

Методы оценивания распределения доходов населения с использованием выборочных обследований и налоговой статистики

Кузин С.С., к.т.н., главный эксперт Центра экономических измерений и статистики НИУ ВШЭ

План

- Проблема оценки распределения доходов, правый хвост распределения
- Методы решения, перечисление, краткая характеристика
- Данные обследования доходов (без внешних источников)
 - Подгонка законов распределения
 - Логнормальный закон
 - Логнормальный закон и закон Парето для правого хвоста
- Данные обследования доходов и макроэкономический показатель доходов домашних хозяйств
 - Логнормальный закон (подгонка на основе комбинации оценок по выборке и СНС)
 - Среднедушевой доход из СНС и дисперсия по выборке
 - Среднедушевой доход из СНС и мода (через среднее и дисперсию) по выборке
- Комбинирование данных выборочного обследования доходов населения и налоговой статистики
 - Взвешивание
 - Замена значений (замена в группах или Парето для правого хвоста на основе налоговых данных)
 - Интерполяция агрегированных налоговых данных с помощью обобщенных кривых Парето
 - Калибровка по макроэкономическому показателю доходов домашних хозяйств
 - Распределительные национальные счета
- Пример методики оценивания распределения доходов населения России по данным обследования и налоговым данным

Проблема оценки хвоста распределения в области высоких доходов

- Основным источником информации о дифференциации населения по доходам являются выборочные обследования домашних хозяйств. Известной проблемой оценивания распределения доходов населения по данным исключительно обследований населения является недостаточная представительность оценок доходов в верхней части распределения. Главной причиной является невозможность опроса интервьюерами домохозяйств с высокими и очень высокими доходами, а также занижение высоких доходов.
- Другим источником данных является административная статистика, главным образом налоговые данные. Проблема оценки «правого хвоста» распределения здесь решается более успешно, поскольку с высоких доходов уплачиваются налоги, но налоговые данные не охватывают или плохо охватывают людей с нулевыми, низкими и средними доходами. Низкие и средние доходы часто оказываются доходами, не облагаемыми налогами.
- Возможным вариантом решения проблемы недооценки доходов в области правого хвоста распределения является комбинирование данных выборочных обследований и административных данных, главным образом налоговой статистики.

Характеристики источников данных о доходах населения

- Данные выборочного обследования
 - Репрезентативность относительно всего населения
 - Регистрация широкого спектра источников доходов, в том числе не облагаемых налогами
 - Достаточно надежные оценки в области низких и средних доходов
 - Относительно небольшие объемы выборки
 - Наблюдение лиц в составе домашних хозяйств
 - Недостаточная представленность высокодоходных слоев населения и информации об их доходах
- Налоговые данные
 - Генеральная совокупность налогоплательщиков
 - Полная информация обо всех налогооблагаемых доходах, независимо от их размера
 - Не включают доходы, не облагаемые налогом, вследствие чего плохо отражают реальные доходы населения в нижней части распределения
 - Статистической единицей является персона, налогоплательщик, а не семья или домашнее хозяйство
- Данные Системы национальных счетов (СНС)
 - Полные данные, гармонизированные на национальном уровне
 - Агрегированные показатели, отсутствует информация о дифференциации доходов

Методы оценивания распределения доходов населения с использованием данных выборочного обследования

- Использование логнормального распределения
- Использование эмпирических данных в нижней и средней части распределения до пороговой величины доходов и замена значений дохода выше этой точки на расчетные значения в соответствии законом распределения, например, законом Парето. Считается, что такой подход может корректировать эффект занижения доходов, но не может корректировать отсутствие в выборке лиц с очень высокими доходами.
- Комбинирование распределений, например, логнормального распределения в области низких и средних доходов и распределения Парето в области высоких доходов. Анализ эмпирических данных по распределению дохода показывает, что средняя часть распределения хорошо аппроксимируется логнормальным законом, тогда как правая часть распределения с высокими доходами лучше аппроксимируется законом Парето [*Clementi F., Gallegati M. (2005). Pareto's law of income distribution: Evidence for Germany, the United Kingdom and the United States. Papers physics/0504217, arXiv.org*].

Логнормальное распределение

- Плотность распределения двухпараметрического логарифмически нормального закона:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-m)^2}{2\sigma^2}}$$

где x – уровень душевого дохода; m – среднее логарифма доходов; σ^2 – дисперсия логарифма дохода.

- Некоторые полезные соотношения между параметрами логнормального распределения. Достаточно иметь любую пару значений из четырех X_{mod} , \bar{X} , μ , σ , чтобы определить другую пару значений [Колмаков И.Б. (2015). *Методология расчета и анализа интегральных оценок показателей поляризации денежных доходов населения. Вопросы статистики. № 2. С. 23-36.*], например:

- $X_{mod} = e^{3\mu}/\bar{X}^2$
- $X_{mod} = \bar{X}\exp(-1,5\sigma^2)$
- $\sigma^2 = \frac{2}{3}(\ln \bar{X} - \ln X_{mod})$
- $\sigma^2 = 2 \ln \bar{X} - 2\mu$
- $\mu = \ln \bar{X} - 0,5\sigma^2$

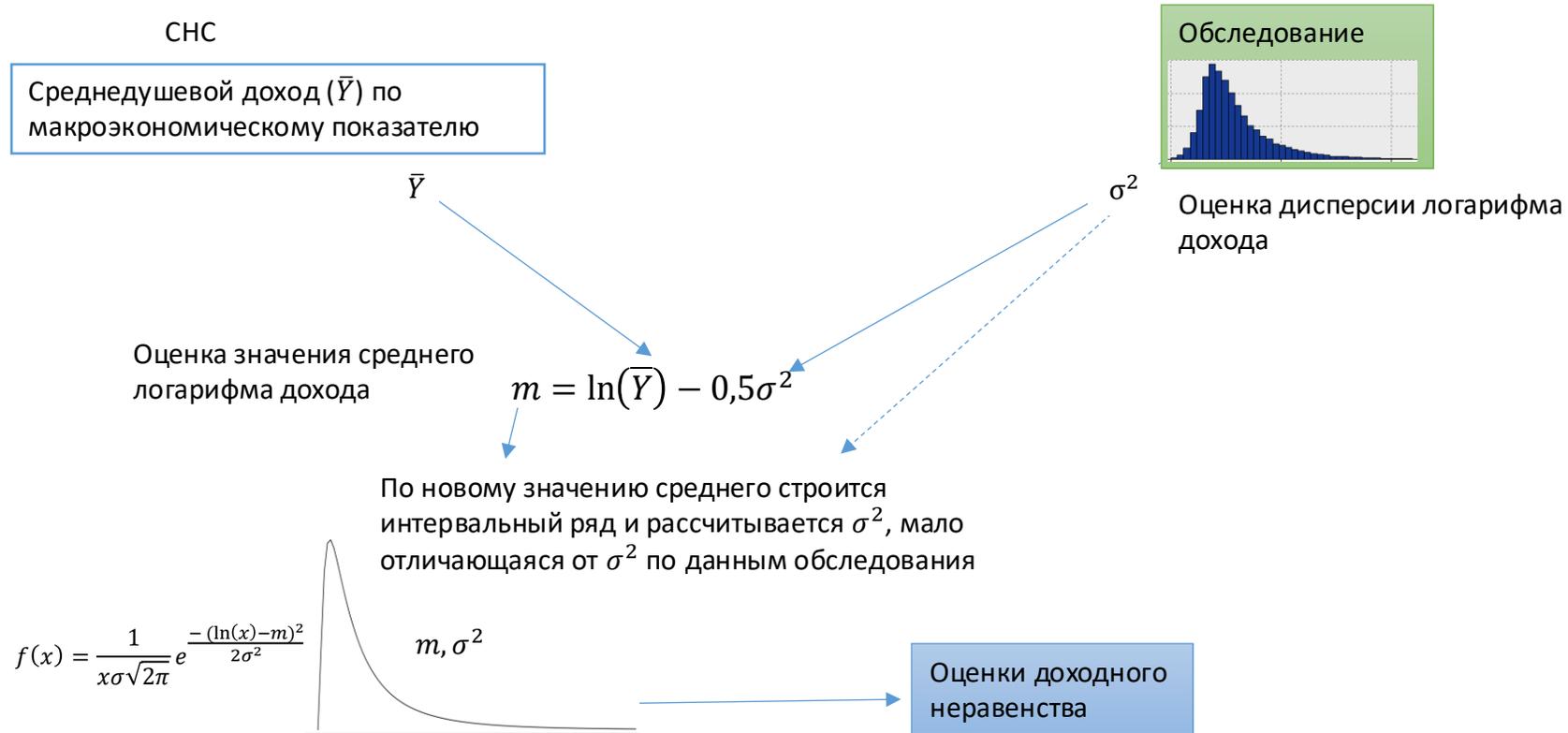
где \bar{X} - среднее душевого дохода; X_{mod} - мода душевого дохода.

- Показатели доходного неравенства для логнормального распределения
 - Индекс Джини $G = 2\Phi(\sigma/\sqrt{2}) - 1$
 - Коэффициент фондов $F \approx \exp(3,56\sigma)$

Варианты оценивания параметров логнормального закона для распределения доходов

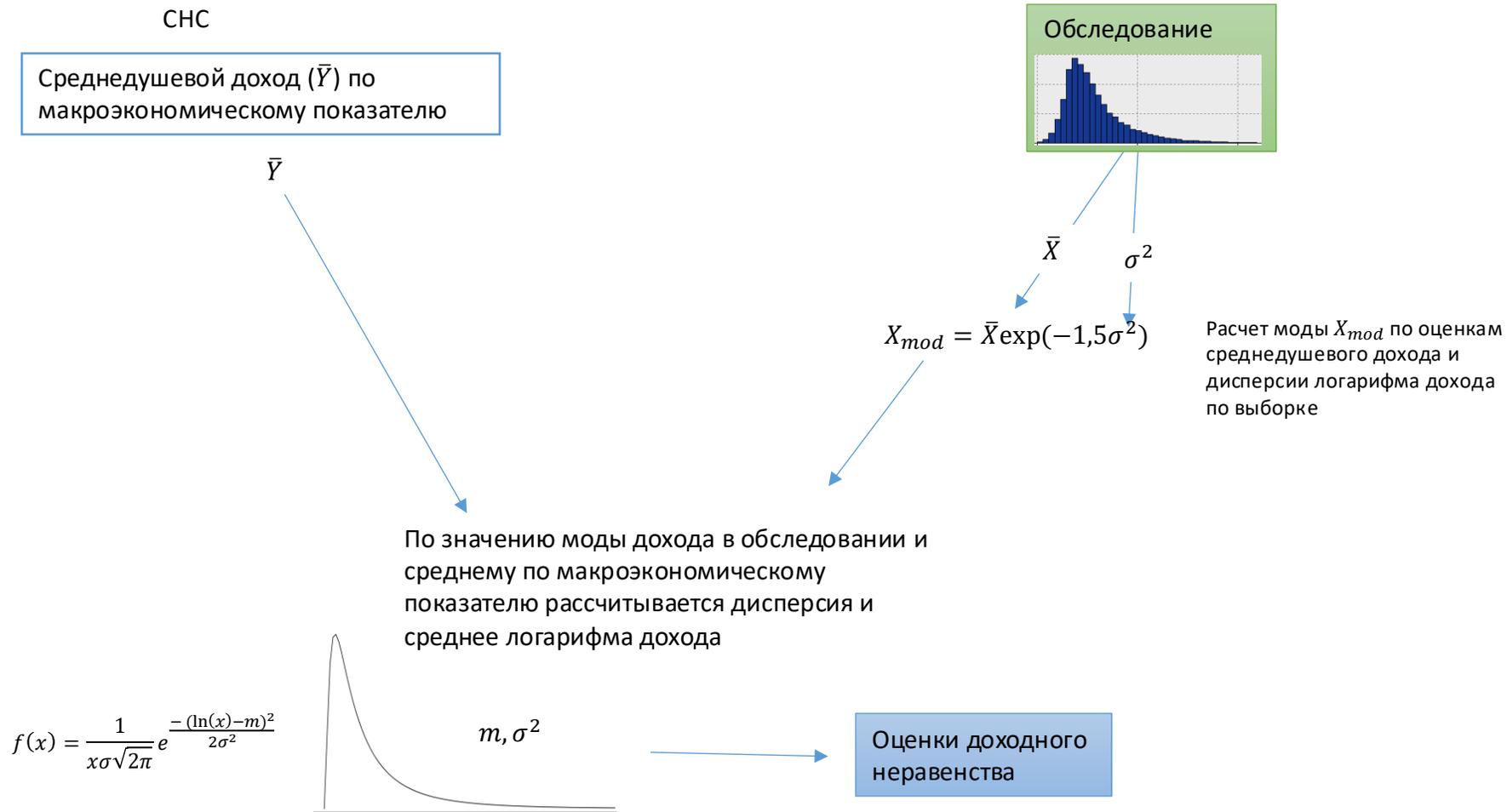
- Данные обследования доходов
 - Непосредственная оценка параметров логнормального распределения по эмпирическим данным обследования доходов.
- Данные обследования доходов и макроэкономический показатель доходов домашних хозяйств [Суворов А.В. (2008). *Проблемы оценки дифференциации доходов населения в современной России // Проблемы прогнозирования. № 2. С. 97-114.; Методологические положения по статистике. Вып. 3. М.: Госкомстат России, 2000.*]:
 - Среднедушевой доход на основе макроэкономического показателя и оценка дисперсии логарифма доходов по выборке
 - Среднедушевой доход на основе макроэкономического показателя и оценка моды (через среднее дохода и дисперсию логарифма дохода) по выборке

Среднедушевой доход из СНС и оценка дисперсии по выборке



Такой вариант оценки параметров логнормального распределения использовался в России ранее, когда различие между оценкой среднедушевого дохода по данным обследования и на основе макроэкономического показателя доходов домашних хозяйств было незначительным, не более 10%. При больших различиях с среднедушевом доходе по данным обследования и макроэкономической статистики было предложено использовать подход на расчете моды доходов по данным обследования и среднедушевого показателя доходов по макроэкономическому показателю.

Среднедушевой доход из СНС и оценка моды (через среднее и дисперсию) по выборке



Закон Парето для правого хвоста распределения доходов

- Закон Парето является одной из наиболее распространенных моделей, наряду с логнормальным законом, применяемых для описания распределения доходов [Atkinson, A. B., (2017). *Pareto and the upper tail of the income distribution in the UK: 1799 to the present*. Oxford: Wiley-Blackwell, ISSN 0013-0427, ZDB-ID 1800-4. - Vol. 84. pp. 129-156.].

- Плотность распределения и закон распределения Парето (тип I):

$$f(x) = \frac{\alpha u^\alpha}{x^{\alpha+1}} \text{ и } F(x) = 1 - \left(\frac{x}{u}\right)^{-\alpha}, \text{ для } x \geq u$$

- Закон Парето можно характеризовать константой $b = \alpha/(\alpha - 1)$ – обратным коэффициентом Парето. Например, если эта константа равна двум, то среднее значение доходов, превышающих 60 тыс. руб. в соответствующем распределении Парето, будет равно 120 тыс. руб., а для доходов более 75 тыс. руб. – 150 тыс. руб.

- Применяется также обобщенный закон Парето (тип II), предоставляющий большую гибкость в выборе формы распределения:

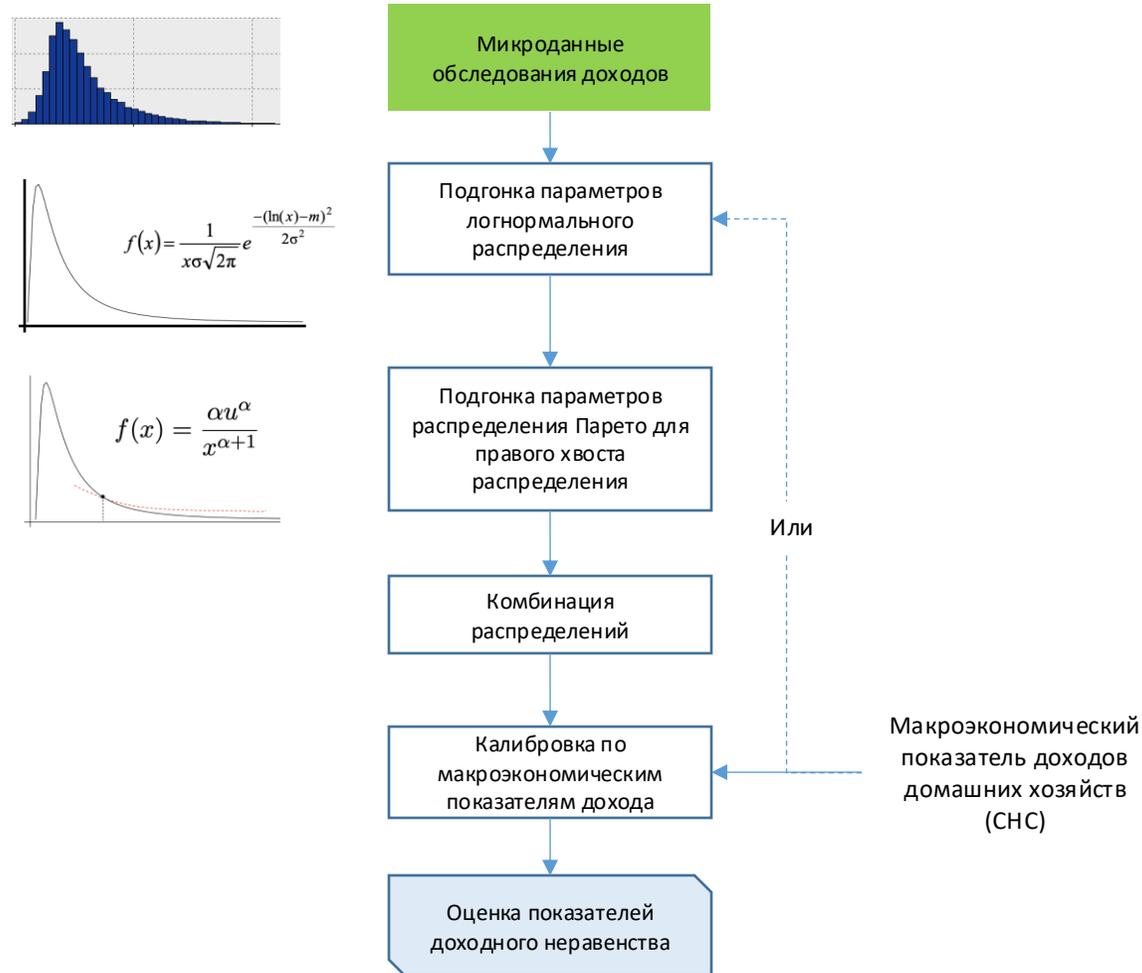
$$F(x) = 1 - \left[1 + \left(\frac{x-u}{\sigma}\right)\right]^{-\alpha}, \text{ для } x \geq u$$

- Закон Парето является частным случаем обобщенного закона Парето при $\sigma = u$.

- Применение закона Парето для описания правого хвоста распределения доходов предполагает выбор порога по доходу. Однозначных рекомендаций по выбору порога нет, а на практике порог часто выбирают в области топ-групп, представляющих менее 5%.

- Применение закона Парето для правого хвоста распределения с использованием только данных выборочного обследования имеет ограниченный эффект, поскольку параметр распределения оценивается по данным обследования в районе самого правого хвоста распределения, который и подлежит корректировке. Более перспективным считается привлечение для этого дополнительных источников данных, главным образом налоговых данных.

Корректировка распределения доходов по данным обследования без использования других данных о распределении

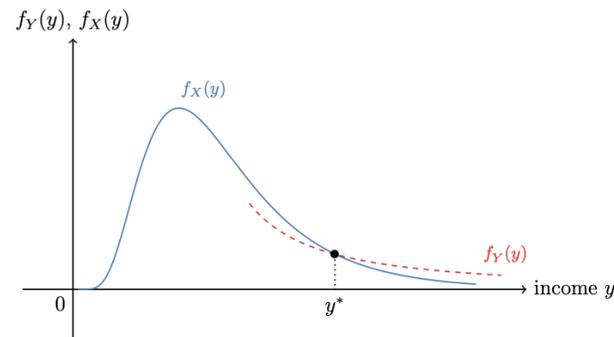


Комбинирование данных выборочного обследования доходов населения и налоговой статистики

- Использование внешних данных в дополнение к данным обследования доходов населения расширяет возможности улучшения оценок распределения в области высоких доходов. В качестве внешних данных могут использоваться макроэкономические данные СНС, а также административные источники данных, главным образом данные налоговой статистики. Комбинирование данных обследования и налоговых данных с последующей калибровкой к макроэкономическим показателям доходов является активно разрабатываемым перспективным направлением:
- *Blanchet, T., I. Flores, M. Morgan (2019). The Weight of the Rich: Improving Surveys Using Tax Data. In: WID.world Working Paper Series 2018/2.*
- *Lustig, N. (2020). The “missing Rich” in Household Surveys: Causes and Correction Approaches, SocArXiv. March 9. doi:10.31235/osf.io/j23pn.;*
- *Jenkins, S.P. (2016). Taking better account of top incomes when measuring inequality levels and trends, 2nd Meeting of Providers of OECD Income Distribution Data, Paris, 19 February 2016.*
- *Webber, D., Tonkin, R.P., Shine, M. (2020) Using tax data to better capture top incomes in official UK income inequality statistics. Working paper 27582, National Bureau of Economic Research, Cambridge, <http://www.nber.org/papers/w27582>.*

Перевзвешивание данных обследования с использованием налоговых данных

- Перевзвешивание заключается в изменении весов в обследовании таким образом, чтобы распределение доходов в обследовании соответствовало распределению в налоговых данных.
- Перевзвешивание может осуществляться, например, путем выделения квантилей в налоговых данных и определение их границ по доходу с последующим формированием соответствующих этим границам групп в выборочном обследовании и приведение их относительных размеров за счет весов в соответствии с размерами групп в налоговых данных. Перевзвешивание по налоговым данным выполняется выше определенной точки в правой части распределения доходов. Эта точка может находиться в окрестностях точки пересечения распределения доходов в обследовании и налоговых данных.
- Само по себе перевзвешивание не может скорректировать эффект отсутствия в выборке лиц с доходами, выходящими за рамки представленных в выборке. В связи с этим в некоторых случаях применяется также сочетание перевзвешивания и замены значений.



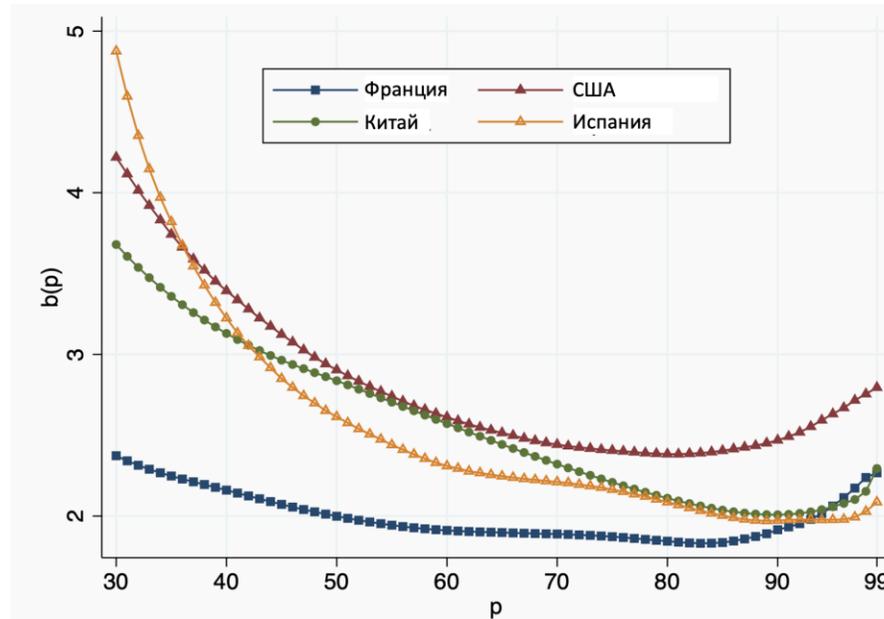
Замена значений правого хвоста распределения доходов с использованием налоговых данных

- Замена значений правого хвоста распределения заключается в замене эмпирических значений дохода в области топ-групп населения с самыми высокими доходами на оценки на основе налоговых данных.
- Параметрический подход заключается в замене значений дохода правого хвоста распределения заменяются в соответствии с распределением Парето или другим параметрическим распределением, параметры которого оцениваются на основе налоговых данных.
- Непараметрический подход предполагает замену значений правого хвоста распределения на средние значения дохода по налоговым данным в соответствующих квантилях.
- В некоторых случаях замену значений рассматривают как дополнение к перевзвешиванию в виде включения дополнительных значений в топ-группу по доходам, поскольку перевзвешивание не может решить проблему отсутствия в выборке лиц с очень высокими доходами.

Обобщенные кривые Парето: непараметрический подход к оцениванию распределения доходов населения

- Помимо параметрических подходов к оцениванию распределения доходов населения существует непараметрический подход, основанный на описании распределения доходов обобщенной кривой Парето.
- Закон Парето предполагает, что обратный коэффициент Парето $b = \alpha / (\alpha - 1)$ является неизменным во всем диапазоне доходов, то есть представляет собой константу для всего распределения. В то же время, если мы рассмотрим поведение эмпирического обратного коэффициента Парето, то он может существенно различаться в области низких, средних и высоких доходов. Использование этого коэффициента удобно, потому что его можно интерпретировать как степень неравенства.
- Использование обобщенных кривых Парето представляет собой более гибкий инструмент оценивания распределения доходов. Обратный коэффициент Парето может интерпретироваться как мера доходного неравенства. Использование закона Парето с постоянным коэффициентом означает, что степень доходного неравенства в области правого хвоста распределения одинакова, независимо от того, какую бы ТОП-группу мы не рассматривали, 5%, 1%, 0,1%. В отличие от этого обобщенная кривая Парето $b(p)$ опирается на эмпирические значения обратных коэффициентов Парето b .
- Подход, основанный на обобщенных кривых Парето предполагает [Blanchet, T., Fournier, J., Piketty, T. (2017). *Generalized Pareto Curves: Theory and Applications*. WID.world Working Paper 2017/03.] использование фактических имеющихся значений обратного коэффициента Парето для доступных точек налоговых данных и подбору наиболее регулярной кривой $b(p)$, которая проходит через имеющиеся точки, то есть интерполяцию агрегированных налоговых данных. Интерполяция обобщенной кривой Парето обеспечивает возможность получения значений среднего дохода для любых квантилей. Это в свою очередь позволяет получить эквивалентные квантили в обследовании и налоговых данных.
- Установлено, что даже по нескольким точкам данных с помощью интерполяции с обобщенными кривыми Парето полное распределение восстанавливается очень хорошо. При этом важным является наличие данные в области правого хвоста распределения.

Примеры обобщенных кривых Парето



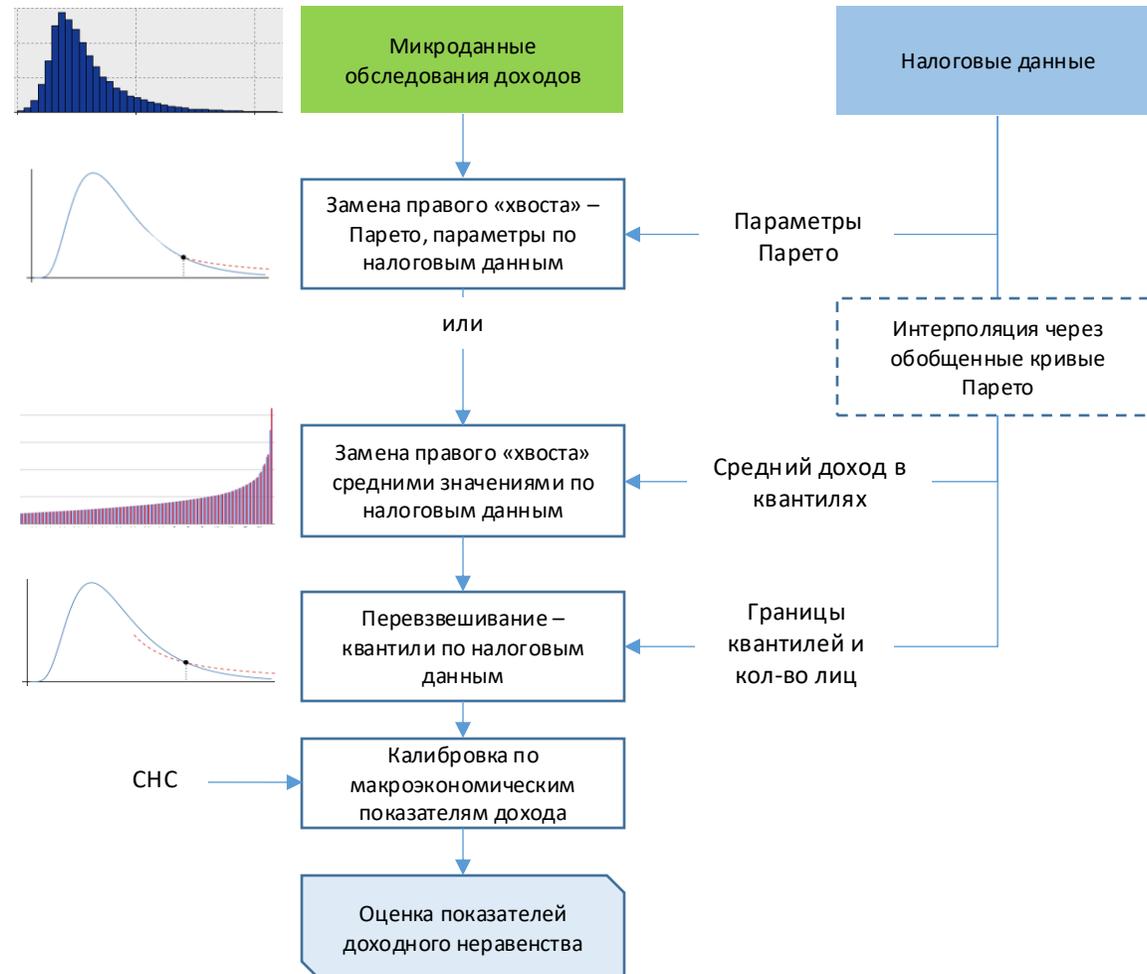
Обобщенные кривые Парето по усредненным за 2000-2014 гг. данным по доходам до вычета налогов и без учета трансфертов, за исключением пенсий и пособий по безработице

Источник: Blanchet T., Garbinti B., Goupille-Lebret J., Martínez-Toledano C. (2018). Applying Generalized Pareto Curves to Inequality Analysis. ASSA Conference.

Пример оценивания распределения доходов с использованием данных обследования и налоговых данных: Распределительные национальные счета (Distributional National Accounts, DINA)

- Идея распределительных национальных счетов заключается в дополнении СНС данными о дифференциации населения по доходам и доходном неравенстве для получения более полной картины на национальном уровне.
- Основные компоненты методики DINA в части оценки распределения доходов населения включают (Blanchet, T., I. Flores, and M. Morgan (2019). “The Weight of the Rich: Improving Surveys Using Tax Data”. In: WID.world Working Paper Series 2018/2):
 - Подготовка налоговых микроданных (если доступны микроданные).
 - Подготовка агрегированных налоговых данных – интерполяция распределения доходов с использованием обобщенных кривых Парето (не обобщенного закона Парето).
 - Подготовка микроданных обследования.
 - Перевзвешивание данных обследования в соответствии с интерполированными налоговыми данными.
 - Калибровка полученного распределения в микроданных обследования с целью их выравнивания с соответствующими агрегатами доходов в СНС.

Варианты использования данных обследования и налоговых данных для оценки распределения доходов



Пример методики оценивания распределения доходов населения
России по данным обследования и налоговым данным

Корректировка правого хвоста распределения доходов в микроданных обследования доходов с использованием агрегированных налоговых данных

- Подход заключается [Кузин С.С., Суринов А.Е. (2025). Оценка доходного неравенства населения России (на основе объединения данных выборочных обследований доходов населения и налоговой статистики) // Вопросы экономики. № 3. С. 97-114. <https://doi.org/10.32609/0042-8736-2025-3-97-114>.] в замене значений индивидуального денежного дохода в верхней части распределения в микроданных ВНДН на средние значения дохода из налоговых деклараций.
- Граница по величине дохода для корректировки верхней части распределения определяется точкой пересечения распределений дохода по данным обследования и налоговым данным.
- Перевзвешивание не применяется. Перевзвешивание имеет смысл, если максимальные значения дохода генеральной совокупности и выборке близки, иначе оно не может скорректировать эффект отсутствия в выборке лиц с доходами, существенно выходящими за рамки доходов в выборке.
- Результаты корректировки индивидуальных доходов используются для получения скорректированных значений доходов домашних хозяйств и расчета среднедушевых доходов.
- Корректировка доходов на уровне микроданных обеспечивает гибкость в получении оценок распределения доходов не только в целом по России, но и для различных групп населения по социально-демографическим, географическим и другим признакам.

Категории дохода в обследовании и налоговых данных

Необходимо обеспечить, насколько это возможно, использование идентичных определений дохода в данных обследования и налоговой статистики. В налоговых данных и данных обследования при корректировке распределения использовались доходы от наемной занятости.

ВНДН

- оплата труда в денежной и натуральной форме в денежном выражении
- сумма заработка за дополнительную работу
- сумма заработка за другую работу

ФНС

- вознаграждение за выполнение трудовых или иных обязанностей в денежной и натуральной формах (в денежном выражении)
- денежное содержание военнослужащих
- выплаты по договорам гражданско-правового характера
- оплата за налогоплательщика работодателями товаров и услуг
- материальная помощь

Формально состав компонентов дохода в источниках отличаются, но фактически сумма перечисленных компонентов дохода в ВНДН и ФНС представляет собой доход от наемной занятости.

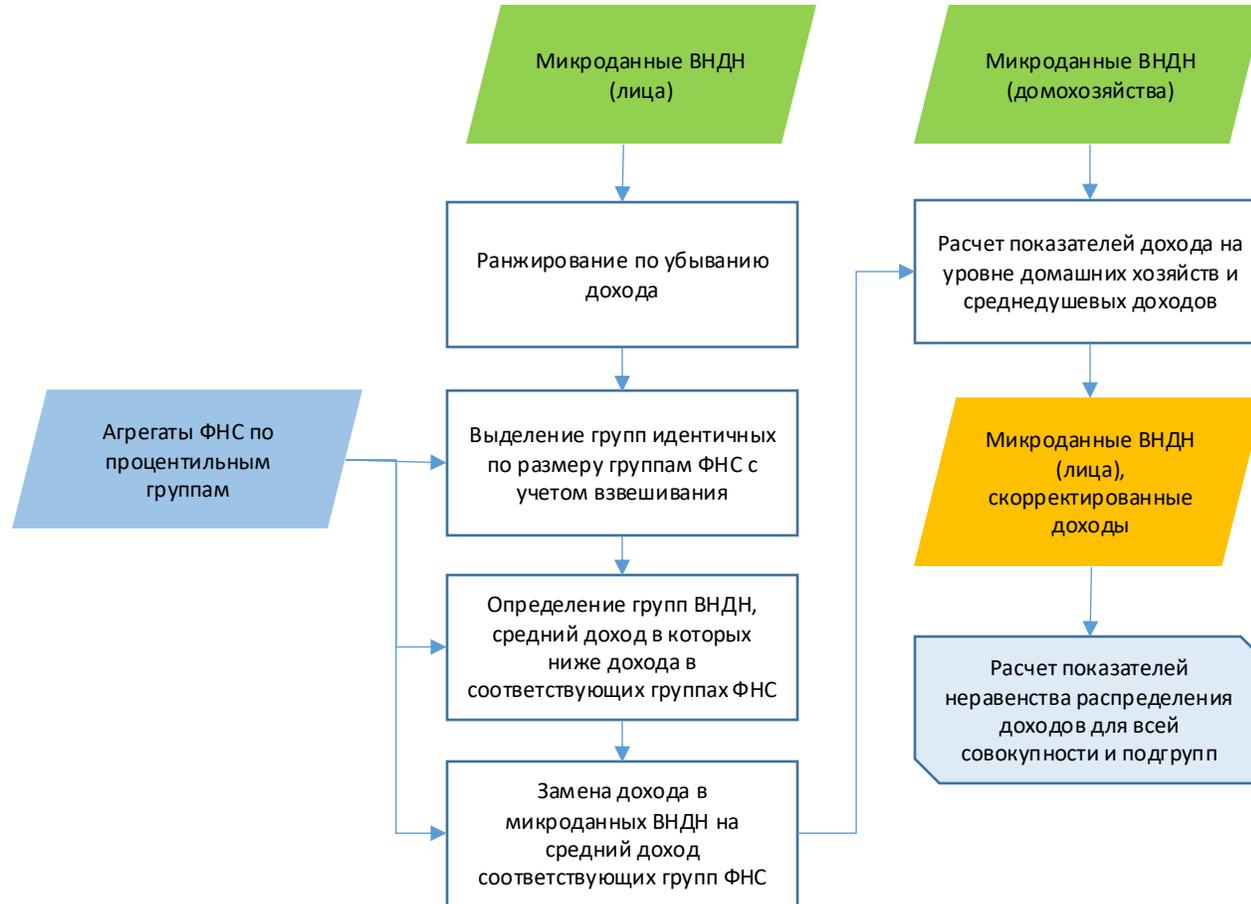
Данные

- Выборочное наблюдение доходов населения и участия в социальных программах (ВНДН)
 - Обследование проводится ежегодно с 2016 года во всех субъектах Российской Федерации с охватом 60 тысяч домашних хозяйств (один раз в три года – 160 тысяч домашних хозяйств).
 - Статистическими единицами являются домашние хозяйства и члены домашних хозяйств.
 - В расчетах используются файлы микроданных ВНДН, проведенного в 2023 году по итогам за 2022 год. Для распространения результатов на генеральную совокупность применяются весовые коэффициенты на уровне лиц и домохозяйств.
- Данные Федеральной налоговой службы (ФНС) по трудовым денежным доходам основаны на сведениях из деклараций о доходах физических лиц, облагаемых налогами:
 - Средние значения дохода по 1-процентным группам налогоплательщиков за 2022 г.
 - Статистической единицей является лицо – налогоплательщик.

Методика оценки неравномерности распределения доходов с использованием данных ВНДН и ФНС

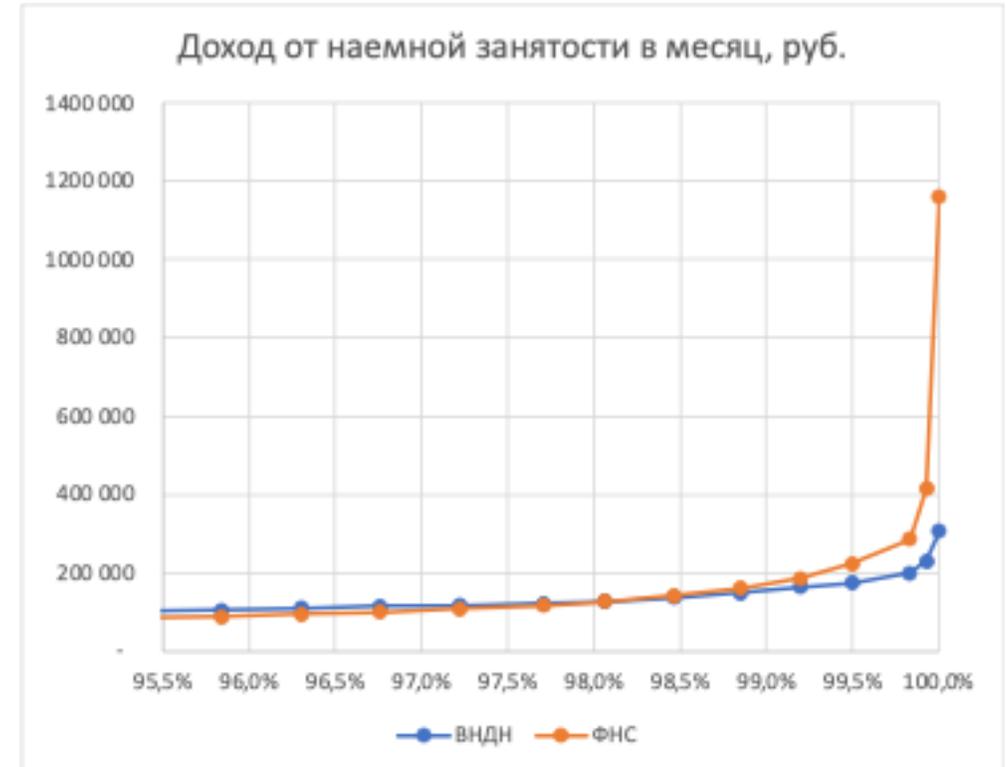
1. Ранжировать респондентов в ВНДН по их трудовому денежному доходу в порядке убывания величины дохода.
2. Сформировать в ВНДН группы, идентичные по размерам (с учетом весовых коэффициентов) группам в налоговых данных.
3. Определить точку пересечения распределения доходов в ВНДН и налоговых данных – пороговую группу, в которой средний доход по налоговым данным превышает средний доход в обследовании.
4. Заменить значения дохода в топ-группах в ВНДН, начиная с пороговой группы, на средний доход соответствующих групп в налоговых данных.
5. Распространить корректировку доходов на уровень домашних хозяйств для получения оценок среднедушевых доходов. Выполняется путем расчета доходов на уровне домохозяйств на основе скорректированных значений индивидуальных трудовых денежных доходов членов домашних хозяйств и расчета среднедушевого дохода в домохозяйстве.
6. Приписать полученные среднедушевые доходы членам домашних хозяйств и использовать индивидуальные весовые коэффициенты для распространения оценок доходов на население России.

Процесс корректировки и результаты



Распределение трудовых доходов по данным ВНДН и ФНС и зона корректировки

- Правая часть распределения денежных доходов на данным ВНДН и ФНС, включающая точку пересечения распределений. Размеры групп ВНДН и ФНС близки, насколько это возможно с учетом того, что шаг изменения размера группы ВНДН определяется величиной коэффициента взвешивания. В группах ВНДН и ФНС лица с их индивидуальным трудовым доходом. Размеры в группах в процентах – по отношению к населению России по данным ВНДН.
- Пороговой группой, в которой доход по данным ФНС начинает превышать доход по данным ВНДН, является группа 98,7%. Это соответствует ситуации в других странах, выполняющих подобные сопоставления – обычно доходы по налоговым данным превышают доходы по данным обследования в зоне гораздо выше 90%, часто в области 98-99%.

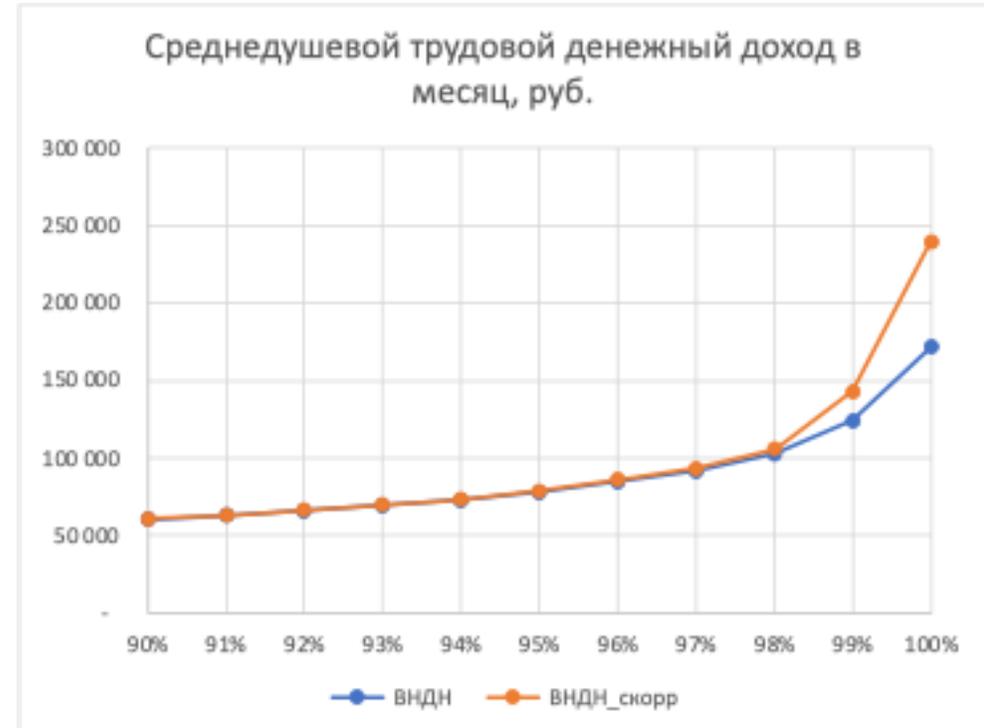


Распространение корректировки индивидуальных доходов на домашние хозяйства и агрегаты доходов более высокого уровня

- Корректировка индивидуальных трудовых доходов в ВНДН на уровне микроданных предоставляет свободу в группировании, как компонентов дохода, так и доходов лиц, входящих в состав домашнего хозяйства. Это позволяет распространять выполненную корректировку доходов ВНДН по налоговым данным на уровень домашнего хозяйства, а затем использовать оценки среднедушевых доходов лиц в домашних хозяйствах для построения распределения населения по доходам в соответствии с рекомендациями справочника по статистике доходов домашних хозяйств.
- Для оценки среднедушевого дохода в домашних хозяйствах часто используется понятие эквивалентного дохода. В представленных далее результатах мы используем средний доход на члена домашнего хозяйства, а не эквивалентный доход.
- Среднедушевой совокупный денежный доход рассчитывается на основе микроданных ВНДН рассчитывается суммированием скорректированного по налоговым данным трудового дохода, добавлением к нему других компонентов дохода, входящих в совокупный доход, агрегированием совокупного индивидуального дохода на уровень домашнего хозяйства и получением среднедушевого дохода.

Результат корректировки совокупного среднедушевого дохода

- Корректировка доходов в ВНДН на уровне микроданных позволяет распространять корректировку на агрегаты денежных доходов более высокого уровня, в частности на среднедушевой совокупный доход.
- Среднедушевой доход в группе ТОП 1% в результате корректировки вырос на 23%. Это эффект от корректировки только одного компонента дохода в составе доходов домохозяйств – дохода от наемной занятости. В дальнейшем целесообразно использовать и другие основные компоненты доходов, отражаемые в налоговой статистике, в первую очередь доходы от самозанятости и от собственности.



Влияние корректировки правого хвоста распределения доходов на показатели доходного неравенства

Среднедушевой совокупный доход

Индекс	ВНДН, эмпирический	ВНДН и корректировка по данным ФНС	Официальный показатель Росстата
Коэффициент Джини	0,348	0,363	0,395
Индекс Тейла	0,210	0,246	-
Коэффициент фондов	9,1	9,9	13,8

- Коэффициент Джини в результате корректировки доходов меняется незначительно, а индекс Тейла оказывается гораздо более чувствительным к корректировке.
- Индекс Тейла является мерой разности энтропии равномерного и фактического распределения доходов, и поэтому чувствителен не только к изменению указанной площади, но и к другим изменениям формы кривой распределения доходов Лоренца. В данном случае большее изменение индекса Тейла обусловлено тем, что корректировка затронула относительно малую группу домашних хозяйств, сконцентрированных в крайней правой части распределения.
- Корректировка эмпирического распределения доходов по данным обследования с использованием налоговых данных выполнялась в данном примере только по одному, хотя и основному компоненту дохода – доходам от наемной занятости. Этим обусловлен умеренный рост показателей доходного неравенства относительно рассчитанных по эмпирическим данным в сравнении с официальными показателями, рассчитываемыми на основе подгонке логнормального распределения с комбинированием эмпирических оценок и среднедушевым доходом по макроэкономическому показателю доходов домашних хозяйств. В дальнейшем целесообразно дополнить корректировку по налоговым данным и другими компонентами доходов – от наемной занятости и от собственности.

Заключение

- Мы рассмотрели наиболее распространенные подходы к оцениванию распределения доходов населения с использованием данных выборочного обследования, макроэкономического показателя доходов домашних хозяйств и налоговых данных, как наиболее информативного источника административных данных.
- Перспективным подходом к оцениванию распределения доходов населения является комбинирование данных выборочного обследования доходов и налоговой статистики. Такой подход позволяет корректировать правый хвост распределения доходов на основе объективных данных налоговой статистики на основе перевзвешивания, замены значений и комбинации перевзвешивания и замены.
- Корректировка правого хвоста распределения с использованием налоговых данных может осуществляться параметрическим способом с использованием закона Парето, параметры которого оцениваются по налоговым данным, или непараметрическим – путем замены значений в доходных топ-группах по данным обследования средними значения дохода в соответствующих топ-группах налоговых данных.
- Налоговые данные могут быть представлены в виде микроданных или агрегатов по процентильным группам или диапазонам дохода. Наиболее распространенным вариантом является доступность агрегированных налоговых данных. Исследования показали, что при наличии даже небольшого количества точек (интервалов), по которым представлены налоговые данные, можно выполнить интерполяцию и с высокой точностью восстановить распределение доходов на основе использования обобщенных кривых Парето – непараметрического способа описания распределения доходов и неравномерности распределения.

Спасибо!

Пожалуйста, вопросы