

А.В. Караулова¹, М.П. Базилевский¹

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

ПОСТРОЕНИЕ АДДИТИВНЫХ СТЕПЕННЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ СО СТЕПЕНЯМИ В ВИДЕ ДИСКРЕТНЫХ ФУНКЦИЙ ВРЕМЕНИ

Аннотация. В статье введена аддитивная степенная регрессионная модель со степенями в виде дискретных функций времени. Рассмотрен её частный случай, для которого разработан алгоритм приближенного оценивания с помощью метода наименьших квадратов. Предложенный алгоритм продемонстрирован на примере построения модели зависимости валового регионального продукта Иркутской области от текущих затрат на научные фундаментальные исследования.

Ключевые слова: регрессионный анализ, аддитивная степенная регрессия со степенями в виде дискретных функций времени, коэффициент детерминации, метод наименьших квадратов.

A.V. Karaulova¹, M.P. Bazilevskiy¹

¹ *Irkutsk State Transport University, Irkutsk, Russia*

CONSTRUCTION OF ADDITIVE POWER REGRESSION MODELS WITH DEGREES IN THE FORM OF DISCRETE TIME FUNCTIONS

Abstract. The article is devoted to the construction of an additive power regression model with degrees in the form of discrete time functions. An algorithm for its approximate estimation using the least squares method is proposed. The proposed estimation algorithm is demonstrated by the example of statistical data of the gross regional product in the Irkutsk region and current expenditures on basic scientific research for the period 2000-2019.

Keywords: regression analysis, additive power regression with degrees in the form of discrete time functions, coefficient of determination, ordinary least squares.

Введение

Признанным инструментом анализа данных является регрессионный анализ [1,2]. Его использование приводит к построению математических моделей влияния одной или нескольких независимых переменных на зависимую переменную. Одним из первоочередных этапов при построении регрессионной модели является выбор её структурной спецификации [2,3], т.е. состава объясняющих переменных и математической формы связи между ними. Проводя регрессионный анализ, чаще всего пользуются линейными моделями, т.к. они легко оцениваются и объясняются. Однако, как показывает практика [4], реальные явления или процессы редко носят линейный характер, поэтому приходится прибегать к построению нелинейных зависимостей.

На сегодняшний день разработано значительное количество известных нелинейных математических форм связи между переменными [5-9]: полиномиальная, степенная и логарифмическая регрессии; регрессии со степенями в виде линейных комбинаций объясняющих переменных и в виде линейных комбинаций их натуральных логарифмов и др. В связи с этим возникают вопросы определения подходящей аппроксимирующей функции, метода оценивания неизвестных параметров, способов интерпретации полученных результатов. Для решения подобных задач в [2] был разработан подход к организации конкурса регрессионных моделей, а в [10] расширена область в построении множества их альтернативных вариантов большой мощности и последующего выбора наиболее приемлемого из них, руководствуясь векторным критерием оценки адекватности каждого варианта. В ряде работ изучаются вопросы, связанные с использованием критериев нелинейности [11-12]. В рамках регрессионного анализа проводятся активные исследования, касающиеся как разработки методов построения конкретных моделей, так и критериев оценки их качества [13-15].

При построении регрессионных уравнений часто описывают динамику некоторого показателя, тенденция которого изменяется с течением времени. От того насколько точно вре-

менное изменение учтено в исследуемой модели, во многом зависит её адекватность. В работе [2] рассматривается зависимость с параметрами в виде дискретных функций времени.

Целью данной статьи является исследование вопроса оценивания с помощью метода наименьших квадратов (МНК) аддитивной степенной регрессии со степенями в виде дискретных функций времени.

1. Аддитивная степенная регрессия со степенями в виде дискретных функций времени

Введем аддитивную степенную регрессию со степенями в виде дискретных функций времени:

$$y_t = \alpha_0 + \sum_{j=1}^l \alpha_j x_{ij}^{\beta_{0j} + \sum_{k=1}^p \beta_{kj} t^k} + \varepsilon_t, \quad t = \overline{t_1, t_n}, \quad (1)$$

где n – объем выборки; $y_t, t = \overline{t_1, t_n}$ – значения зависимой переменной y ; $x_{ij}, t = \overline{t_1, t_n}, j = \overline{1, l}$ – значения независимых переменных $x_1, x_2, \dots, x_l$; $\alpha_0, \alpha_1, \dots, \alpha_l, \beta_{0j}, \beta_{1j}, \dots, \beta_{pj}, j = \overline{1, l}$ – неизвестные параметры; $\varepsilon_t, t = \overline{t_1, t_n}$ – ошибки аппроксимации.

Рассмотрим частный случай модели (1) для одной переменной x при $\beta_{01} = 0, p = 1$:

$$y_t = \alpha_0 + \alpha_1 x_t^{\beta t} + \varepsilon_t, \quad t = \overline{t_1, t_n}. \quad (2)$$

Модель (2) является нелинейной по параметрам и её оценивание с помощью МНК становится трудно решаемой задачей. Однако, если придать параметру β определенное значение, то модель становится линейной по параметрам α_0 и α_1 и их уже можно определить с помощью обычного МНК. Таким образом, перебирая значения параметра β из некоторого интервала, несложно получить приближенные оценки модели (2).

Определим, как нужно назначать ширину интервала для параметра β . Для этого обозначим $\beta t = k$:

$$y_t = \alpha_0 + \alpha_1 x_t^k + \varepsilon_t, \quad t = \overline{t_1, t_n}. \quad (3)$$

Модель (3) является аддитивной степенной регрессией [5]. Если в ней параметр k будет принимать слишком большие по модулю значения, то модель станет в значительной степени нелинейной, следовательно, будет затруднительно использовать её на практике. Поэтому введем ограничения на этот параметр:

$$k_{\text{нижн}}^- \leq k \leq k_{\text{верхн}}^+, \quad (4)$$

где $k_{\text{нижн}}^-, k_{\text{верхн}}^+$ – отрицательная нижняя и положительная верхняя границы параметра k .

Необходимо учесть, что, если параметр k будет принимать близкие к нулю значения, то переменная x^k будет принимать близкие к единице значения. А поскольку в модель (3) входит свободный член α_0 , то возникнет эффект сильной либо полной мультиколлинеарности, из-за которого нельзя будет получить её МНК-оценки. Поэтому скорректируем ограничение (4) на параметр k :

$$k \in [k_{\text{нижн}}^-, k_{\text{верхн}}^-] \cup [k_{\text{нижн}}^+, k_{\text{верхн}}^+], \quad (5)$$

где $k_{\text{нижн}}^-, k_{\text{верхн}}^-, k_{\text{нижн}}^+, k_{\text{верхн}}^+$ – отрицательная нижняя и верхняя, положительная нижняя и верхняя границы параметра k ; символ « $\cup$ » – операция объединения двух множеств, которую с позиции теории множеств следует трактовать как $A \cup B = \{x | x \in A \vee x \in B\}$.

На основе (5) определим отрицательную нижнюю и верхнюю границы параметра β модели (2). Для этого нужно решить задачи линейного программирования с целевыми функциями $\beta \rightarrow \min, \beta \rightarrow \max$ и с линейными ограничениями:

$$k_{\text{нижн}}^- \leq \beta t \leq k_{\text{верхн}}^-, \quad t = \overline{t_1, t_n}. \quad (6)$$

Нетрудно показать, что аналитическое решение задачи линейного программирования при поиске отрицательной нижней границы параметра β имеет вид

$$\beta_{\text{нижн}}^- = \max \left\{ \frac{k_{\text{нижн}}^-}{t_1}, \frac{k_{\text{нижн}}^-}{t_2}, \dots, \frac{k_{\text{нижн}}^-}{t_n} \right\} = \frac{k_{\text{нижн}}^-}{t_n}, \quad (7)$$

а при поиске отрицательной верхней границы параметра β получим

$$\beta_{\text{верхн}}^- = \min \left\{ \frac{k_{\text{верхн}}^-}{t_1}, \frac{k_{\text{верхн}}^-}{t_2}, \dots, \frac{k_{\text{верхн}}^-}{t_n} \right\} = \frac{k_{\text{верхн}}^-}{t_1}. \quad (8)$$

Аналогично определяются положительная нижняя и верхняя граница параметра β модели (2):

$$\beta_{\text{нижн}}^+ = \max \left\{ \frac{k_{\text{нижн}}^+}{t_1}, \frac{k_{\text{нижн}}^+}{t_2}, \dots, \frac{k_{\text{нижн}}^+}{t_n} \right\} = \frac{k_{\text{нижн}}^+}{t_1}, \quad (9)$$

$$\beta_{\text{верхн}}^+ = \min \left\{ \frac{k_{\text{верхн}}^+}{t_1}, \frac{k_{\text{верхн}}^+}{t_2}, \dots, \frac{k_{\text{верхн}}^+}{t_n} \right\} = \frac{k_{\text{верхн}}^+}{t_n}. \quad (10)$$

Тогда для приближенного МНК-оценивания параметров регрессии (2) можно воспользоваться следующим алгоритмом:

1. Разбить каждый из объединенных промежутков $\beta \in [\beta_{\text{нижн}}^-, \beta_{\text{верхн}}^-] \cup [\beta_{\text{нижн}}^+, \beta_{\text{верхн}}^+]$ m точками.
2. В каждой точке найти МНК-оценки параметров α_0 и α_1 регрессии (2).
3. Выбрать модель с наименьшим значением суммы квадратов остатков (с наибольшим значением коэффициента детерминации R^2).

2. Моделирование ВРП Иркутской области

Для демонстрации предложенного алгоритма МНК-оценивания были использованы статистические данные (таблица 1) валового регионального продукта по Иркутской области (y) и текущих затрат на научные фундаментальные исследования (x) за период 2000-2019 гг. (t). Данные были собраны с использованием архивов Росстата (<https://rosstat.gov.ru/>).

Таблица 1

Статистические данные

t	y	x	t	y	x
2000	103013,8	165,9	2010	546141,0	2024,2
2001	120240,0	244,6	2011	634561,4	2154,4
2002	140195,9	322,9	2012	737971,6	2532,5
2003	167927,1	392,3	2013	805197,5	2697,9
2004	213244,2	484,1	2014	916317,5	2835,6
2005	258095,5	545,4	2015	1001717,6	2539,6
2006	330834,3	789,6	2016	1139206,8	2504,2
2007	402654,7	1075,8	2017	1268311,7	2435,2
2008	438852,4	1453,5	2018	1460512,2	2710,8
2009	458774,9	2076,6	2019	1545680,6	3979,1

По статистическим данным таблицы 1 сначала была оценена линейная регрессия с помощью МНК:

$$\tilde{y} = 34178 + 332,4761x, \quad R^2 = 0,8134 \quad (11)$$

Показатель критерия детерминации полученной линейной регрессии достаточно высок (больше 0,8), поэтому модель адекватная.

Для построения аддитивной степенной регрессии со степенями в виде дискретных функций времени (2) была выбрана область изменения параметра $k \in [-3; -0,5] \cup [0,5; 3]$. По формулам (7) – (10) были определены границы изменения параметра β :

$$\beta \in [-0,00148; -0,00025] \cup [0,00025; 0,00148].$$

Затем каждый из этих объединенных диапазонов был разбит на 9 точек. После чего с помощью пакета Gretl в каждой точке находились МНК-оценки параметров a_0 и a_1 регрессии (2) и вычислялось значение коэффициента детерминации. На рис. 1, 2 представлены графики зависимостей коэффициентов детерминации от значений параметра β в каждом диапазоне.

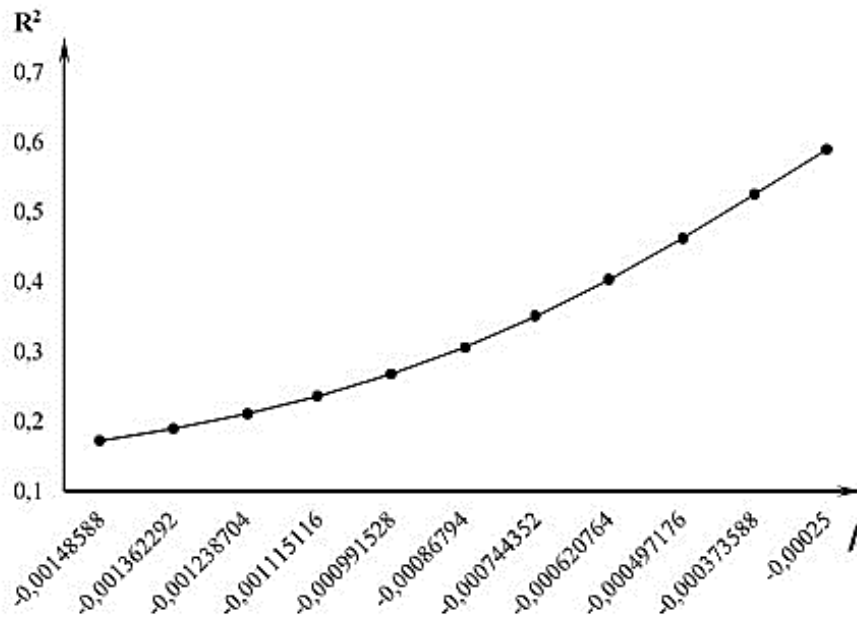


Рис. 1. График зависимости R^2 от значений параметра β на отрезке $[-0,00148; -0,00025]$

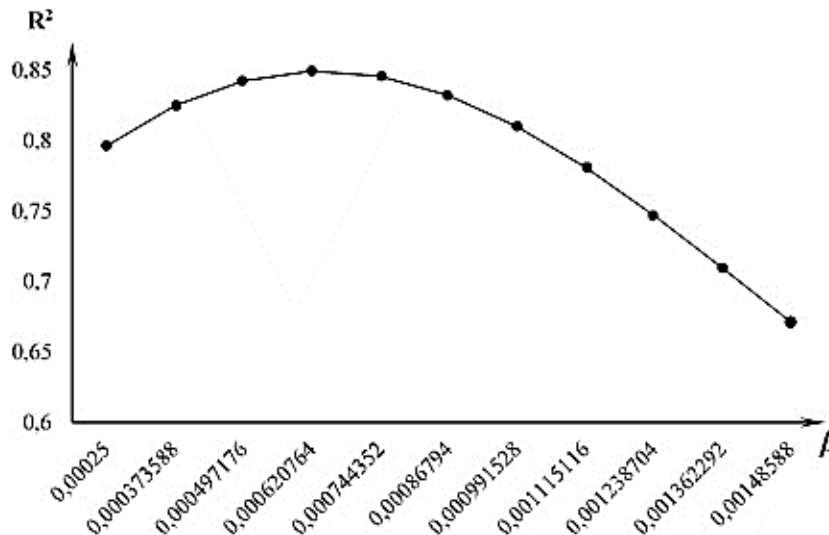


Рис. 2. График зависимости R^2 от значений параметра β на отрезке $[0,00025; 0,00148]$

Как видно, на рис.1 наибольшее значение $R^2 = 0,5892$ достигается в точке $\beta = -0,00025$, а на рис.2 наибольшее значение $R^2 = 0,8495$ достигается в точке $\beta = 0,000620$. Лучшему из этих решений соответствует регрессионное уравнение:

$$\tilde{y} = 85932,3 + 46,97x^{0,00062t}, \quad R^2 = 0,8495 \quad (12)$$

Тем самым, по коэффициенту детерминации качество модели (12) несколько выше, чем у линейной регрессии (11).

На рис. 3 представлен график зависимости степени $\beta t = k$ переменной x в уравнении (12) от времени t .

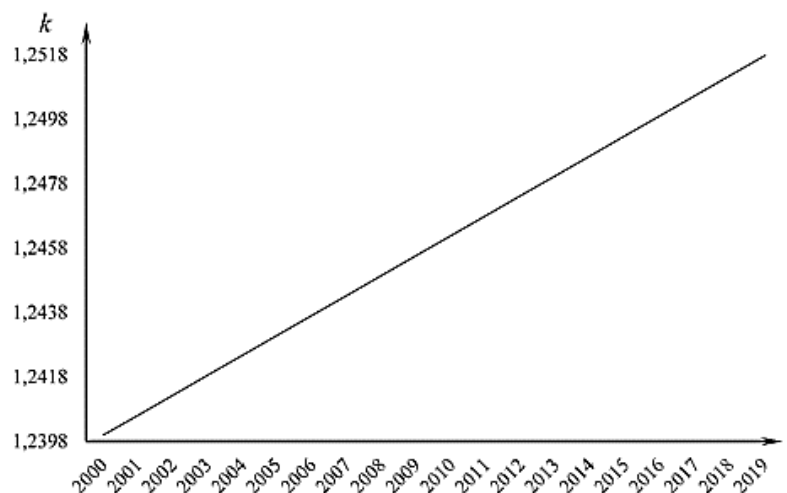


Рис. 3. График зависимости степени от времени t

По графику на рис.3 видно, что с течением времени меняется степень аддитивной степенной регрессии, а значит, и степень её нелинейности [13-15].

Заключение.

В данной работе рассмотрена упрощенная форма аддитивной степенной регрессии со степенями в виде дискретных функций времени. Предложен алгоритм её приближенного оценивания с помощью МНК. Этот алгоритм был применен для моделирования валового регионального продукта Иркутской области.

В дальнейшем планируется автоматизировать процесс построения предложенных моделей с использованием критериев нелинейности.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Pardoe I. Applied Regression Modeling. Wiley, 2020. 336 p.
2. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. - Иркутск: Облформпечать, 1996.- 321 с.
3. Клейнер Г.Б. Производственные функции: теория, методы, применение.- М.: Финансы и статистика. 1986.- 239 с.
4. Караулова А.В., Базилевский М.П. Применение регрессионного анализа при решении реальных задач технического характера // Молодая наука Сибири. - Иркутск: ИрГУПС, 2020.- №3(9).- С. 229-236.
5. Базилевский М.П. Разработка и исследование алгоритмов оценивания параметров аддитивной степенной регрессии // Современные технологии. Системный анализ. Моделирование. -Иркутск: ИрГУПС, 2017.- Т. 56.- №4.-С.131-138.
6. Базилевский М.П. Построение степенно-показательных регрессионных моделей и их интерпретация // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии.- Воронеж: ВГУ, 2020.- № 4.-С. 19-28.
7. Носков С.И., Хоняков А.А. Программный комплекс построения некоторых типов кусочно-линейных регрессий // Информационные технологии и математическое моделирование в управлении сложными системами. -Иркутск: ИрГУПС, 2019. № 3 (4). С. 47—55.
8. Базилевский М.П., Носков С.И. Формализация задачи построения линейно-мультипликативной регрессии в виде задачи частично - булевого линейного программирования

ния // Современные технологии. Системный анализ. Моделирование. -Иркутск: ИрГУПС, 2017.- №3 (55).- С.101—105.

9. Базилевский М.П. Построение степенно-показательных и линейно-логарифмических регрессионных моделей // Проблемы управления.-М.: Институт проблем управления им. В.А. Трапезникова РАН, 2021.- № 3.- С. 25-32.

10. Носков С.И. Реализация конкурса регрессионных моделей с применением критерия согласованности поведения // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии.- Воронеж: ВГУ, 2021. № 2. С. 153-160.

11. Базилевский М.П. Критерии нелинейности квазилинейных регрессионных моделей// Моделирование, оптимизация и информационные технологии.-Воронеж: Воронежский институт высоких технологий, 2018.- № 4.- С. 185-195.

12. Базилевский М. П. Критерии нелинейности многофакторных квазилинейных регрессий // Молодежь и наука: актуальные проблемы фундаментальных и прикладных исследований. Материалы II Всерос. национал. науч. конф. студентов, аспирантов и молодых ученых.-Комсомольск на Амуре: Комсомольский-на-Амуре гос. ун-т, 2019.- С. 210–213.

13. Базилевский М. П., Караулова А. В. Предварительное оценивание степени нелинейности структурных спецификаций квазилинейных регрессий // Матем. методы в технике и технологиях. ММТТ.-Саратов: СГТУ имени Гагарина Ю.А., 2020. - Т. 5.- С. 49–52.

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

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

REFERENCES

1. Pardoe I. Applied Regression Modeling. Wiley, 2020. 336 p.
2. Noskov S.I. Technology of modeling objects with unstable functioning and uncertainty in data. - Irkutsk: Oblinformpechat, 1996.- 321 p.
3. Kleiner G.B. Production functions: theory, methods, application. - M.: Finance and Statistics. 1986.- 239 p.
4. Karaulova A.V., Bazilevsky M.P. Application of regression analysis in solving real problems of a technical nature // Molodaya nauka Sibir. - Irkutsk: IrGUPS, 2020.- №3(9).- pp. 229-236.
5. Bazilevsky M.P. Development and research of algorithms for estimating parameters of additive power regression // Modern technologies. System analysis. The models are different. -Irkutsk: IrGUPS, 2017.- Vol. 56.- No. 4.-pp.131-138.
6. Bazilevsky M.P. Construction of power-exponential regression models and their interpretation // Bulletin of the Voronezh State University. Series: System analysis and information technologies.- Voronezh: VSU, 2020.- No. 4.-pp. 19-28.
7. Noskov S.I., Khonyakov A.A. A software package for constructing some types of piecewise linear regressions // Information technologies and mathematical modeling in the management of complex systems. -Irkutsk: IrGUPS, 2019.- No. 3 (4).- pp. 47-55.
8. Bazilevsky M.P., Noskov S.I. Formalization of the problem of constructing linear multiplicative regression in the form of a partially Boolean linear programming problem // Modern technologies. System analysis. Modeling. -Irkutsk: IrGUPS, 2017.- №3 (55).- pp.101—105.
9. Bazilevsky M.P. Construction of power-exponential and linear-logarithmic regression models // Problems of management.-M.: V.A. Trapeznikov Institute of Management Problems of the Russian Academy of Sciences, 2021.- No. 3.- pp. 25-32.
10. Noskov S.I. Implementation of the competition of regression models using the criterion of consistency of behavior // Bulletin of the Voronezh State University. Series: System Analysis and Information Technologies.- Voronezh: VSU, 2021.- No. 2.- pp. 153-160.

11. Bazilevsky M.P. Criteria of nonlinearity of quasi-linear regression models// Modeling, optimization and information technology.-Voronezh: Voronezh Institute of High Technologies, 2018.- No. 4.- pp. 185-195.
12. Bazilevsky M. P. Criteria of nonlinearity of multifactorial quasi-linear regression // Youth and Science: actual problems of fundamental and applied research. Materials II Vsros. national scientific conference of students, postgraduates and young scientists.-Komsomolsk on Amur: Komso-molsk-on-Amur State University, 2019.- pp. 210-213.
13. Bazilevsky M. P., Karaulova A.V. Preliminary estimation of the degree of non-linearity of the structural specifications of quasi-linear regressions // Matem. methods in engineering and technology. MMTT.-Saratov: Gagarin Yu.A. State Technical University, 2020. - Vol. 5.- pp. 49-52.
14. Bazilevsky M. P., Karaulova A.V. Choosing the optimal ratio between accuracy and non-linearity in the construction of quasi-linear regression models // Bulletin of Cybernetics. – Surgut: SSU, 2021.- №4 (44).- pp.63-70.
15. Karaulova A.V., Bazilevsky M. P. Software package for constructing quasi-linear regres-sions according to accuracy and nonlinearity criteria // Economy. Computer science.- Belgorod: BGNIU, 2022. -Vol. 49.- No. 1.- pp. 121-133.

Информация об авторах

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

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

Authors

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

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

Для цитирования

Караулова А.В., Базилевский М.П. Построение аддитивных степенных регрессионных моделей со степенями в виде дискретных функций времени // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2022. – №2(14). – С.1-7– DOI: 10.26731/2658-3704.2022.2(14).1-7 – Режим доступа: <http://ismm-irgups.ru/toma/214-2022>, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обра-щения: 24.06.2022)

For citations

Karaulova A.V., Bazilevsky M.P. Construction of additive power regression model with de-grees in the form of discrete time // Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal [Information technology and math-ematical modeling in the management of complex systems: electronic scientific journal], 2022. No. 2(14). P. 1-7. DOI: 10.26731/2658-3704.2022.2(14).1-7 [Accessed 24/06/22]