

<https://doi.org/10.17073/2072-1633-2025-2-1393>

Повышение достоверности экономического прогноза за счет проверки нормальности распределения массива данных

Ю.Ю. Костюхин , А.С. Богачев  

Национальный исследовательский технологический университет «МИСИС»,
119049, Москва, Ленинский просп., д. 4, стр. 1, Российская Федерация

 andr.bogachiov@yandex.ru

Аннотация. Одной из ошибок прогнозирования тенденций экономического развития является отсутствие первоначальной проверки нормальности распределения данных как неотъемлемое условие применимости статистических процедур. Применимость данных методов к искаженным данным приводит к неточности и снижению качества экономического прогноза. Цель работы – провести поэтапную проверку нормальности распределения данных для обеспечения более высокой достоверности экономического прогноза на основе тестов симметрии, таких как коэффициент вариации, графики квантилей, среднее абсолютное отклонение, диапазон размаха варьирования и статистика Жарка–Бера. Обработка данных, основанная на распределении валового внутреннего продукта РФ с 2000 по 2020 г., показала наличие нормального распределения массива, что способствует достоверному экономическому прогнозу и оценки перспектив изменений в будущем с целью минимизации ошибок и искажению результатов.

Ключевые слова: коэффициент вариации, среднее абсолютное отклонение, размах варьирования, статистика Жарка–Бера, нормальное распределение

Для цитирования: Костюхин Ю.Ю., Богачев А.С. Повышение достоверности экономического прогноза за счет проверки нормальности распределения массива данных. *Экономика промышленности*. 2025;18(2):275–281. <https://doi.org/10.17073/2072-1633-2025-2-1393>

Increasing the reliability of the economic forecast by checking the normality of the data array distribution

Yu.Yu. Kostyukhin , A.S. Bogachev  

National University of Science and Technology “MISIS”,
4-1 Leninskiy Ave., Moscow 119049, Russian Federation

 andr.bogachiov@yandex.ru

Abstract. One of the errors in forecasting economic development trends is lack of initial normality check of data distribution as an essential condition for the applicability of statistical procedures. The applicability of these methods to distorted data results in inaccuracy and low quality of the economic forecast. The purpose of the study is to carry out a step-to-step normality check of data distribution to ensure a more accurate economic forecast based on the symmetry tests such as the coefficient of variation, quantile graphs, average absolute deviation, range of variation, and Jarque–Bera statistic. Data processing based on distribution of Russia’s gross domestic product from 2000 to 2020 revealed a normal array distribution, which ensures reliable economic forecasting and assessment of prospects for future changes in order to minimize errors and distorted results.

Keywords: coefficient of variation, average absolute deviation, the range of variation, Jarque–Bera statistic, normal distribution

For citation: Kostyukhin Yu.Yu., Bogachev A.S. Increasing the reliability of the economic forecast by checking the normality of the data array distribution. *Russian Journal of Industrial Economics*. 2025;18(2):275–281. <https://doi.org/10.17073/2072-1633-2025-2-1393>

Проверка нормальности распределения данных

Yu.Yu. 科斯秋欣 , A.S. 博加乔夫  

Государственный технический университет «МИСИ», 119049, Российская Федерация, Москва, Ленинский проспект, 40/1

 andr.bogachiov@yandex.ru

Аннотация: Проверка тенденций экономического развития – одна из задач, требующих предварительной проверки, а это статистическая программа. Эти методы применяются к искаженным данным, что приводит к неточности и снижению качества. Целью работы является проверка нормальности распределения данных, чтобы обеспечить надежность коэффициента, диаграммы квантилей, абсолютного отклонения, дисперсии и Jarque-Bera тестов на симметричность. Согласно данным за период с 2000 по 2020 гг. в Российской Федерации, данные имеют нормальное распределение, что способствует надежному прогнозу и оценке, тем самым уменьшая ошибки и искажения.

Ключевые слова: коэффициент, абсолютное отклонение, дисперсия, Jarque-Bera тест, нормальное распределение

Введение

В условиях быстро развивающегося рынка и нестабильных экономических условий основной задачей компании является своевременное и качественное прогнозирование ситуации под воздействием факторов как внутренней, так и внешней среды. Разработанная стратегия развития, обеспечивающая рост экономического потенциала предприятия в будущем, предполагает необходимость объективной и рациональной оценки составляющих [1–3].

Основной проблемой искажения результатов прогнозирования и, как следствие, принятия ошибочных решений является несимметричное распределение набора данных, представленных в виде показателей финансовой эффективности. Это подтверждает необходимость проведения первоначальной проверки на наличие нормального распределения, что необходимо при использовании наиболее распространенных статистических методов [2; 4–6].

Нормальное распределение характеризуется теоретически гладкой гистограммой, функция которой образует так называемый колокол без случайных отклонений с центральным пиком и затухающим распределением с обеих сторон. Эта кривая представляет собой идеально распределенный набор данных с наибольшей концентрацией значений в центральной области диапазона. Построение гистограммы является наиболее простым графическим способом проверки характера распределения, однако наличие условий, таких как минимальное значение набора из 50 точек и одинаковая ширина интервалов, часто показывает неадекватность использования отдельного подхода [7].

С исходной асимметрией можно справиться, заменив каждое значение ряда другим числом,

например логарифмом этого показателя. Логарифм может преобразовать искаженные данные, поскольку он растягивает шкалу значений близкую к нулю, что приводит к распределению небольших значений, сгруппированных вместе. Аналогичная ситуация наблюдается и в области крупных ценностей, которые собираются в более составную часть. Однако условие применимости этого подхода только для положительных чисел приводит к необходимости применения других преобразований [8].

Цель данного исследования – провести поэтапную проверку распределения данных на наличие симметрии, что помогает минимизировать ошибочное распределение вероятностей при дальнейшем прогнозировании.

Проверка по коэффициенту вариации

В качестве примера анализируемых данных взято распределение валового внутреннего продукта (ВВП) в период с 2000 по 2020 г. В **табл. 1** представлены значения изучаемого показателя, расположенные в порядке возрастания, что необходимо в статистических проверках по следующим причинам:

1. Во многих случаях важно не абсолютное значение, а то, как показатели соотносятся между собой и соответствуют друг другу. Присвоение рангов в порядке возрастания помогает установить абстрактное положение каждого наблюдения в массиве данных, что является наиболее информативным для анализа.

2. Ранжирование минимизирует чувствительность к выбросам или экстремальным значениям, которые, в свою очередь, могут исказить результаты анализа.

На **рис. 1** представлена поэтапная схема проверки нормальности распределения массива

данных для обеспечения достоверности прогноза. Предложенные методы оценки допустимо применять в условиях малых выборок (до 30 значений) и отсутствии асимметрии данных. Эмпирическая выборка, представленная в табл. 1, соответствует данным ограничениям.

Проверка нормальности распределения с использованием коэффициента вариации является начальным этапом исследования [4]. Значение рассчитанного коэффициента не должно превышать 33 %, в противном случае нецелесообразно проводить дальнейшую проверку на нормальность, поскольку такие распределения должны

быть преобразованы исходя из дальнейшей минимизации коэффициента вариации [5]. Стандартная мера дисперсии распределения вероятностей рассчитывается по формуле

$$V = \frac{S_n}{\bar{x}}, \quad S_n = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

где V – коэффициент вариации, уд. ед.; S_n – стандартное отклонение, уд. ед.; $\bar{x}$ – среднее значение показателя, млрд долл. США; x_i – i -е значение показателя, млрд. долл. США; n – размер выборки, шт.

Таблица 1 / Table 1

Распределение валового внутреннего продукта РФ

Distribution of Russia's gross domestic product

Год	Валовой внутренний продукт, млрд долл. США	Год	Валовой внутренний продукт, млрд долл. США	Год	Валовой внутренний продукт, млрд долл. США
2000	278,07	2016	1282,66	2010	1632,84
2001	328,27	2008	1309,17	2019	1693,32
2002	369,9	2015	1363,7	2007	1778,39
2003	460,74	2009	1391,68	2011	2044,61
2004	632,76	2020	1491,73	2014	2056,58
2005	817,75	2017	1578,41	2012	2202,67
2006	1059,99	2018	1630,66	2013	2289,24

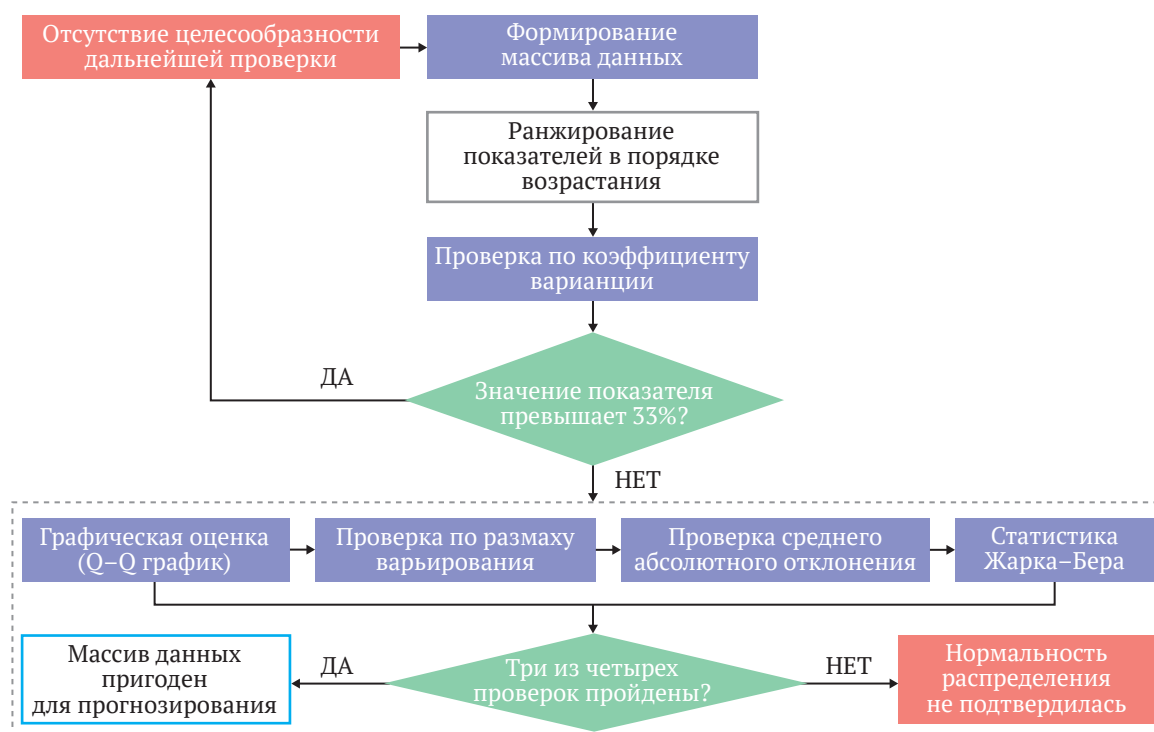


Рис. 1. Схема проверки нормальности распределения массива данных

Fig. 1. Scheme for testing the normality of a data array distribution

Построение графиков квантилей

Одним из этапов проверки характера распределения данных является графическое отображение преобразованного массива с помощью обратной функции стандартного нормального распределения – графики квантилей (Q–Q plots). Данные графики структурированы на основе квантилей распределений, таких как эмпирическое и теоретически ожидаемое наблюдение [6–8]. Прямая линия, выходящая под углом 45 градусов из третьего квадранта в первый, является доказательством нормального распределения (рис. 2). График квантилей служит альтернативой построения гистограмм выраженной формы при оценке небольших по объему совокупностей [9–11].

Проведенная проверка характеризует нормальность распределения данных. Однако необходимо учитывать тот факт, что визуальное обоснование графиков квантилей основывается на многочисленном опыте в виду того, что малые выборки с небольшим числом наблюдений способны исказить результат из-за наличия выбросов [12]. Обоснование этому представлено на рис. 3, который иллюстрирует случайно сгенерированные массивы данных из 20 наблюдений.

В случае массивов данных, превышающих 50 элементов, корректным условием является сравнение функций плотности теоретического нормального распределения и эмпирического. Данная методика основывается на кластеризации точек наблюдений и полигонного распределения частот.

В рамках данной статьи ограничимся использованием проверок нормальности распределения массива данных на основе среднего абсолютного отклонения, размаха варьирования и статистике Жарка–Бера [13–15].

Среднее абсолютное отклонение

Следующим этапом проверки после прохождения коэффициента вариации является критерий среднего абсолютного отклонения (CAO), рассчитываемый как

$$CAO = \frac{\sum_{i=1}^n |x_i - \bar{x}|^2}{n}. \quad (2)$$

Известно, что для теоретического нормального распределения отношение CAO к стандартному отклонению равно $\sqrt{2/\pi}$ [1; 10]. Для

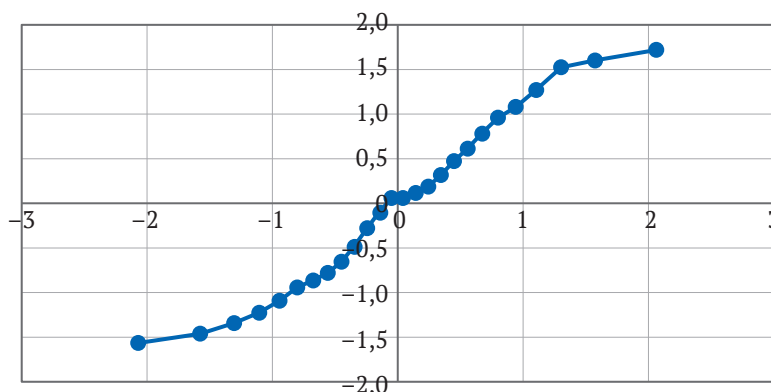


Рис. 2. График квантилей исследуемого массива данных

Fig. 2. Graph of quantiles of the studied data array

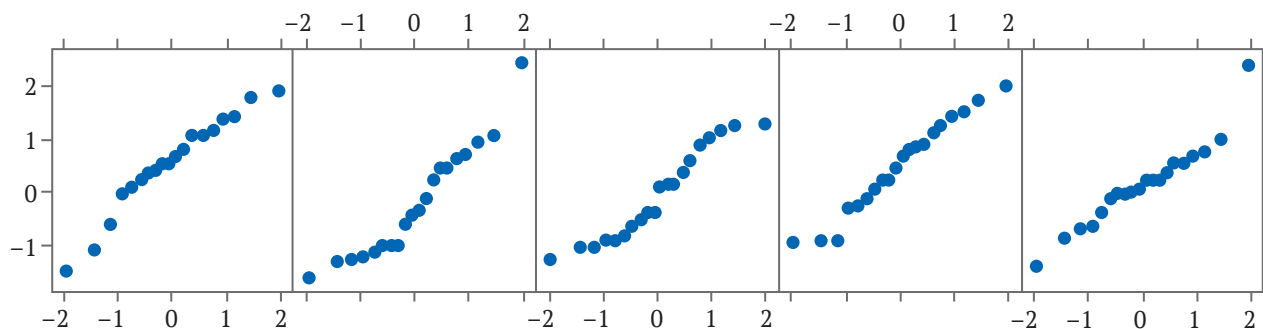


Рис. 3. Отражение отсутствия выраженной прямой на графиках квантилей

Fig. 3. Reflection of the absence of a pronounced straight line on the quantile graphs

выборки, имеющей примерно нормальный закон распределения, должно выполняться следующее соотношение:

$$\left| \frac{CAO}{S_n} = \sqrt{\frac{2}{\pi}} \right| < \frac{0,4}{\sqrt{n}}, \quad (3)$$

где S_n – стандартное отклонение, уд. ед.; n – размер выборки, шт.

Проверка по размаху варьирования

Статистический тест позволяет сделать анализ массивного спектра выборок размером до 1000 наблюдений. Диапазон размаха варьирования (R) определяется как разница максимального и минимального значения выборки [3; 11]. Критериальное отношение (S_r) определяется как

$$S_r = \frac{R}{S_n}. \quad (4)$$

Полученное значение необходимо сопоставить с величиной из критической таблицы соотношений верхнего и нижнего пределов относительно выбранного уровня значимости [3; 12]. Гипотеза о нормальности является удовлетворительной в случае попадания рассчитанного коэффициента в диапазон пределов для статистической значимости в 5 % ($p = 0,05$) [3; 14–20].

Статистика Жарка–Бера

Статистика представляет собой критерий согласия на наличие асимметрии и эксцесса в выборочных данных, соответствующих нормальному распределению, и рассчитывается по формуле

$$JB = \frac{n}{6} \left(S^2 + \frac{(K-3)^2}{4} \right), \quad (5)$$

где JB – эмпирическое значение статистики Жарка–Бера, уд. ед.; n – количество элементов в выборке, шт.; S – выборка асимметрии, уд. ед.; K – выборка эксцесса, уд. ед.

Асимметрия S рассчитывается по следующей формуле:

$$S = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}. \quad (6)$$

Эксцессы K определяются формулой

$$K = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}. \quad (7)$$

Тестовая статистика имеет распределение χ^2 с двумя степенями свободы. Нулевая гипотеза о нормальном распределении остатков рядов отвергается при заданном уровне значимости, если статистика Жака–Бера больше критического значения статистики χ^2 (критическое значение χ^2 для уровня значимости 0,01 равно 9,2103, для значимости уровень 0,05–5,9916).

Результаты

По результатам исследований, представленных выше, коэффициент вариации составил 0,295, или 29,5 %, что не превышает потолок в 33 % и указывает на то, что проверку можно продолжать.

Условия возможности проверки CAO, рассчитанные по формуле (2), также подтвердили наличие нормального распределения данных.

$$\left| \frac{3479,44}{4197,99} - 0,7979 \right| < \frac{0,4}{\sqrt{26}}, \quad (8)$$

$$0,0309 < 0,0784.$$

На уровне значимости $p = 0,1$ гипотеза о размахе варьирования (полученное значение 3,41) удовлетворяется, поскольку интервал при $n = 21$ составляет (3,3–4,35), что говорит о нормальности распределения массива данных в рамках данного теста.

Рассчитанное по формуле (5) значение статистики Жака–Бера больше критического значения распределения (6,82 > 5,99), что также подтверждает нормальное распределение исследуемого показателя

$$JB = \frac{21}{6} \left(0,508^2 + \frac{(0,52-3)^2}{4} \right) = 6,82. \quad (9)$$

Заключение

В реальных условиях руководству компании необходимо принимать рациональные управленческие решения, чтобы минимизировать риск неблагоприятного исхода. Одним из способов прогнозирования микро- и макроэкономических показателей является использование статистических процедур, требующих определенных условий достоверности данных. В их число входит оценка нормальности распределения, отсутствие которой искажает конечный результат распределения вероятностей, что приводит к неэффективному использованию средств и снижению потенциала предприятия.

В статье предложен ряд подходов, позволяющих количественно оценить наличие асимметрии распределения валового внутреннего про-

дукта РФ с 2000 по 2020 г. На основании условий проверки принимается окончательное решение о возможности использования данных в дальнейших расчетах. Все примененные подходы, а именно тестирование по коэффициенту вариации (29,5 %), среднему абсолютному отклонению (0,0309), диапазону вариации (3,41), статистике Жака–Бера (6,82) и построению графиков кван-

тилей, доказали нормальность распределения исследуемого ряда, что подтверждает их пригодность для проведения оценок.

Экономический смысл исследования заключается в использовании на первом этапе проверки массива исходных данных на наличие нормального распределения, что снижает расходы, происходящие по причине ошибок прогнозирования.

Список литературы / References

1. Litimein O., Laksaci A., Ait-Hennani L., Mechab B., Rachdi M. Asymptotic normality of the local linear estimator of the functional expectile regression. *Journal of Multivariate Analysis*. 2024;202(2):10528. <https://doi.org/10.1016/j.jmva.2023.105281>
2. Mohammedi M., Bouzebda S., Laksaci A. The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data. *Journal of Multivariate Analysis*. 2021;181:104673. <https://doi.org/10.1016/j.jmva.2020.104673>
3. Aneiros-Pérez G., Cao R., Ricardo F., Genest Ch., Vieu Ph. Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*. 2019;170:3–9. <https://doi.org/10.1016/j.jmva.2018.11.007>
4. Bellini F., Bignozzi V., Puccetti G. Conditional expectiles, time consistency and mixture convexity properties. *Insurance: Mathematics and Economics*. 2018;82:117–123. <https://doi.org/10.1016/j.insmatheco.2018.07.001>
5. Костюхин Ю.Ю., Богачев А.С. Управление инвестиционной привлекательностью предприятия в период высокой волатильности рынка на основе прогнозирования ожиданий. *Экономика промышленности*. 2024;17(1):20–28. <https://doi.org/10.17073/2072-1633-2024-1-1265>
Kostyukhin Yu.Yu., Bogachev A.S. Managing investment attractiveness of a company during a period of high market volatility based on forecasting expectations. *Russian Journal of Industrial Economics*. 2024;17(1):20–28. (In Russ.). <https://doi.org/10.17073/2072-1633-2024-1-1265>
6. Kara-Zaitri L., Laksaci A., Rachdi M., Vieu Ph. Data-driven KNN estimation in nonparametric functional data analysis. *Journal of Multivariate Analysis*. 2017;153:176–188. <https://doi.org/10.1016/j.jmva.2016.09.016>
7. Litimein O., Laksaci A., Mechab B., Bouzebda S. Local linear estimate of the functional expectile regression. *Statistics & Probability Letters*. 2023;192(2):109682. <https://doi.org/10.1016/j.spl.2022.109682>
8. Ding H., Lu Zh., Zhang J., Zhang R. Semi-functional partial linear quantile regression. *Statistics & Probability Letters*. 2018;142(6):92–101. <https://doi.org/10.1016/j.spl.2018.07.007>
9. Pisciă D., Dammers R., Boersma E., Volovici V. Tents of good practice in regression analysis. A brief tutorial. *World Neurosurgery*. 2022;161:230–239.e6. <https://doi.org/10.1016/j.wneu.2022.02.112>
10. Yurii A.R. A complex approach to evaluating the innovation strategy of a company to determine its investment attractiveness. *Procedia – Social and Behavioral Sciences*. 2013;99:562–571. <https://doi.org/10.1016/j.sbspro.2013.10.526>
11. Shinno H., Yoshioka S., Marpaung S., Hachiga S. Qualitative SWOT analysis on the global competitiveness of machine tool industry. *Journal of Engineering Design*. 2006;17(3):251–258.
12. Barberis N. Investing for the long run when returns are predictable. *Journal of Finance*. 2000;55(1):225–264.
13. Antunes F., Ribeiro B., Pereira F. Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing*. 2017;60:831–843. <https://doi.org/10.1016/j.asoc.2017.06.043>
14. Daemi A., Kodamana H., Huang B. Gaussian process modelling with Gaussian mixture likelihood. *Journal of Process Control*. 2019;81(C):209–220. <https://doi.org/10.1016/j.jprocont.2019.06.007>
15. Dierkes M., Erner C., Zeisberger S. Investment horizon and the attractiveness of investment strategies: A behavioral approach. *Journal of Banking & Finance*. 2010;34(5):1032–1046. <https://doi.org/10.1016/j.jbankfin.2009.11.003>
16. Xiao J., Yu P., Song X., Zhang Z. Statistical inference in the partial functional linear expectile regression model. *Science China Mathematic*. 2022;65(12):2601–2630. <https://doi.org/10.1007/s11425-020-1848-8>
17. Chen C., Guo S., Qiao X. Functional linear regression: Dependence and error contamination. *Journal of Business & Economic Statistics*. 2022;40(1):444–457. <https://doi.org/10.1080/07350015.2020.1832503>
18. Rachdi M., Laksaci A., Al-Kandari N.M. Expectile regression for spatial functional data analysis (sFDA). *Metrika*. 2022;85(5):627–655. <https://doi.org/10.1007/s00184-021-00846-x>
19. Cui X., Lin H., Lian H. Partially functional linear regression in reproducing kernel Hilbert spaces. *Computational Statistics & Data Analysis*. 2020;150(6):106978. <https://doi.org/10.1016/j.csda.2020.106978>
20. Daouia A., Girard S., Stupfler G. Extremiles: A new perspective on asymmetric least squares. *Journal of the American Statistical Association*. 2019;114(527):1366–1381. <https://doi.org/10.1080/01621459.2018.1498348>

Информация об авторах

Юрий Юрьевич Костюхин – д-р экон. наук, профессор, Национальный исследовательский технологический университет «МИСИС», 119049, Москва, Ленинский просп., д. 4, стр. 1, Российская Федерация; ORCID: <https://orcid.org/0000-0003-2108-0241>; e-mail: kostuhinyury@mail.ru

Андрей Сергеевич Богачев – ассистент кафедры промышленного менеджмента, Национальный исследовательский технологический университет «МИСИС», 119049, Москва, Ленинский просп., д. 4, стр. 1, Российская Федерация; ORCID: <https://orcid.org/0009-0008-2915-742X>; e-mail: andr.bogachiov@yandex.ru

Information about the authors

Yuriy Yu. Kostyukhin – Dr.Sci. (Econ.), Professor, National University of Science and Technology “MISIS”, 4-1 Leninskiy Ave., Moscow 119049, Russian Federation; ORCID: <https://orcid.org/0000-0003-2108-0241>; e-mail: kostuhinyury@mail.ru

Andrey S. Bogachev – Assistant of the Department of Industrial Management, National University of Science and Technology “MISIS”, 4-1 Leninskiy Ave., Moscow 119049, Russian Federation; ORCID: <https://orcid.org/0009-0008-2915-742X>; e-mail: andr.bogachiov@yandex.ru

Поступила в редакцию 29.12.2024; поступила после доработки 21.04.2025; принята к публикации 21.05.2025

Received 29.12.2024; Revised 21.04.2025; Accepted 21.05.2025