

Типология данных с использованием технологий разделения смесей вероятностных распределений

Владимир Васильевич Глинский^{а)},
Юлия Николаевна Исмайллова^{а)},
Сергей Евгеньевич Хрущев^{б)},
Людмила Константиновна Серга^{а)}

^{а)} Сибирский институт управления — филиал РАНХиГС, г. Новосибирск, Россия;

^{б)} Институт математики имени С.Л. Соболева Сибирского отделения Российской академии наук, г. Новосибирск, Россия

С использованием информационных ресурсов Федеральной службы государственной статистики, отражающих ряд социально-экономических характеристик российского общества, исследованы неоднородные выборки (по ряду статистических совокупностей), взятые из конечных смесей вероятностных распределений; при этом число смешивающихся распределений (компонент), а также соответствующие им веса и параметры могли быть неизвестными. Содержательно задача разделения (декомпозиции) смесей может быть сведена к определению неизвестных характеристик смешивающихся компонент, а также их весов.

Проанализированы методы решения этой задачи, их преимущества, достоинства и недостатки, условия и область прикладных применений. Осуществлена декомпозиция смесей вероятностных распределений на совокупности субъектов Российской Федерации по трем взаимосвязанным статистическим показателям — уровню безработицы, уровню бедности и уровню насилия, которые можно рассматривать в качестве измерителей достижения целей устойчивого развития для территориального образования и (одновременно) индикаторов социально-экономического «здоровья» территории. Проведены типологии регионов по перечисленным показателям. Выполнено перекрестное сопоставление результатов полученных типологических группировок, позволившее (в том числе) определить субъекты Российской Федерации, в которых прослеживаются как позитивные, так и негативные тенденции в контексте рассматриваемых показателей.

Подчеркивается, что результаты исследования могут быть использованы органами управления для разработки конкретных мер по социально-экономическому развитию России и ее регионов.

Ключевые слова: цели устойчивого развития (ЦУР), межтерриториальная дифференциация, социальная статистика, уровень неравенства, уровень бедности, уровень безработицы, уровень насилия, математико-статистические методы, смеси вероятностных распределений, декомпозиция смесей распределений.

JEL: C02, C12, C13, C38, C40, C60.

doi: <https://doi.org/10.34023/2313-6383-2022-29-6-11-24>.

Для цитирования: Глинский В.В., Исмайллова Ю.Н., Хрущев С.Е., Серга Л.К. Решение проблемы типологии данных с использованием технологий разделения смесей вероятностных распределений. Вопросы статистики. 2022;29(6):11–24.

Data Typology Using Technologies for Separating Mixtures of Probability Distributions

Vladimir V. Glinskiy^{a)},
Yuliya N. Ismailova^{a)},
Sergey E. Khrushchev^{b)},
Lyudmila K. Serga^{a)}

^{a)} The Siberian Institute of Management — Branch of the Russian Presidential Academy of National Economy and Public Administration, Novosibirsk, Russia;

^{b)} Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

The authors studied heterogeneous samples (for a number of statistical populations) taken from finite mixtures of probability distributions, which reflect a number of socio-economic characteristics of the Russian society, using the information resources of the Federal State Statistics Service; in this case the number of mixing distributions (components), as well as their corresponding weights and parameters, may be unknown. In terms of content, the problem of separation (decomposition) of mixtures can be reduced to estimating unknown parameters of mixing distributions and their weights.

The article considers and discusses methods for solving this problem, their advantages and disadvantages, conditions and areas of application. In the study the decomposition of mixtures of probability distributions on the population of subjects of the Russian

Federation was carried out according to three interrelated statistical indicators – the unemployment rate, the poverty rate, the level of violence. These indicators can be considered as statistical measure of achieving sustainable development goals at the regional level and, at the same time, measures of socio-economic «health» of territory. Typologies of subjects of the Russian Federation were carried out according to the listed indicators. A cross-comparison of the results of the obtained typological groupings was performed, which also made it possible to identify the subjects of the Russian Federation that differ in both negative and positive trends in the context of the indicators under consideration.

The authors underline that the results of the study can be used by the authorities to develop specific measures for the socio-economic development of Russia and its regions.

Keywords: Sustainable Development Goals (SDGs), interterritorial differentiation, social statistics, inequality level, poverty level, unemployment rate, level of violence, mathematical and statistical methods, mixtures of probability distributions, decomposition of mixtures of distributions.

JEL: C02, C12, C13, C38, C40, C60.

doi: <https://doi.org/10.34023/2313-6383-2022-29-6-11-24>.

For citation: Glinskiy V.V., Ismaiylola Yu.N., Khrushchev S.E., Serga L.K. Data Typology Using Technologies for Separating Mixtures of Probability Distributions. *Voprosy Statistiki*. 2022;29(6):11–24. (In Russ.)

Введение

Возникновение неравенства как социального или экономического явления можно отнести к самому началу развития человеческого общества. На протяжении всей истории человечества как древние, так и современные государства сталкиваются с проблемой неравенства и неравномерности развития территорий. В России – государстве, в состав которого входит 85 субъектов* и которое занимает 1/8 часть суши, – эта проблема стоит достаточно остро. Следствием этого является интерес научного сообщества к изучению социально-экономического положения и дифференциации (или неравенства) регионов Российской Федерации. Проблемами дифференциации и неравномерности развития территорий занимались такие ученые, как А.Г. Гранберг, Ю.С. Ершов [1 и 2], Н.Н. Михеева, В.В. Кулешов, В.Е. Селиверстов, В.И. Суслов, С.А. Суспицын, П.А. Минакир [2], С.В. Баранов, Т.П. Скуфьина [3], К.П. Глушенко [4], В.В. Глинский, Л.К. Серга, А.М. Булкина [5], Е.С. Губанова, В.С. Клещ [6], Ю.С. Зайцева [7], Е.А. Коломак [8], Е.А. Конопацкая [9], Б.Л. Лавровский [10], М.В. Морошкина [11], Е.А. Погодина, Е.Н. Катаев [12], Т.В. Ускова [13 и 14], Д.И. Усманов [15], Е.Ю. Чемезова [16] и другие.

25 сентября 2015 г. государства – члены ООН, в том числе и Россия, приняли цели в области устойчивого развития (ЦУР) [17]. В их составе 17 целей; в исследовании сделан акцент на трех – повсеместной ликвидации нищеты во всех формах; содействии неуклонному, всеохватному и устойчивому экономическому росту, полной

и производительной занятости и достойной работе для всех; содействии построению миролюбивых и открытых обществ в интересах устойчивого развития, обеспечении доступа к правосудию для всех и создании эффективных, подотчетных и основанных на широком участии учреждений на всех уровнях (1-я, 8-я и 16-я цели соответственно). Для реализации ЦУР в России разработаны и действуют различные национальные проекты, федеральные, региональные и муниципальные программы. В свете вышеизложенного представляет интерес оценка эффекта внедряемых программ: уменьшается ли социально-экономическое неравенство субъектов Российской Федерации или, напротив, растет.

Для измерения такого эффекта были выбраны три показателя – уровень безработицы, уровень бедности и уровень насилия, во-первых, по причине того, что они являются индикаторами выполнения или невыполнения соответствующих целей устойчивого развития, а во-вторых, потому что они выступают важными характеристиками социально-экономического положения регионов. При расчетах использована методика Росстата, в частности: отношение численности безработных к численности рабочей силы (уровень безработицы); численность населения с денежными доходами ниже величины прожиточного минимума в процентах от общей численности населения региона (уровень бедности); число зарегистрированных убийств и покушений на убийство в расчете на 100 000 человек населения территории (уровень насилия). Полученные показатели формируются под воздействием значительного

* На момент написания статьи.

числа разнообразных факторов, как естественных (климат, инфраструктура, половозрастной состав и т. п.), так и искусственных (например, законодательные нормы, предпочтения, льготы, ограничения и т. п.). Очевидно, что регионы дифференцированы в зависимости от значений выбранных индикаторов. Важно отметить, что негативная динамика по данной системе показателей не исключает появления катастрофических исходов, таких, например, как «обезлюдение» территорий, возникновение очагов социальной напряженности и прочих. В этой связи актуализируется традиционная проблема построения типологической группировки в части «наметки» типов, выработки правил и критериев нахождения пороговых значений перехода одного качественного состояния в другое.

Постановка задачи декомпозиции смесей вероятностных распределений

Рассмотрим некоторую, пусть неоднородную, совокупность, которая состоит из k подгрупп. Предположим, что наблюдаемый признак внутри i -ой подгруппы распределен в соответствии с функцией распределения $F_i(t)$, а вероятность того, что случайно выбранный элемент из всей совокупности будет именно из i -ой подгруппы, равна p_i . Тогда вероятность того, что у случайно выбранного из всей совокупности элемента значение наблюдаемого признака будет меньше, чем t , равна

$$F(t) = p_1 F_1(t) + \dots + p_k F_k(t). \quad (1)$$

Будем предполагать, что нам необходимо определить тип подгруппы, к которому принадлежит наблюдение, то есть требуется решить обратную операцию смешивания, а именно операцию разделения (декомпозиции) смеси. И, как следствие, получим типологию рассматриваемой неоднородной совокупности.

Пусть исследуется неоднородная выборка $X = (X_1, X_2, \dots, X_n)$, взятая из конечной смеси вероятностных распределений, тогда в формуле (1) веса $p_i \geq 0$, $i = 1, \dots, k$ — смешивающиеся вероятности, удовлетворяющие условию $p_1 + \dots + p_k = 1$; $F_i(t)$, $i = 1, \dots, k$ — функции распределения (компоненты смеси), соответствующие смешиваемым вероятностным распределениям.

Будем предполагать унимодальность компонент смеси. Предположение об унимодальности распределения упрощает процедуру интерпретации результатов, так как у каждой отдельно взятой компоненты имеется четко выделенное «ядро» распределения.

Если компоненты смеси абсолютно непрерывны и имеют плотность f_i , то смесь также будет абсолютно непрерывной с плотностью

$$f(t) = p_1 f_1(t) + \dots + p_k f_k(t). \quad (2)$$

Выделяют два типа задач разделения смеси распределений: теоретическую и практическую.

Теоретическая задача. Пусть компоненты смеси f_i определяются неким известным параметрическим семейством, то есть $f_i(t) \equiv f_i(t, \theta_i)$, где θ_i — неизвестный параметр (скалярный или векторный). Имеем

$$f(t) = p_1 f_1(t, \theta_1) + \dots + p_k f_k(t, \theta_k). \quad (3)$$

Задачей разделения (декомпозиции) смеси называют проблему нахождения неизвестных параметров $\hat{\theta} = (p_1, \dots, p_k, \theta_1, \dots, \theta_k)$. Отметим, что данная задача не всегда решается однозначно.

Приведем следующий пример. Предположим, что рассматриваемая выборка взята из равномерного распределения на интервале $(0; 1]$, то есть плотность распределения выборки имеет вид $f(t) = I(t \in (0; 1])$, где $I(t \in \Delta)$ — индикатор того, что t принадлежит множеству Δ . Мы также будем предполагать, что компоненты смеси имеют равномерное распределение на некотором неизвестном интервале. Тогда справедливо, что

$$\begin{aligned} I(t \in (0; 1]) &= 0,5 \times 2I(t \in (0; 0,5]) + 0,5 \times 2I(t \in (0,5; 1]) = \\ &= 0,25 \times 4I(t \in (0; 0,25]) + 0,25 \times 4I(t \in (0,25; 0,5]) + \\ &\quad + 0,5 \times 2I(t \in (0,5; 1]). \end{aligned}$$

Получили, что плотность равномерного распределения можно, с одной стороны, представить в виде двухкомпонентной равновесной смеси равномерных распределений на интервалах $(0; 0,5]$ и $(0,5; 1]$. С другой стороны — в виде трехкомпонентной смеси равномерных распределений на интервалах $(0; 0,25]$, $(0,25; 0,5]$ и $(0,5; 1]$ с весами 0,25; 0,25 и 0,5 соответственно, то есть в данном случае нельзя однозначно выделить компоненты смеси. Чтобы избежать неоднозначности, обычно рассматривают идентифицируемые смеси.

Определение. Конечная смесь называется идентифицируемой, если из равенства

$$\sum_{i=1}^k p_i f_i(t, \theta_i) = \sum_{j=1}^m q_j f_j(t, \theta_j) \quad (4)$$

следует, что $k=m$ и для каждого индекса i существует индекс j – такой, что $p_i = q_j$, $f_i(t, \theta_i) \equiv f_j(t, \theta_j)$ [18].

Замечание. Идентифицируемыми являются конечные смеси из следующих распределений: нормальных, гамма, Коши, пуассоновских, биномиальных (с известной вероятностью успеха) [18].

Практическая задача. Пусть компоненты смеси f_i определяются неким известным параметрическим семейством, то есть $f_i(t) \equiv f_i(t, \theta_i)$, где θ_i – неизвестный параметр (скалярный или векторный). Имеем

$$f(t) = p_1 f_1(t, \theta_1) + \dots + p_k f_k(t, \theta_k). \quad (5)$$

В этом случае практической задачей разделения смеси является задача статистического оценивания неизвестных параметров $\bar{\theta} = (p_1, \dots, p_k, \theta_1, \dots, \theta_k)$ по известной реализации выборки. При этом, как правило, предполагается, что рассматриваемая смесь является идентифицируемой.

Данная задача обычно решается одним из двух методов: методом моментов или методом максимального правдоподобия. Кратко отметим некоторые характеристики этих методов (подробнее см. [19]).

Метод моментов (метод К. Пирсона, 1894 г. [20]) в частных случаях позволяет найти оценки в явном виде, но в общем случае их искать сложно. Качество оценок хуже, чем при использовании метода максимального правдоподобия.

Метод максимального правдоподобия. Оценки лучше, чем в методе моментов, но в явном виде их не найти (только в некоторых частных случаях). Поэтому оценки ищут с помощью некоторых специальных итерационных методов. Отметим, что оценка по методу максимального правдоподобия может и не существовать. Для этого приведем следующий классический пример.

Пример. Рассмотрим двухкомпонентную смесь с равными весами произвольного нормального и стандартного нормального распределений:

$$f(t) = 0,5 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\alpha)^2}{2\sigma^2}} + 0,5 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}. \quad (6)$$

Тогда функция максимально правдоподобия имеет вид:

$$\psi(x, a, \sigma^2) = \prod_{i=1}^n \left(0,5 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \alpha)^2}{2\sigma^2}} + 0,5 \frac{1}{\sqrt{2\pi}} e^{-\frac{x_i^2}{2}} \right). \quad (7)$$

Если мы приравняем неизвестный параметр a реализации первого наблюдения x_1 , то получим, что $\psi(x, a=x_1, \sigma^2) \rightarrow \infty$ при $\sigma^2 \rightarrow 0$. Здесь и далее через x будем обозначать реализацию выборки X . В данном случае функция правдоподобия не имеет такого значения неизвестного параметра σ^2 , при котором она достигала бы своего максимального значения. Что и означает отсутствие оценки метода максимального правдоподобия.

При нахождении оценок метода максимального правдоподобия, как правило, пользуются двумя итерационными методами. Кратко отметим некоторые характеристики этих методов (подробнее см. [19]).

ЕМ-алгоритм (А. Демпстер, Н. Лэрд и Д. Рубин, 1977 г. [21]) – не очень быстр, не гарантирует сходимость к глобальному максимуму, обладает неустойчивостью по начальным данным.

SEM-алгоритм (Ж. Селё, Дж. Дилбо, 1985 г. [22]) – быстр, обладает устойчивостью к начальным данным; как правило, сходится к глобальному максимуму.

Добавим, что известны и медианные модификации ЕМ- и SEM-алгоритмов, которые используются для нахождения оценок неизвестных параметров смесей вероятностных распределений [23 и 24].

Определение числа компонент смеси и их границ

Число компонент смеси распределений (типов, классов, групп) может быть заранее неизвестно. Выбор слишком малых значений для числа групп, как правило, может оказать негативное воздействие на качество аппроксимации распределения данных. Неоправданное увеличение числа компонент хоть и будет усиливать правдоподобие, но может привести к ухудшению качества интерпретации результатов. Поэтому желательно соблюдать баланс между неоправданно большим количеством компонент смеси и качеством аппроксимации. Для достижения указанного баланса во многих случаях используют так называемые информационные

критерии (критерий Акаике и байесовский информационный критерий), где мерой качества служит расстояние Кульбака – Лейблера (D_{KL}). Стоит добавить, что для корректности применения таких информационных критериев требуется выполнение некоторых условий регулярности (в частности, конечность функции правдоподобия).

Значение информационного критерия Акаике (Х. Акаике, 1973 г. [25]) определяется следующим образом:

$$AIC = -2\ln\psi(x; \hat{\theta}(x)) + 2d,$$

где d – количество оцениваемых параметров. Чем меньше значение информационного критерия Акаике, тем лучшим качеством обладает аппроксимация распределения данных. Отметим, что $AIC \approx 2D_{KL}$.

Одним из популярных информационных критериев является байесовский (Г. Шварц, 1978 г. [26]), значение которого определяется по формуле

$$BIC = 2\ln\psi(x; \hat{\theta}(x)) - d\ln n.$$

Чем больше значение байесовского информационного критерия, тем лучшим качеством обладает аппроксимация распределения данных.

Для практических целей важно не только определение количества компонент смеси, но и определение их границ. Для этого используют байесовский подход, то есть наблюдение x отнесем к i -ой компоненте смеси, если $p_i f_i(x) > p_j f_j(x)$ для всех $j \neq i$.

Применения технологий декомпозиции смесей на реальных данных

Проведем типологию субъектов Российской Федерации с помощью декомпозиции смесей распределений выбранных нами социально-экономических показателей. Естественно предположить, что выбранные показатели могут описываться смесью распределений.

Проводить декомпозицию смесей рассматриваемых показателей будем по следующему алгоритму (см. также [19, 27 и 28]):

I шаг. С помощью критерия согласия Пирсона определяются периоды, «подозрительные» на наличие в них смесей распределений, то есть те периоды, в которых отвергается гипотеза о принадлежности выборки к однокомпонентному закону распределения.

II шаг. На основе SEM-алгоритма оцениваются параметры смеси из k компонент, где число компонент k определяется с помощью информационного критерия Акаике.

III шаг. С помощью критерия Пирсона подтверждаем гипотезу о принадлежности выборки к k -компонентной смеси распределений. В противном случае делается вывод, что декомпозицию смесей распределений нельзя осуществить.

IV шаг. С применением байесовского подхода определяются границы компонент смеси.

V шаг. Использование определенных на предыдущем шаге границ (пороговых значений) позволяет реализовать разбиение совокупности субъектов Российской Федерации на однокачественные группы (типы).

Типологическая группировка регионов Российской Федерации по уровню безработицы. В анализируемом временном интервале (2005–2020 гг.) уровень безработицы в России имеет достаточно устойчивую тенденцию снижения (см. рис. 1), исключения составляют 2008 и 2009 гг. (последствия экономического кризиса), а также 2020 г. (начало пандемии COVID-19).

При выполнении первого шага представленного выше алгоритма были выявлены периоды, в которых ожидалось наличие смесей вероятностных распределений: 2005 и 2010–2020 гг. (в период с 2006 по 2009 г. с помощью критерия Пирсона на уровне значимости 0,05 подтвердилась гипотеза о том, что данные описываются одним законом распределения вероятностей – гауссианой). Для этих временных интервалов была сформулирована и проверена гипотеза о том, что уровень безработицы описывается смесью двух и более нормальных вероятностных распределений; соответственно, можно предполагать качественные территориальные различия (дифференциацию) по данному показателю. Применение SEM-алгоритма и информационного критерия Акаике для оценки количества компонент смеси для данных лет позволило получить следующие результаты: с 2010 по 2019 г. уровень безработицы описывается двухкомпонентной смесью нормальных распределений вероятностей. Полученные оценки параметров (среднего значения – a , среднего квадратичного отклонения – σ), а также весов (p) компонент вероятностных смесей размещены в таблице 1.

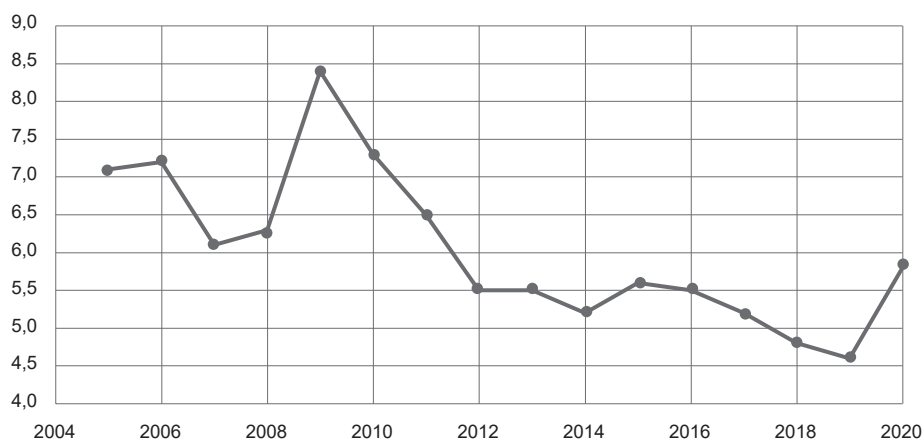


Рис. 1. Динамика уровня безработицы в Российской Федерации, 2005–2020 гг. (в процентах)

Источник: расчеты авторов по данным Росстата.

Таблица 1

Оценки параметров и весов смесей нормальных распределений вероятностей по показателю «Уровень безработицы, 2005 и 2010–2020 годы

Год	p_1	a_1	σ_1	p_2	a_2	σ_2	p_3	a_3	σ_3
2005	0,82	7,66	2,09	0,18	23,37	19,20	-	-	-
2010	0,98	8,10	2,76	0,02	46,50	4,53	-	-	-
2011	0,98	7,23	2,44	0,02	42,70	7,64	-	-	-
2012	0,85	5,83	1,76	0,15	22,05	14,34	-	-	-
2013	0,85	5,75	1,76	0,15	20,92	12,68	-	-	-
2014	0,93	5,64	1,96	0,07	20,85	6,96	-	-	-
2015	0,86	6,07	1,91	0,14	20,33	6,93	-	-	-
2016	0,85	6,07	2,01	0,15	17,80	7,15	-	-	-
2017	0,75	5,18	1,26	0,25	11,43	5,28	-	-	-
2018	0,69	4,81	0,98	0,31	8,78	4,91	-	-	-
2019	0,70	4,55	0,85	0,30	8,71	4,97	-	-	-
2020	0,72	5,36	1,10	0,15	8,32	0,92	0,13	15,37	5,96

Число компонент смеси определяет количество классов для типологизации регионов. Разделение на типы осуществлялось при помощи определения граничного значения (точки перехода одного качественного состояния в другое). Субъекты Российской Федерации с уровнем безработицы,

превышающим пороговое значение, относились к «неблагоприятному» типу по уровню безработицы, а регионы с уровнем безработицы ниже порогового значения – к «благоприятному». Динамика таких границ для двухкомпонентных смесей представлена на рис. 2.

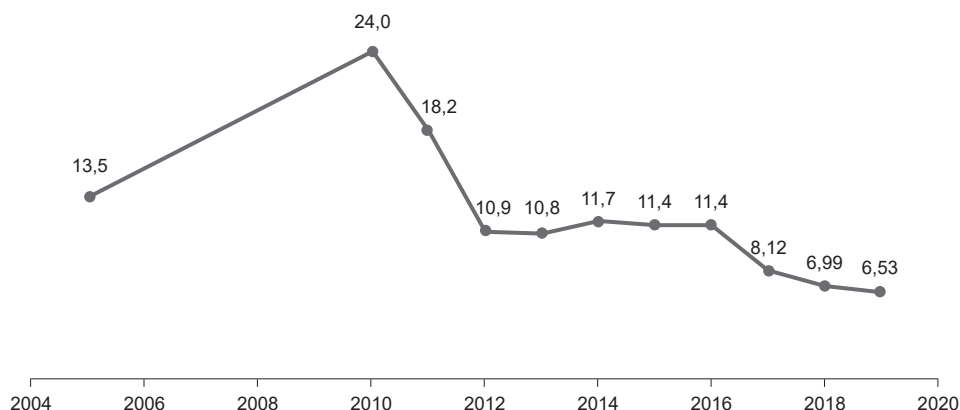


Рис. 2. Динамика порогового значения по уровню безработицы в Российской Федерации, 2005–2019 гг. (в процентах)

Источник: расчеты авторов по данным Росстата.

В таблице 2 представлена типология субъектов Российской Федерации по уровню безработицы (субъекты, не вошедшие в данную таблицу, в течение всего рассматриваемого периода относились к «благоприятному» типу). В табли-

це используются следующие обозначения: «+» — «благоприятный» тип по уровню безработицы, «-» — «неблагоприятный» тип по уровню безработицы, «0» — «промежуточный» тип по уровню безработицы.

Таблица 2

Типология регионов Российской Федерации по уровню безработицы, 2005 и 2010–2020 годы

Субъект Российской Федерации	2005	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Республика Карелия	+	+	+	+	+	+	+	+	-	-	+	0
Республика Коми	+	+	+	+	+	+	+	+	+	-	+	0
Мурманская область	+	+	+	+	+	+	+	+	+	+	+	0
Республика Адыгея	+	+	+	+	+	+	+	+	-	-	+	0
Республика Калмыкия	-	+	+	-	-	+	+	+	-	-	+	0
Астраханская область	+	+	+	+	+	+	+	+	+	-	+	0
Волгоградская область	+	+	+	+	+	+	+	+	+	+	+	0
Республика Дагестан	-	+	+	-	-	+	+	+	-	-	-	-
Республика Ингушетия	-	-	-	-	-	-	-	-	-	-	-	-
Кабардино-Балкарская Республика	-	+	+	+	+	+	+	+	-	-	-	-
Карачаево-Черкесская Республика	-	+	+	+	+	+	-	-	-	-	-	-
Республика Северная Осетия – Алания	+	+	+	+	+	+	+	+	-	-	-	-
Чеченская Республика	+	-	-	-	-	-	-	-	-	-	-	-
Курганская область	+	+	+	+	+	+	+	+	-	-	+	0
Республика Алтай	+	+	+	-	-	+	+	-	-	-	-	-
Республика Тыва	-	+	+	-	-	+	-	-	-	-	-	-
Республика Хакасия	+	+	+	+	+	+	+	+	+	+	+	0
Иркутская область	+	+	+	+	+	+	+	+	-	-	+	0
Омская область	+	+	+	+	+	+	+	+	+	+	+	0
Томская область	+	+	+	+	+	+	+	+	+	+	+	0
Республика Бурятия	+	+	+	+	+	+	+	+	-	-	-	-
Республика Саха (Якутия)	+	+	+	+	+	+	+	+	+	-	+	+
Забайкальский край	+	+	+	+	+	+	+	+	-	-	+	0
Еврейская автономная область	+	+	+	+	+	+	+	+	-	+	+	+

Отдельного внимания заслуживает 2020 г. В этом году уровень безработицы описывается трехкомпонентной смесью нормальных распределений. Тем самым между «благоприятными» и «неблагоприятными» регионами Российской Федерации выделяется новый тип – «промежуточный». Появление нового типа естественным образом объясняется началом пандемии COVID-19 и тяжелым положением российской экономики в этот период.

Группировка субъектов Российской Федерации по уровню бедности. Анализ уровня бедности в субъектах Российской Федерации показал, что в 2005, 2010–2013 и 2015–2020 гг. данный показатель описывается двухкомпонентной смесью нормальных распределений. Это означает, что среди регионов Российской Федерации по уровню бедности выделяются два типа – «благоприятный» и «неблагоприятный». Результаты оценки параметров смеси вероятностных распределений представлены в таблице 3.

Таблица 3

Оценки параметров и весов смесей нормальных распределений вероятностей по показателю «Уровень бедности», 2005, 2010–2013 и 2015–2020 годы

Год	p_1	a_1	σ_1	a_2	σ_2
2005	0,92	20,43	5,74	46,27	9,78
2010	0,97	14,63	3,82	32,65	4,31
2011	0,98	14,66	3,89	33,20	3,68
2012	0,98	12,75	3,51	29,30	1,98
2013	0,98	12,93	3,49	34,40	1,41
2015	0,96	14,42	3,99	34,53	3,40
2016	0,86	13,91	3,40	28,06	6,25
2017	0,87	13,74	3,48	27,76	5,89
2018	0,81	13,00	3,26	20,86	6,18
2019	0,79	12,86	3,15	20,68	6,06
2020	0,80	12,53	3,17	20,52	5,60

Отметим, что начиная с 2011 г. доля «благоприятных» типов субъектов Российской Федерации убывает (с 98 до 80%), а, соответственно, доля «бедных» регионов возрастает.

На рис. 3 представлена динамика порогового значения, различающего два класса регионов Российской Федерации.

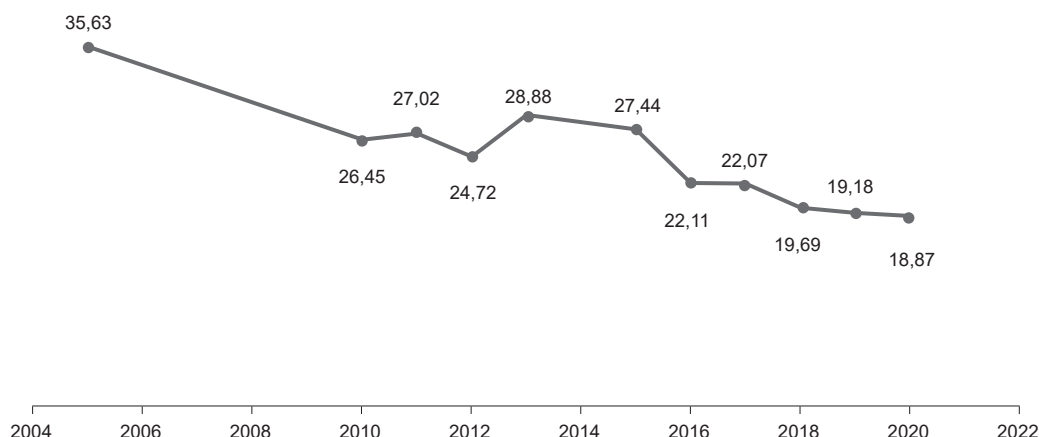


Рис. 3. Динамика порогового значения по уровню бедности в Российской Федерации, 2005–2020 гг. (в процентах)

Источник: расчеты авторов по данным Росстата.

Добавим, что среднее значение показателя уровня бедности в каждой типологической группе в целом уменьшается; также наблюдается снижение порогового значения. Наличие двух компонент свидетельствует в пользу значимых различий выделенных типологических групп.

В таблице 4 представлена типология территорий по показателю уровня бедности (субъекты, не вошедшие в данную таблицу, в течение всего рассматриваемого периода относились к «благоприятному» типу). В таблице используются следующие обозначения: «+» — «благоприятный» тип по уровню бедности, «-» — «неблагоприятный» тип по уровню бедности.

Таблица 4

Типология субъектов Российской Федерации по уровню бедности, 2005, 2010–2013 и 2015–2020 годы

Субъект Российской Федерации	2005	2010	2011	2012	2013	2015	2016	2017	2018	2019	2020
Ивановская область	-	+	+	+	+	+	+	+	+	+	+
Республика Калмыкия	-	-	-	-	-	-	-	-	-	-	-
Республика Ингушетия	-	+	+	+	+	-	-	-	-	-	-
Кабардино-Балкарская Республика	+	+	+	+	+	+	-	-	-	-	-
Карачаево-Черкесская Республика	+	+	+	+	+	+	-	-	-	-	-
Чеченская Республика	+	+	+	+	+	+	+	+	-	-	-
Республика Марий Эл	-	+	+	+	+	+	-	-	-	-	-
Курганская область	+	+	+	+	+	+	+	+	+	-	-
Республика Алтай	-	+	+	+	+	+	-	-	-	-	-
Республика Тыва	-	-	-	-	-	-	-	-	-	-	-
Республика Бурятия	+	+	+	+	+	+	+	+	-	-	-
Забайкальский край	+	+	+	+	+	+	+	+	-	-	-
Еврейская автономная область	+	+	+	+	+	+	-	-	-	-	-

Типология субъектов Российской Федерации по уровню насилия. Решение задачи декомпозиции показателя «Уровень насилия» позволило определить, что в 2007–2017 гг. уровень насилия в субъектах Российской Федерации может

быть описан двухкомпонентной смесью нормальных распределений (результаты оценки параметров смеси представлены в таблице 5), а начиная с 2018 г. уже выделяются три компоненты.

**Оценки параметров и весов смесей нормальных распределений вероятностей
по показателю «Уровень насилия», 2007–2020 годы**

Год	p_1	a_1	σ_1	p_2	a_2	σ_2	p_3	a_3	σ_3
2007	0,73	15,03	4,67	0,27	34,51	10,9	-	-	-
2008	0,98	13,94	4,94	0,02	37,62	12,17	-	-	-
2009	0,86	12,39	3,9	0,14	31,99	11,94	-	-	-
2010	0,79	11,27	3,75	0,21	30,27	13,41	-	-	-
2011	0,77	10,06	3,33	0,23	25,72	7,95	-	-	-
2012	0,78	9,32	3,26	0,22	26,12	9,84	-	-	-
2013	0,69	7,55	1,90	0,31	18,14	6,41	-	-	-
2014	0,79	8,26	3,36	0,21	23,54	8,94	-	-	-
2015	0,79	8,17	3,25	0,21	23,37	7,88	-	-	-
2016	0,77	6,68	2,41	0,23	18,81	6,75	-	-	-
2017	0,74	6,09	2,06	0,26	16,00	5,26	-	-	-
2018	0,77	5,18	1,69	0,10	10,91	0,73	0,13	15,66	5,05
2019	0,40	4,24	1,07	0,48	6,46	2,31	0,12	17,66	7,84
2020	0,62	4,36	1,28	0,30	8,65	2,95	0,07	17,09	7,18

При этом среднее значение показателя уровня насилия в каждой типологической группе в целом снижается; также наблюдается снижение поро-

вого значения. Динамика порогового значения, разделяющего типы в случае двухкомпонентной смеси, представлена на рис. 4.

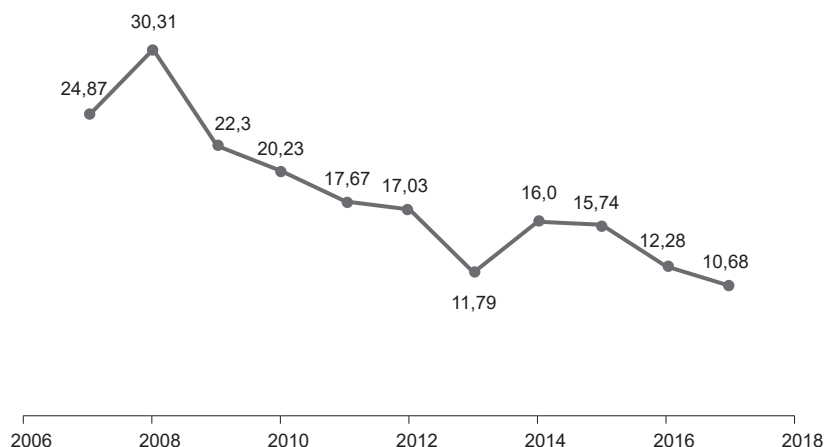


Рис. 4. Динамика порогового значения по уровню насилия в Российской Федерации, 2005–2017 гг. (в процентах)

Источник: расчеты авторов по данным Росстата.

Результаты полученной типологии за период 2007–2020 гг. представлены в таблице 6 (субъекты, не вошедшие в данную таблицу, в течение всего рассматриваемого периода относились к «благоприятному» типу). В таблице используются следующие обозначения: «+» – «благоприятный» тип по уровню насилия, «-» – «неблагоприятный», «0» – «промежуточный» тип региона по уровню насилия.

Отдельного внимания заслуживает период 2018–2020 гг. В этот период уровень насилия описывается трехкомпонентной смесью нормальных распределений, то есть между «благо-

приятными» и «неблагоприятными» субъектами Российской Федерации выделяется новый тип – «промежуточный», что еще больше усиливает дифференциацию регионов по данному показателю.

Таким образом, появляются новые существенные различия между группами субъектов Российской Федерации по показателю уровня насилия. В зависимости от темпов его снижения выделяются регионы-«лидеры» и «аутсайдеры». При этом выделение нового, «промежуточного» типа приходится на два последних года перед началом пандемии COVID-19.

Таблица 6

Типология субъектов Российской Федерации по уровню насилия, 2007–2020 годы

Субъект Российской Федерации	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Орловская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Смоленская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Тверская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Ярославская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Республика Карелия	+	+	+	+	+	+	+	+	+	+	+	+	0	0
Республика Коми	+	+	+	+	+	+	-	+	+	-	+	0	0	0
Архангельская область	+	+	+	+	+	+	+	+	+	-	-	0	0	0
Ленинградская область	+	+	+	+	+	+	+	+	+	+	+	+	0	0
Новгородская область	+	+	+	+	+	+	-	+	+	+	+	+	0	0
Республика Калмыкия	+	+	+	+	+	+	+	+	+	+	+	+	0	0
Карачаево-Черкесская Республика	+	+	+	+	+	+	-	+	+	+	+	+	+	+
Республика Башкортостан	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Республика Марий Эл	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Пермский край	+	+	+	+	+	+	-	+	+	+	+	+	0	+
Кировская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Нижегородская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Оренбургская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Ульяновская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Курганская область	+	+	+	+	+	+	-	+	+	+	-	0	0	0
Свердловская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Челябинская область	+	+	+	+	+	+	+	+	+	+	+	+	0	+
Республика Алтай	+	+	-	-	-	+	-	-	-	-	-	-	0	0
Республика Тыва	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Республика Хакасия	+	+	+	+	+	+	-	+	+	-	-	0	0	0
Алтайский край	+	+	+	+	+	+	+	+	+	+	+	+	0	0
Красноярский край	+	+	+	+	+	+	+	+	+	+	+	0	0	0
Иркутская область	-	+	-	-	-	-	-	-	+	-	-	0	0	0
Кемеровская область	+	+	+	+	+	+	-	+	+	+	-	+	0	0
Республика Бурятия	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Республика Саха (Якутия)	-	+	+	+	+	+	-	+	+	-	-	0	0	0
Забайкальский край	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Камчатский край	-	+	+	+	+	+	+	+	+	+	+	0	0	0
Приморский край	-	+	+	+	+	+	-	-	+	+	+	+	0	0
Хабаровский край	-	+	+	+	+	+	-	+	+	-	-	0	0	0
Амурская область	+	+	+	+	-	-	-	+	+	-	-	-	-	0
Магаданская область	-	-	-	+	-	+	-	-	-	-	-	-	-	0
Сахалинская область	-	+	-	+	+	-	-	+	-	+	-	0	0	0
Еврейская автономная область	+	+	+	+	-	-	-	+	+	-	-	-	-	0
Чукотский автономный округ	-	+	-	+	-	-	-	-	+	-	-	-	-	-

В качестве интегральной оценки социально-экономического положения региона на основе рассмотренных показателей введем следующий коэффициент:

$$K = I_{\text{ур. безработицы}} + I_{\text{ур. бедности}} + I_{\text{ур. насилия}},$$

где

$$I_{\text{ур.}} = \begin{cases} 1 - \text{субъект принадлежит к «благоприятному» типу;} \\ 0 - \text{субъект принадлежит к «промежуточному» типу;} \\ -1 - \text{субъект принадлежит к «неблагоприятному» типу.} \end{cases}$$

Коэффициент К может принимать значения от -3 до 3. Значение коэффициента К = -3 свидетельствует об очень низком социально-экономическом положении субъекта Российской Федерации, а значение К = 3, наоборот, говорит о высоком социально-экономическом положении региона. В таблице 7 представлены типология субъектов Российской Федерации и соответствующие значения коэффициента К, наблюдаемые в 2020 г. (регионы со значение коэффициента К = 3 из таблицы исключены).

Таблица 7

Типология субъектов Российской Федерации по значимым социально-экономическим показателям в 2020 году

Субъект Российской Федерации	Уровень безработицы	Уровень бедности	Уровень насилия	Коэффициент
Республика Тыва	-1	-1	-1	-3
Республика Бурятия	-1	-1	-1	-3
Республика Алтай	-1	-1	0	-2
Забайкальский край	0	-1	-1	-2
Республика Калмыкия	0	-1	0	-1
Республика Ингушетия	-1	-1	1	-1
Кабардино-Балкарская Республика	-1	-1	1	-1
Карачаево-Черкесская Республика	-1	-1	1	-1
Чеченская Республика	-1	-1	1	-1
Курганская область	0	-1	0	-1
Еврейская автономная область	1	-1	0	0
Республика Карелия	0	1	0	1
Республика Коми	0	1	0	1
Республика Дагестан	-1	1	1	1
Республика Северная Осетия – Алания	-1	1	1	1
Республика Марий Эл	1	-1	1	1
Республика Хакасия	0	1	0	1
Иркутская область	0	1	0	1
Чукотский автономный округ	1	1	-1	1
Архангельская область	1	1	0	2
Ленинградская область	1	1	0	2
Мурманская область	0	1	1	2
Новгородская область	1	1	0	2
Республика Адыгея	0	1	1	2
Астраханская область	0	1	1	2
Волгоградская область	0	1	1	2
Алтайский край	1	1	0	2
Красноярский край	1	1	0	2
Кемеровская область	1	1	0	2
Омская область	0	1	1	2
Томская область	0	1	1	2
Республика Саха (Якутия)	1	1	0	2
Камчатский край	1	1	0	2
Приморский край	1	1	0	2
Хабаровский край	1	1	0	2
Амурская область	1	1	0	2
Магаданская область	1	1	0	2
Сахалинская область	1	1	0	2

Данные таблицы свидетельствуют об очень низком социально-экономическом положении четырех субъектов Российской Федерации: республик Тыва, Бурятия, Алтай и Забайкальского края.

Заключение

В работе была решена задача построения типологии субъектов Российской Федерации по значимым социально-экономическим показателям. Установлено, что каждый из рассматриваемых

показателей с 2010 по 2020 г. описывается смесью двух или трех нормальных распределений. В случае двух компонент смеси субъекты Российской Федерации по уровню бедности, уровню безработицы и уровню насилия делятся на два типа: «благоприятный» и «неблагоприятный». В случае трех компонент смеси, помимо представленных выше, выделен «промежуточный» тип субъектов Российской Федерации. Определены пороговые значения – границы перехода из одной типологической группы в другую.

Следует отметить, что количество субъектов Российской Федерации, относящихся к «неблагоприятному» типу, растет. Более того, увеличивается и число регионов этого типа одновременно по двум и трем значимым социально-экономическим показателям.

Особого внимания органов власти, на наш взгляд, заслуживает факт растущей неравномерности развития субъектов Российской Федерации, установленный в результате выделения в последние годы трехкомпонентных смесей распределения субъектов по исследованным социально значимым показателям.

Рост численности субъектов, относящихся к «неблагоприятному» типу, и усиление неравномерности их развития указывают на необходимость принятия управленческих решений для регулирования процессов дифференциации регионов.

Литература

1. Гранберг А.Г., Ершов Ю.С. Краткосрочные последствия влияния кризиса на пространственное развитие экономики РФ // Оптимизация территориальных систем / под ред. д. э. н. С.А. Суспицына. Новосибирск: ИЭОПП СО РАН, 2010. С. 206–214.
2. Гранберг А.Г. и др. Воздействие мирового кризиса на стратегию пространственного социально-экономического развития Российской Федерации // Регион: экономика и социология. 2009. № 4. С. 69–101.
3. Баранов С.В., Скуфына Т.П. Сравнительная динамика экономического роста и межрегиональная дифференциации территории российского Севера // Вопросы статистики. 2015. № 11. С. 69–77.
4. Глушенко К.П. Об оценке межрегионального неравенства // Пространственная экономика. 2015. № 4. С. 39–58. doi: <https://doi.org/10.14530/se.2015.4.039-058>.
5. Глинский В.В., Серга Л.К., Булкина А.М. Дифференциация муниципальных образований как фактор экономического развития территорий // Вопросы статистики. 2016. № 8. С. 46–52.
6. Губанова Е.С., Клещ В.С. Преодоление социально-экономического неравенства как условие устойчивого сбалансированного пространственного развития региона // Экономические и социальные перемены: факты, тенденции, прогноз. 2019. Т. 12. № 5. С. 44–57. doi: <https://doi.org/10.15838/esc.2019.5.65.3>.
7. Зайцева Ю. Межрегиональная дифференциация в странах БРИК: возможности оценки // Мировая экономика и международные отношения. 2010. № 5. С. 44–51. doi: <https://doi.org/10.20542/0131-2227-2010-5-44-51>.
8. Коломак Е.А. Межрегиональное неравенство в России: экономический и социальный аспекты // Пространственная экономика. 2010. № 1. С. 26–35. doi: <https://doi.org/10.14530/se.2010.1.026-035>.
9. Конопатская Е.А. Дифференциация регионов Российской Федерации по уровню развития человеческого капитала // Экономика, статистика и информатика. Вестник УМО. 2011. № 2. С. 178–184.
10. Лавровский Б.Л. Территориальная дифференциация и подходы к ее ослаблению в Российской Федерации // Экономический журнал Высшей школы экономики. 2003. Т. 7. № 4. С. 524–537.
11. Морошкина М.В. Межрегиональная дифференциация субъектов Российской Федерации // Экономический анализ: теория и практика. 2014. Т. 13. № 45. С. 20–28.
12. Погодина Е.А., Катаев Е.Н. Оценка уровня и глубины дифференциации экономического и социального развития регионов Приволжского федерального округа // Национальные интересы: приоритеты и безопасность. 2014. Т. 10. № 21. С. 30–37.
13. Ускова Т.В. Пространственное развитие территорий: состояние, тенденции, пути снижения рисков // Проблемы развития территории. 2015. № 1(75). С. 7–15.
14. Ускова Т.В. Устойчивость развития территорий и современные методы управления // Проблемы развития территории. 2020. № 2(106). С. 7–18. doi: <https://doi.org/10.15838/ptd.2020.2.106.1>.
15. Усманов Д.И. Неравенство социально-экономического развития регионов России // European Social Science Journal («Европейский журнал социальных наук»). 2014. № 5–1(44). С. 476–480.
16. Чemezова Е.Ю. Типологии субъектов РФ по уровню социально-экономического развития // Вестник НГУЭУ. 2010. № 1. С. 171–176.
17. ООН. Преобразование нашего мира: Повестка дня в области устойчивого развития на период до 2030 года. Резолюция, принятая Генеральной Ассамблеей 25 сентября 2015 года. ООН, 2015. 45 с. URL: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/291/92/PDF/N1529192.pdf?>
18. Айвазян С.А. и др. Прикладная статистика: Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
19. Исмаиловой Ю.Н., Хрушев С.Е. Типологическая группировка на основе декомпозиции смесей вероятностных распределений // Вестник НГУЭУ. 2020. № 1. С. 255–267. doi: <https://doi.org/10.34020/2073-6495-2020-1-255-267>.
20. Pearson K. Contributions to the Mathematical Theory of Evolution // Philosophical Transactions of the Royal Society of London. 1894. Vol. 185. P. 71–110.
21. Dempster A.P., Laird N.M., Rubin D.B. Maximum Likelihood from Incomplete Data Via the EM Algorithm (with Discussion) // Journal of the Royal Statistical Society B. 1977. Vol. 39. Iss. 1. P. 1–38. doi: <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
22. Celeux G., Diebolt J. The SEM Algorithm: A Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem // Computation Statistics Quarterly. 1985. Vol. 2. No 1. P. 73–82.

23. **Королёв В.Ю.** ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений. Теоретический обзор. М.: ИПИ РАН, 2007. 94 с.

24. **Горшенин А.К., Королёв В.Ю., Турсунбаев А.М.** Медианные модификации ЕМ- и SEM-алгоритмов для разделения смесей вероятностных распределений и их применение к декомпозиции волатильности финансовых временных рядов // Информатика и ее применения. 2008. Т. 2. № 4. С. 12–47.

25. **Akaike H.** Information Theory and an Extension of the Maximum Likelihood Principle // Petrov B.N., Csaki F. (eds) Proc. of the 2nd Int. Symposium on Information Theory. Budapest: Akademiai Kiado, 1973. P. 267–281.

26. **Schwartz G.** Estimating the dimension of a model // The Annals of Statistics. 1978. Vol. 6. Iss. 2. P. 461–464. doi: <https://doi.org/10.1214/aos/1176344136>.

27. **Глинский В.В., Исмайлова Ю.Н.** Вероятностные смеси в измерениях межтерриториальной дифференциации // Вопросы статистики. 2020. Т. 27. № 3. С. 53–64. doi: <https://doi.org/10.34023/2313-6383-2020-27-3-53-64>.

28. **Исмайлова Ю.Н.** Типология данных на основе декомпозиции смесей вероятностных распределений // дисс. ... канд. экон. наук: 08.00.12. Новосибирск: НГУЭУ, 2022. 172 с.

Информация об авторах

Глинский Владимир Васильевич — д-р экон. наук, заведующий научно-исследовательской лабораторией «Устойчивое развитие социально-экономических систем», Сибирский институт управления — филиал РАНХиГС. 630102, г. Новосибирск, ул. Нижегородская, д. 6. E-mail: s444@ngs.ru. ORCID: <https://orcid.org/0000-0002-7149-3020>.

Исмайлова Юлия Николаевна — канд. экон. наук, доцент кафедры бизнес-аналитики и статистики, Сибирский институт управления — филиал РАНХиГС. 630102, г. Новосибирск, ул. Нижегородская, д. 6. E-mail: ismaiylova@gmail.com. ORCID: <https://orcid.org/0000-0003-1706-6973>.

Хрущев Сергей Евгеньевич — канд. физ.-мат. наук, доцент, старший научный сотрудник лаборатории прикладных обратных задач, Институт математики имени С.Л. Соболева СО РАН. 630090, г. Новосибирск, пр-т акад. Коптюга, д. 4. E-mail: hushew@rambler.ru. ORCID: <https://orcid.org/0000-0002-0281-5022>.

Серга Людмила Константиновна — канд. экон. наук, заведующий кафедрой бизнес-аналитики и статистики, Сибирский институт управления — филиал РАНХиГС. 630102, г. Новосибирск, ул. Нижегородская, д. 6. E-mail: serga-lk@ranepa.ru. ORCID: <https://orcid.org/0000-0002-8934-8876>.

Финансирование

Исследование выполнено при финансовой поддержке РНФ, проект № 22–28–00629.

References

1. **Granberg A.G., Ershov Yu.S.** Short-Term Consequences of the Impact of the Crisis on the Spatial Development of the Russian Economy. In: Suspitsyn S.A. (ed.) *Optimization of Territorial Systems*. Novosibirsk: IEOPP SB RAS; 2010. Pp. 206–214. (In Russ.)

2. **Granberg A.G. et al.** World Crisis and Russian Spatial Socio-Economic Development Strategy. *Region: Economics and Sociology*. 2009;(4):69–101. (In Russ.)

3. **Baranov S.V., Skufina T.P.** Comparative Dynamics of Economic Growth and Interregional Differentiation of the Northern Region of Russia. *Voprosy Statistiki*. 2015;(11):69–77. (In Russ.)

4. **Glushchenko K.P.** On Estimation of Inter-Regional Inequality. *Prostranstvennaya Ekonomika = Spatial Economics*. 2015;(4):39–58. (In Russ.) Available from: <https://doi.org/10.14530/se.2015.4.039-058>.

5. **Glinitsky V.V., Serga L.K., Bulkina A.M.** Differentiation of Municipal Entities as a Factor of Economic Development of Territories. *Voprosy Statistiki*. 2016;(8):46–52. (In Russ.)

6. **Gubanova E.S., Kleshch V.S.** Overcoming Socio-Economic Inequality as a Condition for Sustainable and Balanced Spatial Development of the Region. *Economic and Social Changes: Facts, Trends, Forecast*. 2019;12(5):44–57. (In Russ.) Available from: <https://doi.org/10.15838/esc.2019.5.65.3>.

7. **Zaytseva Y.** Interregional Differentiation in BRIC Countries: Possibilities for Assessment. *World Economy and International Relations*. 2010;(5):44–51. (In Russ.) Available from: <https://doi.org/10.20542/0131-2227-2010-5-44-51>.

8. **Kolomak E.A.** Interregional Disparities in Russia: Economic and Social Aspects. *Prostranstvennaya Ekonomika = Spatial Economics*. 2010;(1):26–35. (In Russ.) Available from: <https://doi.org/10.14530/se.2010.1.026-035>.

9. **Konopatskaya E.A.** Differentiation of Regions of the Russian Federation According to the Level of Development of Human Capital. *Economics, Statistics and Informatics. Vestnik UMO*. 2011;(2):178–184. (In Russ.)

10. **Lavrovsky B.L.** Regional Asymmetry in Russian Federation: Estimation and Approaches to its Reduction. *HSE Economic Journal*. 2003;7(4):524–537. (In Russ.)

11. **Moroshkina M.V.** Trans-Regional Differentiation of Subjects of the Russian Federation. *Economic Analysis: Theory and Practice*. 2014;13(45):20–26. (In Russ.)
12. **Pogodina E.A., Kataev E.N.** An Assessment of Differentiation Degree of Economic and Social Development of the Volga Federal District's Regions. *National Interests: Priorities and Security*. 2014;10(21):30–37. (In Russ.)
13. **Uskova T.V.** Spatial Development of Territories: State, Trends, Ways of Reducing Risks. *Problems of Territory's Development*. 2015;1(75):7–15. (In Russ.)
14. **Uskova T.V.** Territories' Sustainable Development and Modern Management Methods. *Problems of Territory's Development*. 2020;2(106):7–18. (In Russ.) Available from: <https://doi.org/10.15838/ptd.2020.2.106.1>.
15. **Usmanov D.I.** Inequality in the Socio-Economic Development of Russian Regions. *European Social Science Journal*. 2014;5-1(44):476–480. (In Russ.)
16. **Chemezova E.Yu.** Typologies of Subjects of the Russian Federation According to the Level of Socio-Economic Development. *Vestnik NSUEU*. 2010;(1):171–176. (In Russ.)
17. UN. *Transforming Our World: The 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015*. UN, 2015. 45 p. Available from: <https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/291/89/PDF/N1529189.pdf?>
18. **Aivazyan S.A.** et al. *Applied Statistics: Classification and Dimension Reduction*. Moscow: Finansy i Statistika Publ.; 1989. 607 p. (In Russ.)
19. **Ismailylova Yu.N., Khrushchev S.E.** Typological Grouping Based on Decomposition of Probability Distributions Mixtures. *Vestnik NSUEU*. 2020;(1):255–267. (In Russ.) Available from: <https://doi.org/10.34020/2073-6495-2020-1-255-267>.
20. **Pearson K.** Contributions to the Mathematical Theory of Evolution. *Philosophical Transactions of the Royal Society of London*. 1894;(185):71–110.
21. **Dempster A.P., Laird N.M., Rubin D.B.** Maximum Likelihood from Incomplete Data Via the EM Algorithm (with Discussion). *Journal of the Royal Statistical Society B*. 1977;39(1):1–38. Available from: <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>
22. **Celeux G., Diebolt J.** The SEM Algorithm: A Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem. *Computation Statistics Quarterly*. 1985;2(1):73–82.
23. **Korolev V.Yu.** *EM-Algorithm, Its Modifications and Their Application to the Problem of Separation of Mixtures of Probability Distributions. Theoretical Review*. Moscow: IPI RAN; 2007. 94 p. (In Russ.)
24. **Gorshenin A.K., Korolev V.Yu., Tursunbaev A.M.** Median Modification of Em- and Sem-Algorithms for Separation of Mixtures of Probability Distributions and Their Application to the Decomposition of Volatility of Financial Time Series. *Informatics and Applications*. 2008;2(4):12–47. (In Russ.)
25. **Akaike H.** Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov B.N., Csaki F. (eds) *Proc. of the 2nd Int. Symposium on Information Theory*. Budapest: Akademiai Kiado; 1973. Pp. 267–281.
26. **Schwartz G.** Estimating the Dimension of a Model. *The Annals of Statistics*. 1978;6(2):461–464. Available from: <https://doi.org/10.1214/aos/1176344136>.
27. **Glinskiy V.V., Ismailylova Yu.N.** Probabilistic Mixtures in Measurements of Interterritorial Differentiation. *Voprosy Statistiki*. 2020;27(3):53–64. (In Russ.) Available from: <https://doi.org/10.34023/2313-6383-2020-27-3-53-64>.
28. **Ismailylova Yu.N.** *Data Typology Based on the Decomposition of Mixtures of Probability Distributions*. Cand. Economy Sci. Diss. 08.00.12. Novosibirsk: National State University of Economics; 2022. 172 p. (In Russ.)

About the authors

Vladimir V. Glinskiy – Dr. Sci. (Econ.), Professor, Head, Research Laboratory «Sustainable Development of Social and Economic Systems», Siberian Institute of Management – Branch of RANEPa. 6, Nizhegorodskaya Str., Novosibirsk, 630102, Russia. E-mail: s444@ngs.ru. ORCID: <https://orcid.org/0000-0002-7149-3020>.

Yuliya N. Ismailylova – Cand. Sci. (Econ.), Associate Professor, Department of Business Analytics and Statistics, Siberian Institute of Management – Branch of RANEPa. 6, Nizhegorodskaya Str., Novosibirsk, 630102, Russia. E-mail: ismailylova@gmail.com. ORCID: <https://orcid.org/0000-0003-1706-6973>.

Sergey E. Khrushchev – Cand. Sci. (Phys.-Math.), Associate Professor, Senior Researcher, Laboratory of Applied Inverse Problems, Sobolev Institute of Mathematics of the SB RAS. 4, Acad. Koptyuga Ave., Novosibirsk, 630090, Russia. E-mail: hreshew@rambler.ru. ORCID: <https://orcid.org/0000-0002-0281-5022>.

Lyudmila K. Serga – Cand. Sci. (Econ.), Associate Professor, Head, Department of Business Analytics and Statistics, Siberian Institute of Management – Branch of RANEPa. 6, Nizhegorodskaya Str., Novosibirsk, 630102, Russia. E-mail: serga-lk@ranepa.ru. ORCID: <https://orcid.org/0000-0002-8934-8876>.

Funding

The study was funded by the Russian Science Foundation, project no. 22-28-00629.