

ЧЕБЫШЕВСКИЙ СБОРНИК

Том 26. Выпуск 1.

УДК 519.2

DOI 10.22405/2226-8383-2025-26-1-149-156

О формировании норм на основе регрессионных моделей по малым выборкам¹

Р. А. Жуков

Жуков Роман Александрович — доктор экономических наук, кандидат физико-математических наук, доцент, Тульский филиал Финансового университета при Правительстве РФ (г. Тула).

e-mail: pluszh@mail.ru

Аннотация

Представлено расширение метода формирования нормативов для структурных элементов сложной системы, основанное на методологии байесовских интеллектуальных измерений и эконометрического моделирования на малых выборках. Реализация метода продемонстрирована на примере сельского хозяйства Тульской области.

Ключевые слова: плотность распределения вероятностей, байесовские интеллектуальные измерения, нормы, эконометрическое моделирование, регрессия, сложная система.

Библиография: 17 названий.

Для цитирования:

Жуков, Р. А. О формировании норм на основе регрессионных моделей по малым выборкам // Чебышевский сборник, 2025, т. 26, вып. 1, с. 149–156.

CHEBYSHEVSKII SBORNIK

Vol. 26. No. 1.

UDC 519.2

DOI 10.22405/2226-8383-2025-26-1-149-156

On the formation of norms based on regression models for small samples

R. A. Zhukov

Zhukov Roman Aleksandrovich — doctor of economic sciences, candidate of physical and mathematical sciences, associate professor, Financial University under the Government of the Russian Federation (Tula Branch) (Tula).

e-mail: pluszh@mail.ru

¹Статья подготовлена по результатам исследований, выполненных за счет бюджетных средств по государственному заданию Финуниверситета

Abstract

The extension of the method of forming norms for structural elements of a complex system based on the methodology of Bayesian intelligent measurements and econometric modeling in small samples is presented. The implementation of the method is demonstrated using the example of agriculture in the Tula region.

Keywords: probability distribution density, Bayesian intelligent measurements, norms, econometric modeling, regression, complex system.

Bibliography: 17 titles.

For citation:

Zhukov, R. A. 2025, "On the formation of norms based on regression models for small samples", *Chebyshevskii sbornik*, vol. 26, no. 1, pp. 149–156.

1. Введение

Для оценки соответствия функционирования сложных систем заданным режимам, которые идентифицируются набором норм на основе стандартов [1], экспертных оценок [2] или специальных методик [3], в том числе на основе экономико-математического моделирования [4], возникает необходимость учета условий (факторов), которые влияют на результативность таких систем. При этом характер и степень влияния факторов, представленных малыми выборками, и которые обладают значительной информационной неопределенностью, существенно усложняют процесс построения адекватных норм. В этом аспекте применение байесовского подхода [5, 6] и мягких норм [7] в рамках регуляризующего байесовского подхода [8] является целесообразным, в том числе при учете характера и степени влияния факторов как на результаты функционирования структурных элементов системы, так и на их нормативы. Следует отметить, что данные обладают относительно небольшой глубиной по времени, а, следовательно, выборка оказывается малой, что не позволяет методологически обоснованно применять методы математической статистики для построения регрессионных моделей [6], которые могут быть использованы для конструирования нормативов для результатов функционирования структурных элементов системы.

2. Методология и материалы исследования

В исследовании будем опираться на методологию байесовских интеллектуальных измерений (БИИ) [8], а также на результаты, полученные в [9] и их расширение на нормы различных типов [7]

Пусть имеется k_i структурный элемент системы, характеризуемый в период времени t ($t = 1, \dots, T \in N$) значениями результативного $y_{k,i}(t)$ и факторных $x_{k,i,j}(t)$ признаков – условий функционирования k_i – влияющих факторов ($k = 1, \dots, K_k \in N$, $i = 1, \dots, I \in N$, $j = 1, \dots, J \in N$). Каждые результативный и факторные признаки характеризуются наборами пар чисел $(h_{(\cdot),l}(t); p_{(\cdot),l}^{ap}(t))$: $h_{(\cdot),l}(t)$ – значение признака $(\cdot)(t)$, которое соответствует положению репера на числовой шкале с диапазоном динамических ограничений ($l = 1, \dots, L_r \in N$) и положению на сопряженной с числовой – лингвистической шкале ($l = 1, \dots, L_c \in N$, $c = 9$ – классы: от предельно ниже нормы до предельно выше нормы); $p_{(\cdot),l}^{ap}(t)$ – вероятность, того, что значение $(\cdot)(t) = h_{(\cdot),l}(t)$. В данном случае $(\cdot)(t)$ есть $y_{k,i}(t)$ или $x_{k,i,j}(t)$.

Тогда норма для результативного признака может быть определена набором пар чисел $(h_{(\cdot),l}(t); \bar{p}_{(\cdot),l}^{ap}(t))$, где апостериорная вероятность $\bar{p}_{(\cdot),l}^{ap}(t)$ определяется по формуле [10], аналогичной [7]:

$$\bar{p}_{(\cdot),l}^{ap}(t) = \frac{P_{y_{k,i},l}^a(h_{y_{k,i},l}(t-1))P(\tilde{h}_{y_{k,i},l}(t))}{\sum_{j=1}^L P_{y_{k,i},j}^a(h_{y_{k,i},j}(t-1))P(\tilde{h}_{y_{k,i},j}(t))}, \quad (1)$$

где $h_{y_{k,i},l}(t-1)$, $\tilde{h}_{y_{k,i},l}(t)$ – значение положения репера (класса) и его оценка соответственно; a – обозначение априорной вероятности; ap – обозначение апостериорной вероятности; $t-1$ означает, что априорные значения вероятности получены в предыдущем периоде.

Априорная вероятность $P_{y_{k,i},l}^a(h_{y_{k,i},l}(t-1))$ на первом шаге принимается одинаковой $\forall l$ и определяется как:

$$P_{y_{k,i},l}^a(h_{y_{k,i},l}(t-1)) = \frac{1}{L_r}, P_{y_{k,i},l}(h_{y_{k,i},l}(t-1)) = \frac{1}{L_c}. \quad (2)$$

Вероятность $P(\tilde{h}_{y_{k,i},l}(t))$ можно вычислить по рекуррентному соотношению [10], аналогичному модифицированной байесовской свертке [11]:

$$P(\tilde{h}_{y_{k,i},l}(t)) = \frac{1}{J} \sum_{j=1}^J P_{x_{k,i,j},l}^{ap}(h_{x_{k,i,j},l}(t) | \bigcup_{s=1}^J h_{x_{k,i,s},l}(t)). \quad (3)$$

Для двух влияющих факторов апостериорную вероятность $P_{x_{k,i,j},l}^{ap}(h_{x_{k,i,j},l}(t) | \bigcup_{s=1}^J h_{x_{k,i,s},l}(t))$ можно представить как:

$$P_{x_{k,i,j},l}^{ap}(h_{x_{k,i,1},l}(t) | h_{x_{k,i,2},l}(t)) = P(\tilde{h}_{y_{k,i},l}(t)) = \frac{\sum_{q=1}^Q P^{ap}(H_{x_{k,i,1},l}|S) P^{ap}(H_{x_{k,i,2},q}|S)}{\sum_{l=1}^L \sum_{q=1}^Q P^{ap}(H_{x_{k,i,1},l}|S) P^{ap}(H_{x_{k,i,2},q}|S)}, \quad (4)$$

где L, Q – значимые альтернативные значения для первого и второго фактора соответственно; S – событие – совместное появление оценок $\tilde{h}_{(\cdot),l}(t)$ для $(\cdot)$; $H_{(\cdot),l}$ – набор гипотез или альтернативных решений.

Для преобразования мягкой нормы к единственному значению рассчитывается средневзвешенное значение [10]:

$$\bar{y}_{k,i} = \sum_{l=1}^L \tilde{h}_{y_{k,i},l}(t) \cdot \bar{p}_{y_{k,i},l}^{ap}(t). \quad (5)$$

Применение формул (1)–(5) учитывает влияние факторов, характеризующих условия функционирования структурного элемента системы, но не учитывает степень влияния каждого из них на результативный признак. С целью учета такого влияния формулу (3) можно дополнить весовыми коэффициентами:

$$P(\tilde{h}_{y_{k,i},l}(t)) = \frac{\sum_{j=1}^J w_{k,i,j} \cdot P_{x_{k,i,j},l}^{ap}(h_{x_{k,i,j},l}(t) | \bigcup_{s=1}^J h_{x_{k,i,s},l}(t))}{\sum_{j=1}^J w_{k,i,j}}, \quad (6)$$

где $w_{k,i,j}$ – весовой коэффициент для значений фактора $x_{k,i,j}(t)$.

Связь между результативным признаком и фактором в простейшем случае можно представить в виде:

$$y_{k,i}^*(t) = w_{i,1} \cdot x_{k,i,1}^*(t) + \epsilon_{k,i,1}^*(t), \quad (7)$$

где $w_{i,1}$ – параметры функции $f_i(\cdot) = \hat{y}_i^*$; “*” – стандартизованные (центрированные и нормированные) значения результативного и факторного признаков.

Из эконометрики известно, что в случае наличия только одного факторного признака, $w_{i,1}$ – есть не что иное, как парный коэффициент корреляции между переменными $y_{k,i}$ и $x_{k,i,1}$.

Поэтому, с целью учета силы влияния факторов, в первом приближении можно использовать корреляционные коэффициенты в качестве весов в формуле (6).

В случае наличия нескольких факторов в качестве весовых коэффициентов могут выступать коэффициенты стандартизированной регрессионной модели, представляющей собой преобразованную производственную функцию (ПФ) [4]:

$$\hat{y}_{k,i}^* = \sum_{j=1}^J w_{i,j} \cdot x_{k,i,j}^* \quad (8)$$

Тогда весовые коэффициенты будут соответствовать стандартизованным коэффициентам ПФ, найденным по малым выборкам [12].

Для получения выборки достаточного объема (число точек должно быть не менее 7000 [6]) для каждого из 9 классов сопряженной лингвистической шкалы генерируется выборка по треугольному распределению для каждого из значений влияющих факторов $x_{k,i,j}(t)$ периода(ов) t , представленных набором пар чисел $(h_{(\cdot),i}(t); p_{(\cdot),i}^{ap}(t))$ на сопряженной числовой шкале. Для результативного признака генерация может осуществляться в двух вариантах: 1) по мягким нормам, полученным с помощью (1)–(4); 2) по представлениям $y_{k,i}(t)$. Сгенерированная таким образом выборка по малым выборкам будет обладать достаточной мощностью, чтобы методологически обоснованно применить методы математической статистики для нахождения параметров модели, в частности метод наименьших квадратов (МНК), для расчета норм по формуле (6).

В случае, если коэффициент $w_{i,j} < 0$, то осуществляется реверсия вероятностей значений фактора $x_{k,i,j}(t)$ относительно нормы $(h_{(\cdot),i}(t); rev_{norm}(p_{(\cdot),i}^{ap}(t)))$, а в формулу (8) подставляется модуль $|w_{i,j}|$.

3. Результаты

Представленный метод был апробирован на примере Тульской области. В качестве результативного признака был выбран объем валового регионального продукта (ВРП) для раздела А (сельское хозяйство) по ОКВЭД 2. В качестве факторных признаков были использованы стоимость основных фондов ($x_{1,1}$), среднегодовая численность занятых ($x_{1,2}$), инвестиции в основной капитал ($x_{1,3}$) по соответствующему разделу, а также внесение минеральных (x_3) и органических удобрений (x_4). Источником данных послужила информация Росстата, полученная из открытых источников [13] за 2017–2022 гг. Стоимостные показатели были скорректированы на уровень инфляции [14] и приведены к 2017 году. Для построения мягких норм была использована программная платформа «Инфоаналитик 2.0» [15]; применены модуль генерации выборки с заданным законом распределения [16] (генерировались данные для каждого периода, $N_{i,j}=7000$ наблюдений), а также программный комплекс «ЭФРА» [17] для поиска параметров стандартизованных моделей (использовались выборки для каждого периода и объединенная по t выборка). В качестве ПФ была выбрана степенная мультипликативная функциональная форма модели.

Результаты представлены на рис. 1.

Первое значение в ячейках соответствует коэффициентам, полученным по выборке, построенной по мягким нормам, вычисленным с помощью (1)–(4), второе значение – по выборке, построенной по представлениям $y_{k,i}(t)$. Последний столбец соответствует объединенной выборке за 2017–2022 гг. Последние 2 строки включают значения коэффициента детерминации R^2 и среднюю относительную ошибку аппроксимации E_{otn} соответственно, которые свидетельствуют о значимости коэффициентов (значения близки к 1, модели качественные) и высокой точности моделей.

Параметры / Период выборки	2017	2018	2019	2020	2021	2022	2017-2022
$x_{1,1}$	0,140/-0,017	0,058/0,441	0,294/0,182	0,198/0,045	0,242/0,040	0,135/0,045	0,260/0,558
$x_{1,2}$	0,198/0,401	0,295/0,016	0,307/0,057	0,243/0,474	0,175/0,048	0,106/0,605	0,309/-0,129
$x_{1,3}$	0,131/0,030	0,234/0,063	0,120/0,100	0,125/0,223	0,248/0,538	0,352/0,080	0,232/-0,134
x_3	0,137/0,180	0,031/0,429	0,078/0,449	0,305/0,116	0,19/0,259	0,088/-0,001	0,195/0,456
x_4	0,396/0,357	0,380/0,034	0,195/0,210	0,137/0,129	0,146/0,102	0,311/0,259	0,267/0,113
R^2	0,922/0,852	0,924/0,923	0,917/0,941	0,928/0,905	0,924/0,914	0,901/0,921	0,918/0,897
$E_{tot}, \%$	5,566/5,527	3,178/3,647	3,740/2,258	3,963/2,671	3,600/2,320	5,156/2,551	4,769/6,453

Рис. 1: Значения стандартизованных коэффициентов с базовыми статистическими характеристиками

Полученные значения стандартизованных коэффициентов были подставлены в «Инфоаналитик 2.0», что позволило сформировать мягкие нормы для результативного признака (ВРП, раздел А), для каждой из которых по формуле (5) были получены единственные значения норм.

При фактическом значении скорректированного объема ВРП 47750,899 млн руб. в 2022 г. норма по первому варианту (нормативная регрессия) составила 46532,881 млн руб.; по второму варианту (фактическая регрессия) – 43002,843 млн руб.; по регрессиям, построенным по объединенным выборкам значения соответственно составили 46356,170 млн руб. и 42710,388. То есть расчет по первому варианту дает более высокие нормы по сравнению со вторым вариантом. В случае использования коэффициентов корреляции в качестве весовых коэффициентов значение в 2022 году оказалось равным 46280,667 млн руб., без учета весовых коэффициентов (формула (4)) – 45862,205 млн руб., что ниже, чем норма, полученная по первому варианту.

На рис. 2 представлена оценка фактического значения скорректированного ВРП (раздел А) по норме, рассчитанной по первому варианту на основе построенной нормативной регрессии по сгенерированным данным для 2022 г.

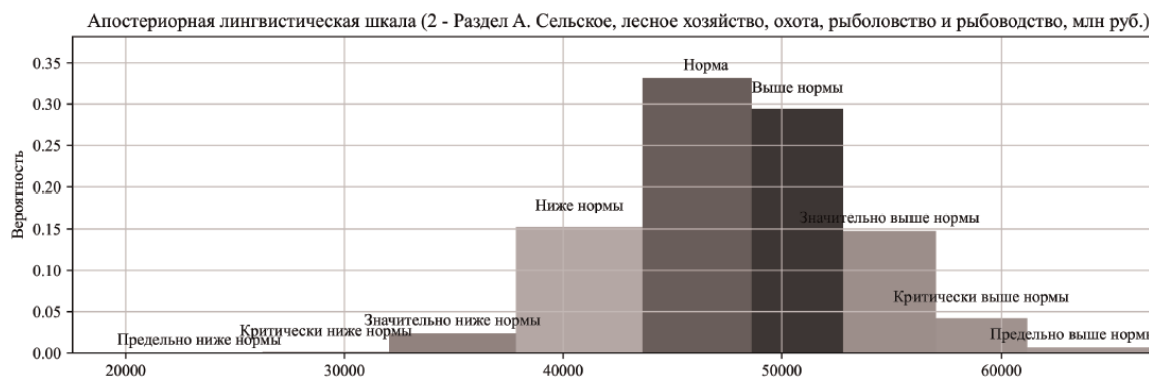


Рис. 2: Оценка ВРП Раздел А Тульской области в 2022 г. на основе нормативной регрессии, млн руб.

Из рис. 2 видно, что функционирование элемента системы, характеризуемого скорректированным объемом ВРП по разделу А, находится в классе «норма» с вероятностью 0,332; в классе «выше нормы» с вероятностью 0,294; «ниже нормы» – 0,153; «значительно выше нормы» – 0,148 (что ниже заданного порога значимости 0,150).

Таким образом, представленный метод учитывает силу влияния факторов – условий функционирования структурных элементов системы, что расширяет возможности ранее разработанного метода, широко используемого в подобных исследованиях.

4. Заключение

Представленный и апробированный на примере Тульской области метод формирования норм на основе регрессионных моделей, построенных по малым выборкам, может выступать в качестве составляющей оценки результативности функционирования элемента сложной системы, действующей в условиях информационной неопределенности. Метод может быть применен для систем различного типа, в том числе технических, информационных и социо-эколого-экономических систем.

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

1. Техническое регулирование: технические регламенты и стандартизация : учебное пособие / сост. И. Ю. Матушкина, Л. А. Онищенко. — Екатеринбург : Изд-во Урал. ун-та, 2018. 208 с.
2. Мадера А. Г. Моделирование и оптимизация бизнес-процессов и процессных систем в условиях неопределенности // Бизнес-информатика. 2017. № 4. С. 74–82.
3. Trung D. D. Development of data normalization methods for multi-criteria decision making: Applying for MARCOS method // Manufacturing Review. 2022. Vol. 9. No. 22.
4. Жуков Р. А. Метод оценки результатов функционирования иерархических социально-экономических систем на основе агрегированной производственной функции // Экономика и математические методы. 2021. Т. 57. № 3. С. 17-31.
5. Айвазян С. А. Байесовский подход в эконометрическом анализе // Прикладная эконометрика. 2008. № 1 (9). С. 93-130.
6. Прокопчина С. В. Байесовские интеллектуальные технологии в задачах моделирования закона распределения в условиях неопределенности: монография. М.: ИД «НАУЧНАЯ БИБЛИОТЕКА», 2020. 292 с.
7. Прокопчина С. В. Основы теории шкалирования в экономике: учебное пособие. — М.: ИД «НАУЧНАЯ БИБЛИОТЕКА», 2021. 272 с.
8. Прокопчина С. В. Байесовские интеллектуальные измерения. — М.: ИД «НАУЧНАЯ БИБЛИОТЕКА», 2021. 495 с.
9. Жуков Р. А., Прокопчина С. В. О Формировании мягких норм для оценки функционирования сложных систем // Чебышевский сборник. 2024. Т. 25. № 3 (94). С. 351-358.
10. Жуков Р. А., Прокопчина С. В., Плинская М. А., Желуница М. А. Построение системы динамических нормативов для оценки функционирования сложных систем на примере субъектов Центрального федерального округа // Бизнес-информатика. 2024. Т. 18. № 4. С. 46-60.
11. Прокопчина С. В., Щербаков Г. А., Ефимов Ю. В. Моделирование социально-экономических систем в условиях неопределенности: учебное пособие. — М.: ИД «НАУЧНАЯ БИБЛИОТЕКА», 2019. 508 с.
12. Жуков Р. А., Прокопчина С. В., Плинская М. А., Желуница М. А. Моделирование функциональных связей региональных экономических систем по малым выборкам на основе байесовских интеллектуальных измерений // Journal of Applied Economic Research. 2024. Т. 23. № 3. С. 721-750.

13. Федеральная служба государственной статистики Российской Федерации.
URL: <https://gks.ru>.
14. Таблицы уровней инфляции. URL: <https://уровень-инфляции.рф/таблицы-инфляции>
15. Жуков Р. А., Прокопчина С. В. Программный комплекс «Инфоаналитик 2.0». Свидетельство о государственной регистрации программы для ЭВМ No 2024617544 от 03.04.2024.
URL: <https://elibrary.ru/item.asp?id=65627372>
16. Жуков Р. А., Прокопчина С. В., Гиниятов И. А., Николина Е. М. Применение библиотеки «Байесовская математическая статистика» в программном комплексе «Инфоинтегратор» // Мягкие измерения и вычисления. 2022. Т. 54. № 5. С. 99-108.
17. Жуков Р. А. Подход к оценке функционирования иерархических социально-экономических систем и принятию решений на базе программного комплекса «ЭФРА» // Бизнес-информатика. 2020. Т. 14. № 3. С. 82-95.

REFERENCES

1. Matushkina, I. Y., Onishchenko L. A., 2018, "Technical regulation: technical regulations and standardization: a textbook", *Yekaterinburg, Ural University Publishing House*, 208 p.
2. Madera, A. G. 2017, "Modeling and optimization of business processes and process systems under conditions of uncertainty", *Business Informatics*, no. 4, pp. 74–82.
3. Trung, D. D. 2022, "Development of data normalization methods for multi-criteria decision making: Applying for MARCOS method", *Manufacturing Review*, vol. 9, no. 4, pp. 15.
4. Zhukov, R. A. 2021, "Method for assessing the results of hierarchical socio-economic systems' functioning based on the aggregated production function", *Economics and Mathematical Methods*, vol. 57, no. 3, pp. 17-31.
5. Ayvazyan, S. A. 2008, "Bayesian approach in econometric analysis", *Applied Econometrics*, no. 1 (9), pp. 93-130.
6. Prokopchina, S. V. 2020, "Bayesian intelligent technologies in problems of modeling the distribution law under uncertainty: monograph", *M., Publishing house "SCIENTIFIC LIBRARY"*, 292 p.
7. Prokopchina, S. V. 2021, "Fundamentals of the theory of scaling in economics: textbook", *M., Publishing house "SCIENTIFIC LIBRARY"*, 272 p.
8. Prokopchina, S. V. 2021, "Bayesian intellectual measurements", *M., Publishing house "SCIENTIFIC LIBRARY"*, 495 p.
9. Zhukov, R. A., Prokopchina, S. V. 2024, "On the formation of soft norms for evaluating the functioning of complex systems", *Chebyshevskii Sbornik*, vol. 25, no. 3 (94), pp. 351-358.
10. Zhukov, R. A., Prokopchina, S. V., Plinskaya, M. A., Zhelunitsina, M. A. 2024, "Building a system of dynamic norms for evaluating the functioning of complex systems on the example of the regions of the Central Federal District", *Business Informatics*, vol. 18, no. 4, pp. 46-60.
11. Prokopchina, S. V., Shcherbakov, G. A., Efimov, Yu. V. 2019, "Modeling of socio-economic systems in conditions of uncertainty", *M., Publishing house "SCIENTIFIC LIBRARY"*, 508 p.

12. Zhukov, R. A., Prokopchina, S. V., Plinskaya, M. A., Zhelunitsina, M. A. 2024, ‘Modeling of functional relationships of regional economic systems based on small samples based on Bayesian intelligent measurements’, *Journal of Applied Economic Research*, vol. 23, no. 3, pp. 721-750.
13. Federal State Statistics Service of the Russian Federation. URL: <https://gks.ru>.
14. Tables of inflation rates. URL: <https://xn--ctbjnaatncev9av3a8f8b.xn-p1ai/>
15. Zhukov R.A., Prokopchina S.V. The Infoanalyst 2.0 software package. Certificate of state registration of the computer program No. 2024617544 dated 04/03/2024.
URL: <https://elibrary.ru/item.asp?id=65627372>
16. Zhukov, R. A., Prokopchina, S. V., Giniatov, I. A., Nikolina, E. M. 2022, “Application of the Bayesian Mathematical Statistics Library in the Infointegrator Software Package”, *Soft Measurement and Computing*, vol. 54, no. 5, pp. 99–108.
17. Zhukov R. A. 2020, “An approach to assessing the functioning of hierarchical socio-economic systems and decision-making based on the EFRA software package”, *Business Informatics*, vol. 14, no. 3, pp. 82-95.

Получено: 19.12.2024

Принято в печать: 10.03.2025