1 УДК 519.862.6

© *М. П. Базилевский*[1], *А. В. Караулова*[1]

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

ОЦЕНКА СТЕПЕНИ НЕЛИНЕЙНОСТИ ПОЛИНОМИАЛЬНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Аннотация. Данная статья посвящена разработке подхода к оценке степени нелинейности полиномиальных регрессионных моделей. Предложенные ранее авторами критерии нелинейности «по площади» ограничены тем, что справедливы только для функций, не имеющих ни экстремумов, ни перегибов, поэтому при моделировании измерить степень нелинейности можно было только для моделей, особо не отличающихся от линейных. В данной работе для полиномиальных регрессионных моделей предложен векторный критерий нелинейности. Большое количество близких к единице компонент этого вектора позволяет делать вывод о значительной нелинейности полинома. Прямая линия характеризуется вектором из одной нулевой компоненты. Если вектор состоит из нескольких нулей, то функция регрессии представляет собой ломаную. Предложенный подход успешно продемонстрирован на конкретном примере.

Ключевые слова: полиномиальная регрессия, критерий нелинейности, метод наименьших квадратов.

© *M. P. Bazilevskiy*[1], *A. V. Karaulova*[1]

[1] Irkutsk State Transport University, Irkutsk, Russian Federation

ESTIMATING THE DEGREE OF NONLINEARITY FOR POLYNOMIAL REGRESSION MODELS

Abstract. This article is devoted to the development of an approach to estimating the degree of nonlinearity for polynomial regression models. The non-linearity «over the area» criteria proposed earlier by the authors are limited by the fact that they are valid only for functions that do not have either extrema or inflections, therefore, when modeling, it was possible to measure the degree of non-linearity only for models that do not differ much from linear ones. In this paper, for polynomial regression models, a vector criterion for nonlinearity is proposed. A large number of components of this vector close to unity allows us to conclude that the polynomial is significantly non-linear. A straight line is characterized by a vector of one zero component. If the vector consists of several zeros, then the regression function is a broken line. The proposed approach has been successfully demonstrated on a specific example.

Keywords: polynomial regression, non-linearity criterion, ordinary least squares

Введение

В настоящее время методы регрессионного анализа не просто находят применение при решении различных прикладных задач анализа данных, но и продолжают активно развиваться. Так, в [1] предложен метод оценивания линейных регрессий с применением непрерывной формы метода максимальной согласованности поведения фактических и расчетных значений зависимой переменной, в [2] разработан способ построения линейной свёртки критериев адекватности регрессионных моделей, в [3] предложен метод смешанного оценивания параметров линейной регрессии, в [4] исследованы критерии адекватности модели регрессии Деминга, в [5] разработан программный комплекс регрессионного моделирования с учётом критерия согласованности поведения, в [6] задача отбора информативных регрессоров в линейной регрессии, оцениваемой с помощью метода наименьших квадратов (МНК), формализована в виде задачи частично-булевого линейного программирования и т.д.

В [7] для оценки степени нелинейности квазилинейных регрессий предложены критерии нелинейности «по площади», которые для каждой объясняющей переменной находятся по формуле [8]:

$$NC_{x_j} = \left| \frac{f_{\omega_j}(x^j_{\max}) + f_{\omega_j}(x^j_{\min})}{f_{\omega_j}(x^j_{\max}) - f_{\omega_j}(x^j_{\min})} - \frac{2\int\limits_{x^j_{\min}}^{x^j_{\max}} f_{\omega_j}(x_j)dx_j}{\left(x^j_{\max} - x^j_{\min}\right)\left(f_{\omega_j}(x^j_{\max}) - f_{\omega_j}(x^j_{\min})\right)} \right|, \quad j = \overline{1,m}, \qquad (1)$$

где m – число объясняющих переменных; $x^j_{\min}$, $x^j_{\max}$ – минимальное и максимальное по выборке значение j-й объясняющей переменной; $f_k(x)$, $k = \overline{1,q}$ – набор непрерывных и монотонных на отрезках $\left[x^j_{\min}, x^j_{\max}\right]$ элементарных функций; ω_j – номер элементарной функции для преобразования j-й объясняющей переменной.

Критерии (1) принимают значения от 0 до 1. Если $NC_j = 0$, то на отрезке $\left[x^j_{\min}, x^j_{\max}\right]$ переменная x_j преобразована линейно. Если $NC_j \to 1$, то преобразование переменной в значительной степени нелинейно.

В работах [9,10] с помощью критериев (1) сформулирована задача, состоящая в выборе спецификации квазилинейной регрессии одновременно по двум критериям – точности и нелинейности. К сожалению, критерии (1) справедливы только тогда, когда функции $f_k(x)$ на заданных отрезках не имеют экстремумов и перегибов. Из-за этого при решении в [9,10] двухкритериальной задачи приходилось ограничиваться лишь самыми простыми квазилинейными моделями, иногда не особо отличающимися по качеству от линейных. Целью данной работы является демонстрация возможности оценки степени нелинейности полиномиальных регрессионных моделей с помощью критериев (1).

1. Оценка нелинейности полиномиальных регрессий

Полиномиальная регрессионная модель имеет вид:

$$y_i = \alpha_0 + \sum_{j=1}^{q} \alpha_j x_i^j + \varepsilon_i, \qquad i = \overline{1,n}, \qquad (2)$$

где n – объем выборки; q – степень полинома; y_i, $i = \overline{1,n}$ – значения объясняемой переменной; x_i, $i = \overline{1,n}$ – значения объясняющей переменной; α_0, α_j, $j = \overline{1,q}$ – неизвестные параметры; ε_i, $i = \overline{1,n}$ – ошибки аппроксимации.

Пусть оцененная с помощью МНК модель (2) имеет вид:

$$\tilde{y} = \tilde{\alpha}_0 + \sum_{j=1}^{q} \tilde{\alpha}_j x_i^j, \qquad (3)$$

где $\tilde{\alpha}_0$, $\tilde{\alpha}_j$, $j = \overline{1,q}$ – оценки параметров.

Обозначим минимальное и максимальное значение переменной x на выборке $x_{\min}$ и $x_{\max}$ соответственно. Непрерывная функция (3) на отрезке $\left[x_{\min}, x_{\max}\right]$ может иметь экстремумы и перегибы, поэтому сразу оценивать её нелинейность с помощью критериев (1) будет ошибочным действием. Сначала нужно все возможные экстремумы и перегибы идентифицировать. Это несложно сделать, используя необходимые и достаточные условия экстремума и перегиба. Пусть s – общее количество точек экстремума и перегиба функции (3). Обозначим их как x_1^*, x_2^*, ..., x_s^*. Очевидно, что эти точки разбивают отрезок $\left[x_{\min}, x_{\max}\right]$ на $(s+1)$ отрезков: $\left[x_{\min}, x_1^*\right]$, $\left[x_1^*, x_2^*\right]$, ..., $\left[x_s^*, x_{\max}\right]$. При этом на каждом из этих отрезков функция (3) не имеет экстремумов и перегибов, следовательно, на каждом из них справедливы критерии нелинейности (1), которые обозначим $NC_{x,1}$, $NC_{x,2}$, ..., $NC_{x,s+1}$.

Таким образом, степень нелинейности полинома (3) можно охарактеризовать таблицей 1.

Таблица 1.

Характеристика степени нелинейности полинома (3)

Интервал	$\left[x_{\min}, x_1^*\right]$	$\left[x_1^*, x_2^*\right]$	...	$\left[x_s^*, x_{\max}\right]$
Критерий NC_x	$NC_{x,1}$	$NC_{x,2}$	...	$NC_{x,s+1}$

Более компактной характеристикой степени нелинейности полинома (3), по сравнению с таблицей 1, может служить вектор критериев нелинейности переменной x:

$$VNC_x = \left(NC_{x,1}, NC_{x,2}, ..., NC_{x,s+1}\right). \tag{4}$$

Для обоих характеристик справедливы следующие утверждения.

1. Чем больше компонент в векторе (4) (столбцов в таблице 1), тем больше экстремумов и перегибов имеет функция (3) на отрезке $\left[x_{\min}, x_{\max}\right]$, тем сложнее её поведение с точки зрения нелинейности. Если вектор (4) содержит одну компоненту (таблица 1 содержит один столбец), то функция (3) на отрезке $\left[x_{\min}, x_{\max}\right]$ монотонна.

2. Если все компоненты вектора (4) (ячейки второй строки таблицы 1) равны 0, то функция (3) представляет собой ломаную, состоящую из $(s+1)$ звеньев, которую легко интерпретировать. Если при этом вектор (4) содержит одну компоненту (таблица 1 содержит один столбец), то функция (3) является линейной. Чем выше значения компонент вектора (4) (ячеек второй строки таблицы 1), тем труднее интерпретировать модель.

2. Пример

Для демонстрации предложенного подхода были использованы статистические данные, представленные в таблице 2.

Таблица 2.

Статистические данные

y	1	5	6	5	3	7	10	4	2	5
x	1	2	3	4	5	6	7	8	9	10

По этим данным с помощью МНК была оценена полиномиальная регрессия степени 6:

$$\tilde{y} = -4{,}7 + 2{,}194557x + 6{,}86086x^2 - 4{,}34479x^3 + 0{,}995913x^4 - 0{,}09854x^5 + 0{,}003542x^6 \tag{5}$$

Коэффициент детерминации регрессии (5) равен 0,805, что подтверждает её адекватность.

График функции (5) изображен на рис. 1 сплошной линией.

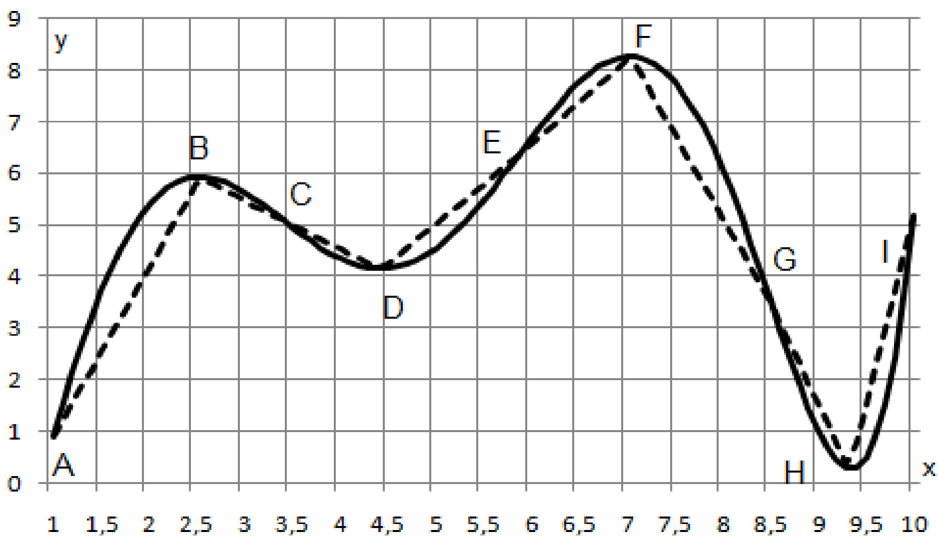

Рис. 1. График функции (5)

Как видно, функция (5) на отрезке $[1,10]$ имеет 2 локальных максимума (в точках $B(2.53,5.93)$ и $F(7.04,8.25)$), 2 локальных минимума (в точках $D(4.42,4.14)$ и $H(9.33,0.27)$) и 3 перегиба (в точках $C(3.41,5.08)$, $E(5.88,6.37)$ и $G(8.5,3.51)$). Итого суммарное число точек экстремума и перегиба $s=7$. Эти точки разбивают отрезок $[1,10]$ на 8 отрезков: $[1,2.53226]$, $[2.53226,3.40545]$, $[3.40545,4.42241]$, $\left[4.42241,5.87945\right]$, $\left[5.87945,7.04048\right]$, $[7.04048,8.50362]$, $[8.50362,9.32894]$, $[9.32894,10]$. На каждом из этих отрезков пунктиром (рис. 1) построены прямые, соединяющие концы графика функции, и с помощью формул (1) вычислены критерии нелинейности, из которых был сформирован вектор:

$$VNC_x=\left(0.347,0.251,0.261,0.276,0.309,0.359,0.057,0.234\right). \tag{6}$$

Наличие в векторе (6) восьми компонент указывает на высокую степень нелинейности функции (5) на отрезке $[1,10]$. Самая большая компонента этого вектора равна 0,359 на отрезке $[7.04,8.5]$, на котором график функции (5) имеет наиболее сильное отклонение от пунктирной прямой, а самая маленькая – 0,057 на отрезке $[8.5,9.33]$, на котором график функции (5) практически не отличается от пунктирной прямой.

Заключение

В статье предложен подход к оценке степени нелинейности полиномиальных регрессионных моделей. Полученные результаты могут быть использованы в процессе построении нелинейных регрессионных моделей с целью контроля степени их нелинейности.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Носков С.И., Бычков Ю.А. Вычислительные эксперименты с непрерывной формой метода максимальной согласованности в регрессионном анализе // Вестник Воронежского государственного технического университета. - Воронеж, 2022.- Т. 18.- № 2.- С. 7-12.

2. Носков С.И. Построение свертки критериев адекватности регрессионных моделей // Модели, системы, сети в экономике, технике, природе и обществе.-Пензенский государственный университет, 2022.- № 1 (41).- С. 73-81.

3. Носков С.И. Метод смешанного оценивания параметров линейной регрессии: особенности применения // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии.-Воронеж, 2021.- № 1.- С. 126-132.

4. Базилевский М.П. Аналитические зависимости для некоторых критериев адекватности модели регрессии Деминга // Вестник Иркутского государственного технического университета.- Иркутск, 2016.- Т. 20.- № 10 (117).- С. 81-89.

5. Базилевский М.П., Носков С.И. Программный комплекс построения линейной регрессионной модели с учётом критерия согласованности поведения фактической и расчетной траекторий изменения значений объясняемой переменной // Вестник Иркутского государственного технического университета.-Иркутск, 2017.- Т. 21.- № 9 (128).- С. 37-44.

6. Базилевский М.П. Сведение задачи отбора информативных регрессоров при оценивании линейной регрессионной модели по методу наименьших квадратов к задаче частично-булевого линейного программирования // Моделирование, оптимизация и информационные технологии.-Воронеж, 2018.- Т. 6.- № 1 (20).- С. 108-117.

7. Базилевский М.П. Критерии нелинейности квазилинейных регрессионных моделей // Моделирование, оптимизация и информационные технологии.-Воронеж, 2018.- Т. 6.- № 4 (23).- С. 185-195.

8. Базилевский М.П., Караулова А.В. Предварительное оценивание степени нелинейности структурных спецификаций квазилинейных регрессий // Математические методы в технике и технологиях - ММТТ.- Саратов, 2020.- Т. 5. - С. 49-52.

9. Базилевский М.П., Караулова А.В. Выбор оптимального соотношения между точностью и нелинейностью при построении квазилинейных регрессионных моделей // Вестник кибернетики.- Сургут, 2021.- № 4 (44).- С. 63-70.

10.Караулова А.В., Базилевский М.П. Программный комплекс построения квазилинейных регрессий по критериям точности и нелинейности // Экономика. Информатика.- Белгород, 2022.- Т. 49.- № 1.- С. 121-133.

REFERENCES

1. Noskov S.I., Bychkov Yu.A. Vychislitel'nye eksperimenty s nepreryvnoy formoy metoda maksimal'noy soglasovannosti v regressionnom analize [Computational experiments with the continuous form of the maximum consistency method in regression analysis] // Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta – Bulletin of Voronezh state technical University. 2022. Vol. 18. No. 2. Pp. 7-12.

2. Noskov S.I. Postroenie svertki kriteriev adekvatnosti regressionnykh modeley [Construction of convolution criteria for regression models] // Modeli, sistemy, seti v ekonomike, tekhnike, prirode i obshchestve – Models, systems, networks in economics, technology, nature and society. 2022. No. 1 (41). Pp. 73-81.

3. Noskov S.I. Metod smeshannogo otsenivaniya parametrov lineynoy regressii: osobennosti primeneniya [Method of mixed estimation of linear regression parameters: application features] // Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyy analiz i informatsionnye tekhnologii – Proceedings of Voronezh State University. Series: systems analysis and information technologies. 2021. No. 1. Pp. 126-132.

4. Bazilevskiy M.P. Analiticheskie zavisimosti dlya nekotorykh kriteriev adekvatnosti modeli regressii Deminga [Analytical dependences for some adequacy criteria of Deming regression model] // Vestnik Irkutskogo gosudarstvennogo tekhnicheskogo universiteta – Proceedings of Irkutsk State Technical University. 2016. Vol. 20. No. 10 (117). Pp. 81-89.

5. Bazilevskiy M.P., Noskov S.I. Programmnyy kompleks postroeniya lineynoy regressionnoy modeli s uchetom kriteriya soglasovannosti povedeniya fakticheskoy i raschetnoy traektoriy izmeneniya znacheniy ob"yasnyaemoy peremennoy [Program complex for linear regression model construction considering behavior consistency criterion of actual and calculated trajectories of explained variable value change] // Vestnik Irkutskogo gosudarstvennogo tekhnicheskogo universiteta – Proceedings of Irkutsk State Technical University. 2017. Vol. 21. No. 9 (128). Pp. 37-44.

6. Bazilevskiy M.P. Svedenie zadachi otbora informativnykh regressorov pri otsenivanii lineynoy regressionnoy modeli po metodu naimen'shikh kvadratov k zadache chastichno-bulevogo lineynogo programmirovaniya [Reduction the problem of selecting informative regressors when estimating a linear regression model by the method of least squares to the problem of partial-Boolean linear programming] // Modelirovanie, optimizatsiya i informatsionnye tekhnologii – Modeling, optimization and information technology. 2018. Vol. 6. No. 1 (20). Pp. 108-117.

7. Bazilevskiy M.P. Kriterii nelineynosti kvazilineynykh regressionnykh modeley [Nonlinear criteria of quasilinear regression models] // Modelirovanie, optimizatsiya i informatsionnye tekhnologii – Modeling, optimization and information technology. 2018. Vol. 6. No. 4 (23). Pp. 185-195.

8. Bazilevskiy M.P., Karaulova A.V. Predvaritel'noe otsenivanie stepeni nelineynosti strukturnykh spetsifikatsiy kvazilineynykh regressiy [Preliminary estimation of non-linearity degree of quasilinear regressions structural specifications] // Matematicheskie metody v tekhnike i tekhnologiyakh - MMTT – Mathematical methods in technics and technology. 2020. Vol. 5. Pp. 49-52.

9. Bazilevskiy M.P., Karaulova A.V. Vybor optimal'nogo sootnosheniya mezhdu tochnost'yu i nelineynost'yu pri postroenii kvazilineynykh regressionnykh modeley [Selecting the optimum relationship between accuracy and non-linearity in constructing quasi-linear regression models] // Vestnik kibernetiki – Proceedings in Cybernetics. 2021. No. 4 (44). Pp. 63-70.

10.Karaulova A.V., Bazilevskiy M.P. Programmnyy kompleks postroeniya kvazilineynykh regressiy po kriteriyam tochnosti i nelineynosti [Software complex for constructing quasi-linear regressions according to the criteria of accuracy and non-linearity] // Ekonomika. Informatika – Economics. Information technologies. 2022. Vol. 49. No. 1. Pp. 121-133.

Информация об авторах

Михаил Павлович Базилевский – к. т. н., доцент, доцент кафедры «Математика», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: mik2178@yandex.ru

Анна Витальевна Караулова – аспирант кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, email: anuta_kav@mail.ru

Authors

Mikhail Pavlovich Bazilevskiy – Ph. D. in Engineering Science, Associate Professor, the Subdepartment of Mathematics, Irkutsk State Transport University, Irkutsk, e-mail: mik2178@yandex.ru

Anna Vitalievna Karaulova – postgraduate student of the Department "Information Systems and Information Protection", Irkutsk State University of Railways, Irkutsk, email: anuta_kav@mail.ru

Для цитирования

Базилевский М.П., Караулова А.В. Оценка степени нелинейности полиномиальных регрессионных моделей // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2022. – №3(15). – С.1-6– DOI: 10.26731/2658-3704.2022.3(15).1-6 – Режим доступа: http://ismm-irgups.ru/toma/315-2022, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 15.10.2022)

For citations

Bazilevsky M.P., Karaulova A.V. Estimating the degree of nonlinearity for polynomial 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], 2022. No. 3(15). P. 1-6. DOI: 10.26731/2658-3704.2022.3(15).1-6 [Accessed 15/10/22]