как явные, так и скрытые тенденции в их функционировании и развитии. Цель исследования состоит в решении задачи вычисления компромиссной оценки параметров линейной регрессионной модели, возникающей при ее построении несколькими альтернативными методами идентификации. *Материалы и методы.* Для достижения поставленной цели применялся математический аппарат решения задач линейного программирования. *Результаты.* В общем случае сформулированная задача сводится к весьма сложной в вычислительном отношении задаче нелинейного программирования. При некоторых упрощающих исходную постановку допущениях исходную задачу удалось свести к задаче линейного программирования. При этом был использован популярный в теории принятия решений метод уступок. *Выводы.* Описанный в работе алгоритм позволяет избежать альтернативности при оценивании параметров регрессионных моделей.

Ключевые слова: регрессионная модель, функция потерь, методы оценивания параметров, задача линейного программирования, альтернативность, метод уступок

Для цитирования: Носков С. И. Способ разрешения альтернативности в оценках параметров регрессионных моделей // Модели, системы, сети в экономике, технике, природе и обществе. 2023. № 4. С. 154–162. doi: 10.21685/2227-8486-2023-4-10

METHOD FOR RESOLUTION OF ALTERNATIVENESS IN ESTIMATES OF PARAMETERS OF REGRESSION MODELS

S.I. Noskov

Irkutsk State Transport University, Irkutsk, Russia
sergey.noskov.57@mail.ru

Abstract. *Background.* Regression analysis is a very effective means of mathematical modeling of complex systems of various nature and scale, allowing you to identify both explicit and hidden trends in their functioning and development. The purpose of the study is to solve the problem of calculating a compromise estimate of the parameters of a linear regression model that arises when it is built by several alternative identification methods. *Materials and methods.* To achieve this goal, a mathematical apparatus for solving linear programming problems was used. *Results.* In the general case, the formulated problem is reduced to a computationally complex problem of nonlinear programming. Under some assumptions simplifying the original formulation, the original problem was reduced to a linear programming problem. At the same time, the method of concessions, popular in the theory of decision making, was used. *Conclusions.* The algorithm described in the paper

© Носков С. И., 2023. Контент доступен по лицензии Creative Commons Attribution 4.0 License / This work is licensed under a Creative Commons Attribution 4.0 License.

makes it possible to avoid alternativeness when estimating the parameters of regression models.

Keywords: regression model, loss function, parameter estimation methods, linear programming problem, alternativeness, concession method

For citation: Noskov S.I. Method for resolution of alternativeness in estimates of parameters of regression models. *Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve* = *Models, systems, networks in economics, technology, nature and society*. 2023;(4):154–162. (In Russ.). doi: 10.21685/2227-8486-2023-4-10

Введение

Регрессионный анализ является весьма эффективным средством математического моделирования сложных систем различного характера и масштаба, позволяет выявлять как явные, так и скрытые тенденции в их функционировании и развитии. Так, в работе [1] с помощью смешанной регрессионной модели разработан прогноз годового стока реки Хуанхэ в окрестности гидроэлектростанции Санмэнся. При этом в качестве независимых факторов использованы осадки, температура воздуха и потребление воды. В работе [2] представлена нелинейная регрессионная модель лазерных триангуляционных зондов, эффективно используемая при их проектировании. Статья [3] посвящена разработке модели регрессии со случайными коэффициентами для оценки средних переменных издержек конкретной производственной деятельности предприятия в совокупности фирм с несколькими видами производимой продукции. В исследовании [4] описана регрессионная модель оценки безопасности перевозочного процесса железнодорожного транспорта на региональном уровне.

В работе [5] представлено сочетание использования бинарной логистической регрессии и стохастического граничного анализа для оценки оперативной эффективности координационных центров береговой охраны (морской спасательной службы) Великобритании за период с 1995 по 1998 г. В частности, изучается обоснование принятого в 1999 г. решения правительства о закрытии ряда координационных центров. Регрессионные модели, описанные в статье, представляют собой систему измерения эффективности, которая значительно более реалистична и сложна, чем та, которая в настоящее время используется правительством Великобритании. В работе [6] используется регрессионная модель Вейбулла со случайными эффектами для оценивания среднего времени наработки на отказ и среднего времени до ремонта при техническом обслуживании истребительной авиации, что является важным вопросом управления военно-воздушными силами. В работе [7] многомерная кусочно-линейная регрессионная модель и модель логистической регрессии используются для планирования контрповстанческой деятельности в республике Сомали при оценке вероятности четырех типов разрешения вероятных конфликтов на 15-летнем временном горизонте.

В работе [8] исследуется применение многомерных адаптивных регрессионных сплайнов для разработки как краткосрочных, так и долгосрочных прогнозов спроса на природный газ бытовых пользователей для операторов распределительных систем. Статья [9] посвящена проведению эмпирического исследования эффективности пяти популярных непараметрических регрессионных моделей для оценки стоимости жизненного цикла продукта (включая его создание, использование и утилизацию) в различных моделируемых

средах. Они устанавливаются путем изменения количества факторов (независимых переменных), размера выборочных данных, степени шума выборочных данных и степени их смещения. В работе [10] показано, как регрессионные модели могут оказаться полезными для оценки внутреннего и внешнего диаметров в процессе экструзии труб.

Цель настоящей работы состоит в разработке подхода к построению компромиссного в некотором заданном смысле вектора оценок параметров регрессионной модели по отношению к совокупности его альтернативных вариантов.

Материалы и методы

Рассмотрим обязательный элемент практически любой регрессионной модели – линейное уравнение (зависимость) вида

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

где y – зависимая, а x_i – i -я независимая переменные; α_i – i -й подлежащий оцениванию параметр; ε_k – ошибки аппроксимации; k – номер наблюдения; n – число наблюдений (длина выборки).

Представим уравнение (1) в векторной форме:

$$y = X\alpha + \varepsilon, \quad (2)$$

где $y = (y_1, \dots, y_n)^T$, $\alpha = (\alpha_1, \dots, \alpha_m)^T$, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^T$, $X = (n \times m)$ – матрица с компонентами x_{ki} .

Будем считать все переменные и ошибки аппроксимации в модели (2) детерминированными.

Основная проблема построения регрессионных моделей связана с оцениванием вектора неизвестных параметров α . Для этого в рамках регрессионного анализа разработан весьма значительный и постоянно расширяемый арсенал методов (см., например, работы [11–15]). Так, только в программном комплексе GRETL [16] для эконометрического моделирования реализовано около 30 таких методов.

Весьма широкий класс методов оценивания параметров уравнения (2) связан с вычислением так называемых L_v -оценок посредством минимизации функций потерь вида [17, 18]

$$J_v(\alpha) = \sum_{k=1}^n |\varepsilon_k|^v.$$

Каждая из них различается своей реакцией на выбросы – наблюдения, не согласованные с выборкой в целом. Чем больше значение числа v , тем в большей степени L_v -оценка на них реагирует. В регрессионном анализе методы оценивания параметров, малочувствительные к выбросам или вообще их игнорирующие, называют робастными [17].

Методом оценивания параметров уравнения (2), соответствующим $v = 2$, является самый популярный в регрессионном анализе в силу своей простоты и хорошей интерпретируемости метод наименьших квадратов (МНК). Задание

числа v равным единице ведет к применению метода наименьших модулей (МНМ), соответствующего манхэттенскому (городскому) расстоянию между расчетными и фактическими значениями зависимой переменной в (1). При $v \rightarrow \infty$ построение L_v -оценки производится с помощью метода антиробастного оценивания (МАО) [19, 20], соответствующего расстоянию Чебышева. Отметим при этом, что имеет место соотношение [20]

$$J_{\infty}(\alpha) = \max_{k=1,n} |\varepsilon_k|.$$

При оценивании параметров регрессии (1) часто возникает проблема альтернативности. Действительно, пусть при этом используются s методов идентификации (не все из которых могут быть связаны с L_v -оценками), каждому из которых соответствует своя функция потерь I^j , $j = \overline{1, s}$. Обозначим полученные при этом векторы оценок параметров через α^j .

Иногда трудно исходя из формальных и (или) содержательных соображений выбрать из всех векторов α^j , $j = \overline{1, s}$ только один для последующей реализации. В этом случае лучше воспользоваться каким-либо обоснованным приемом, позволяющим построить некий компромисс между всеми этими векторами, который бы отчасти отражал присущие им свойства.

Результаты

Для этого воспользуемся популярным в теории принятия решений методом уступок [21], который в данном случае может быть использован следующим образом.

Пусть исследователь для достижения указанного компромисса готов пойти на некоторую уступку ΔI^j , на которую он может пойти по j -й функции потерь (т.е. на ее увеличение на величину ΔI^j).

Определим средневзвешенную оценку:

$$\bar{\alpha} = \frac{1}{s} \sum_{j=1}^s \alpha^j.$$

Тогда компромиссная оценка α^* может быть вычислена посредством решения задачи:

$$\alpha^* = \arg \min_{\alpha \in A} \rho(\alpha, \bar{\alpha}), \quad (3)$$

где

$$A = \{ \alpha \in R^m \mid I^j(\alpha) \leq I^j(\alpha^j) + \Delta I^j, j = \overline{1, s} \}. \quad (4)$$

Здесь $\rho(\alpha, \bar{\alpha})$ – расстояние между векторами α и $\bar{\alpha}$.

В общем случае задача (3), (4) представляет собой весьма сложную задачу нелинейного программирования. Однако после некоторых упрощений ее можно свести к существенно более простой в вычислительном отношении задаче линейного программирования (ЛП). Действительно, воспользуемся для

этого приемами, позволяющими свести задачу оценивания параметров уравнения (1) с помощью МНМ к задаче ЛП (см., например, работы [22–24]).

Введем в рассмотрение неотрицательные действительные переменные u_k , v_k , выражающие соответственно положительные и отрицательные части ошибок аппроксимации ϵ_k в (1):

$$u_k = \begin{cases} y_k - \sum_{i=1}^m \alpha_i x_{ki}, & y_k > \sum_{i=1}^m \alpha_i x_{ki} \\ 0, & \text{в противном случае.} \end{cases}$$

$$v_k = \begin{cases} -y_k + \sum_{i=1}^m \alpha_i x_{ki}, & y_k < \sum_{i=1}^m \alpha_i x_{ki} \\ 0, & \text{в противном случае.} \end{cases}$$

При этом для всех k очевидны следующие равенства:

$$\epsilon_k = u_k - v_k, u_k v_k = 0, |\epsilon_k| = u_k + v_k. \quad (5)$$

Тогда уравнение (1) можно представить в виде равенств

$$\sum_{i=1}^m \alpha_i x_{ki} + u_k - v_k = y_k, k = \overline{1, n}. \quad (6)$$

Наложим на u_k , v_k естественные условия неотрицательности:

$$u_k \geq 0, v_k \geq 0, k = \overline{1, n}. \quad (7)$$

Используем в качестве метрики ρ в (3) городское расстояние

$$\rho(\alpha, \bar{\alpha}) = \sum_{i=1}^m |\alpha_i - \bar{\alpha}_i|. \quad (8)$$

Ослабим задающие множество A условия, приняв в качестве базовой только функцию потерь $J_1(\alpha)$:

$$A_1 = \{ \alpha \in R^m \mid J_1(\alpha) \leq J_1(\alpha^1) + \Delta I^1 \}. \quad (9)$$

Для последующей замены в (8) модулей по аналогии с (5) введем неотрицательные переменные p_i , r_i :

$$\alpha_i + p_i - r_i = \bar{\alpha}_i, i = \overline{1, m}. \quad (10)$$

$$p_i \geq 0, r_i \geq 0, i = \overline{1, m}. \quad (11)$$

Сформируем ограничение, задающее множество A_1 в (9):

$$\sum_{k=1}^n (u_k + v_k) \leq J_1(\alpha^1) + \Delta I^1. \quad (12)$$

Тогда задача (3) примет вид задачи ЛП с целевой функцией:

$$\sum_{i=1}^m (p_i + r_i) + \delta \sum_{k=1}^n (u_k + v_k) \rightarrow \min \quad (13)$$

и ограничениями (6), (7), (10)–(12), где δ – заранее заданная малая положительная константа. Присутствие в (13) второго слагаемого вызвано

необходимостью выполнения равенств $u_k v_k = 0$ в (5), следующих из определения переменных u_k и v_k .

Обсуждение

Проанализируем полученный результат на простом численном примере. Пусть исходная выборка данных имеет вид

$$X = \begin{pmatrix} 129 \\ 423 \\ 719 \\ 914 \end{pmatrix}, y = \begin{pmatrix} 17 \\ 38 \\ 8 \\ 49 \end{pmatrix}.$$

Требуется построить линейное регрессионное уравнение

$$y_k = \alpha_1 x_{k1} + \alpha_2 x_{k2} + \varepsilon_k, k = \overline{1, 4}.$$

Будем оценивать его параметры с помощью МНМ, МНК и МАО (т.е. $s = 3$) с использованием специализированного программного комплекса [25]. В результате получим

$$\alpha^1 = (4.789, 0.421), \alpha^2 = (3.208, 0.512), \alpha^3 = (2.834, 0.352).$$

При этом $J_1(\alpha^1) = 42.684$.

Рассчитаем средневзвешенную оценку: $\bar{\alpha} = (3.610, 0.428)$.

Назначим уступку $\Delta\Gamma^1$, составляющую 10 % от $J_1(\alpha^1)$: $\Delta\Gamma^1 = 4.2684$.

После решения задачи ЛП (6), (7), (10)–(12), (13) получим компромиссную оценку: $\alpha^* = (3.996, 0.448)$.

Заключение

В работе решена проблема вычисления компромиссной оценки параметров линейной регрессионной модели, возникающая при ее построении несколькими альтернативными методами идентификации. В общем случае эта проблема сводится к весьма сложной в вычислительном отношении задаче нелинейного программирования. При некоторых упрощающих исходную постановку допущениях проблема сводится к задаче линейного программирования. Приведен простой численный пример.

Список литературы

1. Jiang X., Liu Ch., Wang Y., Wang H. The study on Sanmenxia annual flow forecasting in the Yellow River with mix regression model // Science in China Series E: Technological Sciences volume. 2004. Vol. 47. P. 118–126.
2. Dong Ch. A regression model for analysing the non-linearity of laser triangulation probes // The International Journal of Advanced Manufacturing Technology. 2012. Vol. 59. P. 691–695.
3. Hornbaker H., Dixon B., Sonka S. Estimating Production Activity Costs for Multioutput Firms with a Random Coefficient Regression Model Arrow // American Journal of Agricultural Economics. 1989. Vol. 71, № 1. P. 167–177.

4. Носков С. И., Оленцевич В. А., Базилевский М. П. Математическая модель оценки безопасности перевозочного процесса на региональном уровне // Транспортная инфраструктура Сибирского региона. 2014. Т. 1. С. 537–542.
5. Van der Meer R. B., Quigley J., Storbeck J. E. Using regression analysis to model the performance of UK Coastguard centres // Journal of the Operational Research Society. 2005. Vol. 56. P. 630–641.
6. Sohn S. Y., Yoon K. B. Dynamic preventive maintenance scheduling of the modules of fighter aircraft based on random effects regression model // Journal of the Operational Research Society. 2010. Vol. 61. P. 974–979.
7. King M. L., Galbreath D. R., Newman A. M. Combining regression and mixed-integer programming to model counterinsurgency // Annals of Operations Research. 2020. Vol. 292. P. 287–320.
8. Özmen A. Sparse regression modeling for short- and long-term natural gas demand prediction // Annals of Operations Research. 2023. Vol. 322. P. 921–946.
9. Liu H., Gopalkrishnan V., Kim Thi Nhu Quynh, Wee-Keong Ng. Regression models for estimating product life cycle cost // Journal of Intelligent Manufacturing. 2009. Vol. 20. P. 401–408.
10. García V., Sánchez S., Rodríguez-Picón L. A. [et al.]. Using regression models for predicting the product quality in a tubing extrusion process // Journal of Intelligent Manufacturing. 2019. Vol. 30. P. 2535–2544.
11. Pardoe I. Applied Regression Modeling. Wiley, 2020. 336 p.
12. Доугерти К. Введение в эконометрику. М. : ИНФРА, 2009. 465 с.
13. Montgomery D. C., Peck E. A., Vining G. G. Introduction to Linear Regression Analysis. Wiley, 2012. 672 p.
14. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия. 3-е изд. М. : Диалектика, 2007. 912 с.
15. Радченко С. Г. Методология регрессионного анализа. К. : Корнийчук, 2011. 376 с.
16. Lampis F., Díaz-emparanza I., Banerjee A. How to use setar models in GRETL // Computational Economics. 2015. Т. 46, № 2. С. 231–241.
17. Демиденко Е. З. Линейная и нелинейная регрессии. М. : Финансы и статистика, 1981. 302 с.
18. Носков С. И. L-множество в многокритериальной задаче оценивания параметров регрессионных уравнений // Информационные технологии и проблемы математического моделирования сложных систем. 2004. № 1. С. 64–71.
19. Носков С. И. Метод антиробастного оценивания параметров линейной регрессии: число максимальных по модулю ошибок аппроксимации // Южно-Сибирский научный вестник. 2020. № 1. С. 51–54.
20. Noskov S. I. Compromise pareto estimates of linear regression parameters // Mathematical Models and Computer Simulations. 2021. Vol. 13, № 4. P. 586–590.
21. Растринин Л. А. Системы экстремального управления. М. : Наука, 1974. 632 с.
22. Баенхаева А. В., Базилевский М. П., Носков С. И. Моделирование валового регионального продукта Иркутской области на основе применения методики множественного оценивания регрессионных параметров // Фундаментальные исследования. 2016. № 10-1. С. 9–14.
23. 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.
24. Носков С. И. Метод максимальной согласованности в регрессионном анализе // Известия Тульского государственного университета. Технические науки. 2021. № 10. С. 380–385.
25. Носков С. И., Перфильева К. С. Программный комплекс построения линейных регрессионных моделей методом смешанного оценивания // Южно-Сибирский научный вестник. 2021. № 3. С. 38–42.

References

1. Jiang X., Liu Ch., Wang Y., Wang H. The study on Sanmenxia annual flow forecasting in the Yellow River with mix regression model. *Science in China Series E: Technological Sciences volume*. 2004;47:118–126.
2. Dong Ch. A regression model for analysing the non-linearity of laser triangulation probes. *The International Journal of Advanced Manufacturing Technology*. 2012;59:691–695.
3. Hornbaker N., Dixon B., Sonka S. Estimating Production Activity Costs for Multioutput Firms with a Random Coefficient Regression Model Arrow. *American Journal of Agricultural Economics*. 1989;71(1):167–177.
4. Noskov S.I., Olentsevich V.A., Bazilevskiy M.P. A mathematical model for assessing the safety of the transportation process at the regional level. *Transportnaya infrastruktura Sibirskogo regiona = Transport infrastructure of the Siberian region*. 2014;1: 537–542. (In Russ.)
5. Van der Meer R.B., Quigley J., Storbeck J.E. Using regression analysis to model the performance of UK Coastguard centres. *Journal of the Operational Research Society*. 2005;56:630–641.
6. Sohn S.Y., Yoon K.B. Dynamic preventive maintenance scheduling of the modules of fighter aircraft based on random effects regression model. *Journal of the Operational Research Society*. 2010;61:974–979.
7. King M.L., Galbreath D.R., Newman A.M. Combining regression and mixed-integer programming to model counterinsurgency. *Annals of Operations Research*. 2020;292:287–320.
8. Özmen A. Sparse regression modeling for short- and long-term natural gas demand prediction. *Annals of Operations Research*. 2023;322:921–946.
9. Liu N., Gopalkrishnan V., Kim Thi Nhu Quynh, Wee-Keong Ng. Regression models for estimating product life cycle cost. *Journal of Intelligent Manufacturing*. 2009;20:401–408.
10. García V., Sánchez S., Rodríguez-Picón L. A. et al. Using regression models for predicting the product quality in a tubing extrusion process. *Journal of Intelligent Manufacturing*. 2019;30:2535–2544.
11. Pardoe I. *Applied Regression Modeling*. Wiley, 2020:336.
12. Dougerti K. *Vvedenie v ekonometriku = Introduction to econometrics*. Moscow: INFRA, 2009:465. (In Russ.)
13. Montgomery D.C., Peck E.A., Vining G.G. *Introduction to Linear Regression Analysis*. Wiley, 2012:672.
14. Dreyper N., Smit G. *Prikladnoy regressionnyy analiz. Mnozhestvennaya regressiya. 3-e izd. = Applied regression analysis. Multiple regression. 3rd ed.* Moscow: Dialektika, 2007:912. (In Russ.)
15. Radchenko S.G. *Metodologiya regressionnogo analiza = Methodology of regression analysis*. Kyiv: Korniychuk, 2011:376.
16. Lampis F., Díaz-emparanza I., Banerjee A. How to use setar models in GRETL. *Computational Economics*. 2015;46(2):231–241.
17. Demidenko E.Z. *Lineynaya i nelineynaya regressii = Linear and nonlinear regressions*. Moscow: Finansy i statistika, 1981:302. (In Russ.)
18. Noskov S.I. L-set in a multicriteria problem of estimating parameters of regression equations. *Informatsionnye tekhnologii i problemy matematicheskogo modelirovaniya slozhnykh system = Information technologies and problems of mathematical modeling of complex systems*. 2004;(1):64–71. (In Russ.)
19. Noskov S.I. The method of anti-robust estimation of linear regression parameters: the number of maximum modulo approximation errors. *Yuzhno-Sibirskiy nauchnyy vestnik = South Siberian Scientific Bulletin*. 2020;(1):51–54. (In Russ.)

20. Noskov S.I. Compromise pareto estimates of linear regression parameters. *Mathematical Models and Computer Simulations*. 2021;13(4):586–590.
21. Rastrigin L.A. *Sistemy ekstremal'nogo upravleniya = Extreme control systems*. Moscow: Nauka, 1974:632. (In Russ.)
22. Baenkhaeva A.V., Bazilevskiy M.P., Noskov S.I. Modeling of the gross regional product of the Irkutsk region based on the application of the methodology of multiple estimation of regression parameters. *Fundamental'nye issledovaniya = Fundamental research*. 2016;(10-1):9–14. (In Russ.)
23. Noskov S.I. Method of mixed estimation of linear regression parameters: application features. *Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyy analiz i informatsionnye tekhnologii = Bulletin of the Voronezh State University. Series: System analysis and Information Technology*. 2021;(1):126–132. (In Russ.)
24. Noskov S.I. Method of maximum consistency in regression analysis. *Izvestiya Tul'skogo gosudarstvennogo universiteta. Tekhnicheskie nauki = Proceedings of Tula State University. Technical sciences*. 2021;(10):380–385. (In Russ.)
25. Noskov S.I., Perfil'eva K.S. Software package for constructing linear regression models by the method of mixed estimation. *Yuzhno-Sibirskiy nauchnyy vestnik = South Siberian Scientific Bulletin*. 2021;(3):38–42. (In Russ.)

Информация об авторах / Information about the authors

Сергей Иванович Носков

доктор технических наук, профессор,
профессор кафедры информационных
систем и защиты информации,
Иркутский государственный
университет путей сообщения
(Россия, г. Иркутск, ул. Чернышевского, 15)
E-mail: sergey.noskov.57@mail.ru

Sergey I. Noskov

Doctor of technical sciences, professor,
professor of the sub-department
of information systems
and information protection,
Irkutsk State Transport University
(15 Chernyshevskogo street, Irkutsk, Russia)

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

Поступила в редакцию/Received 21.05.2023

Поступила после рецензирования/Revised 14.09.2023

Принята к публикации/Accepted 05.10.2023