



Munich Personal RePEc Archive

## **Endogenous household classification: Russian regions**

Nartikoev, Alan and Peresetsky, Anatoly

National Research University "Higher School of Economics",  
National Research University "Higher School of Economics"

11 September 2020

Online at <https://mpra.ub.uni-muenchen.de/104351/>  
MPRA Paper No. 104351, posted 26 Nov 2020 07:58 UTC

# **Endogenous household classification: Russian regions.**

**Alan Nartikoev**

National Research University Higher School of Economics, Moscow, Russian Federation;  
alan.nartikoev@gmail.com.

**Anatoly Peresetsky**

National Research University Higher School of Economics, Moscow, Russian Federation;  
aperesetsky@hse.ru

## **Abstract**

In order to study the structure of society, sociologists usually distinguish several homogeneous social groups, or classes. The most common division consists of three groups: upper, middle and lower classes. Such a partition is traditionally based on a subjective (exogenous) criteria adopted by a particular researcher. In this paper, the distribution of households in Russian federal districts is modeled as a mixture of three lognormal distributions. The mixing proportions (probabilities) of the mixture components and the corresponding distribution parameters are modeled as functions of the individual characteristics of households. The result is an endogenous decomposition of household sample into three clusters (lower, middle, upper). This classification allows to analyze the difference between regions and the patterns of intergroup dynamics in the period 2014–2018. The approach used in this work demonstrated great flexibility in analyzing the distribution of income, the dynamics of this distribution over time, as well as a migration between relatively homogeneous clusters. The use of mixture density function with endogenously determined probabilities allows for precise detection of the effects of the income heterogeneity determinants within each cluster.

## **Эндогенная классификация домохозяйств в регионах России**

Для изучения социальной структуры общества социологи выделяют группы (классы), которые характеризуются некоторой однородностью входящих в них индивидов. Наиболее часто встречается разбиение на три класса: высший класс, средний класс и низший класс. Критерии такого разбиения традиционно основаны на субъективных (экзогенных) критериях, принимаемых тем или другим исследователем. В данной работе распределение домохозяйств в федеральных округах России моделируется в виде смеси трех логнормальных распределений. Смешивающие доли (вероятности) компонент смеси, и параметры среднего трех логнормальных распределений моделируются как функции от индивидуальных характеристик домохозяйств. В результате получается эндогенное разбиение домохозяйств на три кластера (нижний, средний, верхний). Такое разбиение позволяет анализировать различие между Федеральными округами и закономерности межгрупповой динамики в период 2014–2018 гг. Использованный в работе подход продемонстрировал большую гибкость в задачах анализа распределения доходов, динамики этого распределения во времени, а также миграции между относительно однородными кластерами. Модель смеси распределений с эндогенно определяемыми вероятностями позволяет более тонко регистрировать эффекты детерминант неоднородности доходов внутри каждого кластера.

**Ключевые слова:** распределение по доходам; смесь распределений; Россия; средний класс.

**JEL classification:** C14; C15; C46; D31; I32; R20.

### **1. Введение**

При изучении социальной структуры общества социологами принято выделять сравнительно однородные группы (классы). Наиболее часто встречается разбиение на три класса: высший класс, средний класс, низший класс.

Подобная классификация позволяет анализировать гетерогенность общества, направление его развития, реакцию на экономические шоки. Многие исследователи используют более детальные классификации, предполагающие наличие шести-семи страт. При этом огромное внимание в литературе уделяется среднему классу как классу, обеспечивающему стабильность развития общества.

---

<sup>1</sup> **Нартиков Алан Ревазович** — Национальный исследовательский университет «Высшая школа экономики», Москва, alan.nartikoev@gmail.com.

**Пересецкий Анатолий Абрамович** — Национальный исследовательский университет «Высшая школа экономики», Москва, aperesetsky@hse.ru.

Обзор различных подходов к определению среднего класса можно найти, например, в (Тихонова, Мареева, 2009; Тихонова, 2016). В этих работах выделяются четыре подхода:

- Первый (*маркетинговый*) связан с уровнем среднедушевого дохода, со сравнительно высоким жизненным стандартом.
- Второй (*субъективный*) основан на самоидентификации индивида;
- Третий (*ресурсный*) основан на типе и структуре ресурсов, которыми располагает тот или иной индивид;
- Четвертый (*неовеберлианский*) связан с попыткой комплексного применения в условиях России критериев выделения среднего класса, традиционных для так называемой многомерной стратификации (определенные характеристики социально-профессионального статуса, образование, имущественные и доходные характеристики, а также иногда к ним добавляется и самоидентификация).

Все эти критерии основаны на субъективных (экзогенных) критериях, принимаемых исследователем. Соответственно, в разных работах приводятся существенно различающиеся оценки численности среднего класса в России: 20% (Малева, 2008); 59% — оценка Всемирного банка по состоянию на 2010 г. (Доклад..., 2014); 44% в 2015 г. (Тихонова, 2016); 86% в 2020 г. по данным опроса Левада-Центра; 58% в 2015 г., 61% в 2018 (Аникин, 2020); 42% в 2014 г. (Аналитический...2014), 58% в 2018 г. (Тихонова, 2020).

Можно ли предложить способ разбиения населения России на классы по уровню дохода без применения извне задаваемых экзогенных порогов? В рамках первого (*маркетингового*) подхода, это возможно сделать с помощью модели смеси распределений. При таком подходе распределение домохозяйств по доходам представляется в виде смеси распределений, с вероятностями принадлежности той или иной компоненте, зависящими от различных характеристик домохозяйств.

Оценивание параметров такой модели позволяет получить «оптимальное» *эндогенное* разбиение, при этом не требуется заранее устанавливать *экзогенные* пороговые значения переменных для классификации. При этом полученные компоненты

распределения можно ассоциировать с теми или иными социальными группами, поскольку каждая компонента разбиения будет относительно однородна в терминах индивидуальных характеристик входящих в нее домохозяйств. Такой подход реализован в ряде работ, отметим некоторые из них.

Pittau, Zelli (2006) применяли смесь нормальных распределений для изучения распределения покупательной способности подушевого дохода в 110 регионах стран EU-12 в период 1977–1996 гг. Компоненты смеси соответствуют богатым и бедным регионам, а вероятности компонент соответствуют пропорциям соответствующих регионов. Количество компонент (1, 2, 3) не фиксировалось заранее, а определялось по бутстрапированному тесту отношения правдоподобия (LR). Дело в том, что, несмотря на то, что модель с  $k$  компонентами смеси является ограничением на параметры модели с  $k + 1$  компонентами, статистика LR-теста в данном случае не имеет асимптотического хи-квадрат распределения, поскольку параметры модели с ограничениями не являются внутренней точкой допустимого множества параметров модели без ограничений. Качество подгонки модели смеси распределений определялось также и по сравнению с непараметрической оценкой плотности. Cowell, Flachaire (2015) отмечают, что использование модели смеси  $k$  распределений является «полупараметрическим компромиссом» между простым параметрическим подходом ( $k = 1$ ) и непараметрическим подходом ( $k = n$ ), при котором модель смеси эквивалентна ядерной оценке плотности.

Pittau et al. (2010) моделируют межстрановое распределение населения по доходам на данных таблиц Penn World в 1960–1995–2000 гг., с помощью смеси нормальных распределений. В результате применения LR-теста они приходят к смеси из трех распределений, таким образом, для каждого года выделяя 3 группы (богатые-средние-бедные) из 102 стран. Они изучают переходы стран между группами. Мобильность стран между группами оказалась довольно низкой. Из 816 возможных переходов реализовались только 51, при этом 64 страны из 102 оставались в одной группе в течение всего периода.

В работе (Flachaire, Nunez, 2007) предложена модель смеси логнормальных распределений, вероятности принадлежности которым определяются logit-моделью упорядоченного множественного выбора. В качестве факторов этой logit-модели, определяющих вероятность принадлежности той или иной группе, выбраны следующие характеристики домохозяйства: наличие пенсионера во главе домохозяйства, семейное положение главы домохозяйства, статус занятости взрослых членов домохозяйства, число детей в домохозяйстве. Эта идея моделирования межгруппового различия была также реализована в работе (Cowell, Flachaire, 2015), в которой авторы используют модель смеси трех логнормальных распределений с вероятностями, зависящими от состава домохозяйства. Кроме того, Cowell, Flachaire (2015) на основе данных о доходах в Великобритании в 1992 и 1999 гг. рассматривают также и другую модель, в которой параметры логнормальных распределений смеси зависят от двух факторов: пол главы домохозяйства и факт членства главы домохозяйства в профсоюзе. Такая модель позволяет моделировать внутригрупповую неоднородность.

Модель смеси гамма-распределений для использована в работе (Chotikapanich, Griffiths, 2008) для анализа распределения по доходам в Канаде в 1978 г.. Оптимальной оказалось смесь, состоящая из двух компонент. В работе полученное параметрическое распределение применяется для оценки и анализа мер неравенства и бедности. Выбор смеси гамма распределений был обусловлен тем, что использование других распределений (Beta-2, логнормальное, Dagum, Singh-Maddala) давало большее отклонение модельной функции распределения от выборочной функции распределения, измеренного как RMSE.

В работе (García-Fernández et al., 2013) авторы применяют смесь логнормальных распределений для изучения динамики распределения домохозяйств по доходам в Израиле в 1997–2008 гг. Вероятности принадлежности группам задаются логистической моделью дискретного множественного выбора (multinomial logit) с несколькими группами факторов: 1) демографические (принадлежность одной из социокультурных групп); 2) характеристики главы домохозяйства (возраст, образование,

количество детей, количество работающих взрослых членов домохозяйства); 3) степень зависимости от социальных пособий. Авторы получили, что оптимальное количество компонент смеси было равно четырём в 1997 г. и трём — в последующие годы. Принадлежность к ультра-ортодоксальным евреям, возраст и зависимость от социальных пособий повышают вероятность принадлежности к низкой доходной группе. Участие на рынке труда и малый размер семьи повышают вероятность принадлежности к группе с высокими доходами.

Lubrano, Ndoye (2016) применяют смесь логнормальных распределений для моделирования распределения по доходам.

Смесь логнормальных распределений применена для моделирования распределения по доходам в работе (Lubrano, Ndoye, 2016). В качестве эмпирической базы они использовали данные по Великобритании (UK Family Expenditure Survey, FES) за период 1979–1996 гг. По критериям BIC и Chib в большинстве периодов времени они получают модель с тремя компонентами в качестве оптимальной. Авторы интерпретируют каждую компоненту смеси как социальную группу, однако указывают на то, что нет гарантии, что в разные периоды компоненты смеси соответствуют одним и тем же социальным группам. Для полной идентификации межгрупповой мобильности необходимо иметь в распоряжении панельные данные. Использование модели смеси распределений позволяет разложить индекс неравенства (который аналитически выражается через параметры модели смеси) на межгрупповую и внутригрупповую компоненты.

В работе (Pittau et al., 2016) авторы тщательно изучают различные критерии выбора количества компонент смеси, используя симуляционное моделирование. Применяя модель смеси нормальных распределений для анализа распределения стран по доходам (данные Penn World Table, PWT, за период 1960–2010) они показывают, что большинство критериев указывают на смесь из трех распределений.

В работе (Fourrier-Nicolai, Lubrano, 2020) смесь логнормальных распределений используется для изучения эволюции детской бедности в Германии после 2002 г.

В данной работе изучается динамика распределения российских домохозяйств по доходам с использованием смеси логнормальных распределений<sup>2</sup>. Количество компонент смеси — три — заранее фиксировано. Такой выбор обусловлен тем, что такое количество компонент смеси получалось оптимальным (по формальным статистическим критериям), в большинстве цитированных выше работ. Кроме того, деление общества на три класса наиболее часто встречается в социологии.

В работе модель смеси оценивается отдельно на данных по каждому из восьми федеральных округов России на основе данных Выборочного наблюдения доходов населения (ВНДН). В определенном смысле это соответствует второму (*субъективному*) подходу к определению среднего класса: понятие высоких (низких, средних) доходов может быть разным в разных федеральных округах. Таким образом, в работе учитываются региональные различия в распределении доходов. Отметим, что Gontmakher, Ross (2015) отмечают необходимость учета региональных различий, размера города проживания, распределения населения в регионе по возрасту при определении среднего класса.

Для моделирования вероятностей компонент смеси используется модель дискретного множественного выбора (multinomial logit) с набором факторов, аналогичных используемым в (García-Fernández et al., 2013). Эти факторы отвечают за межгрупповую гетерогенность. Также рассматривается вторая модель, в которой параметры компонент смеси представляются в виде линейной комбинации характеристик домохозяйств, аналогично работе Flachaire, Nunez (2007). Такая модель позволяет учесть как межгрупповую, так и внутригрупповую неоднородность.

Представление распределения домохозяйств по доходам в виде смеси трех логнормальных распределений позволяет оценить для каждого домохозяйства вероят-

---

<sup>2</sup> В монографии (Айвазян, 2012) логнормальное распределение широко используется для моделирования распределения населения России по доходам.



ность его принадлежности к той или иной компоненте смеси. Это позволяет разделить домохозяйства на три кластера, соответствующих трем компонентам смеси. Естественно отнести домохозяйство к кластеру, которому соответствует наибольшая вероятность. Таким образом, получается эндогенное разбиение домохозяйств в каждом федеральном округе на три кластера. При этом критерии принадлежности к каждому из кластеров различны в разных федеральных округах и определяются распределением характеристик домохозяйств и их доходов в каждом из федеральных округов.

Во втором разделе работы приведено описание двух применяемых моделей смеси распределений. Третий раздел посвящен описанию и первичному анализу выборочных данных. В четвертом разделе обсуждаются результаты оценивания моделей и дается интерпретация получившемуся разбиению на три кластера, а также проводится анализ мобильности между выделенными кластерами в период с 2014 по 2018 гг., который характеризовался экономической рецессией. Показано, что в предложенных моделях классические детерминанты дохода демонстрируют ожидаемые эффекты на межгрупповую и внутригрупповую динамику, а также продемонстрированы различия в численности среднего кластера и динамике переходов между кластерами по федеральным округам России.

## **2. Модели**

Для каждого региона (федерального округа) и для каждого года (2014, 2018) Рассматриваются две модели распределения доходов.

### *Модель 1. Кластеризация. Смесь логнормальных распределений*

Плотность распределения доходов является смесью трех логнормальных распределений, причем вероятности принадлежности домохозяйства одному из кластеров задаются мультиномиальной логит моделью.

$$f_1(x; w, \lambda, \mu, \sigma^2) = \sum_{k=1}^3 \pi_k(w, \lambda) \psi(x; \mu_k, \sigma_k^2), \quad (1)$$

где  $x$  — доход домохозяйства,  $\mu = \{\mu_1, \mu_2, \mu_3\}$ ,  $\sigma^2 = \{\sigma_1^2, \sigma_2^2, \sigma_3^2\}$  — параметры трех

логнормальных распределений с плотностями  $\psi(x, \mu_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi} \sqrt{\sigma_k^2} x} e^{-\frac{(\ln x - \mu_k)^2}{2\sigma_k^2}}$

(предполагается  $\mu_1 \leq \mu_2 \leq \mu_3$ ), а параметры смеси  $\pi_k$  (вероятности принадлежности домохозяйства одному из кластеров), определяются факторами  $w$  в соответствии с мультиномиальной логит моделью (первый кластер взят за базис)

$$\pi_1(w; \lambda) = \frac{1}{1 + e^{w'\lambda_2} + e^{w'\lambda_3}}, \quad \pi_2(w; \lambda) = \frac{e^{w'\lambda_2}}{1 + e^{w'\lambda_2} + e^{w'\lambda_3}}, \quad \pi_3(w; \lambda) = \frac{e^{w'\lambda_3}}{1 + e^{w'\lambda_2} + e^{w'\lambda_3}}, \quad (2)$$

Здесь  $\lambda = \{\lambda_2, \lambda_3\}$ , каждый из векторов  $\lambda_2, \lambda_3$  имеет размерность вектора факторов  $w$ , определяющих разбиение домохозяйств на кластеры. Факторы  $w$  отвечают за межгрупповую гетерогенность.

Подобная модель использовалась в (García-Fernández et al., 2013); в работе (Flachaire, Nunez, 2007), в которой параметры смеси  $\pi_k$  определяются в соответствии с моделью упорядоченного выбора (ordered logit).

*Модель 2. Кластеризация. Смесь логнормальных распределений с регрессионной моделью для параметра  $\mu$  логнормального распределения в каждой компоненте смеси.*

Модель 2 отличается от модели (1) тем, что дополнительно предполагается зависимость параметра  $\mu$  логнормального распределения от факторов  $z$ :

$$f_2(x; w, z, \lambda, \beta, \sigma^2) = \sum_{k=1}^3 \pi_k(w, \lambda) \psi(x; \mu(z, \beta_k), \sigma_k^2), \quad (3)$$

где  $\mu(z, \beta_k) = z'\beta_k$ , здесь  $\beta = \{\beta_1, \beta_2, \beta_3\}$  и  $\beta_k$  — вектор параметров, размерность которого равна размерности вектора факторов  $z$ . Модель 1 является частным случаем модели 2. Вероятности принадлежности кластерам следуют мультиномиальной

логит модели (2). Факторы  $w$  отвечают за межгрупповую, а факторы  $z$  — за внутригрупповую гетерогенность.

Подобная модель с учетом внутригрупповой неоднородности использовалась, например, в работе (Cowell, Flachaire, 2015).

### *Метод оценивания*

Параметры моделей 1 и 2 оцениваются методом максимального правдоподобия.

Для модели 1 функция правдоподобия имеет вид

$$L_1(\lambda, \mu, \sigma^2) = \prod_{i=1}^n f_1(x_i; w_i, \lambda, \mu, \sigma^2), \quad (3)$$

где  $i = 1, \dots, n$  — номера домохозяйств. Соответственно, для модели 2 функция правдоподобия имеет вид:

$$L_2(\lambda, \beta, \sigma^2) = \prod_{i=1}^n f_2(x_i; w_i, z_i, \lambda, \beta, \sigma^2). \quad (4)$$

Стандартные ошибки полученных оценок коэффициентов получаются из общей асимптотической теории оценок максимального правдоподобия. Оценки рассчитываются отдельно для каждого из 8 федеральных округов, по каждому из двух периодов (2014 г., 2018 г.).

В моделях 1 и 2 число оцениваемых параметров велико, функция правдоподобия имеет сложный вид. В данной работе (как и во многих работах цитированных выше), для оценки параметров этих моделей применяется EM-алгоритм (Dempster et al., 1977), который, в данном случае, обладает преимуществом в быстрой работе и робастности по сравнению с традиционными алгоритмами поиска максимума функции.

Заметим, что по модели 1 получаются оценки параметров  $(\hat{\mu}_k, \hat{\sigma}_k^2)$  трех лог-нормальных распределений доходов домохозяйств по трем кластерам, и оценки ве-

роятности того, что данное домохозяйство  $i$  принадлежит кластеру  $k$  :  
 $\hat{\pi}_k = \pi_k(w_i, \hat{\lambda}_k)$ .

Апостериорные вероятности принадлежности домохозяйства с доходом  $x_i$  и факторами  $w_i$  к кластеру  $k$  рассчитываются в модели 1 по формуле Байеса:

$$\hat{\pi}_{ik} = \frac{\pi_k(w_i, \hat{\lambda}_k) \psi(x_i; \hat{\mu}_k, \hat{\sigma}_k^2)}{\sum_{m=1}^3 \pi_m(w_i, \hat{\lambda}_m) \psi(x_i; \hat{\mu}_m, \hat{\sigma}_m^2)}. \quad (5)$$

Отсюда, в соответствии с идеологией метода максимального правдоподобия можно получить *эндогенное* разбиение домохозяйств на 3 кластера: домохозяйство  $i$  относится к кластеру  $k$ , если

$$\hat{\pi}_{ik} = \max\{\hat{\pi}_{i1}, \hat{\pi}_{i2}, \hat{\pi}_{i3}\}.$$

Упорядочив полученные кластеры, например, по  $\hat{\mu}_k$ , можно их условно назвать «нижний кластер», «средний кластер» и «верхний кластер».

В модели 2 апостериорные вероятности принадлежности домохозяйства с доходом  $x_i$  и факторами  $w_i, z_i$  к кластеру  $k$  рассчитываются, соответственно, по аналогичной формуле:

$$\hat{\pi}_{ik} = \frac{\pi_k(w_i, \hat{\lambda}_k) \psi(x_i; \mu(z_i, \hat{\beta}_k), \hat{\sigma}_k^2)}{\sum_{m=1}^3 \pi_m(w_i, \hat{\lambda}_m) \psi(x_i; \mu(z_i, \hat{\beta}_m), \hat{\sigma}_m^2)}. \quad (6)$$

Аналогичные разбиения домохозяйств на кластеры можно получить с использованием модели 2. Однако в этом случае разбиение на кластеры происходит не на основе близости значений  $\ln x_i$ , а по принципу близости моделей  $\ln x_i = z_i' \beta_k$ . (см. *метод динамических сгущений*, (Дидэ,1985)). В этом случае труднее присвоить названия трем кластерам, разве что упорядочив их средним значениям дохода в каждом кластере.

### 3. Данные

Модели распределения по доходам (1) и (2) оценивались по микроданным располагаемых совокупных доходов домохозяйств во всех областях и федеральных

округах России в 2014 и 2018 годах. Данные взяты из Выборочного наблюдения доходов населения (ВНДН), ежегодно проводимого Федеральной службой государственной статистики, начиная с 2014 года.

Различные домохозяйства имеют разные вероятности включения в программу обследования, что нарушает классические предпосылки проведения статистического оценивания. Поэтому в выборочных исследованиях к данным традиционно применяется взвешивание по обратной вероятности (англ. *inverse probability sampling*) — каждому домохозяйству присваивается коэффициент взвешивания, обратно пропорциональный вероятности включения домохозяйства в выборку. В случае ВНДН весовые коэффициенты дважды корректируются с учётом демографической структуры населения, известной из переписи, а также с учётом соответствия выборочных показателей доходов данным макроэкономической статистики<sup>3</sup>. Сумма всех весовых коэффициентов равна численности всех домохозяйств в рассматриваемом субъекте федерации в соответствующем году. Взвешивание численных характеристик домохозяйств позволяет распространить выборочные статистики на генеральную совокупность, и далее расчет выборочных статистик и оценивание параметров будет производиться с учетом взвешивания.

Для учета неоднородности домохозяйств по своему социально-демографическому составу, обеспечения их сопоставимости следует рассматривать нормированный доход — в пересчете на одного условного члена домохозяйства, так называемый эквивалентный доход. Показатель эквивалентного дохода (*equivalized income*) обычно используется в работах по анализу неравенства по доходам (Cowell, van Kerm, 2015).

В целях обеспечения сопоставимости, величины располагаемых доходов домохозяйств приведены к ценам 2014, а также были заранее нормированы в соответствии с модифицированной шкалой эквивалентности ОЭСР (англ. *OECD modified*

---

<sup>3</sup> Процедура взвешивания подробнее описана в официальной документации исследования: [https://rosstat.gov.ru/free\\_doc/new\\_site/vndn-2018/index.html](https://rosstat.gov.ru/free_doc/new_site/vndn-2018/index.html)

*equivalence scale*), предложенной в работе (Hagenaars et al. 1994). Шкала назначает каждому члену домохозяйства весовые коэффициенты в зависимости от возраста:

- глава домохозяйства (взрослый старше 14 лет) — 1;
- каждый последующий взрослый — 0.5;
- ребёнок до 14 лет — 0.3.

Следовательно, величина эквивалентного дохода  $i$ -го домохозяйства определяется формулой

$$x_i = \frac{d_i}{1 + 0.5(a_i - 1) + 0.3k_i}, \quad (7)$$

где  $d_i$  — располагаемый годовой денежный доход  $i$ -го домохозяйства,  $a_i$  — количество взрослых (старше 14 лет),  $k_i$  — количество детей (до 14 лет).

В таблице 1 приведена динамика выборочных статистик распределения эквивалентного дохода по регионам. Распределения скошены вправо: отношение среднего к медиане для ЦФО равно 1.26–1.31, а в остальных округах лежит в диапазоне 1.06–1.21. ЦФО выделяется наибольшей скошенностью распределения вправо, т.е. долей населения с доходами, существенно превышающими медианный доход. Это также можно наблюдать по ядерным оценкам плотности, представленным на рис.1: ЦФО характеризуется наиболее «тяжёлым» правым хвостом, эта особенность характерна также для СЗФО, УФО и ДВФО. Графики оценок плотностей распределения в ПФО, СФО, ЮФО, СКФО имеют более симметричный профиль и малую величину разброса (см. IQR в таблице 1), отчего кривые кажутся островершинными.

Также во всех федеральных округах заметна общая тенденция изменения формы графика плотности, а именно уменьшение асимметрии вместе с уменьшением дисперсии, отчего в период с 2014 по 2018 гг. кривые приобретают сравнительно все более островершинные силуэты. Это косвенно свидетельствует о сближении уровня доходов между представителями нижней и верхней частей распределения в результате общего снижения среднего и медианного дохода — аналогичный вывод получен в работе (Нарतिकоев, Пересецкий, 2019).

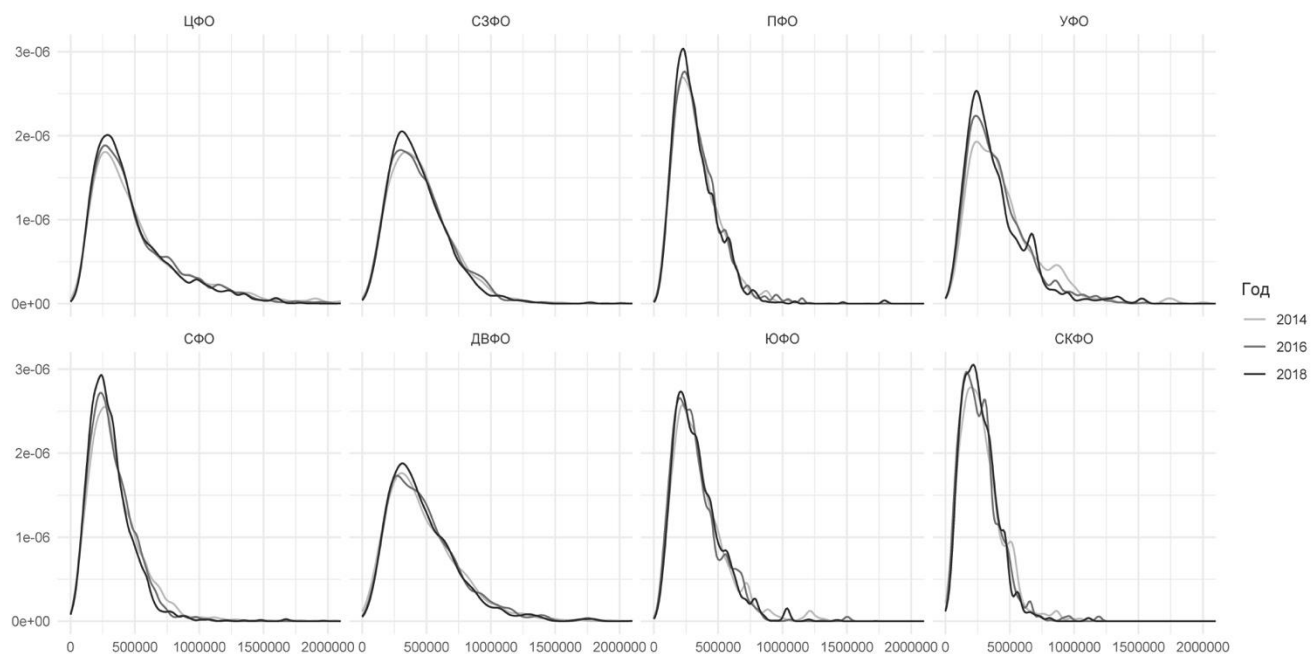
Чтобы подтвердить это наблюдение численно, рассмотрим динамику медианы эквивалентного дохода по регионам (в ценах 2014 г.), представленную на рис. 2. Заметно падение медианного дохода, что есть закономерное следствие экономического кризиса. В большинстве регионов падение продолжается до 2017 г, а в 2018 г. наблюдается начало восстановления. В то же время на рис. 3 представлена динамика отношения IRQ к медианному доходу — эта статистика отражает концентрацию значений величины вокруг медианы. Все графики показывают снижение этого показателя по всем регионам, начиная с 2015 г., что соответствует наблюдаемому на рис. 1 изменению формы кривых в сторону уменьшения правосторонней скошенности и более острого пика.

Значения статистики IQR/медиана в Центральном федеральном округе выше, чем в других регионах, и в то же время наименьшие значения соответствуют Приволжскому и Южному федеральным округам. Четыре федеральных округа, которым соответствуют наиболее островершинные графики на рис. 1, — Приволжский, Сибирский, Южный и Северо-Кавказский — имеют существенно меньшие значения относительного интерквартильного размаха (таблица 1) и медианного дохода. Отсюда можно сделать вывод, что наименее экономически благополучные регионы характеризуются более островершинным распределением доходов, то есть для таких регионов ожидается большая концентрация вероятностной массы компонент смеси в левом хвосте распределения, в отличие от более «богатых» федеральных округов с «тяжелыми» правыми хвостами графиками плотности распределения доходов.

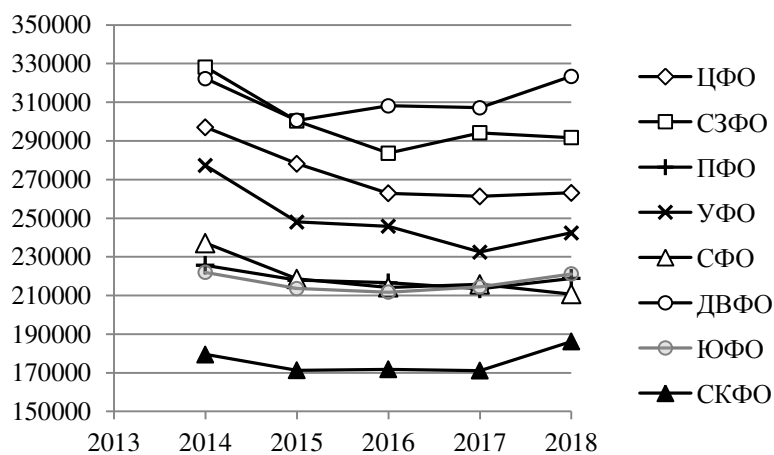
**Таблица 1.** Описательные статистики реального эквивалентного располагаемого дохода, 2014–2018 гг. (в ценах 2014 г.)

ФО	Q1	Медиана	Среднее	Q3	IQR	IRQ/медиана
<b>ЦФО</b>						
2014	201580	297110	381830	489400	287820	0.97
2015	190630	278200	364720	461090	270460	0.97
2016	185090	262900	338780	419190	234100	0.89
2017	184900	261330	341440	431470	246570	0.94
2018	187780	263130	331110	415400	227620	0.87
<b>СЗФО</b>						
2014	220030	328240	375380	476460	256430	0.78
2015	204290	300290	345140	440770	236480	0.79
2016	195560	283640	328470	417460	221900	0.78
2017	208480	294170	334180	419080	210600	0.72
2018	203720	291720	329660	414030	210310	0.72
<b>ПФО</b>						
2014	163590	225600	245900	300810	137220	0.61
2015	161090	217920	237710	288100	127010	0.58
2016	162440	216590	233270	283810	121370	0.56
2017	159940	213420	230000	277000	117060	0.55
2018	168960	218820	235540	281740	112780	0.52
<b>УФО</b>						
2014	196940	277290	324020	399240	202300	0.73
2015	171920	248050	290340	355050	183130	0.74
2016	178420	245790	290120	352680	174260	0.71
2017	166670	232460	272950	329530	162860	0.70
2018	178260	242470	283290	340020	161760	0.67
<b>СФО</b>						
2014	163160	237150	269600	329790	166630	0.70
2015	151580	218610	249290	310860	159280	0.73
2016	148730	214160	237420	291370	142640	0.67
2017	153900	215760	239310	292280	138380	0.64
2018	150750	210700	231520	283650	132900	0.63
<b>ДВФО</b>						
2014	222430	322250	389210	483860	261430	0.81
2015	207570	300630	364740	451460	243890	0.81
2016	215020	308150	365300	456820	241800	0.78
2017	214720	307220	357370	444340	229620	0.75
2018	221810	323420	367810	458950	237140	0.73
<b>ЮФО</b>						
2014	161330	221990	245110	302490	141160	0.64
2015	153330	213630	235500	289680	136350	0.64
2016	154350	211710	229350	282290	127940	0.60
2017	160840	214440	232200	281320	120480	0.56
2018	166390	221140	238930	290400	124010	0.56
<b>СКФО</b>						
2014	118730	179510	198180	249570	130840	0.73
2015	119140	171350	187450	231930	112790	0.66
2016	119950	171820	185680	230650	110700	0.64
2017	120100	171170	184220	227080	106980	0.62
2018	132960	186180	197790	247360	114400	0.61





**Рис. 1.** Ядерные оценки плотности распределения доходов в 2014, 2016 и 2018 гг. по федеральным округам



**Рис. 2.** Динамика медианы дохода по регионам.

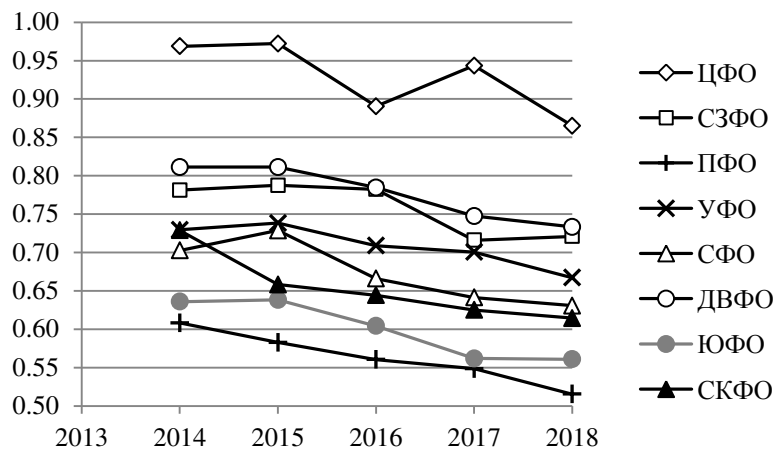


Рис. 3. Динамика показателя IRQ/медиана доходов по регионам.

## 4. Результаты

### 4.1. Кластеризация. Смесь логнормальных распределений

В данном разделе приводятся результаты оценивания модели 1 (1) для каждого федерального округа России в 2014 и 2018 гг. В качестве факторов  $w$ , определяющих эндогенное разбиение на кластеры и отвечающих за межгрупповую неоднородность, рассматривались различные наборы индивидуальных характеристик домохозяйств. Были выбраны следующие (относительно простой набор, при котором получаются интерпретируемые результаты) <sup>4</sup>:

- пол главы домохозяйства ( $male = 0$  для женщин и 1 для мужчин),
- проживание домохозяйства в городе или в сельской местности ( $urban = 1$  в городе, 0 иначе)
- доля работающих взрослых лиц в домохозяйстве ( $empl.ratio$ )

Согласно методологии ВДН, главой домохозяйства называется человек, вносящий наибольший вклад в бюджет домохозяйства. В работе (Flachaire, Nunez, 2007) применялся схожий по содержанию набор вспомогательных переменных.

<sup>4</sup> Рассматривались также такие факторы как наличие пенсии у главы домохозяйства, наличие инвалидности у главы домохозяйства, но они оказались незначимы.

В таблице 2 приведены результаты оценивания — оценки векторов  $\lambda_2, \lambda_3$  и параметров компонент логнормальной смеси:  $\mu_i, \sigma_i^2, i = 1, \dots, 3$ . Кластеры (компоненты смеси) во избежание неоднозначности решений перенумерованы по возрастанию оценок параметра среднего  $\mu$  логнормальных распределений. В качестве базисной взята компонента с наименьшим значением параметра  $\mu$ .

Средние значения дохода по полученным кластерам (точнее – математические ожидания соответствующих распределений) ведут себя по-разному в результате кризиса (при переходе от 2014 г. к 2018 г.) в различных округах:

- в ЦФО, СФО, СКФО, ЮФО средние значения уменьшаются во всех кластерах;
- в УФО среднее значение уменьшается в верхнем и среднем кластерах и несколько увеличиваются в нижнем;
- в ДВФО среднее значение уменьшается в верхнем кластере и увеличивается в нижнем и среднем.

Знаки и величины коэффициентов в таблице 2 показывают, что в городе (*urban*) вероятность принадлежности к нижнему кластеру ниже, а к верхнему кластеру — выше, чем в сельской местности; увеличение доли работающих членов домохозяйства (*empl. ratio*) также снижает вероятность принадлежности к нижнему кластеру и повышает вероятность принадлежности к верхнему кластеру во всех ФО, за исключением ПФО в 2014 г. — в последнем увеличивается вероятность верхнего кластера и уменьшается вероятность среднего. Во всех федеральных округах домохозяйство, во главе которого мужчина, имеет больше шансов принадлежать к верхнему кластеру, за исключением СКФО, в котором ситуация обратная.

В результате кризиса в 2018 г. по сравнению с 2014 г.: в ЦФО значительно уменьшились средние доходы в верхнем кластере, и уменьшилась разница между средними доходами верхнего и нижнего кластеров; в СЗФО и ДВФО увеличились средние доходы среднего кластера; доходы нижнего кластера несколько увеличи-

лись в УФО и ДВФО. Все это может быть объяснено изменением границ между кластерами. Ниже переходы между кластерами исследованы подробнее.

Значение параметра разброса  $\sigma$  после кризиса снижается во всех компонентах, за исключением нижней компоненты в УФО и СКФО и средней и верхней компоненты в ДВФО. Этот факт подтверждает динамику усиления «островершинности» кривых плотности на рис. 1, т.е. уменьшением разброса доходов внутри кластеров.

Таблица 2. Оценки параметров модели 1\*

	ЦФО			СЗФО			ПФО			УФО		
	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3
$\mu$	12.292	12.960	13.845	11.951	12.568	13.238	11.975	12.381	12.932	12.059	12.571	13.244
	12.285	12.886	13.650	12.094	12.710	13.280	11.918	12.355	12.963	12.105	12.414	13.085
$\sigma$	0.516	0.406	0.401	0.810	0.404	0.374	0.565	0.341	0.398	0.549	0.360	0.446
	0.466	0.333	0.345	0.542	0.352	0.337	0.548	0.327	0.350	0.669	0.323	0.436
<i>среднее</i>	248945	461586	1116231	215195	311627	602143	186174	252501	447179	200757	307280	623565
	241199	417169	899303	207037	352362	619259	174146	244998	453155	226133	259333	529468
const	0	-3.774	-21.979	0	0.619	-6.103	0	-0.265	-5.194	0	-0.982	-9.354
	0	-3.262	-12.487	0	-1.230	-8.831	0	0.322	-5.662	0	-0.809	-7.366
urban	0	2.632	16.299	0	1.636	4.818	0	1.609	2.405	0	1.490	4.400
	0	1.899	7.783	0	2.183	4.742	0	1.212	2.583	0	1.373	3.294
empl. ratio	0	4.242	8.706	0	1.166	6.920	0	-1.724	6.395	0	1.856	10.503
	0	4.306	7.666	0	2.020	8.635	0	0.278	7.400	0	0.095	8.747
male	0	0.415	0.719	0	-0.087	0.753	0	0.500	0.674	0	0.329	0.780
	0	-0.085	0.232	0	0.329	0.840	0	-0.166	0.223	0	0.430	0.336

	СФО			ДВФО			ЮФО			СКФО		
	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3
$\mu$	11.813	12.432	13.014	12.143	12.731	13.403	11.870	12.433	13.017	11.376	12.222	12.830
	11.800	12.334	12.914	12.390	12.901	13.356	11.933	12.431	13.039	11.243	12.083	12.773
$\sigma$	0.664	0.359	0.435	0.688	0.394	0.386	0.694	0.434	0.419	0.809	0.480	0.382
	0.671	0.340	0.366	0.521	0.459	0.413	0.509	0.357	0.353	0.839	0.424	0.310
<i>среднее</i>	168338	267441	492882	237892	365315	713455	181829	275665	491282	120943	227990	401345
	166805	240848	434127	275210	445381	687966	173254	267015	489413	108574	193486	369796
const	0	-0.458	-6.557	0	-1.455	-9.345	0	0.358	-7.150	0	1.024	-4.908
	0	-0.020	-6.944	0	-10.483	-43.083	0	-0.349	-7.916	0	1.871	-4.867
urban	0	1.936	3.915	0	1.804	4.979	0	2.200	4.539	0	1.492	3.471
	0	1.636	3.553	0	2.510	5.096	0	2.016	3.273	0	2.196	4.472
empl. ratio	0	0.448	6.791	0	2.941	9.621	0	1.136	8.720	0	4.710	13.028
	0	0.250	8.397	0	18.952	51.112	0	1.922	10.698	0	5.888	15.283
male	0	0.308	1.188	0	0.361	0.348	0	-0.148	0.925	0	-1.039	-1.256
	0	0.015	0.368	0	1.523	1.481	0	-0.725	0.120	0	-1.724	-1.551

[\*] Верхние значения ячеек соответствуют значениям 2014 года, нижние — 2018 года. Все коэффициенты значимы на 1%-ном уровне.

$\text{среднее} = \exp(\mu + \sigma^2 / 2)$  — математическое ожидание логнормальной случайной величины с параметрами  $(\mu, \sigma^2)$

#### 4.2. Кластеризация. Смесь логнормальных распределений с регрессионной моделью для параметра среднего $\mu$ .

В данном разделе приводятся результаты оценивания модели 2 для каждого округа в 2014 и 2018 гг. В качестве переменных  $w$  определяющих эндогенное разбиение на кластеры взят тот же набор факторов, что и в модели 1. В качестве факторов, определяющих внутригрупповую неоднородность, определяющих значение параметра среднего  $\mu$ , взяты<sup>5</sup>:

- возраст главы домохозяйства (лет), *Age*;
- уровень образования главы домохозяйства: (две фиктивных переменных: *secondary* — среднее, *tertiary* — высшее, (базисная категория — начальное);
- семейное положение главы домохозяйства: *married* — женат/замужем; *divorced* — разведен(-а), *widow(er)* — вдова/вдовец, (базисная категория — не состоял в браке);
- население местности, в которой проживает домохозяйство: *Urban: 0–250k* (города с населением менее 250 тыс.), *Urban: 250k–1000k* (города с населением от 250 тыс. до 1 млн.), *urban: 1000k+* — (города с населением более 1 млн.), (базисная категория — проживание в сельской местности).

Результаты оценивания приведены в таблице 3. Исходя из полученных оценок, можно сделать следующие выводы о влиянии факторов на показатель среднего внутри каждой компоненты.

---

<sup>5</sup> Для проверки робастности результатов, был оценен ряд моделей с альтернативной спецификацией переменных. В большинстве случаев проявлялись закономерности, описанные ниже.

- Возраст главы домохозяйства — положительный предиктор дохода во всех случаях, кроме верхней компоненты в СКФО;
- Образование увеличивает доход по сравнению с базисным уровнем, во всех случаях, кроме верхней компоненты в 2018 г. в СЗФО, ЮФО и средней компоненты в ЮФО.
- Домохозяйство с главой, не состоящим(-ей) в браке, имеет больший доход, чем другие категории в среднем кластере, за редкими исключениями в 2014 г. (СФО, УФО). Для СЗФО, ЦФО, ДВФО характерно, что в нижнем кластере состоящий в браке глава домохозяйства имеет больший доход, чем не состоящий в браке; в ЦФО, ПФО, СФО, ДФО, СКФО — в верхнем кластере, однако после кризиса, в 2018 г. это преимущество уменьшается
- Жители села проигрывают городским жителям, за редкими исключениями: например, ЮФО, верхний кластер — в 2018 г.

Таблица 3. Оценки параметров модели 2.

	ЦФО			СЗФО			ПФО			УФО		
	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3
const	10.839	12.066	12.828	9.744	11.836	12.560	10.854	11.896	12.673	9.439	11.511	12.490
	10.990	11.910	12.628	10.966	12.253	12.913	10.405	11.641	12.556	10.184	11.871	12.190
Age	0.016	0.006	-0.000	0.024	0.008	0.005	0.018	0.010	0.001	0.024	0.011	0.006
	0.016	0.011	0.002	0.015	0.010	0.001	0.023	0.013	0.001	0.024	0.010	0.010
secondary	0.063	0.187	0.122	0.275	0.115	0.190	0.090	0.149	0.165	0.057	0.137	0.160
	0.122	0.154	0.078	0.139	0.154	-0.086	0.075	0.191	0.094	0.123	0.040	0.161
tertiary	0.298	0.479	0.376	0.302	0.304	0.411	0.238	0.451	0.276	0.152	0.386	0.417
	0.281	0.467	0.302	0.340	0.398	0.093	0.242	0.434	0.297	0.302	0.360	0.433
married	0.246	-0.118	0.014	0.670	0.032	-0.042	-0.013	-0.142	0.122	0.108	0.015	-0.001
	0.037	-0.210	0.034	0.151	-0.166	0.192	0.126	-0.167	0.148	0.225	-0.188	0.206
divorced	0.110	-0.094	0.048	0.325	-0.016	0.014	-0.104	-0.247	0.043	-0.041	-0.122	-0.101
	0.048	-0.162	-0.021	0.018	-0.251	0.126	-0.018	-0.186	0.114	0.090	-0.333	0.022
widow(er)	0.156	-0.248	-0.110	0.544	-0.132	-0.195	-0.111	-0.249	-0.055	-0.031	-0.165	-0.196
	-0.031	-0.369	-0.186	0.012	-0.263	0.014	-0.023	-0.394	0.019	0.065	-0.506	-0.194
Urban: 0-250k	0.124	0.180	0.237	0.164	-0.002	0.081	0.188	0.115	0.012	1.009	0.301	0.331
	0.101	0.148	0.279	0.144	-0.043	0.067	0.124	0.034	0.106	0.171	0.184	0.127
Urban: 250k-1000k	0.236	0.114	-0.088	0.361	0.204	0.104	0.265	0.196	0.006	1.232	0.466	0.175
	0.187	0.098	0.105	0.297	0.029	0.197	0.138	0.047	-0.018	0.203	0.269	0.093
Urban: 1000k+	0.757	0.754	0.847	0.572	0.551	0.325	0.436	0.447	0.112	1.209	0.518	0.406
	0.663	0.729	0.800	0.640	0.318	0.327	0.256	0.264	0.239	0.283	0.395	0.069
$\sigma$	0.409	0.351	0.370	0.573	0.313	0.338	0.398	0.460	0.294	0.319	0.346	0.386
	0.366	0.379	0.323	0.358	0.329	0.293	0.336	0.349	0.318	0.349	0.384	0.380
<i>Межгрупповые факторы W</i>												
const	0	-2.015	-7.253	0	0.473	-4.976	0	-3.682	-14.875	0	2.345	-2.787
	0	-3.367	-13.756	0	-3.601	-45.575	0	-1.707	-9.518	0	-1.037	-8.092
urban	0	0.657	0.810	0	0.944	2.418	0	0.084	0.814	0	-2.856	-3.059
	0	0.060	-0.369	0	0.857	3.569	0	0.204	1.083	0	0.217	0.896
empl. ratio	0	4.710	10.714	0	2.283	8.265	0	9.299	20.433	0	5.434	12.642
	0	9.648	21.013	0	6.571	45.956	0	7.204	15.601	0	6.417	14.622
male	0	0.512	1.030	0	0.014	0.884	0	1.289	2.101	0	0.446	1.366
	0	0.722	1.906	0	0.759	2.626	0	0.350	1.332	0	-0.090	0.651



Таблица 3. Оценки параметров модели 2 (продолжение)

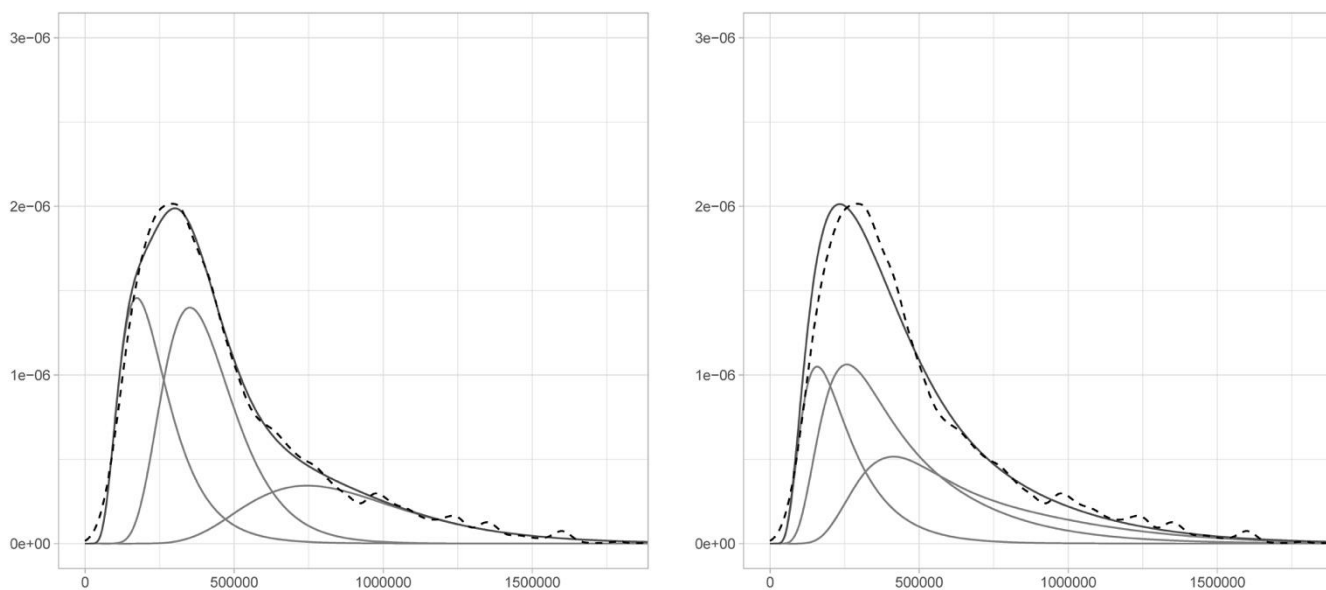
	СФО			ДВФО			ЮФО			СКФО		
	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3	Комп. 1	Комп. 2	Комп. 3
const	9.753	11.777	12.580	10.212	11.984	12.479	9.247	11.441	12.540	9.808	11.833	12.638
	9.647	11.379	12.361	11.149	12.098	12.816	11.027	11.983	12.670	9.589	11.574	12.663
Age	0.027	0.009	0.003	0.026	0.008	0.004	0.040	0.014	0.002	0.033	0.008	-0.002
	0.034	0.013	0.004	0.016	0.011	0.006	0.017	0.011	0.008	0.019	0.010	-0.002
secondary	0.204	0.147	0.147	0.088	0.087	0.120	0.107	0.071	0.147	0.265	0.041	0.072
	0.257	0.120	0.097	0.108	0.279	0.113	0.089	0.199	-0.003	0.076	-0.017	0.005
tertiary	0.561	0.452	0.381	0.382	0.305	0.365	0.316	0.325	0.387	0.598	0.261	0.203
	0.344	0.356	0.282	0.343	0.584	0.449	0.278	0.505	0.224	0.220	0.238	0.214
married	0.414	-0.040	0.052	0.317	-0.081	0.083	0.137	-0.129	-0.010	-0.439	-0.212	0.209
	-0.072	-0.093	0.029	0.035	-0.137	0.050	-0.129	-0.282	0.027	-0.696	-0.216	0.067
divorced	0.250	-0.144	-0.059	-0.074	-0.166	-0.010	-0.043	-0.110	-0.081	-0.153	-0.185	-0.058
	-0.259	-0.169	-0.081	-0.044	-0.165	-0.053	-0.135	-0.300	0.006	-0.542	-0.133	-0.097
widow(er)	0.228	-0.209	-0.037	0.138	-0.337	-0.097	-0.264	-0.237	-0.083	-0.243	-0.292	0.016
	-0.150	-0.280	-0.158	-0.085	-0.162	-0.155	-0.152	-0.324	-0.158	-0.727	-0.217	0.027
Urban: 0–250k	0.219	0.092	0.120	0.185	0.218	0.379	0.351	0.148	-0.040	0.306	0.158	0.097
	0.055	0.159	0.175	0.102	0.008	0.044	0.116	-0.131	-0.206	1.924	0.551	0.169
Urban: 250k–1000k	0.393	0.214	0.057	0.402	0.415	0.508	0.664	0.319	0.344	0.331	0.418	0.260
	0.067	0.212	0.178	0.315	0.140	-0.002	0.307	0.039	-0.020	1.998	0.574	0.224
Urban: 1000k+	0.253	0.282	0.163	—	—	—	0.135	0.300	0.267	—	—	—
	0.051	0.362	0.243	—	—	—	0.327	0.044	-0.100	—	—	—
$\sigma$	0.484	0.347	0.409	0.523	0.323	0.401	0.822	0.392	0.369	0.589	0.407	0.316
	0.544	0.317	0.352	0.388	0.361	0.369	0.371	0.325	0.303	0.407	0.379	0.266
<i>Межгрупповые факторы W</i>												
const	0	-0.696	-6.755	0	-0.336	-3.811	0	1.801	-3.541	0	0.182	-4.263
	0	0.900	-3.555	0	-3.745	-16.051	0	-3.320	-13.597	0	1.966	-3.106
urban	0	1.245	2.476	0	0.707	-0.282	0	0.958	1.609	0	-0.188	-0.598
	0	1.029	1.267	0	0.718	1.473	0	0.585	2.947	0	-5.565	-7.583
empl. ratio	0	2.533	8.647	0	2.342	9.331	0	2.199	9.733	0	6.785	13.855
	0	1.846	8.703	0	8.492	20.537	0	8.004	17.225	0	11.172	20.576
male	0	0.501	1.584	0	0.209	0.558	0	0.046	0.903	0	-0.156	-0.226
	0	0.361	0.756	0	0.857	1.912	0	0.273	1.852	0	0.500	0.085

Примечание: Верхние значения ячеек соответствуют значениям 2014 года. нижние — 2018 года. Все коэффициенты, кроме набранных курсивом значимы на 1%-ом уровне.

### 4.3. Качество подгонки.

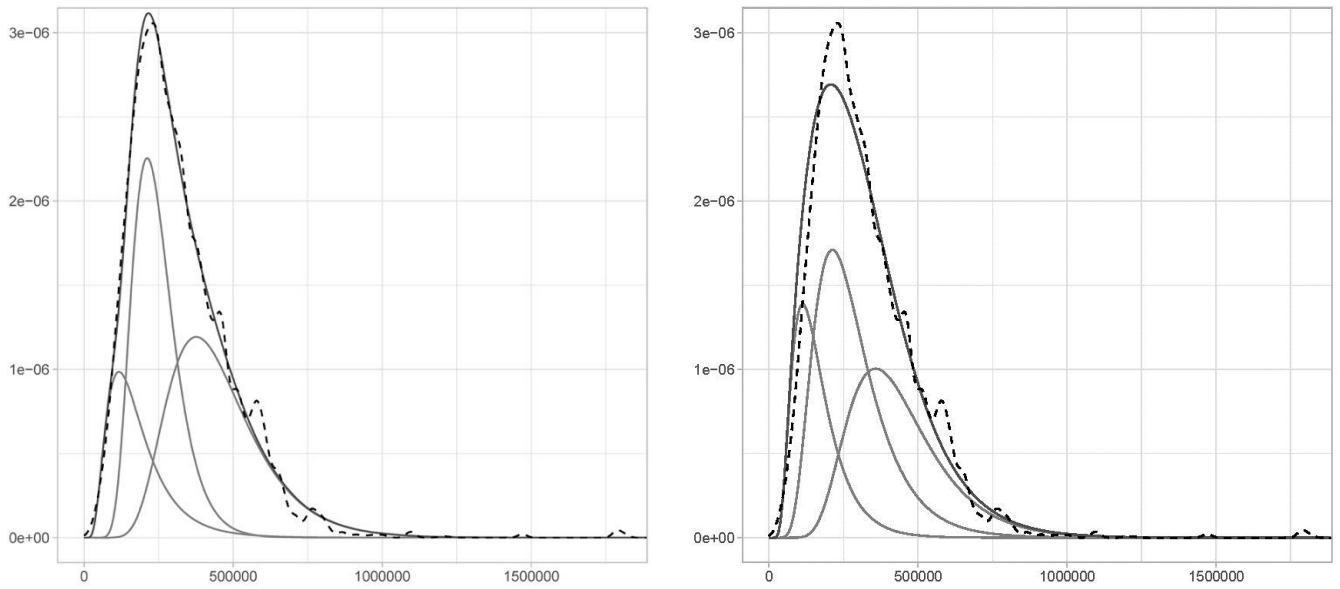
В качестве примера качества аппроксимации истинного распределения смесью трех распределений приведем графики для двух округов с наибольшим количеством наблюдений (ЦФО и ПФО) по данным 2018 г. Графики представлены на рисунках 4 (ЦФО) и 5 (ПФО).

Пунктирная черная линия — ядерная оценка плотности распределения по доходам (состоятельная непараметрическая оценка), черная сплошная линия — плотность смеси распределений (параметрическая оценка плотности), на рисунках так же представлены графики плотности трех компонент смеси (с учетом их вероятностей).



**Рис. 4.** ЦФО. Ядерная оценка плотности распределения доходов и плотность смеси распределений.

График слева – модель 1, график справа — модель 2.



**Рис.5.** ПФО. Ядерная оценка плотности распределения доходов и плотность смеси распределений.

График слева – модель 1, график справа — модель 2.

При интерпретации этих рисунков следует учитывать, что на них нанесены *маржинальные* плотности. Например, на графика справа отмечена маргинальная плотность  $f_2(x)$ :

$$\begin{aligned}
 f_2(x) &= \iint f_2(x; w, z, \lambda, \beta, \sigma^2) dw dz = \iint \sum_{k=1}^3 \pi_k(w, \lambda) \psi(x; \mu(z, \beta_k), \sigma_k^2) dw dz = \\
 &= \sum_{k=1}^3 \iint \pi_k(w, \lambda) \psi(x; \mu(z, \beta_k), \sigma_k^2) dw dz = \sum_{k=1}^3 \left( \int \pi_k(w, \lambda) dw \right) \left( \int \psi(x; \mu(z, \beta_k), \sigma_k^2) dz \right) = \\
 &= \sum_{k=1}^3 \bar{\pi}_k \psi_k(x), \tag{8}
 \end{aligned}$$

На графиках также представлены три компоненты смеси с соответствующими вероятностями:  $\bar{\pi}_k \psi_k(x)$ . Здесь  $\bar{\pi}_k$  — усредненная по  $w$  вероятность принадлежности  $k$ -ой компоненте смеси, а  $\psi_k(x)$  — маргинальная плотность  $k$ -ой компоненты смеси.

Из графиков видно, что распределение по доходам в кластерах, построенных по модели 1, значительно отличается от распределения по доходам в кластерах, построенных по модели 2. Это ожидаемо, поскольку модель 2 берёт в расчет неоднородность распределения внутри каждого кластера.

#### **4.4. Мобильность между кластерами. Сходство и различие двух классификаций**

Рассмотрим, какая доля домохозяйств в каждом ФО попадет в нижний, средний и верхний кластер в соответствии с классификацией по модели 1 (кластеризация) и модели 2 (кластеризация с регрессией). Результаты представлены в таблице 4. Доля домохозяйств, отнесенных в средний кластер, в основном лежит в диапазоне 40–70% — величины, похожие на приведенные в работах социологов, но было бы ошибочно непосредственно отождествлять эндогенно выделенные кластеры с социологическими классами. Следует отметить, что выделяемый в рамках предложенной модели средний кластер в большинстве округов представляет собой наиболее многочисленную группу, что соответствует наивному представлению среднего класса как «моста» между наименее и наиболее материально обеспеченными группами.

В результате кризиса доля среднего кластера, в соответствии с моделью 1, в 2018 снизилась в ЦФО, СЗФО, УФО, ДВФО, ЮФО и повысилась в ПФО, СФО, СКФО. За исключением ПФО, аналогичные изменения доли среднего кластера получены и по модели 2 (по классификации по модели 2 доля среднего кластера несколько уменьшилась).

**Таблица 4.** Доли кластеров (1 — нижний, 2 — средний, 3 — верхний) в федеральных округах.

		модель 1		модель 2	
		2014	2018	2014	2018
<b>ЦФО</b>	<b>1</b>	44%	53%	36%	38%
	<b>2</b>	46%	36%	40%	41%
	<b>3</b>	10%	11%	24%	21%
<b>СЗФО</b>	<b>1</b>	5%	22%	5%	42%
	<b>2</b>	57%	56%	53%	41%
	<b>3</b>	39%	22%	42%	17%
<b>ПФО</b>	<b>1</b>	25%	16%	41%	41%
	<b>2</b>	44%	63%	43%	40%
	<b>3</b>	31%	21%	15%	19%
<b>УФО</b>	<b>1</b>	20%	26%	8%	32%
	<b>2</b>	51%	43%	55%	40%
	<b>3</b>	29%	30%	38%	28%
<b>СФО</b>	<b>1</b>	20%	19%	18%	6%
	<b>2</b>	53%	59%	56%	63%
	<b>3</b>	27%	22%	26%	30%
<b>ДВФО</b>	<b>1</b>	21%	43%	12%	37%
	<b>2</b>	55%	36%	43%	42%
	<b>3</b>	24%	21%	45%	21%
<b>ЮФО</b>	<b>1</b>	10%	24%	2%	51%
	<b>2</b>	69%	59%	67%	37%
	<b>3</b>	21%	17%	31%	13%
<b>СКФО</b>	<b>1</b>	10%	3%	17%	10%
	<b>2</b>	72%	77%	68%	76%
	<b>3</b>	18%	20%	15%	14%

Разбиения на кластеры по моделям 1 и 2 различаются. В модели 1 разбиение происходит по близости параметра среднего  $\mu$ , а в модели 2 — по близости линейных форм  $z'\beta$ . Насколько близки друг другу разбиения на три кластера по моделям 1 и 2? Для данных 2018 года соответствие разбиений приведено в таблице 5. В этой таблице в клетке  $(i, j)$ ,  $i = 1, 2, 3$ ,  $j = 1, 2, 3$  представлена доля домохозяйств, относящихся к кластеру  $i$  по модели 1 и к кластеру  $j$  — по модели 2 в 2018 г. В последней колонке — сумма диагональных элементов, т.е. доля домохозяйств, одинаково классифицированных обеими моделями. Наибольшая доля совпадений наблюдается в ДВФО, наименьшая в ПФО, при этом доля совпадений для всех федеральных округов превышает 55%.

Можно заметить, что наибольшее расхождение между двумя моделями заключается в распределении домохозяйств между кластерами 1 и 2, а именно значитель-

ная доля домохозяйств, по модели 1 относящихся к среднему кластеру, относится к нижнему кластеру по классификации, соответствующей модели 2 в СЗФО, ПФО, ЦФО, ЮФО, СКФО. В определенном смысле прогнозы модели 2 более «консервативны».

**Таблица 5.** Соответствие двух разбиений на кластеры в 2018 г.

	Модель 1	Модель 2			
		1	2	3	
ЦФО	1	37%	15%	1%	69%
	2	2%	24%	11%	
	3	0%	2%	8%	
СЗФО	1	21%	2%	0%	68%
	2	23%	32%	1%	
	3	0%	6%	16%	
ПФО	1	10%	6%	0%	57%
	2	32%	30%	1%	
	3	0%	4%	17%	
УФО	1	13%	9%	1%	63%
	2	22%	25%	0%	
	3	0%	5%	25%	
СФО	1	5%	14%	0%	78%
	2	0%	52%	7%	
	3	0%	1%	21%	
ДВФО	1	37%	6%	0%	86%
	2	1%	32%	2%	
	3	0%	4%	16%	
ЮФО	1	22%	3%	0%	63%
	2	30%	28%	1%	
	3	0%	4%	13%	
СКФО	1	1%	2%	0%	58%
	2	27%	46%	4%	
	3	0%	9%	11%	

В первой модели кластеры формируются в значительной степени по близости доходов. Во второй по близости линейных форм  $z' \beta_k$  (иначе — доходов, но с учетом факторов  $z$ , предикторов среднего дохода для домохозяйства). Насколько различаются ожидаемые средние доходов в кластерах, построенных по моделям 1 и 2? В таблице 6 приведены ожидаемые средние доходы в кластерах, рассчитанные по моделям 1 и 2 в 2014 г. и в 2018 г. (в т.р.) и — для сравнения этих средних — отношения ожидаемых средних доходов, рассчитанных по модели 2 к доходам, рассчитанным по модели 1. (Для модели 1 эти данные приведены в таблице 2). Удивительно,

но в большинстве случаев средние отличаются незначительно. Случаи отличия более чем на 20% выделены полужирным. Это СКФО (нижний кластер), ПФО (средний кластер, 2014 г.) СЗФО (средний кластер, 2018 г.) и ЮФО (нижний кластер, 2019 г.). Эти несовпадения соответствуют результатам таблицы 5.

**Таблица 6.** Средние доходы в разбиении на кластеры по моделям 1 и 2

	Кластер	модель 1		модель 2		m2/m1	
		2014	2018	2014	2018	2014	2018
ЦФО	1	249	241	271	266	1.09	1.10
	2	462	417	439	437	0.95	1.05
	3	1116	899	823	682	0.74	0.76
СЗФО	1	215	207	178	272	0.83	1.31
	2	312	352	324	437	1.04	<b>1.24</b>
	3	602	619	595	630	0.99	1.02
ПФО	1	186	174	219	209	1.18	1.20
	2	253	245	337	267	<b>1.33</b>	1.09
	3	447	453	460	458	1.03	1.01
УФО	1	201	226	222	214	1.11	0.95
	2	307	259	288	300	0.94	1.16
	3	624	529	614	551	0.99	1.04
СФО	1	168	167	163	99	0.97	0.59
	2	267	241	278	221	1.04	0.92
	3	493	434	507	420	1.03	0.97
ДВФО	1	238	275	195	250	0.82	0.91
	2	365	445	331	440	0.91	0.99
	3	713	688	631	705	0.88	1.02
ЮФО	1	182	173	143	208	<b>0.79</b>	1.20
	2	276	267	249	305	0.90	1.14
	3	491	489	461	508	0.94	1.04
СКФО	1	121	109	167	182	<b>1.38</b>	<b>1.68</b>
	2	228	193	235	223	1.03	1.15
	3	401	370	414	369	1.03	1.00

Как велика мобильность между кластерами? Точно ответить на этот вопрос можно было бы при наличии панельных данных. Данные, используемые в этой статье – Выборочные наблюдения доходов населения (ВНДН) — не являются панельными. Однако можно косвенным образом оценить эту мобильность применяя следующий подход. Возьмем домохозяйство из наблюдений 2014 г., с помощью модели 2014 г. определим, к какому кластеру оно принадлежало в 2014 г., далее рассмотрим модель 2018 г. и определим, в какой кластер попало бы подобное (с теми же индивидуальными характеристиками и тем же приведенном доходе с учетом инфляции) домохозяйство в 2018 г.

Альтернативный вариант — проделать те же вычисления с домохозяйством из наблюдений 2018 г., и ретроспективно «вменить» им классификацию в 2014 году с помощью модели, оцененной по данным 2014 года. Как показали расчеты, матрицы переходов из кластера в кластер при этих двух подходах отличаются несущественно. Похожая идея, но с применением более простой линейной модели классификации, была успешно реализована на примере выборочных обследований в Перу в работе (Winkelried, Torres, 2019).

В таблице 7 приведены матрицы переходов, рассчитанные для наблюдений 2014 г., с кластеризацией, полученной в модели 1 (кластеризация) и модели 2 (кластеризация с регрессией). В каждой матрице значения на главной диагонали показывают доли устойчивых в своем классе наблюдений, значения в нижнем треугольнике (под главной диагональю) соответствуют перемещению из более высокого в более низкий класс, тогда как значения в верхнем треугольнике — перемещения в более высокие классы.

Отметим особенности миграции домохозяйств, отнесенных в 2014 г. к 2-му кластеру по модели 1. В ЦФО и СКФО более 90% остаются в 2018 г. во 2-м кластере. В СФЗО, УФО, ДВФО, ЮФО 33–67% перемещаются в кластер 1. В ПФО и СФО напротив, 16 и 23% перемещаются из 2-го кластера в верхний.

Верхний кластер (3-й) наиболее устойчив в ПФО, СФО, СКФО, поскольку там не наблюдается миграции вниз, тогда как существенное перемещение из верхнего кластера в средний наблюдается в СЗФО, ДВФО, ЮФО. В то же время значительное перемещение из нижнего кластера в средний — в СФО и СКФО.

В целом для СЗФО, ДВФО, ЮФО наблюдается «ухудшение», т.е. перемещения в стоящие ниже кластеры. Противоположная ситуация в СФО и ПФО, где наблюдается больше перемещений снизу вверх.



Мобильность, оцененная по модели второго типа, несколько отличается. В ЦФО, ПФО, УФО, ДВФО, ЮФО, СКФО наблюдается «встречная мобильность» — переходы из 1-го и 3-го кластеров в средний, т.е. выравнивание доходов в результате кризиса. Это наблюдение косвенно подтверждает выводы об уменьшении номинального неравенства, полученные в (Нартикоев, Пересецкий, 2019). «Одностороннее» движение наблюдается в СЗФО (средний и верхний кластеры ухудшают свои позиции) и в СФО (нижний и средний кластеры улучшают свои позиции).

**Таблица 7.** Мобильность между кластерами 2014 и 2018 гг.

		модель 1					модель 2			
		2014	2018				2014	2018		
			1	2	3			1	2	3
ЦФО	1	1	92%	8%	0%	ЦФО	1	70%	28%	2%
	2	2	5%	90%	6%		2	6%	65%	30%
	3	3	0%	13%	87%		3	0%	23%	77%
СЗФО	1	1	97%	3%	0%	СЗФО	1	93%	7%	0%
	2	2	58%	42%	1%		2	41%	57%	1%
	3	3	0%	53%	47%		3	0%	28%	72%
ПФО	1	1	86%	14%	0%	ПФО	1	61%	38%	0%
	2	2	1%	83%	16%		2	50%	48%	2%
	3	3	0%	0%	100%		3	0%	19%	81%
УФО	1	1	94%	6%	0%	УФО	1	57%	39%	3%
	2	2	33%	65%	2%		2	46%	53%	1%
	3	3	0%	12%	88%		3	0%	17%	83%
СФО	1	1	31%	68%	0%	СФО	1	27%	72%	2%
	2	2	0%	77%	23%		2	1%	88%	12%
	3	3	0%	0%	100%		3	0%	4%	96%
ДВФО	1	1	98%	2%	0%	ДВФО	1	86%	14%	0%
	2	2	53%	47%	0%		2	4%	90%	6%
	3	3	0%	53%	47%		3	0%	21%	79%
ЮФО	1	1	98%	0%	2%	ЮФО	1	89%	11%	0%
	2	2	67%	33%	0%		2	51%	48%	1%
	3	3	0%	55%	45%		3	0%	25%	75%
СКФО	1	1	59%	41%	0%	СКФО	1	35%	65%	0%
	2	2	5%	95%	1%		2	35%	60%	5%
	3	3	0%	7%	93%		3	0%	45%	55%

*Примечание:* например, согласно модели 1 (левая панель), из домохозяйств, находящихся в кластере 1 в 2014 г., 92% останутся в кластере 1, а 8% перейдут в кластер 2. Согласно модели 2 (правая панель), из домохозяйств, находящихся в кластере 1 в 2014 г., 70% останутся в кластере 1, а 28% и 2% перейдут, соответственно, в кластеры 2 и 3.

## 5. Заключение

В работе модель смеси трех логнормальных распределений применяется для анализа динамики распределения домохозяйств по доходам по каждому из федеральных округов РФ. В качестве данных используются данные выборочного наблюдения

доходов населения (ВНДН), ежегодно проводимого Федеральной службой государственной статистики, за 2014 и 2018 гг.. Несмотря на то, что используемые данные не являются панельным исследованием, предложенный метод позволяет косвенно оценить направление и масштаб социальной мобильности, что является большим преимуществом в условиях отсутствия репрезентативных панельных данных.

Рассмотрены две модели: первая учитывает факторы межгруппового различия (пол главы домохозяйства, доля работающих взрослых, тип местности проживания), вторая дополнительно к этому учитывает внутригрупповую гетерогенность (семейный статус, уровень образования и возраст главы домохозяйства, размер населенного пункта). С помощью каждой из моделей получены эндогенные распределения домохозяйств по трем кластерам (нижний, средний, верхний). При этом классификация домохозяйств зависит не только от уровня их дохода, но и от перечисленных выше индивидуальных характеристик.

Показано, что в результате кризиса в 2018 г. по сравнению с 2014 г. в ЦФО, СФО, СКФО, ЮФО средние значения дохода уменьшаются во всех кластерах. В УФО среднее значение уменьшается в верхнем и среднем кластерах, но несколько увеличиваются в нижнем; тогда как в ДВФО среднее значение уменьшается в верхнем кластере и увеличивается в нижнем и среднем.

Получено, что домохозяйство, проживающее в городской местности, при прочих равных, имеет меньшую вероятность принадлежности к нижнему кластеру и большую — к верхнему. К такому же эффекту приводит увеличение доли работающих членов домохозяйства (кроме ПФО в 2014 г., в котором увеличивается вероятность верхнего кластера и уменьшается вероятность среднего).

Во всех федеральных округах домохозяйство, во главе которого мужчина имеет больше шансов принадлежать к верхнему кластеру, за исключением СКФО, в котором ситуация обратная. Однако этот эффект требует дополнительного, более глубокого анализа.

В большинстве кластеров оценка параметра разброса  $\sigma$  уменьшается, т.е. в результате негативных макроэкономических шоков в кластерах наблюдается концентрация доходов домохозяйств вокруг среднего. Аналогичный эффект по домохозяйствам Испании был отмечен в (López Rodríguez, Varas, 2019).

Возраст главы домохозяйства и его/ее уровень образования положительно связан с доходами (исключение составляют верхний и средний кластеры в ЮФО и верхний кластер в СЗФО в 2018 г.).

В среднем кластере домохозяйство с главой, не состоящим(-ей) в браке, имеет при прочих равных больший подушевой доход, чем другие категории семейного статуса, за редкими исключениями (СФО, УФО в 2014 г.).

Показано, что для разных кластеров классические детерминанты дохода (образование, возраст, население города, семейное положение) различаются по силе влияния. Этот факт может иметь применение в маркетинге и разработке социальной политики.

Использованный в работе подход продемонстрировал большую гибкость и эффективность для задач анализа распределения доходов, а также межгрупповой социальной динамики во времени. Выделение относительно однородных по своим характеристикам кластеров позволяет более тонко регистрировать эффекты детерминант неоднородности доходов внутри каждого кластера.

## Список литературы

Айвазян С. А. (2012). Анализ качества и образа жизни населения. Москва: Наука.

Аникин В.А. (2020). Социальные классы новой России — неравные и разные. Социологические исследования, 2020 (2), 31–42.

Дидэ Э. (1985). Методы анализа данных. Подход, основанный на методе динамических сгущений. М.: Финансы и статистика.

Доклад об экономике России (2014) Всемирный банк. Представительство в России. Март, № 31.

Малева Т. М. (2008). Формирование массового среднего класса: желаемое и действительное: Доклад на III Всероссийском социологическом конгрессе. М.: ГУ ВШЭ, 22 октября 2008 г.

Нартикоев, А. Р., Пересецкий А. А. (2019). Моделирование динамики распределения доходов в России. Прикладная эконометрика, 54, 105–126.

Средний класс в современной России: 10 лет спустя. (2014). Аналитический доклад. М.: Институт социологии РАН.

Тихонова Н. Е. (2016). Российский средний класс в условиях кризиса. *Общественные науки и современность*, 2016, (4), 48–54.

Тихонова Н. Е. (2020) Особенности идентичностей и мировоззрения основных страт современного российского общества. *Мир России*, 29 (1), 6–30. DOI: 10.17323/1811-038X-2020-29-1-6-30.

Тихонова Н. Е., Мареева С. В. (2009). *Средний класс: теория и реальность*. М.: Альфа-М, 2009.

Chotikapanich D., Griffiths W. (2008). Estimating income distributions using a mixture of gamma densities. In: Chotikapanich, D. (Ed.), *Modelling Income Distributions and Lorenz Curves*. Springer, New York, Chapter 16, pp. 285–302.

Cowell F. A., Flachaire E. (2015). Statistical Methods for Distributional Analysis. In *Handbook of Income Distribution*, Volume 2A, Ch. 6., pp. 359-465. <http://dx.doi.org/10.1016/B978-0-444-59428-0.00007-2>.

Cowell F. A., van Kerm P. (2015). Wealth inequality: A survey. *Journal of Economic Surveys*, 29 (4), 671–710.

Dempster A. P., Laird N. M., Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B.* 39 (1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.

Flachaire E., Nunez O. (2007). Estimation of income distribution and detection of subpopulations: an explanatory model. *Computational Statistics and Data Analysis*, 51 (7), 3368–3380.

Fourrier-Nicolai E., Lubrano M. (2020). Bayesian inference for TIP curves: an application to child poverty in Germany. *The Journal of Economic Inequality*. 18 (1), 91–111. <https://doi.org/10.1007/s10888-019-09426-6>.

García-Fernández R. M., Gottlieb G., Palacios-González F. (2013). Polarization, growth and social policy in the case of Israel, 1997–2008. *Economics: The Open-Access, Open-Assessment E-Journal*, 7 (2013–15). <http://dx.doi.org/10.5018/economics-ejournal.ja.2013-15>.

Gontmakher T., Ross C. (2015). The middle class and democratisation in Russia. *Europe-Asia studies*, 67 (2), 269–284.

Hagenaars A. J. M., de Vos K., Zaidi M. A. (1994), *Poverty statistics in the late 1980s: Research based on micro-data*. Office for official publications of the European Communities. Luxembourg.

López Rodríguez M. I., Barac M. (2019). Inequality of Spanish household expenditure for the 2006–2016 period — Are we converging? *Economy of Region*, 15 (3), 780–790.

Lubrano M., Ndoeye A. A. J. (2016) Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. *Computational Statistics and Data Analysis*, 100, 830–846.

Pittau M. G., Zelli R. (2006). Empirical evidence of income dynamics across EU regions. *Journal of Applied Econometrics*, 21 (5), 605–628.

Pittau M. G., Zelli R., Johnson P. A. (2010). Mixture models, convergence clubs and polarization. *Review of Income Wealth*, 56 (1), 102–122.

Pittau M. G., Zelli R., Massari R. (2016). Evidence of convergence clubs using mixture models. *Econometric Reviews*, 35 (7), 1317–1342.

Winkelried D., Torres J. (2019) Economic mobility along the business cycle. The case of Peru, *Applied Economics*, 51 (18), 1894-1906.

Nartikoev A., Peresetsky A. Endogenous household classification: Russian regions.

**Alan Nartikoev**

National Research University Higher School of Economics, Moscow, Russian Federation;  
alan.nartikoev@gmail.com.

**Anatoly Peresetsky**

National Research University Higher School of Economics, Moscow, Russian Federation;  
aperesetsky@hse.ru

**Abstract**

In order to study the structure of society, sociologists usually distinguish several homogeneous social groups, or classes. The most common division consists of three groups: upper, middle and lower classes. Such a partition is traditionally based on a subjective (exogenous) criteria adopted by a particular researcher. In this paper, the distribution of households in Russian federal districts is modeled as a mixture of three lognormal distributions. The mixing proportions (probabilities) of the mixture components and the corresponding distribution parameters are modeled as functions of the individual characteristics of households. The result is an endogenous decomposition of household sample into three clusters (lower, middle, upper). This classification allows to analyze the difference between regions and the patterns of intergroup dynamics in the period 2014–2018. The approach used in this work demonstrated great flexibility in analyzing the distribution of income, the dynamics of this distribution over time, as well as a migration between relatively homogeneous clusters. The use of mixture density function with endogenously determined probabilities allows for precise detection of the effects of the income heterogeneity determinants within each cluster.

**Keywords:** mixture models; Russia; income distribution; middle class.

**JEL classification:** C14; C15; C46; D31; I32; R20.

**References**

- Aivazian S. A. (2012). *Analiz kachestva i obraza zhizni naselenija*. Moscow: Nauka. (in Russian).
- Anikin V. A. (2020). Social classes of the new Russia: Unequal and different. *Sociological Studies*, 2020. 2.31–42. DOI: 10.31857/S013216250008492-4. (in Russian).
- Didé E. (1985). *Metody analiza dannyh. Podhod, osnovannyj na metode dinamicheskikh sgushhenij*. M.: Finansy i statistika. (in Russian).
- Russia economic report (2014). The World Bank in the Russian Federation. №31, March 2014.

Maleva T. M. (2008). Formirovanie massovogo srednego klassa: zhelaemoe i dejstvitel'noe: Doklad na III Vserossijskom sociologicheskom kongresse. M.: HSE, October, 22, 2008. (in Russian).

Nartikoev A. R., Peresetsky A. A. (2019). Modeling the dynamics of income distribution in Russia. *Applied Econometrics*, 54, 105–126.

Srednij klass v sovremennoj Rossii: 10 let spustja. (2014). Analiticheskij doklad. Institute of Sociology of the Russian Academy of Sciences (ISRAS). (in Russian).

Tikhonova N. E. (2016). The impact of crisis on the life of Russian middle class. *Social Sciences and Contemporary World*, 2016, (4), 48–54. (in Russian).

Tikhonova N. E., Mareeva S. V. (2009). *Srednij klass: teorija i real'nost'*. M.: Alpha-M.

Tikhonova N. E. (2020). The worldviews and identities of the mass strata of modern Russian society. *Mir Rossii*, 29 (1), 6–30 (in Russian). DOI: 10.17323/1811-038X-2020-29-1-6-30

Chotikapanich D., Griffiths W. (2008). Estimating income distributions using a mixture of gamma densities. In: Chotikapanich, D. (Ed.), *Modelling Income Distributions and Lorenz Curves*. Springer, New York, Chapter 16, pp. 285–302.

Cowell F. A., Flachaire E. (2015). Statistical Methods for Distributional Analysis. In *Handbook of Income Distribution*, Volume 2A, Ch. 6., pp. 359-465. <http://dx.doi.org/10.1016/B978-0-444-59428-0.00007-2>.

Cowell F. A., van Kerm P. (2015). Wealth inequality: A survey. *Journal of Economic Surveys*, 29 (4), 671–710.

Dempster A. P., Laird N. M., Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B.* 39 (1), 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.

Flachaire E., Nunez O. (2007). Estimation of income distribution and detection of subpopulations: an explanatory model. *Computational Statistics and Data Analysis*, 51 (7), 3368–3380.

Fourrier-Nicolai E., Lubrano M. (2020). Bayesian inference for TIP curves: an application to child poverty in Germany. *The Journal of Economic Inequality*. 18 (1), 91–111. <https://doi.org/10.1007/s10888-019-09426-6>.

García-Fernández R. M., Gottlieb G., Palacios-González F. (2013). Polarization, growth and social policy in the case of Israel, 1997–2008. *Economics: The Open-Access, Open-Assessment E-Journal*, 7 (2013–15). <http://dx.doi.org/10.5018/economics-ejournal.ja.2013-15>.

Hagenaars A. J. M., de Vos K., Zaidi M. A. (1994), *Poverty statistics in the late 1980s: Research based on micro-data*. Office for official publications of the European Communities. Luxembourg.

López Rodríguez M. I., Barac M. (2019). Inequality of Spanish household expenditure for the 2006–2016 period — Are we converging? *Economy of Region*, 15 (3), 780–790.

Lubrano M., Ndoye A. A. J. (2016) Income inequality decomposition using a finite mixture of log-normal distributions: A Bayesian approach. *Computational Statistics and Data Analysis*, 100, 830–846.

Pittau M. G., Zelli R. (2006). Empirical evidence of income dynamics across EU regions. *Journal of Applied Econometrics*, 21 (5), 605–628.

Pittau M. G., Zelli R., Johnson P. A. (2010). Mixture models, convergence clubs and polarization. *Review of Income Wealth*, 56 (1), 102–122.

Pittau M. G., Zelli R., Massari R. (2016). Evidence of convergence clubs using mixture models. *Econometric Reviews*, 35 (7), 1317–1342.

Winkelried D., Torres J. (2019) Economic mobility along the business cycle. The case of Peru, *Applied Economics*, 51 (18), 1894-1906.