Keywords: equations of mixed-composite type, Tricomi problem, fractional integration operator, difference operator.

References

1. Samko S.G., Kilbas A.A., Marichev O.I. (1987) Integraly` i proizvodny`e drobnogo poriadka i nekotory`e ikh prilozheniia [Integrals and derivatives of a fractional order and some of their applications]. Minsk: Science and technology.
2. Gradstein I.S., Ryzhik I.M. (1971) Tablitcy` integralov, summ, riadov i proizvedenii` [Tables of integrals, sums, series and products]. Moscow: Science.
3. Kilbas A.A., Srivastava H.M., Trujillo J.J. (2005) Theory and Applications of Fractional Differential Equations. Amsterdam-Tokyo: Elsevier.
4. Agranovich M.S. (2008) Obobshchenny`e funktcii [The generalized functions]. M.: MTsIMO.

УДК 519.862.6

МУЛЬТИПЛИКАТИВНЫЙ КРИТЕРИЙ ДЕТЕРМИНАЦИИ–АВТОКОРРЕЛЯЦИИ В РЕГРЕССИОННОМ АНАЛИЗЕ

Михаил Павлович Базилевский
к.т.н., доцент
mik2178@yandex.ru
г. Иркутск

Иркутский государственный университет путей сообщения

Аннотация. При построении регрессионной модели одной из главных проблем является выбор её структурной спецификации. Поскольку каждая регрессия может быть охарактеризована множеством различных критериев адекватности, то такая задача часто оказывается многокритериальной. Вопросами решения многокритериальных задач занимается теория принятия решений. Актуальной задачей теории принятия решений является построение свёртки локальных критериев. В работе предложены две формы двухкритериального мультипликативного критерия детерминации – автокорреляции, одновременно характеризующего точность регрессии и автокорреляцию в её остатках. Для демонстрации потенциала предложенных критериев решена задача моделирования грузооборота Красноярской железной дороги. При этом выбор спецификации регрессионной модели осуществлялся как по мультипликативным критериям детерминации – автокорреляции, так и по их аддитивному аналогу, а также по методу «идеальной» точки. Достоинством одного из предложенных критериев является возможность его интеграции в виде целевой функции в задачу отбора информативных регрессоров.

Ключевые слова: регрессионная модель, критерий детерминации, критерий Дарбина-Уотсона, мультипликативный критерий детерминации-автокорреляции, грузооборот, теория принятия решений.

В рамках регрессионного анализа разработана емкая система критериев оценки адекватности регрессионных моделей [5]. Наиболее распространенными из этих критериев являются: критерий детерминации, Фишера, Дарбина-Уотсона, смещения, согласованности поведения, Акаике, Шварца и др. Они позволяют оценивать самые различные качественные стороны модельного описания исследуемого объекта или процесса. С помощью критериев адекватности осуществляется выбор структурной спецификации регрессионной модели, под которой понимается связь между переменными в математической форме. При этом такой выбор можно проводить как по одному критерию адекватности, так и решать многокритериальную задачу. В последнем случае целесообразно реализовывать алгоритмическую схему «конкурса» регрессионных моделей [3, 4, 5]. В теории принятия решений актуальной задачей является построение свертки локальных критериев (глобального критерия). Зачастую такая свертка представляет собой линейную комбинацию локальных критериев с заданными коэффициентами, т.е. каждый из этих критериев входит в свёртку аддитивно. Целью данной работы является исследование мультипликативного критерия оценки адекватности регрессионных моделей, представляющего из себя произведение двух локальных критериев: критерия детерминации и Дарбина-Уотсона. Рассмотрим линейную регрессионную модель:

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

где n – объем выборки; m – количество независимых переменных; y_i и x_{ij}, $i = \overline{1,n}$, $j = \overline{1,m}$ – известные значения зависимой и независимых переменных соответственно; α_i, $i = \overline{0,m}$ – подлежащие оцениванию параметры; ε_i, $i = \overline{1,n}$ – ошибки аппроксимации. Будем считать, что оценки параметров линейной регрессии (1) находятся с помощью метода наименьших квадратов (МНК). Тогда её адекватность можно охарактеризовать с помощью критерия детерминации R^2 и Дарбина-Уотсона DW.

Критерий детерминации R^2 находится по формуле:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{2}$$

где $\hat{y}_i$, $i = \overline{1,n}$ – вычисленные значения зависимой переменной y; $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – среднее значение y; $e_i = y_i - \hat{y}_i$, $i = \overline{1,n}$ – остатки модели. Критерий детерминации R^2 показывает, какой процент значимых факторов учтен в уравнении регрессии. В силу своего определения имеет место включение $R^2 \in [0,1]$. Если $R^2 = 0$, то это значит, что переменные x_j, $j = \overline{1,m}$, не улучшают качество предсказания y по сравнению с тривиальным предсказанием $\hat{y}_i = \bar{y}$. Если же $R^2 = 1$, то все остатки равны нулю. Чем ближе значение R^2 к 1, тем выше качество регрессии. Модели с коэффициентом детерминации выше 0.8 можно считать приемлемыми.

Дробь в выражении (2) будем называть обратным критерием детерминации:

$$\overline{R^2} = 1 - R^2 = \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}, \tag{3}$$

область значений которого $\overline{R^2} \in [0,1]$, наилучшее значение равно 0, а наихудшее – 1.

Критерий Дарбина-Уотсона DW находится по формуле:

$$DW = \frac{\sum_{i=2}^{n}\left(e_i - e_{i-1}\right)^2}{\sum_{i=1}^{n} e_i^2}. \tag{4}$$

Критерий Дарбина-Уотсона указывает на наличие или отсутствие автокорреляции (положительной или отрицательной) остатков e_i. Критерий DW принимает значения на отрезке [0,4]. Идеальное его значение, указывающее на полное отсутствие автокорреляции остатков, равно двум. Для того чтобы придать критерию Дарбина-Уотсона однородный по отношению к критерию $\overline{R^2}$ вид, преобразуем его следующим образом:

$$DW^* = \frac{1}{2}\left|2 - DW\right| = \frac{1}{2}\left|2 - \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}\right|. \tag{5}$$

Область значений критерия $DW^* \in [0,1]$, наилучшее значение равно 0, а наихудшее – 1. Отметим, что этот критерий уже не позволяет идентифицировать знак автокорреляции остатков. Используя формулы (3) и (5), рассмотрим аддитивный критерий детерминации – автокорреляции [1]:

$$DA_{add} = w_1 \overline{R^2} + w_2 DW^*, \tag{6}$$

где w_1 и w_2 – «веса» локальных критериев. Если приоритеты по локальным критериям $\overline{R^2}$ и DW^* отсутствуют, то $w_1 = w_2 = 0{,}5$. В этом случае область значений критерия $DA_{add} \in [0,1]$, наилучшее значение равно 0, а наихудшее – 1.

Введем мультипликативный критерий детерминации – автокорреляции:

$$DA_{mult} = \overline{R^2} \cdot DW^* = \frac{1}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\left|\sum_{i=1}^{n} e_i^2 - \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{2}\right|. \tag{7}$$

Область значений критерия $DA_{mult} \in [0,1]$, наилучшее значение равно 0, а наихудшее – 1. Можно ввести еще одну форму мультипликативного критерия детерминации – автокорреляции:

$$DA^*_{mult} = R^2 \cdot \left(1 - DW^*\right) = \left(1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right) \cdot \left(1 - \left|1 - \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{2\sum_{i=1}^{n} e_i^2}\right|\right). \tag{8}$$

Область значений критерия $DA^*_{mult} \in [0,1]$, наилучшее значение равно 1, а наихудшее – 0. Для демонстрации свойств предложенных критериев адекватности решалась задача моделирования грузооборота Красноярской железной дороги (КЖД). Для этого была использована статистическая информация из работы [2] за период с 2000 г. по 2015 г. по следующим показателям КЖД: y – грузооборот, млн. т. км; x_1 – прием порожних вагонов, штук; x_2 – динамическая нагрузка, т. км / км; x_3 – среднесуточный пробег локомотива, км; x_4 – эксплуатируемый парк локомотивов, штук; x_5 – техническая скорость локомотивов, км / час. По исходным статистическим данным были построены все возможные двухфакторные линейные регрессионные модели, общее количество которых равно 10. Для каждой регрессии определялись критерии R^2, DW, DA_{add}, DA_{mult} и DW^*_{mult}. Результаты представлены в таблице 1.

Таблица 1

Результаты моделирования

Модель	Регрессоры	R^2	DW	DA_{add}	DA_{mult}	DW^*_{mult}
1	1, 2	0,8804	1,4155	0,2059	0,03495	0,6231
2	1, 3	0,8878	1,4520	0,1931	0,03074	0,6445
3	1, 4	0,9362	0,9252	0,3006	0,03429	0,4331
4	1, 5	0,8958	1,4887	0,1799	0,02664	0,6668
5	2, 3	0,8184	1,5674	0,1989	0,03928	0,6414
6	2, 4	0,8891	0,7796	0,3606	0,06767	0,3466
7	2, 5	0,8451	1,6528	0,1642	0,02689	0,6984
8	3, 4	0,9722	2,2773	0,0832	0,00385	0,8374
9	3, 5	0,6009	1,0204	0,4444	0,19548	0,3066
10	4, 5	0,9515	1,8027	0,0736	0,00478	0,8577

Используя данные таблицы 1, построенные регрессии были упорядочены по убыванию значений критерия детерминации R^2: 8, 10, 3, 4, 6, 2, 1, 7, 5, 9. С использованием критерия DW регрессии были упорядочены по убыванию эффекта автокорреляции остатков: 10, 8, 7, 5, 4, 2, 1, 9, 3, 6. Таким образом, лучшей по критерию детерминации является модель:

$$\hat{y} = -68066{,}4 + 82{,}8528x_3 + 209{,}510x_4,$$

а по критерию Дарбина-Уотсона модель:

$$\hat{y} = -245720 + 177{,}487x_4 + 4904{,}57x_5.$$

Затем осуществлялось упорядочивание регрессий одновременно по двум критериям: детерминации и Дарбина-Уотсона. Для этого применялся метод «идеальной» точки. Сначала значения критериев R^2 и DW нормировались по правилу:

$$\tilde{K}_i = \frac{K_i - \min(K)}{\max(K) - \min(K)}, \qquad i = \overline{1, n}.$$

После чего построенные регрессии были упорядочены по убыванию суммы значений нормированных критериев R^2 и DW: 10, 8, 7, 4, 2, 1, 5, 3, 6, 9. И, наконец, построенные регрессии были упорядочены по возрастанию предложенных критериев детерминации – автокорреляции. По критерию DA_{add}: 10, 8, 7, 4, 2, 5, 1, 3, 6, 9; по критерию DA_{mult}: 8, 10, 4, 7, 2, 3, 1, 5, 6, 9. По убыванию критерия DW^*_{mult} регрессии упорядочены следующим образом: 10, 8, 7, 4, 2, 5, 1, 3, 6, 9. Таким образом, по критериям DA_{add} и DW^*_{mult} регрессии упорядочены абсолютно одинаково. По методу «идеальной» точки получен практически такой же результат, с единственным отличием в том, что модель 1 оказалась лучше регрессии 5. А вот порядок регрессий по критерию DA_{mult} оказался наиболее отличным от альтернативных вариантов. Полученные результаты демонстрируют сходство предложенных критериев адекватности, но всё же при решении конкретных задач спецификации каждый из этих критериев может приводить к построению разных регрессионных моделей. На наш взгляд, на практике стоит отдавать предпочтение мультипликативному критерию детерминации – автокорреляции DA_{mult}, достоинствами которого является компактность его аналитического выражения и возможность интеграции его в виде целевой функции в задачу отбора информативных регрессоров [5].

Список литературы

1. Базилевский М.П. (2017) Оценивание параметров регрессионных моделей со стохастическими переменными по критерию детерминации-автокорреляции // Транспортная инфраструктура Сибирского региона. Иркутск. Т.1. С.382-386.
2. Базилевский М.П., Врублевский И.П., Носков С.И., Яковчук И.С. (2016) Среднесрочное прогнозирование эксплуатационных показателей функционирования Красноярской железной дороги // Фундаментальные исследования. Москва. №10(3). С.471-476.
3. Базилевский М.П., Носков С.И. Методические и инструментальные средства построения некоторых типов регрессионных моделей // Системы. Методы. Технологии. Братск, 2012. №1 (13). С.80-87.
4. Базилевский М.П., Носков С.И. (2009) Технология организации конкурса регрессионных моделей // Информационные технологии и проблемы математического моделирования сложных систем. Иркутск. Вып. 7. С. 77-84.
5. Носков С.И. (1996) Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. Иркутск: РИЦ ГП «Облинформпечать». 321 с.

MULTIPLICATIVE CRITERION OF DETERMINATION-AUTOCORRELATION IN REGRESSION ANALYSIS

M.P. Bazilevskiy
Cand. Sci. (Eng.), associate professor
mik2178@yandex.ru
Irkutsk

Irkutsk State Transport University
(Irkutsk, Russia)

Summary. When constructing a regression model, one of the main problems is the choice of its structural specification. Since each regression can be characterized by a variety of different adequacy criteria, such a task often turns out to be multi-criteria. The problems of solving multicriteria problems are solved by decision theory. The actual problem of the theory of decision-making is the construction of the convolution of local criteria. In article two forms of a two-criteria multiplicative criterion for determination - autocorrelation, simultaneously characterizing the accuracy of regression and autocorrelation in its residues, is proposed. To demonstrate the potential of the proposed criteria, the task of modeling the freight turnover of the Krasnoyarsk Railway is solved. At the same time, the choice of the regression model specification was carried out both by the multiplicative determination criteria - autocorrelation, and by their additive counterpart, and also by the method of the «ideal» point. The advantage of one of the proposed criteria is the possibility of its integration as an objective function in the task of «subset selection in regression».

Keywords: regression model, determination criterion, Darbin-Watson criterion, multiplicative criterion of determination-autocorrelation, freight turnover, decision theory.

References

1. Bazilevskii` M.P. (2017) Ocenivanie parametrov regressionny`kh modelei` so stohasticheskimi peremenny`mi po kriteriiu determinatcii-avtokorreliatcii [Estimation of parameters of regression model with stochastic variables parameter determination-autocorrelation] Transport infrastructure of Siberian region. Irkutsk. T. 1, pp. 382-386.

2. Bazilevskii` M.P., Vrublevskii` I.P., Noskov S.I., Iakovchuk I.S. (2016) Srednesrochnoe prognozirovanie e`kspluatatcionny`kh pokazatelei` funktcionirovaniia Krasnoiarskoi` zheleznoi` dorogi [Medium-term forecasting of operational performance of the Krasnoyarsk railway] Basic research. Moscow. №10, pp. 471-476.
3. Bazilevskii` M.P., Noskov S.I. (2012) Metodicheskie i instrumental`ny`e sredstva postroeniia nekotory`kh tipov regressionny`kh modelei` [Methodological and instrumental means to build certain types of regression models] The system. Methods. Technology. Bratsk. №1, pp. 80-87.
4. Bazilevskii` M.P., Noskov S.I. (2009) Tekhnologiia organizatcii konkursa regressionny`kh modelei` [Technology competition regression models] Information technologies and problems of mathematical modeling of complex systems. Irkutsk. №7, pp. 77-84.
5. Noskov S.I. (1996) Tekhnologiia modelirovaniia ob``ektov s nestabil`ny`m funktcionirovaniem i neopredelennost`iu v danny`kh. [Technology modeling objects with unstable functioning of and uncertainty in the data.] Irkutsk: RITC GP «Oblinformpechat`», 1996. 321 p.

УДК 511.14

О ДОСТАТОЧНЫХ УСЛОВИЯХ РАЦИОНАЛЬНОСТИ СУММ СЛОЖНЫХ КУБИЧЕСКИХ РАДИКАЛОВ

Николай Николаевич Волотов
к.ф.-м.н., доцент
volotovnn132@yandex.ru
г. Липецк
Липецкий государственный педагогический университет имени П.П. Семенова-Тян-Шанского

Аннотация. В работе символами $N, Z, Q, Q^+, Q^-, J, R, R^+$ обозначены множества натуральных, целых, рациональных, рациональных положительных, рациональных отрицательных, иррациональных, всех вещественных и положительных вещественных чисел, соответственно. Суммы сложных кубических радикалов, т.е. числа вида

$$\sqrt[3]{a+\sqrt{b}}+\sqrt[3]{a-\sqrt{b}}\left(a \in R\backslash\{0\}, b \in R^+ : \sqrt{b}, \sqrt[3]{a \pm \sqrt{b}} \in J\right), (*)$$

появляются [Курош, 2004: 234-239] при решении кубических уравнений методом Кардано: уравнение $x^3+3px+2q=0\ (p,q \in R)$ в случае, когда число $D=q^2+p^3 \in R^+$, имеет два сопряженных комплексных корня и один вещественный корень, определяемый формулой Кардано через его коэффициенты при помощи квадратных и кубических радикалов:

$$x=\sqrt[3]{-q+\sqrt{q^2+p^3}}+\sqrt[3]{-q-\sqrt{q^2+p^3}}. \quad (1)$$

Под рационализацией алгебраических выражений понимается приведение их к выражениям, содержащим меньшее число алгебраических операций, которые следует произвести над входящими в них величинами для нахождения их значений. Задачи на рационализацию сумм сложных кубических радикалов при отдельных значениях параметровaи b не новы, но при этом отсутствуют алгоритмы нахождения таких a и b, при которых числа вида (*) допускают рационализацию. В частности, в работах [2]-[5], [7-11] приведены доказательства равенств сумм $(*)$ для упорядоченных пар чисел $(a; b)$: $(2;\ 5), (7;\ 50), (9;\ 80), (20;\ 392), (45;\ 1682), (54;\ 2700), \left(6; \frac{847}{27}\right)$ числам 1, 2, 3, 4,