УДК 519.852 ISSN 1995-5499
DOI: https://doi.org/10.17308/sait/1995-5499/2022/3/15-21
Поступила в редакцию 14.06.2022
Подписана в печать 30.09.2022

КУСОЧНО-ЛИНЕЙНАЯ СВЕРТКА ВАРИАНТОВ РЕГРЕССИОННОЙ МОДЕЛИ ОБЪЕКТА

© 2022 С. И. Носков✉

Иркутский государственный университет путей сообщения
ул. Чернышевского, 15, 664074 Иркутск, Российская Федерация

Аннотация. В работе дан краткий обзор результатов по применению методов регрессионного анализа при исследовании сложных систем: мощности ветряных турбин в Бразилии и прогнозировании скорости ветра, оценке теплотворной способности брикетов биомассы при их использовании в качестве эффективного горючего топлива, изучении проблемы утилизации растительных остатков на севере Китая, анализе развития малого и среднего предпринимательства в Республике Казахстан, выработки необходимых мер для стимулирования роста производства пшеницы в Южной Африке. Рассмотрен случай, когда по разным причинам при построении регрессионной модели сложного объекта построено несколько ее альтернативных вариантов, каждый из которых приемлем как по знакам оцениваемых параметров, так и по значениям критериев адекватности. К числу этих причин могут, в частности, относиться: использование различных видов аппроксимирующей функции, применение нескольких способов идентификации модельных параметров, варьирование набора независимых переменных, в том числе путем их преобразований — обратных, возведения в степень, мультипликативных, логарифмических, экспоненциальных, тригонометрических, логистических и т. д. Предложено правило выбора варианта модели из нескольких альтернативных, формализованное путем разработки алгоритма построения кусочно-линейной свертка этих вариантов в виде функции риска, задача оценивания параметров которой сведена к задаче линейно-булевого программирования. С помощью данного алгоритма построена кусочно-линейная сверка трех вариантов регрессионной модели грузооборота Красноярской железной дороги — одной линейной и двух линейно-мультипликативных. В качестве зависимой переменной принят грузооборот дороги, независимыми же переменными являются прием груженых вагонов, прием порожних вагонов, динамическая нагрузка, передача по стыкам поездов.

Ключевые слова: регрессионная модель, кусочно-линейная свертка, линейно-булевое программирование, функция риска, критерии адекватности.

ВЕДЕНИЕ

Применение методов регрессионного анализа для исследования сложных систем предполагает использование самых различных подходов и приемов в зависимости от свойств изучаемых объектов. В работе [1] для моделирования мощности ветряных турбин в Бразилии и для прогнозирования скорости ветра используется логистическая регрессия. Полученные при этом результаты служит инструментом для решения вопросов продажи энергии и планирования обслуживания турбин. В [2] регрессионный анализ применяется для оценки теплотворной способности брикетов биомассы при их использовании в качестве эффективного горючего топлива. В статье [3] логистические модели и модели множественной линейной регрессии положены в основу изучения проблемы утилизации растительных остатков на севере Китая, которая стала существенным препятствием для устойчивого развития овощеводства в регионе. Эмпирические результаты показали, что политика субсидирования, время подкорм-

✉ Носков Сергей Иванович
e-mail: sergey.noskov.57@mail.ru

ки, возраст, масштабы, инвестиции в орошение и чистый доход предприятий оказывают значительное положительное влияние на эффективность использования компоста. При этом государственные субсидии являются необходимой предпосылкой для реализации программы компостирования. Работа [4] посвящена анализу развития малого и среднего предпринимательства (МСП) в Республике Казахстан, которое является эффективным рычагом обеспечения экономического роста страны. Для более полного представления о влиянии малых и средних предприятий на экономику Республики Казахстан была построена регрессионная модель, описывающая влияние количества действующих МСП и объемов кредитов, выдаваемых им банками, на валовый внутренний продукт страны. В [5] регрессионные модели применены для выработки необходимых мер для стимулирования роста производства пшеницы в Южной Африке путем смягчения воздействия ценовых шоков, исходящих от мировых рынков пшеницы. В работе [6] описана двухфакторная модель полносвязной линейной регрессии динамики валового внутреннего продукта России.

ПОСТАНОВКА ЗАДАЧИ

Пусть при анализе некоторого сложного объекта произвольной природы исследователь с помощью методов регрессионного анализа изучает характер влияния некоторого набора независимых (входных) переменных (факторов, показателей) $x_1,\ x_2,\dots,\ x_m$ на зависимую (выходную) переменную y. Это влияние может быть формализовано посредством построения в общем случае P альтернативных вариантов регрессионной модели (уравнения) вида:

$$y_k = F^j(\alpha^j, x_k) + \varepsilon_k,\ k=\overline{1,n},\ j=\overline{1,P}. \quad (1)$$

Здесь $x_k = (x_{k1}, x_{k2}, \dots, x_{km})$, k — номер наблюдения, n — их количество (длина выборки), F^j — j-я аппроксимирующая функция, α^j — вектор оцениваемых параметров, ε_k — ошибки аппроксимации. Будем вести изложение в рамках так называемого аппроксимационного (логико-алгебраического) подхода к анализу данных (см., например, [7]), не предполагающего наличия у ошибок ε_k каких-либо вероятностных свойств, а трактующего их как погрешности модельного приближения и только.

Указанная альтернативность при описании влияния переменных $x_1,\dots,x_m$ на y может объясняться многими причинами.

Во-первых, функции F^j могут иметь различный вид. Так, если при моделировании экономического объекта переменная y представляет собой выпуск продукции, а переменные x_i, $i=\overline{1,m}$ — ресурсные показатели, то аппроксимирующая связь в модели (1) трактуется как производственная функция (ПФ). В эконометрике известно большое количество наиболее часто используемых видов ПФ. Для динамических рядов наблюдений показателей (где k означает момент времени) это, в частности:

– линейная ПФ

$$y_k = \alpha_0 + \sum_{i=1}^{m} \alpha_i x_{ki},$$

– ПФ Кобба — Дугласа

$$y_k = a_0 \prod_{i=1}^{m} x_{ki}^{\ a_i},$$

– ПФ Тинбергена

$$y_k = a_0 e^k \prod_{i=1}^{m} x_{ki}^{\ a_i},$$

– ПФ функция с постоянной эластичностью замещения

$$y_k = \left(\sum_{i=1}^{m} a_i x_{ki}^{-\rho} \right)^{-\gamma/\rho},$$

– многорежимная ПФ с различными параметрами крутизны

$$y_k = \left(\sum_{i=1}^{m} a_i x_{ki}^{-\rho_1} \right)^{-\gamma_1/\rho_1} \left(\sum_{i=1}^{m} \beta_i x_{ki}^{-\rho_2} \right)^{-\gamma_2/\rho_2},$$

– ПФ Солоу

$$y_k = \left(\sum_{i=1}^{m} a_i x_{ki}^{\beta_i} \right)^{\gamma},$$

– ПФ Аллена

$$y_k = \sum_{i>j} a_{ij} x_{ki} x_{kj} - \sum_{s=1}^{m} \beta_s x_{ks}^2,$$

– ПФ Сато

$$y_k = a_0 \prod_{i=1}^{m} x_{ki}^{a_i} \left(\sum_{j=1}^{m} \beta_j x_{kj}^{-\rho} \right)^{-\gamma/\rho},$$

– ПФ Кокса — Бокса

$$\frac{y_k^{\lambda} - 1}{\lambda} = a_0 + \sum_{i=1}^{m} a_i \frac{x_{ki}^{\lambda} - 1}{\lambda},$$

– ПФ с постоянными пропорциями

$$y_k = \min\{\alpha_1 x_{k1}, \alpha_2 x_{k2}, ..., \alpha_m x_{km}\}.$$

Каждая из этих ПФ обладает своими весьма информативными характерными свойствами.

Во-вторых, множественность представлений (1) может достигаться значительным арсеналом разработанных в рамках современного анализа данных методов оценивания векторов параметров α^j, $j = \overline{1, P}$. Например, в размещенном в Internet в свободном доступе прикладном программном пакете GRETL для регрессионного анализа [8] реализовано порядка тридцати таких методов.

Наконец, в-третьих, при построении альтернативных вариантов модели (1) исследователь может для разных функций F^j использовать разные поднаборы независимых переменных в рамках исходного набора $x_1, x_2, \ldots, x_m$, а также расширять его путем введения широкого круга их преобразований — обратных, возведения в степень, мультипликативных, логарифмических, экспоненциальных, тригонометрических, логистических и т. д.

Задача состоит в построении некоего формального правила, позволяющего для каждой конкретной ситуации выбирать один и только вариант из P возможных и не выражающегося в каком-либо комбинировании, включая линейное, этих вариантов.

ПОСТРОЕНИЕ ФУНКЦИИ РИСКА ДЛЯ ВЫБОРА ВАРИАНТА МОДЕЛИ (1)

Введем в рассмотрение новые переменные, которые являются расчетными значениями зависимой переменной для каждого варианта модели (1):

$$y_k^j = F^j(\alpha^j, x_k), \; k = \overline{1, n}, \; j = \overline{1, P}.$$

Тогда указанное выше правило может заключаться в построении кусочно-линейной свертки всех построенных вариантов, имеющей, например, вид функции риска [9, 10]:

$$y_k = \max\{\beta_1 y_k^1, ..., \beta_P y_k^P\} + \delta_k, \; k = \overline{1, n}, \quad (2)$$

где δ_k — ошибки аппроксимации кусочно-линейного представления (2).

Построение функции риска (2) предполагает оценивание параметров β_i, $i = \overline{1, P}$ посредством минимизации выбранной функции потерь, например, соответствующей методу наименьших модулей (МНМ) [9]:

$$\sum_{k=1}^{n} |\delta_k| \to \min. \quad (3)$$

Для решения задачи (3) воспользуемся алгоритмом [9], позволяющим свести ее к задаче линейно-булевого программирования (ЛБП).

Введем в рассмотрение переменные z_k по правилу:

$$z_k = \max\{\beta_1 y_k^1, ..., \beta_P y_k^P\}, k = \overline{1, n}.$$

Введем также переменные u_k и v_k следующим образом:

$$u_k = \begin{cases} y_k - z_k, \; y_k > z_k \\ 0, \; \textit{в пр. случае} \end{cases},$$

$$v_k = \begin{cases} -y_k + z_k, \; y_k < z_k \\ 0, \; \textit{в пр. случае} \end{cases}.$$

Тогда равенства (2) представимы в виде:

$$z_k + u_k - v_k = y_k, \; k = \overline{1, n}, \quad (4)$$

а из определения z_k следуют неравенства

$$z_k \geq \beta_i y_k^i, \;\; k = \overline{1, n}, \;\; i = \overline{1, P}, \quad (5)$$

и при этом по крайней мере одно из них должно обращаться в равенство. Для выполнения этого условия введем nP булевых переменных

$$\sigma_{ki} \in \{0,1\}, \; k = \overline{1, n}, \; i = \overline{1, P} \quad (6)$$

и сформируем ограничения

$$\beta_i y_k^i - z_k \geq (\sigma_{ki} - 1)M, k = \overline{1, n}, \; i = \overline{1, P}, \quad (7)$$

$$\sum_{i=1}^{P} \sigma_{ki} = 1, \; k = \overline{1, n}, \quad (8)$$

где M — заранее назначенное большое положительное число.

Тогда задача (3) сводится к задаче ЛБП с ограничениями (4)–(8) и целевой функцией

$$\sum_{k=1}^{n} (u_k + v_k) \to \min. \quad (9)$$

Задача ЛБП (4)–(9) содержит $Pn+3n+P$ переменных (из которых Pn — булевы) и $2(Pn+n)$ ограничений.

МОДЕЛИРОВАНИЕ ГРУЗООБОРОТА ЖЕЛЕЗНОДОРОЖНОГО ТРАНСПОРТА

Проиллюстрируем реализацию предлагаемого подхода к выбору варианта регрессионной модели сложного объекта на проблеме построения регрессионной модели грузооборота Красноярской железной дороги. Введем в рассмотрение переменные:

y — грузооборот (млн. т.км);

x_1 — прием груженых вагонов (ваг.);

x_2 — прием порожних вагонов (ваг.);

x_3 — динамическая нагрузка (т.км/км);

x_4 — передача по стыкам (поездов).

В качестве информационной базы используем статистическую информацию по указанным показателям за 2001–2019 гг. [11].

Построим три варианта регрессионной модели грузооборота Красноярской железной дороги — линейный с полным набором независимых переменных и два линейно-мультипликативных [12]. Адекватность каждого из них будем оценивать с помощью известных в регрессионном анализе критериев: коэффициента множественной детерминации R, критерия Фишера F, средней относительной ошибки аппроксимации E.

1. Линейная модель.

$$y^1 = -24856.1 + 11.89x_1 + 9.58x_2 + 358.22x_3 + 142.73x_4,$$

$$R = 0.994,\ F = 573,\ E = 1.65.$$

2. Линейно-мультипликативная модель с переменными $x_1,\ x_2$.

$$y^2 = 18494.7 + 9.08x_1 + 0.00249x_1x_2, \quad R = 0.998,\ F = 4121,\ E = 1.06. \qquad (11)$$

3.Линейно-мультипликативная модель с переменными $x_1,\ x_2,\ x_4$.

$$y^3 = 13923.4 + 11.8x_1 + 0.0744x_2x_4, \quad R = 0.997,\ F = 3334,\ E = 1.17. \qquad (12)$$

Следует обратить внимание, что знаки параметров всех трех моделей соответствуют содержательному смыслу входящих в их состав независимых переменных — с ростом значений каждой из последних значение зависимой переменной возрастает. Отметим также высокую адекватность моделей, на что указывают значения всех трех используемых критериев.

В табл. 1 представлены фактические и расчетные (по каждой из моделей (10)–(12)) значения зависимой переменной.

Таблица 1. Фактические и расчетные значения зависимой переменной
[Table 1. Actual and calculated values of the dependent variable]

y	y^1	y^2	y^3
54558	52821.62	53421.71	53912.33
59462	57700.86	58839.86	57910.36
58280	58729.77	58938.84	58953.63
64041	63046.62	64117.2	64141.32
67865	68504.14	69362.08	69386.17
68816	68501.82	69419.35	69479.27
73086	71471.91	71634.71	71764.59
78777	79144.51	79066.5	78978.56
85347	84762.92	84265.93	84048.63
79232	82916.82	80763.32	80878.61
85194	87683.19	86300.58	86305.57
91580	90862.6	89793.78	89650.9
95099	95566.72	94187.54	94463.4
97070	98820.46	97566.77	97676.33
104363	106133.5	105871.4	105779.5
109552	111194.5	109728.7	110221.9
118366	118184.8	117938.1	118451.3
124450	123173.7	124471	124751.2
132700	128617.5	132150.7	131084.4

Из табл. 1 следует близость всех компонент векторов $y,\ y^1,\ y^2,\ y^3$, несколько различающаяся для разных наблюдений.

Построенная с помощью решения задачи ЛБП (4)–(9) кусочно-линейная свертка вариантов модели (10)–(12) имеет вид

$$y_k = \max\{0.9731y^1, 0.9984y^2, 0,9976y^3\} + \delta_k, \qquad (13)$$

$$k = \overline{1,19},\quad E = 0.99\ \%.$$

Вызывает интерес то, какая из моделей «сработала» на каждом из 19 наблюдений выборки. С этой целью рассчитаем так называемый вектор срабатываний [10] $\lambda = (\lambda_1, ..., \lambda_n)$, компоненты которого рассчитаны по правилу: если $\lambda_k = s$, то $\max\left\{\beta_1 y_k^1, \beta_2 y_k^2, \beta_3 y_k^3\right\} = \beta_s y_k^s$.

По отношению к данной задаче вектор срабатываний имеет вид:

$$\lambda = (3,\ 2,\ 2,\ 2,\ 2,\ 3,\ 3,\ 2,\ 2,\ 1,$$
$$2,\ 2,\ 3,\ 3,\ 2,\ 3,\ 3,\ 3,\ 2).$$

Таким образом, на 10 из 19 наблюдениях в соответствии со сверткой (13) «сработала» модель (11), в 8 — модель (12) и лишь в одном — модель (10).

При использовании свертки (3) в режиме прогнозирования следует в нее вместо y^1, y^2, y^3 подставить соответствующие им представления (10)–(12).

ЗАКЛЮЧЕНИЕ

В работе рассмотрен случай, когда по разным причинам при построении регрессионной модели сложного объекта построено несколько ее альтернативных вариантов, каждый из которых приемлем как по знакам оцениваемых параметров, так и по значениям критериев адекватности. Предложен алгоритм построения кусочно-линейной свертки этих вариантов в виде функции риска, задача оценивания параметров которой сведена к задаче линейно-булевого программирования. Алгоритм апробирован на решении залачи построения регрессионной модели грузооборота Красноярской железной дороги.

КОНФЛИКТ ИНТЕРЕСОВ

Автор декларирует отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Reboucas Filho P. P., De Medeiros Mendonca E Nascimento N., Araujo Alves S.S., Luz Gomes S., Marques De Sa Medeiros C.* Estimation of the energy production in a wind farm using regression methods and wind speed forecast // Brazilian Conference on Intelligent Systems, BRACIS 2018. 8575593. – P. 79–84.

2. *Deshannavar U. B., Hegde P. G., Dhalayat Z., Patil V., Gavas S.* Production and characterization of agro-based briquettes and estimation of calorific value by regression analysis: An energy application // Materials Science for Energy Technologies. – 2018. – № 1(2). – P. 175–181.

3. *Zhou Y., Zhou Q., Gan S., Wang L.* Factors affecting farmers' willingness to pay for adopting vegetable residue compost in North China // Acta Ecologica Sinica. – № 6(38). – P. 401–411.

4. *Beisengaliyev B., Khishauyeva Z., Lesbayeva G., Tasbulatova D., Turekulova D.* Impact of small and medium enterprises on the economy // Journal of Applied Economic Sciences. – 2018. – № 8(13). – P. 2437–2445.

5. *Motengwe C., Pardo A.* Major International Information Flows Across the Safex Wheat Market // South African Journal of Economics. – 2016. – № 4(84). – P. 636–653.

6. *Базилевский М. П.* Двухфакторная модель полносвязной линейной регрессии динамики ВВП России // Математическое и компьютерное моделирование в экономике, страховании и управлении рисками. – 2019. – № 4. – С. 8–12.

7. *Айвазян С. А., Енюков И. С, Мешалкий Л. Д.* Прикладная статистика. Основы моделирования и первичная обработка данных. – М. : Финансы и статистика. – 1983. – 472 с.

8. https://soft.mydiv.net/win/download-gretl.html.

9. *Носков С. И.* Идентификация параметров кусочно-линейной функции риска // Транспортная инфраструктура Сибирского региона. – 2017. – Т. 1. – С. 417–421.

10. *Носков С. И., Хоняков А. А.* Применение функции риска для модельного описания колебания цен на рынке недвижимости // Инженерно-строительный вестник Прикаспия. – 2021. – № 3 (37). – С. 77–82.

11. *Носков С. И., Врублевский И. П.* Анализ регрессионной модели грузооборота железнодорожного транспорта // Вестник транспорта Поволжья. – 2020. – № 1 (79). – С. 86–90.

12. *Базилевский М. П., Носков С. И.* Алгоритм построения линейно-мультипликативной регрессии // Современные технологии. Системный анализ. Моделирование. – 2011. – № 1. – С. 88–92.

Носков Сергей Иванович — д-р техн. наук, проф., профессор кафедры «Информационные системы и защита информации» Иркутского государственного университета путей сообщения.
E-mail: sergey.noskov.57@mail.ru
ORCID iD: https://orcid.org/0000-0003-4097-2720

DOI: https://doi.org/ 10.17308/sait/1995-5499/2022/3/15-21 ISSN 1995-5499
Received 14.06.2022
Accepted 30.09.2022

PIECE-LINEAR CONVOLUTION OF OPTIONS OF THE OBJECT REGRESSION MODEL

© 2022 S. I. Noskov[✉]

Irkutsk State Transport University
15, Chernyshevskogo Street, 664074 Irkutsk, Russian Federation

Annotation. The paper provides a brief overview of the results on the application of regression analysis methods in the study of complex systems: the power of wind turbines in Brazil and wind speed forecasting, the assessment of the calorific value of biomass briquettes when they are used as an efficient combustible fuel, the study of the problem of utilization of plant residues in northern China, the analysis development of small and medium-sized businesses in the Republic of Kazakhstan, development of necessary measures to stimulate the growth of wheat production in South Africa. The case is considered when, for various reasons, when constructing a regression model of a complex object, several of its alternative variants are constructed, each of which is acceptable both in terms of the signs of the estimated parameters and in terms of the values of the adequacy criteria. These reasons may, in particular, include: the use of various types of approximating functions, the use of several methods for identifying model parameters, varying the set of independent variables, including through their transformations - inverse, exponentiation, multiplicative, logarithmic, exponential, trigonometric, logistics, etc. A rule is proposed for choosing a model variant from several alternative ones, formalized by developing an algorithm for constructing a piecewise linear convolution of these variants in the form of a risk function, the problem of estimating the parameters of which is reduced to a linear Boolean programming problem. With the help of this algorithm, a piecewise-linear reconciliation of three variants of the regression model of the cargo turnover of the Krasnoyarsk railway is built - one linear and two linear-multiplicative. As a dependent variable, the freight turnover of the road is taken, while the independent variables are the reception of loaded cars, the reception of empty cars, the dynamic load, and the transfer at the joints of trains.
Keywords: regression model, piecewise linear convolution, linear Boolean programming, risk function, adequacy criteria.

CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

✉ Носков Сергей Иванович
e-mail: sergey.noskov.57@mail.ru

REFERENCES

1. *Reboucas Filho P. P., De Medeiros Mendonca E Nascimento N., Araujo Alves S. S., Luz Gomes S. and Marques De Sa Medeiros C.* (2018) Estimation of the energy production in a wind farm using regression methods and wind speed forecast. *Brazilian Conference on Intelligent Systems, BRACIS 2018*. 8575593. P. 79–84.

2. *Deshannavar U. B., Hegde P. G., Dhalayat Z., Patil V. and Gavas S.* (2018) Production and char-

acterization of agro-based briquettes and estimation of calorific value by regression analysis: An energy application. *Materials Science for Energy Technologies*. No 1(2). P. 175–181.

3. *Zhou Y., Zhou Q., Gan S. and Wang L.* Factors affecting farmers' willingness to pay for adopting vegetable residue compost in North China. *Acta Ecologica Sinica.* No 6(38). P. 401–411.

4. *Beisengaliyev B., Khishauyeva Z., Lesbayeva G., Tasbulatova D. and Turekulova D.* (2018) Impact of small and medium enterprises on the economy. *Journal of Applied Economic Sciences*. No 8(13). P. 2437–2445.

5. *Motengwe C. and Pardo A.* (2016) Major International Information Flows Across the Safex Wheat Market. *South African Journal of Economics*. No 4(84). P. 636–653.

6. *Bazilevsky M. P.* (2019) Two-factor model of a fully connected linear regression of the Russian GDP dynamics. *Mathematical and computer modeling in economics, insurance and risk management*. No 4. P. 8–12.

7. *Aivazyan S. A., Enyukov I. S. and Meshalky L. D.* (1983) Applied statistics. Fundamentals of modeling and primary data processing. *Moscow : Finance and statistics*. 472 p.

8. https://soft.mydiv.net/win/download-gretl.html.

9. *Noskov S. I.* (2017) Identification of the parameters of the piecewise linear risk function. *Transport infrastructure of the Siberian region*. Vol. 1. P. 417–421.

10. *Noskov S. I. and Khonyakov A. A.* (2021) Application of the risk function for a model description of price fluctuations in the real estate market. No 3 (37). P. 77–82.

11. *Noskov S. I. and Vrublevsky I. P.* (2021) Analysis of the regression model of freight turnover of railway transport. *Bulletin of Transport of the Volga Region*. No 1 (79). P. 86–90.

12. *Bazilevsky M. P. and Noskov S. I.* (2011) Algorithm for constructing linear-multiplicative regression. *Modern technologies. System analysis. Modeling*. No 1. P. 88–92.

Noskov Sergey Ivanovich — Doctor of Technical Sciences, Professor, Professor of the Department of Information Systems and Information Security, Irkutsk State University of Railways.
E-mail: sergey.noskov.57@mail.ru
ORCID iD: https://orcid.org/0000-0003-4097-2720