

**МЕТОДЫ ВОССТАНОВЛЕНИЯ ГЕНЕРАЛЬНОГО РАСПРЕДЕЛЕНИЯ ДУШЕВЫХ ДЕНЕЖНЫХ ДОХОДОВ НАСЕЛЕНИЯ НА ОСНОВЕ ВЫБОРОЧНЫХ ДАННЫХ БОЛЬШОГО ОБЪЕМА\***

**В.С. Жаромский,  
А.М. Рудберг,  
С.А. Тер-Акопов**

*В статье дается анализ популярных методов коррекции выборочного распределения душевых доходов населения и описывается методика оценки параметров логнормального распределения денежных доходов, учитывающая неравномерное представительство респондентов с разным уровнем доходов, в доходных децилях, которое зависит от дизайна выборки и доли отказов респондентов от участия в опросе. Ключевая идея работы состоит в выборе критерия для оценки параметров теоретического распределения, основанного не на сравнении групповых частот, а на сравнении границ и средних значений в теоретических децилях выборочных и искомым распределений. Средние значения в интервалах доходов, при достаточном количестве наблюдений, значительно меньше зависят от частоты представительности респондентов в интервалах. Это особенно важно, когда частотные характеристики выборки существенно отличаются от аналогичных показателей генеральной совокупности.*

*Рассмотрены два критерия оптимального выбора параметров логнормального распределения. Первый близок по методике, используемой в российской статистике, но при этом не требуется предварительная информация о том, какую часть составляет бедное население. Параметры оцениваются исходя из условия выполнения равенства среднего генерального дохода и совпадения выборочных и теоретических границ первого децильного интервала. Во втором случае минимизируется сумма отклонений выборочных и теоретических средних, рассчитанных в теоретических децилях. Оба критерия построены без учета гипотезы о возрастании вероятности отказов населения от участия в опросе в связи с ростом дохода домохозяйства, что позволяет сделать оценки представительности генеральных данных в выборочной совокупности.*

*Результаты конкретных расчетов показывают наибольшее представительство в средней части распределения и недостаточность наблюдений на концах, то есть и в группе бедных домохозяйств.*

*Ключевые слова:* логнормальное распределение, оценка параметров, выборочное обследование, отказы от участия в опросе, средний душевой денежный доход, генеральная совокупность, децильные интервалы денежных доходов населения.

*JEL:* C83, D31.

## **Введение**

Публикуемая статья посвящена «вечной» проблеме - проблеме переноса результатов выборочных оценок на генеральную совокупность. В данном случае рассматриваются выборочное распределение душевых денежных доходов и методики реконструкции на его основе генерального распределения. Необходимость реконструкции вытекает из того факта, что

средний душевой доход, рассчитанный на основе выборочных данных, может существенно отличаться от его оценки, полученной на основе макроэкономических данных: суммарной заработной платы, стипендий, пенсий, пособий, налогов, денежных льгот и данных о расходах. Как правило, макрооценка оказывается выше выборочной оценки денежного душевого дохода. Процент отличия сильно варьирует от региона к региону.

*Жаромский Владислав Спиридонович (zharomskyvs@gmail.com) - канд. экон. наук, ст. научн. сотрудник, Центр анализа доходов и уровня жизни Института управления социальными процессами НИУ ВШЭ.*

*Рудберг Алексей Михайлович (arudberg@hse.ru) - научн. сотрудник, Центр анализа доходов и уровня жизни Института управления социальными процессами НИУ ВШЭ.*

*Тер-Акопов Сергей Александрович (sterakopov@hse.ru) - научн. сотрудник, Центр анализа доходов и уровня жизни Института управления социальными процессами НИУ ВШЭ.*

\* Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ.

Авторы выражают благодарность сотрудникам Росстата и рецензентам статьи за предоставленные замечания, позволившие сделать текст более кратким и понятным.

Существует множество причин, из-за которых это происходит. Однако главные из них следующие:

- изначальная социально-демографическая структура населения, представленная в выборке, не вполне адекватна действительной;

- неравномерное распределение по территории России наиболее бедных и наиболее богатых домохозяйств;

- определенная форма сопротивления со стороны населения, которая проявляется на этапе самого опроса (отказы от участия в обследовании, искажение или сокрытие данных о доходах и расходах).

*Первая причина* связана с отсутствием текущих данных о социально-демографической структуре населения. Как правило, используются данные последней переписи населения. Поэтому если дизайн выборки осуществлялся в 2002 г., то к моменту следующей переписи структура населения могла значительно измениться, особенно в регионах. Проблема усугубляется длительностью обработки данных переписи, поэтому в 2002 г. при разработке дизайна выборки учитывались структурные данные еще более раннего периода времени. В случае, когда имеют место только количественные изменения, а качественная структура сохраняется, можно провести коррекцию представительности домохозяйств разного типа с помощью дополнительного взвешивания. Сложности возникают, когда происходят качественные изменения в структуре населения. Например, в выборку не попадают семьи, которые имеют право на материнский капитал. В этом случае имеющиеся макроданные о суммарном выделении средств на выплату материнского капитала не могут быть отражены в выборке, так как нет семей, которые его получили. Поэтому любое взвешивание может только перераспределить этот вид дохода среди других типов семей, но не скорректировать его, и поэтому результаты распределительных отношений будут искажены.

*Поляризация общества по денежным доходам проявляется неравномерно в географическом плане*, а ограничение на объем выборки не позволяет при существующем способе отбора домохозяйств, используемом для их включения в выборку, адекватно отразить распределение доходов, в результате чего происходит заметный сдвиг в оценках [5]. Надо также иметь в виду, что включение или исключение нескольких «очень богатых» домохо-

зяйств из выборки может на десятки рублей изменить средний душевой доход в регионе (объем выборки не превышает 500-800 домохозяйств). Этого не происходит, если пропущено домохозяйство с доходом из диапазона массовых значений (средняя часть распределения).

И наконец, *отказы от участия в обследовании являются реальным фактом*. Происходит это по разным причинам, как объективным (болезнь, переезд, отдых вне дома и т. д.), так и субъективным (нежелание пускать в дом чужих людей, показывать несоответствие между заявляемым доходом и видимой частью расходов и т. д.). И это касается не только богатых, но и бедных домохозяйств, которые могут скрывать неофициальные или нерегулярные доходы.

В последние годы для оценки успешности социальной политики все больше внимания уделяется показателям, характеризующим уровень бедности, неравенства и поляризации, которые рассчитываются на основе распределения населения по душевым денежным доходам. Более того, эффективность отдельных программ и проектов в области социальной политики все чаще оценивается на основе перечисленных показателей. Сравнивая значения показателей бедности и неравенства до и после реализации программ, можно оценить их эффективность. На стадии формирования программ и выработки предложений разработчики ориентируются на страны, в которых неравенство и бедность находятся на приемлемом уровне. Появляется возможность получить стоимостную оценку достижения европейских стандартов качества жизни не только на уровне страны, но и для разных социально-демографических групп населения. Однако это оценочное направление требует серьезного усовершенствования методологии анализа и моделирования уровня жизни населения. К этому кругу вопросов относится и задача реконструкции генерального распределения душевых доходов на основе данных выборочных обследований.

### **Анализ используемых методов реконструкции**

В силу того что не известно реальное распределение населения по денежному душевому доходу, а известны только суммарные макроэкономические оценки (общий, полученный населением доход и численность населения), оцененные Росстатом, приходится использовать выборочные дан-

ные большого объема (около 45000 домашних хозяйств), которые поступают из регулярных квартальных обследований бюджетов домашних хозяйств (ОБДХ). Вполне естественно решить задачу более полного совместного использования генеральных и выборочных данных для реконструкции генерального распределения населения по доходам.

В настоящее время в практической работе Росстата используется методика реконструкции, которая предполагает, что распределение душевых денежных доходов населения относится к классу логарифмически нормальных. Поэтому для формального описания генерального распределения достаточно оценить два параметра, которые жестко связаны с его статистическими характеристиками. В качестве первого показателя берется макроэкономическая оценка среднего душевого денежного дохода, а для определения второго параметра используется доля населения с душевым доходом ниже прожиточного минимума или другой величины, например ниже 1500 или 7000 рублей. Выбор конкретного значения этой границы очень важен, и этот вопрос мы обсудим позже более подробно, а сейчас кратко рассмотрим популярные методы и модели реконструкции генерального распределения.

**Метод 1.** В работе [8] предлагается схема взвешивания выборочных данных, направленная на уменьшение структурных отличий в представительности домохозяйств с разным уровнем дохода, возникающих в результате отказа респондентов от участия в обследовании. При этом принимается гипотеза *о монотонном возрастании вероятности* отказа домохозяйства от участия в опросе вместе с ростом денежного душевого дохода домохозяйства. Точнее, предполагается, что доля домохозяйств в выборке по сравнению с их долей в генеральной совокупности тем меньше, чем выше их душевой доход. Само предположение базируется на статистике отказов домашних хозяйств от участия в обследовании. К сожалению, эта статистика не полна и не точна и поэтому не может быть использована для прямой коррекции данных обследований. В силу указанного предположения средний выборочный душевой доход должен получаться ниже, чем душевой доход, рассчитанный на основе макроданных о доходах и расходах населения. Поэтому в работах [6, 7] авторами была предложена методика построения корректирующих весов наблюдений, после умно-

жения на которые взвешенное выборочное среднее значение совпадает с генеральным, а сами веса минимально различаются между собой. Важным следствием этого подхода явилась возможность получения оценок децилей (правые границы децильных интервалов) для *генеральной* совокупности и соответствующих средних доходов в децильных интервалах, которые используются для построения различных индексов бедности, неравенства и поляризации. Предполагается, что без такой коррекции оценки выборочных децилей получают заметные смещения, особенно в правой части распределения доходов, где располагается наиболее состоятельная часть населения. В основе этой методики лежат следующие соображения. Если бы представители групп населения с разным уровнем душевых доходов попадали в выборку бюджетных обследований с равной вероятностью, то значение  $k$ -й децили  $q(k)$  (правая граница  $k$ -го децильного интервала) на выборочных данных можно было оценивать следующим образом:

$$q(k) = y \left( \left[ \frac{kn}{10} \right] + 1 \right), \quad k = \{1, \dots, 9\}, \quad (1)$$

где  $y(i)$  - значение душевого денежного дохода  $i$ -го домохозяйства, а  $n$  - размер выборки.

При этом все децильные интервалы содержали бы равное (или почти равное) количество наблюдений, а именно  $\frac{n}{10}$ . Предполагается, что все наблюдения  $y(i)$  предварительно сортированы в порядке возрастания душевого денежного дохода. Если же выполняется гипотеза монотонного роста отказов, то условие равной представительности нарушено и для оценки децилей выражение (1) использовать неправомерно. Авторы метода предложили оригинальную схему построения оценок. Если обозначить  $p(k)$  вероятность попадания наблюдений из  $k$ -го децильного интервала генеральной совокупности в выборку [полагая при этом  $p(1) > p(2) > \dots > p(10)$ ], то «след» истинной  $k$ -й децили генеральной совокупности может быть оценен как

$$q(k) = y \left( \left[ \frac{s(k)n}{10} \right] + 1 \right), \quad k = \{1, \dots, 9\}, \quad (2)$$

где

$$s(k) = \frac{\sum_{j=1}^k p(j)}{\sum_{j=1}^{10} p(j)}, \quad k = 1, \dots, 9.$$

Величина  $s(k)$  представляет собой долю выборки, в которую попали респонденты с доходами не выше  $k$ -й децили для распределения доходов в генеральной совокупности. То есть это правая граница суммарного «следа»  $k$  первых генеральных децильных интервалов. К сожалению, формула (2) полезна только в том случае, если известны сами величины  $p(j)$ . Однако такой статистики нет, и поэтому была предложена конструкция, позволяющая оценить доли выборки  $s(k)$  с помощью назначения всем наблюдениям  $y(i)$  таких весов  $w(i)$ , при которых выполняются следующие условия:

$$\sum_{i=1}^n w(i) \log\left(\frac{1}{w(i)}\right) \rightarrow \max; \quad (3)$$

$$\sum_{i=1}^n w(i) = 1; \quad (4)$$

$$\sum_{i=1}^n w(i)y(i) = M, \quad (5)$$

где  $M$  - средний душевой доход, являющийся макроэкономической оценкой душевого денежного дохода в генеральной совокупности.

Смысл этих выражений понятен. Критерий максимизации энтропии (3) реализует требование максимальной равномерности весов, то есть близости их к  $\frac{1}{n}$ . Ограничение (4) является условием нормировки весов, а выражение (5) требует точного совпадения, взвешенного выборочного душевого дохода с его макроэкономической оценкой. После такого взвешивания величины  $s(k)$  определяются как  $s(k) = \frac{n(k)}{n}$ , где  $n(k)$  - число наблюдений в выборке, определяемое из уравнения:

$$\sum_{i=1}^{n(k)} w(i) = \frac{k}{10}, \quad k = \{1, \dots, 9\}. \quad (6)$$

Фактически веса  $w(i)$  меняют частоту включения домохозяйств с доходом  $y(i)$  из генеральной совокупности в выборку. При этом наборы наблюдений  $\{1 \dots n(1)\}, \{n(1) + 1 \dots n(2)\}, \dots, \{n(9) + 1 \dots n\}$  будут иметь, вообще говоря, разные размеры и представлять собой «следы» генеральных децильных интервалов в рассматриваемой выборке. Задача нахождения весов легко решается, так как система (3) - (5) имеет аналитическое решение, представимое в простом виде:

$$w(i) = ae^{by(i)}, \quad (7)$$

где параметры  $a$  и  $b$  определяются из условий (4) и (5) и могут быть рассчитаны с помощью некоторой итерационной процедуры.

Из вида решения (7) следует, что параметр  $b > 0$  и поэтому веса  $w(i)$  действительно возрастают с ростом дохода  $y(i)$ . Если выборочный средний душевой доход меньше генерального  $M$ , то из требований баланса (5) и условия минимизации (3) следует, что наблюдения с более высоким доходом будут иметь относительно большие веса для компенсации этого превышения. Критерий так построен, что он принципиально не может давать высокие оценки домохозяйствам с низким уровнем доходов. Хотя это возможно в случае, если генеральный доход  $M$  ниже выборочного, однако на практике это не наблюдается. В то же время при таком подходе не учитывается подмеченный экономистами факт частого отказа от участия в обследовании и той части населения, которая имеет низкие душевые доходы. О причине этого явления было сказано выше. Таким образом, модель с монотонным ростом весов, решая задачу выхода на уровень генерального среднего, может внести дополнительные искажения в оценку показателей неравенства и бедности, если условие монотонности не выполняется.

Большим достоинством реконструкции выборочного распределения с помощью метода взвешивания является его независимость от вида распределения населения по доходам. Однако статистические характеристики выборки используются в рамках этого подхода минимально (выход на генеральное среднее). Остается также открытым вопрос о выборе самого критерия. На самом деле, существуют другие критерии качества, которые при ограничении (4) также достигают оптимума в случае равенства всех весов. Например, вместо критерия (3) можно взять критерий

$$\sum_{i=1}^n \log(w(i)) \rightarrow \max, \quad (8)$$

который вместе с условиями (4) - (5) приводит к решению

$$w(i) = \frac{1}{a + by(i)}. \quad (9)$$

Еще одним примером такого критерия является сумма квадратов весов, то есть

$$\sum_{i=1}^n w(i)^2 \rightarrow \min, \quad (10)$$

для которого вид решения является линейной функцией от дохода

$$w(i) = a + by(i). \quad (11)$$

Наши расчеты, проведенные на реальной выборке 44530 домохозяйств из обследования НОБУС 2002 г., показали, что хотя решения оказываются близкими, но их поведение на краях

распределения неодинаково. Энтропийный критерий более сильно занижает веса для малых доходов и, наоборот, увеличивает веса для больших доходов.

Таблица 1

Сравнение коэффициентов взвешивания по двум критериям

Группа	Выборка	Критерий $\sum w(i) \times \log\left(\frac{1}{w(i)}\right)$			Критерий $\sum \log(w(i))$		
		Средние весовые коэффициенты по группе	Произведение средних весовых коэффициентов на средний душевой денежный доход	Коэффициенты коррекции исходных долей	Средние весовые коэффициенты по группе	Произведение средних весовых коэффициентов на средний душевой денежный доход	Коэффициенты коррекции исходных долей
А	1	2	3	4	5	6	7
1	839,6	0,029	25,1	0,74	0,032	27,4	0,81
2	2905,6	0,031	91,3	0,78	0,033	97,4	0,83
3	3999,4	0,032	128,9	0,80	0,034	136,3	0,85
4	4814,8	0,032	157,6	0,81	0,034	166,0	0,86
5	5481,7	0,033	182,3	0,83	0,034	190,9	0,87
6	6058,0	0,033	204,2	0,84	0,035	212,7	0,87
7	6616,3	0,034	225,9	0,85	0,035	234,3	0,88
8	7147,8	0,034	247,1	0,86	0,035	255,2	0,89
9	7689,4	0,035	269,2	0,87	0,036	276,8	0,90
10	8210,6	0,035	291,0	0,88	0,036	297,9	0,90
11	8703,1	0,035	312,0	0,89	0,036	318,2	0,91
12	9231,4	0,036	335,0	0,90	0,036	340,3	0,92
13	9798,8	0,036	360,4	0,91	0,037	364,5	0,93
14	10414,4	0,037	388,6	0,93	0,037	391,2	0,93
15	11019,5	0,037	417,0	0,94	0,037	418,0	0,94
16	11708,0	0,038	450,2	0,96	0,038	449,1	0,95
17	12566,3	0,039	493,0	0,98	0,038	488,9	0,97
18	13460,6	0,040	539,2	1,00	0,039	531,5	0,98
19	14504,4	0,041	595,3	1,02	0,040	583,0	1,00
20	15749,0	0,042	665,5	1,05	0,041	646,8	1,02
21	17315,4	0,043	758,9	1,09	0,042	731,1	1,05
22	19336,3	0,045	888,4	1,14	0,043	847,2	1,09
23	22188,5	0,049	1089,5	1,22	0,046	1026,8	1,15
24	26818,0	0,054	1467,1	1,36	0,050	1365,6	1,27
25	47592,6	0,088	4227,1	2,22	0,092	4408,9	2,31
Средние	12166,8		14809,6			14806,1	
КД*	<b>12,47</b>		<b>13,43</b>			<b>13,75</b>	

\* КД - коэффициент дифференциации.

В средней части распределения веса примерно одинаковы. В таблице 1 приведены расчеты для групповых весов (всего 25 групп), полученных по двум критериям. Последний столбец содержит их отношение, которое более наглядно демонстрирует отличия решений на концах распределения. Первый столбец содержит выборочные оценки средних душевых доходов для 4% групп. Столбцы 2 и 5 содержат веса, рассчитанные по двум критериям, а столбцы 4 и 7 - коэффициенты коррекции

исходных долей для каждого критерия, которые равны  $\frac{w(i)}{0,04}$ . Они показывают, во сколько раз нужно изменить исходные равные веса, чтобы средний взвешенный душевой доход стал равным генеральному (14800). Из-за особенностей критериев заметно различие в коэффициентах дифференциации (2,5%). Точнее, приведенные КД являются отношением средних значений последних 12% наблюдений к первым 12%, а не к 10%, как это обычно принято.

На концах распределения различия для корректирующих весов особенно заметны, и поэтому остается открытым вопрос: какой критерий предпочтительнее? Возможно, требуется какое-то дополнительное условие, которое позволит сделать правильный выбор критерия, который в большей степени будет учитывать статистические характеристики выборки, например дисперсию или другую меру разброса данных.

Дисперсия выборки при описанном подходе учитывается лишь косвенно при расчете весов и не очевидным образом. Можно, конечно, получить оценку дисперсии с учетом полученных весов и принять ее за оценку генеральной дисперсии. Но при этом не очень ясно, с чем ее сравнивать, так как в отличие от генерального среднего значения для генеральной дисперсии такой внешней оценки нет. Вообще, подобные модели корректируют данные только с одной целью - выйти на уровень генерального среднего. И на основании этой коррекции делаются все дальнейшие расчеты показателей и выводы относительно генеральной совокупности. Игнорируя тот факт, что нет реальной статистики, подтверждающей верность «гипотезы монотонности» для децилей из среднего диапазона доходов, то есть децилей с номерами 4-8 или в рассмотренном случае процентиля с номерами от 9 до 20.

Можно указать еще на одну особенность рассмотренной методики взвешивания. Если выполняется предположение о логнормальности распределения доходов, то это позволяет пролонгировать его действие на более широкий класс наблюдений - фактически на всю генеральную совокупность. В то время как метод простого взвешивания переводит выборочное распределение в распределение, которое ничего не может сказать о весах значений, лежащих вне интервала наблюдения.

**Метод 2.** Существует интересный подход Айвазяна-Каленикова [1], в котором предполагается, что плотность распределения случайной величины душевого дохода будет описываться моделью смеси логнормальных законов распределения однородных социально-экономических страт населения. Поэтому результирующее генеральное распределение как бы склеено из нескольких типов распределений. При этом расчет параметров распределений также нацелен на достижение генерального среднего дохода и плавности перехода одного распределения в другое. Возникает дополнительная, довольно сложная

итерационная задача. Кроме того, необходимо предварительно выделить однородные по доходу страты, число которых во многом определяется «вкусом» исследователя.

**Метод 3.** Известен еще один широко используемый метод, с помощью которого производится взвешивание выборочных данных, названный авторами CALMAR [13, 14, 15]. По своим достоинствам он близок к методу Шевякова-Кируты и хорошо «работает», когда выборочные значения показателей имеют небольшие отклонения от генеральных. При заметных отклонениях наиболее популярная линейная модель метода CALMAR может приводить к появлению отрицательных корректирующих весов. Для исправления этого дефекта вводятся дополнительные ограничения на величину изменения исходных весов, выбор которых требует обоснований. В указанных работах дается полное описание метода и приведены ссылки на его реализацию в пакетах SPSS и SAS. В работе [13] рассматривается вариант взвешивания, учитывающий статистические характеристики отказов.

### Предлагаемая модель реконструкции

Учитывая приведенные выше соображения и факт широкого использования в практике Росстата логнормального распределения при анализе доходов населения, авторы данной работы пришли к идее совместного использования теоретических оценок децилей и децильных средних для выбора параметров логнормального распределения. Мы также полагаем, что наибольшее сходство выборочного и теоретического распределений должно наблюдаться в средней части распределения, где действуют массовые распределительные процессы. Существуют и теоретические обоснования выбора типа распределения, которые были получены в работе [4] при анализе разных схем стимулирования работников. Еще ранее в работе [2] были проведены исследования о природе распределения заработной платы работников и доходов домохозяйств, которые указали на их логарифмически нормальный вид. Из более поздних работ, нацеленных на выбор наиболее подходящего вида распределения доходов, можно отметить работу [9] на данных Австралии (14000 домохозяйств) и [11] на данных Италии (10000 домохозяйств). Вопрос выбора типа распределения пока не получил окончательного решения.

Так, известный бразильский экономист К. Дагум долгое время занимался исследованием распределения доходов и заработной платы. Он предложил вариант трехпараметрического распределения, которое, по его мнению, более точно описывает характер распределения по доходам [12]. Также трехпараметрическое логнормальное распределение рассмотрено в работе [10], где предложена методика оценки параметров с помощью метода  $L$ -моментов. Фактически третий параметр вводится для того, чтобы исключить из рассмотрения домохозяйства со слишком низкими доходами. Однако доходы ниже 500 рублей в месяц могут наблюдаться в сельских домохозяйствах, где денежный доход оценивается на основе продаж продуктов из личного подсобного хозяйства.

Главная проблема реконструкции заключается в нахождении таких критериев, которые не приводят к решениям, плохо согласующимся с существующими представлениями о текущих уровнях бедности и богатстве в обществе. Так, практическое использование в качестве параметра генерального распределения доли населения с доходом ниже 1500 рублей (эти данные относятся к 2003 г.!) отражает наше представление о бедности и ее глубине. Спорным моментом такого выбора является не само значение границы, а **оценка доли** населения с доходами ниже границы. Дело в том, что по результатам анализа выборочных данных большого объема можно определить тот рубеж, ниже которого домохозяйство не в состоянии удовлетворять свои самые насущные потребности (состояние бедности). Например, оно не может купить товары и оплатить услуги, включенные в минимальную потребительскую корзину. Однако при этом остается открытым вопрос о способе оценки доли людей с душевым денежным доходом ниже этого порога, так как эта доля зависит от используемой модели реконструкции распределения. Если представительность бедных в выборке совпадает с реальной, тогда эту долю можно принять и использовать для оценки параметров генерального распределения. А если этого не происходит, то необходимо использовать дополнительные соображения для оценки этой доли.

В публикуемой статье мы предлагаем подход, который статистические характеристики выборки учитывает в большей степени, чем энтропийный критерий при решении задачи взвешивания. Как отмечалось выше, для задания логнормального распределения достаточно знать два показателя,

которые его полностью определяют, например среднее значение и дисперсию или среднее значение и медиану. В силу того что между показателями и параметрами логнормального распределения существуют точные аналитические связи:

$$M = e^{\frac{\mu + \sigma^2}{2}}; \quad (12)$$

$$D = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1); \quad (13)$$

$$Med = e^{\mu}; \quad (14)$$

$$Mod = e^{\mu - \sigma^2}, \quad (15)$$

где  $M$  и  $D$  обозначают среднее значение и дисперсию;  $Med$  и  $Mod$  - медиану и моду соответственно;  $\mu$  и  $\sigma^2$  - среднее значение и дисперсию логарифмов душевых денежных доходов.

Для реконструкции генерального распределения можно использовать любую их пару. Однако для практического применения они не равноценны. Из всех этих величин только средний душевой доход  $M$  имеет макроэкономическую оценку, другие показатели имеют только выборочные оценки. Правда, в работе [3] в качестве второго параметра используется мода и предполагается, что в большой выборке она практически близка к моде генерального распределения. Если бы это было так, то параметры логнормального распределения легко определялись на основе выборки. Мы в качестве второго показателя выбрали дисперсию, оптимальное значение которой рассчитывается с помощью не вполне аналитической процедуры, в том смысле, что оценка получается итерационно и на каждом шаге итерации необходимо использовать все выборочные данные.

Основная идея нашего подхода состоит в нахождении такой оценки для дисперсии, при которой наблюдаются минимальные расхождения между теоретическими средними значениями в каждом дециле и выборочными оценками. Как мы уже отмечали выше, при достаточной представительности выборки в децилях средние значения более устойчивы, чем выборочные частоты. В некотором смысле предлагается рассматривать задачу «сверху - вниз», то есть от генерального среднего к выборочному среднему, а не наоборот, как это делается обычно. Фактически рассматривается «след» выборочного распределения в генеральном. И по этой причине стремление к равной наполненности децильных интервалов

выборочными наблюдениями не является необходимым и правильным, поскольку частотные характеристики выборки заведомо отличаются от генеральных.

**Описание модели.** Мы рассмотрели два варианта критерия оптимального выбора дисперсии. В первом случае ищется такое значение дисперсии, при котором выборочные и теоретические значения первых децилей (правые границы) совпадали. Этот вариант близок к методике Росстата, но отличается выбором уровня границы бедности. Вторым подходом склоняется к выбору такого значения дисперсии, при котором достигается минимум суммы квадратов отклонений средних децильных выборочных значений душевого дохода от теоретических (генеральных). При этом в расчетах участвуют все децильные средние кроме крайних, то есть со 2-го дециля по 9-й. Для формализации задачи введем следующие обозначения. Пусть  $s$  - дисперсия выборочного распределения, а  $m$  - выборочное среднее. Тогда при заданном генеральном среднем душевом доходе  $M$  и принятой начальной дисперсии  $D$  по логнормальной модели определим:

- границы децильных интервалов  $d(i)$  и средние значения теоретических доходов  $M(i)$ ;
- размеры  $n(i)$  «следов» выборки в генеральных децилях, то есть число элементов выборки, попавших в  $i$ -й децильный интервал, и среднее выборочное значение дохода  $m(i)$  в этом интервале.

Для среднего душевого выборочного дохода можно записать следующее равенство:

$$m = \sum_{i=1}^{10} p(i)m(i), \quad (16)$$

где  $p(i) = \frac{n(i)}{n}$ , а  $n$  - объем выборки.

Чтобы связать выборочные оценки (13) с теоретическими, рассмотрим схему независимых статистических испытаний, результатом которых являются выборки, полученные следующим образом:

- пусть  $\xi$  - случайная величина, распределенная по логарифмически нормальному закону с параметрами  $M$  и  $D$ ;
- генерируем значение величины  $\xi$  и определяем номер децильного интервала  $i$ , в который оно попадает;
- с вероятностью  $q(i)$  попадания в  $i$ -й дециль внесим это значение в выборку, иначе его отвергаем;
- процесс продолжается до тех пор, пока размер выборки не будет равен  $n$ .

В рамках этой схемы, если генеральное распределение доходов является логарифмически нормальным, наша исходная выборка может рассматриваться как один из опытов в серии. Легко видеть, что математическое ожидание среднего душевого выборочного дохода  $m(i)$  в  $i$ -м генеральном дециле будет равно  $M(i)$ , и поэтому математическое ожидание  $\underline{M}$  самого выборочного дохода  $m$  будет равно:

$$\underline{M} = \sum_{i=1}^{10} p(i)M(i), \quad (17)$$

где  $p(i) = \frac{q(i)}{\sum qk}$ .

Использовать в формуле (17) сами величины  $q(i)$  нельзя, так как их сумма не равна 1, что требуется в процедуре расчета среднего значения. Полученное среднее выборочное значение  $m$  является оценкой для  $\underline{M}$ , и поэтому модуль их разности может быть использован в качестве критерия для выбора дисперсии логарифмически нормального распределения, наиболее согласованного с выборочными данными. Однако представляется более интересным ориентироваться не на одну величину  $\underline{M}$ , а на целый ряд величин  $M(i)$ , которые при выполнении предположения о согласованности генерального и выборочного распределений (в смысле схемы независимых испытаний) должны быть равными при достаточном объеме выборки. Дело в том, что при большом размере выборки наполнение каждого децильного интервала элементами выборки будет представительным, а оценка среднего дохода в дециле - значимой. Частоты  $n(i)$ , конечно, будут разные, но близкие к своим средним значениям, зависящим от структурных отличий, заложенных в дизайн выборки, а также от интенсивности отказов от опросов. Именно по причине изначального отличия генеральных и выборочных частот попадания в децили совершенно неправильно искать решение, которое минимизирует эти частотные отклонения. Более полезными свойствами обладают средние значения доходов в децилях. Ошибки больших отличий здесь подчиняются закону больших чисел, то есть с ростом числа наблюдений, попадающих в дециль, вероятность большой ошибки отклонения быстро уменьшается.

Действительно, если в децильный интервал должно попасть 10000 наблюдений, а реально попало только 6500 наблюдений (или 13000), то отличие частот будет очень заметно, а значения



средних будут отличаться незначительно. Фактически здесь действует принцип, используемый при определении необходимого объема выборки, достаточного для оценки среднего значения с заданной точностью. Из теории известно, что ошибка отклонения выборочного среднего от генерального пропорциональна дисперсии и обратно пропорциональна корню квадратному из объема выборки. А так как дисперсия наблюдений в дециле значительно ниже, чем во всей выборке, то вероятность больших отклонений становится еще меньше.

### Расчеты и оценки

Все расчеты проводились на актуализированных (на 2010 г.) данных 2003 г. из обследования НОБУС по России в целом, объем которых - 44529 домохозяйств. Это обследование содержало данные не только о расходах, но и о доходах населения. После актуализации данных средний душевой денежный доход был равен 12167 рублям, а генеральный был принят равным 14800 рублям. Ниже представлены результаты расчетов для двух альтернативных критериев выбора значения дисперсии, используемых для реконструкции логнормального распределения, то есть условно генерального распределения.

В таблице 2 приведены характеристики выборки и трех модельных логарифмически нормальных распределений, построенных с использованием выборочных параметров. Так, в первом столбце таблицы приведены расчетные данные для самих доходов и их логарифмов. Чтобы понять, в какой степени логнормальные распределения, построенные на базе исходных данных и их логарифмов, в столбцах 2 и 3 приведены альтернативные параметры, рассчитанные по формулам (12) и (13). Выбранная пара в таблице выделена жирным шрифтом. Можно заметить, что построенные на разных парах логнормальные распределения и рассчитанные вторые пары параметров существенно различаются между собой. Особенно это касается дисперсии. Если рассчитать значение дисперсии на основе выборочных логарифмических параметров, то она возрастает почти в семь раз - с 175092098 до 1212047801. Распределение, восстановленное по логарифмическим параметрам, дает среднее, большее на 40% (18406 против 12167). На первый взгляд кажется, что использование гипотезы о логарифмически нормальном

распределении доходов в данном случае неправомерно, но дальнейший анализ покажет, как это противоречие можно объяснить и разрешить.

Таблица 2

Статистические характеристики исходной выборки и модельных распределений

Характеристики	Выборочные значения	Реконструкция на основе выборочных параметров	Реконструкция на основе выборочных логарифмических параметров
А	1	2	3
Среднее значение душевого денежного дохода	12167	<b>12167</b>	18406
Дисперсия	175092098	<b>175092098</b>	1212047801
Средний логарифм	9,060	8,823	<b>9,060</b>
Дисперсия логарифма	1,521	1,168	<b>1,521</b>

**Критерий 1.** В качестве первого критерия для выбора дисперсии рассмотрим подход, применяемый в Росстате, но с отличием, которое коснулось оценки доли бедного населения. В таблице 3 приведены результаты расчетов, где величина дисперсии логарифмов доходов выбиралась из условия совпадения первых децилей (выборочная граница - 4002 рубля, а генеральная - 4003 рубля). Отличие в 1 рубль связано с дискретностью выборочных значений. Таблица содержит четыре группы данных:

- границы и средние значения выборочных децильных интервалов;
- генеральные средние значения и доли, рассчитанные для выборочных децильных интервалов по логнормальному закону;
- границы и средние значения для генеральных децильных интервалов;
- выборочные средние значения и доли выборки, попавшие в генеральные децильные интервалы.

В таблице 3 выделены два столбца (2-й и 4-й), которые показывают, что различие между выборочными средними и расчетными (генеральными) не превышают 0,23% для интервалов (2-9) и равны 22% для 1-го и 10-го интервалов. Если бы было полное совпадение распределений, то в столбце 4 все доли были бы равны 0,1 (10% выборки). Фактически графики средних значений совпадают везде, кроме концов. Представляется, что население с «не крайними» доходами представлено в выборке более *полно* и поэтому хорошо покрывается выборкой. Крайние группы «богатых» и «бедных» домохозяйств распределены

не совсем равномерно, или доступ к ним более затруднен (отказы по разным причинам). Однако нам представляется, что на концах распределения перестают действовать обычные (массовые) формы получения дохода, такие, как заработная плата, премии, надбавки, пенсии и т. д., и начинают

действовать законы и причины, не характерные для основной массы населения (бонусы, доходы от ценных бумаг и т. д.). Аналогичные замечания касаются наиболее бедной части населения, доходы которой могут быть нерегулярны и тоже не соответствовать трудовым усилиям.

Таблица 3

Значение дисперсии  $s^2$ , обеспечивающее совпадение первых децилей

$s^2 = 0,734$	Исходная выборка (децили)		Логнормальное распределение в выборочных децилях		Логнормальное распределение оптимальное (децили)		Выборка в логнормальных децилях		
	Децили	Граница	Среднее	Доля	Среднее	Граница	Среднее	Доля	Среднее
	А	1	2	3	4	5	6	7	8
	1	4002	<b>2253</b>	0,010	<b>2943</b>	4003	2943	0,1001	2254
	2	5781	<b>4964</b>	0,118	<b>4913</b>	5529	4784	0,0832	4824
	3	7159	<b>6474</b>	0,095	<b>6467</b>	6978	6251	0,1032	6266
	4	8460	<b>7815</b>	0,085	<b>7801</b>	8514	7734	0,1170	7747
	5	9790	<b>9104</b>	0,078	<b>9113</b>	10254	9362	0,1275	9344
	6	11348	<b>10563</b>	0,080	<b>10549</b>	12349	11263	0,1197	11232
	7	13456	<b>12352</b>	0,090	<b>12361</b>	15067	13639	0,1083	13620
	8	16477	<b>14843</b>	0,097	<b>14880</b>	19017	16894	0,0960	16842
	9	22067	<b>18928</b>	0,111	<b>18981</b>	26264	22167	0,0829	22104
	10	1000000	<b>34373</b>	0,148	<b>43549</b>	1000000	52961	0,0621	40654
	СДД*		12167		14800		14800		12167
	КД		15,3		14,8		18,0		18,0

\* СДД - среднедушевой денежный доход.

Если теперь посмотреть на предпоследний столбец, содержащий доли выборки, попадающие в генеральные децильные интервалы, то можно заметить, что вид этого распределения совершенно не похож на тот, который предполагает гипотеза о росте отказов при росте доходов [6]. Никакой монотонности в поведении весов не наблюдается. Кроме того, видно, что средняя часть распределения представлена более полно в выборке, а недостаток наблюдается на краях, то есть и среди бедных.

**Критерий 2.** В качестве второго критерия оптимальности была выбрана сумма квадратов отклонений выборочных средних от теоретических средних. В таблице 4 приведены результаты расчетов, подобные рассмотренным выше. Как и в таблице 3, здесь также выделены два столбца (6-й и 8-й), которые показывают, что различие между выборочными средними и расчетными (генеральными) не превышают 0,23% для интервалов (2-9) и равны 22% для 1-го и 10-го интервалов.

Таблица 4

Значение дисперсии  $s^2$ , минимизирующее отличия средних в децилях

$s^2 = 0,6994$	Исходная выборка (децили)		Логнормальное распределение в выборочных децилях		Логнормальное распределение оптимальное (децили)		Выборка в логнормальных децилях		
	Децили	Граница	Среднее	Доля	Среднее	Граница	Среднее	Доля	Среднее
	А	1	2	3	4	5	6	7	8
	1	4002	2253	0,085	3006	4257	<b>3171</b>	0,110	<b>2422</b>
	2	5781	4964	0,114	4922	5791	<b>5044</b>	0,091	<b>5060</b>
	3	7159	6474	0,096	6470	7230	<b>6510</b>	0,106	<b>6521</b>
	4	8460	7815	0,087	7802	8739	<b>7973</b>	0,117	<b>8006</b>
	5	9790	9104	0,082	9114	10433	<b>9565</b>	0,117	<b>9550</b>
	6	11348	10563	0,084	10549	12455	<b>11408</b>	0,115	<b>11352</b>
	7	13456	12352	0,094	12361	15055	<b>13691</b>	0,103	<b>13666</b>
	8	16477	14843	0,101	14878	18794	<b>16788</b>	0,093	<b>16744</b>
	9	22067	18928	0,115	18970	25565	<b>21746</b>	0,081	<b>21645</b>
	10	1000000	34373	0,142	43698	1000000	<b>52105</b>	0,069	<b>39278</b>
	СДД		12167		14800		14800		12167
	КД		15,3		14,5		16,4		16,2

Результаты, представленные в таблицах 3 и 4, показывают, что для обоих критериев средние доходы в крайних децилях корректируются в большую сторону. Это связано с тем, что имеется недостаток наблюдений с высокими доходами в последнем дециле.

В таблице 5 приведены децильные процентные отличия по обоим критериям, которые выражены в процентах к выборочным значениям, что можно описать следующей формулой:  $100 \times \frac{M(i) - m(i)}{m(i)}$ , где  $M(i)$  и  $m(i)$  - теоретическое и выборочное средние в  $i$ -м децильном интервале.

Таблица 5

Степень отличия средних в выделенных столбцах, оцененных по двум критериям (в процентах)

Децили	Критерий 1		Критерий 2	
	$G \rightarrow V$	$V \rightarrow G$	$G \rightarrow V$	$V \rightarrow G$
А	1	2	3	4
1	30,578	30,637	30,889	33,450
2	0,840	1,030	0,326	0,841
3	0,226	0,115	0,169	0,068
4	0,167	0,181	0,409	0,161
5	0,187	0,096	0,161	0,106
6	0,283	0,136	0,495	0,130
7	0,140	0,072	0,181	0,072
8	0,312	0,249	0,262	0,237
9	0,288	0,280	0,467	0,223
10	30,272	26,695	32,658	27,129

Из данных таблицы 5 хорошо видно, что между обоими решениями наблюдается явное сходство. Различие лишь в том, что по первому критерию ошибки в 1-м столбце в среднем меньше, чем в 3-м. Аналогично по второму критерию, средняя ошибка в 4-м столбце меньше, чем во 2-м. По второму критерию генеральные коэффициенты дифференциации ближе к выборочным значениям. Здесь символы  $G \rightarrow V$  и  $V \rightarrow G$  означают направление расчетов долей и средних, либо доли и средние генерального распределения в выборочных децильных интервалах, либо наоборот.

Полученные результаты еще раз показывают, что распределение душевых доходов в основной своей части близко к логарифмически нормаль-

ному закону, что подтверждает утверждение, что на доходы населения различные факторы оказывают мультипликативное воздействие. То есть воздействие на их рост и вариацию (тарифы, премии, проценты по вкладам, акциям) или падение (налоги, вычеты, потери от падения курсов акций и т. д.) носит процентный характер, поэтому и их суммарное влияние тоже носит процентный характер. Именно по этой причине основная часть доходов населения имеет логнормальное распределение. Этот закон, порождаемый множеством *небольших* мультипликативных воздействий, нарушается для особо бедных и особо богатых домохозяйств. Постоянный низкий или нерегулярный доход, с одной стороны, и чрезвычайно высокие проценты, премии и бонусы (которые нельзя назвать малыми) - с другой, нарушают логнормальность распределения доходов на концах. Однако с этим надо считаться. Фактически правильно согласованное логнормальное распределение показывает и измеряет недоучет числа богатых и бедных в составе выборки, а также корректирует среднее значение в этих децилях в большую сторону. Предложения Государственной Думы об ограничении размера «парашютов» для топ-менеджеров, а также предложение ограничить заработную плату руководителей корпораций с большим государственным участием лишь подчеркивают неэкономическую природу формирования доходов указанной категории работников.

К сожалению, многие показатели неравенства и бедности учитывают только доходы этих полярных групп населения, что приводит к серьезным спорам о глубине бедности, степени неравенства и поляризации. Любая недооценка или переоценка средних доходов на концах распределения существенно влияет на показатели бедности и неравенства.

Полезный результат, полученный в рамках предлагаемой методики, касается выбора и оценки границы бедности, которую можно использовать при построении генерального логнормального распределения. В качестве дополнительного условия (кроме среднего генерального душевого дохода), позволяющего оценить второй параметр распределения (дисперсию), может быть использовано равенство первого выборочного дециля (правая граница первого децильного интервала) и первого генерального дециля (теоретического).

Нам представляется, что этот путь практичнее, чем гадание на тему: «Какая доля населения живет ниже прожиточного минимума?». В нашем случае и выборочный, и теоретический уровни бедности совпадают и соответствуют 10% населения. При этом само значение первого дециля не зависит от методики расчета величины ПМ (прожиточного минимума), а зависит только от представительности выборочной совокупности домашних хозяйств, получаемой из результатов ОБДХ, проводимых Росстатом. Эти обследования являются регулярными и самыми представительными. Поэтому более полной и точной информации о социально-экономическом состоянии населения и результатах влияния распределительных процессов, включающих все источники доходов, на социально-демографическую структуру населения пока получить невозможно.

\* \*  
\*

Рассмотренные в статье подходы к реконструкции генерального распределения по доходам населения на основе выборки являются первым этапом при анализе распределительных отношений внутри социально-демографической структуры населения России. Следующие шаги требуют более тонкой балансировки и калибровки выборочных данных на уровне отдельных групп домохозяйств, чтобы получить детальную картину распределения всех источников дохода среди важнейших референтных групп населения: семьи с детьми, пенсионеры, малообеспеченные семьи и т. д.

Важность рассматриваемой в статье проблемы переноса выборочных оценок на уровень генеральной совокупности возрастает при переходе с общероссийского уровня на уровень региональный, когда количество наблюдений по отдельным регионам становится критическим. В этом случае отсутствие общей модели (типа) распределения приводит к результатам, которые не соответствуют или даже противоречат действительным наблюдениям, а наличие отказов усугубляет проблему измерения.

Значения, приведенные в таблицах 3, 4 и 5, показывают, что концепция возрастания доли отказов при росте доходов приводит к недооценке доходов в самой бедной децильной группе. Особенно это проявляется при расчетах показателей

бедности и неравенства для большинства регионов России. Понятно, что никакое взвешивание не сможет дать адекватной картины распределения доходов в регионе, если выборка состоит из 500 домохозяйств. Здесь чрезвычайно велика вероятность провалов в данных как среди бедных, так и среди богатых. Так, в Тюменской области «географическое» распределение доходов очень неравномерно, и отразить эту неравномерность посредством анализа 600 семей практически невозможно, если не придерживаться определенных теоретических положений, зарекомендовавших себя на более высоких уровнях (Россия и федеральные округа).

### Литература

1. Айвазян С.А., Калеников С.О. Качество жизни, уровень бедности и дифференциации по расходам населения России. Промежуточный отчет по гранту РПЭИ, декабрь, 1999.
2. Балансы доходов и потребления населения. Вопросы методологии и статистического анализа / под ред. А.Х. Карапетяна и Н.М. Римашевской. - М.: Статистика, 1969.
3. Колмаков И.Б. Методы и модели прогнозирования показателей дифференциации и поляризации денежных доходов населения: дисс... д-ра экон. наук. М., 2008.
4. Поманский А.Б. Анализ модели стимулирования и логарифмически нормальное распределение доходов // Экономика и математические методы. 1985. Т. XXI. Вып. 3.
5. Суворов А.В. Проблемы оценки дифференциации доходов населения в современной России // Проблемы прогнозирования. 2008. № 2.
6. Шевяков А.Ю., Кирута А.Я. Измерение экономического неравенства и бедности. - М.: Лето, 2002.
7. Шевяков А.Ю., Кирута А.Я. Неравенство, экономический рост и демография: неисследованные взаимосвязи. - М.: М-студия, 2009.
8. Шевяков А.Ю., Кирута А.Я. Экономическое неравенство, уровень жизни и бедность населения России и ее регионов в процессе реформ; методы измерения и анализ причинных зависимостей. Заключительный отчет по гранту РПЭИ. - М.: ЭПИКОН, 1999.
9. Banerjee A., Yakovenko V.M., Di Matteo T. A study of the personal income distribution in Australia // Physica A: statistical mechanics and its applications. 2006. Vol. 370.
10. Bilkova D. Lognormal distribution and using L-moment method for estimating its parameters // International journal of mathematical models and methods in applied sciences. 2012. Issue 1. Vol. 6.

11. **Clementi F., Gallegati M.** Power law tails in the Italian personal income distribution // *Physica A: statistical mechanics and its applications*. 2005. Vol. 350.
12. **Dagum C.** A systemic approach to the generation of income distribution models // *Journal of Income Distribution*. 1997. Vol. 6. No 1.
13. **Deville J.-C.** Generalized calibration and application to waiting for non-response // *COMPSTAT: Proceedings in Computational Statistics 14th Symposium held in Utrecht, The Netherlands, 2000*. - Heidelberg: Physica-Verl., 2000.
14. **Vanderhoeft C.** Generalised Calibration at Statistics Belgium: SPSS Module g-CALIB-S and Current Practices // *Statistics Belgium Working Paper*. 2001. No 3.
15. **Vanderhoeft C., Museux J.-M., Waeeyens E.** g-DESIGN and g-CALIB-S: SPSS modules for Generalized Calibration // *The Survey Statistician. IASS Newsletter*, 2000. No. 43.

#### METHODS FOR RESTORING THE PER-CAPITA INCOME DISTRIBUTION IN LARGE SAMPLES TO GENERALIZED POPULATION LEVELS\*

*Vladislav Zharomskiy*

*Author affiliation:* National Research University - Higher School of Economics (Moscow, Russia). E-mail: zharomskyvs@gmail.com.

*Alexey Rudberg*

*Author affiliation:* National Research University - Higher School of Economics (Moscow, Russia). E-mail: arudberg@hse.ru.

*Sergey Ter-Akopov*

*Author affiliation:* National Research University - Higher School of Economics (Moscow, Russia). E-mail: sterakopov@hse.ru.

This paper provides an analysis of popular methods for correcting sample distribution of income per-capita and proposes a methodology for evaluating the parameters of a lognormal income distribution, taking into account unequal response rates between individuals with different income levels, income deciles - a result of survey design and the survey non-response rate. The authors propose the fitting of a lognormal distribution on the basis of comparing mean and boundary income levels for defined population intervals between the sample and general distribution, instead of the more common approach of frequency analysis between the two. The mean income value of a given interval, with enough observations, is less volatile than the individual frequencies on the interval. This is especially important in situations where individual frequencies in the sample distribution significantly differ from the population distribution itself.

The authors examine two different criteria for estimating the optimal lognormal distribution parameters. The first method is similar to the methodology used in Russian statistics, and does not require preliminary information on the share of the poor population. The parameters are estimated using the condition of equality between the sample and population mean income, and the right-income boundary of the first income deciles. The second criterion is based on minimizing the squared sum of deviations between the mean income levels for the middle eight income deciles of the sample and population mean values. Neither of the two criteria uses the hypothesis of non-response rates increasing with households' income growth, which allows one to assess the representative-value of the sample survey.

The results of the calculations show that the method achieves the highest parity between sample and population distributions in the middle-part of the lognormal distribution, but suffers from underrepresentation in the lower part of the distribution, i.e. for poor households and individuals.

*Keywords:* lognormal distribution, parameter estimation, sample survey, non-response rate, mean income, general equilibrium, income deciles.

*JEL:* C83, D31.

#### Acknowledgements

The authors express their gratitude to the staff of the Federal State Statistics Service and to the referees of this article for their notes, which allowed the text to be more concise and clear.

\* The research was carried out as part of the HSE Program of Fundamental Studies.

#### References

1. **Ayvazyan S.A., Kalenikov S.O.** *Kachestvo zhizni, uroven' bednosti i differentsiatsii po raskhodam naseleniya Rossii*. Promezhutochnyy otchet po grantu RPEI, dekabr', 1999 [Quality of life, poverty, and the differentiation of expenditures of the Russian population. Interim report for the RFEI grant, December, 1999]. (In Russ.).
2. **Karapetyan A.Kh., Rimashevskaya N.M.** (eds.). *Balansy dokhodov i potrebleniya naseleniya. Voprosy metodologii i statisticheskogo analiza*. [The balance of income and consumption. Methodological issues and statistical analysis]. Moscow, Statistika Publ., 1969. (In Russ.).
3. **Kolmakov I.B.** *Metody i modeli prognozirovaniya pokazateley differentsiatsii i polyarizatsii denezhnykh dokhodov naseleniya*. Diss. dokt. ekon. nauk [Methods and models of forecasting of indicators of differentiation and polarization of incomes of the population: Dr. econ. sci. diss.]. Moscow, 2008. (In Russ.).

4. Pomanskiy A.B. Analiz modeli stimulirovaniya i logarifmicheski normal'noye raspredeleniye dokhodov [Analysis of the model of stimulation and lognormal distribution of income]. *Ekonomika i matematicheskiye metody*, 1985, vol. XXI (3). (In Russ.).
5. Suvorov A.V. Problemy otsenki differentsiatsii dokhodov naseleniya v sovremennoy Rossii [Problems of evaluation of income differentiation in modern Russia]. *Problemy prognozirovaniya*, 2008, no. 2. (In Russ.).
6. Shevyakov A.Yu., Kiruta A.Ya. *Izmereniye ekonomicheskogo neravenstva i bednosti* [The measurement of economic inequality and poverty]. Moscow, Leto Publ., 2002. (In Russ.).
7. Shevyakov A.Yu., Kiruta A.Ya. *Neravenstvo, ekonomicheskiy rost i demografiya: neissledovannyye vzaimosvyazi* [Inequality, economic growth and demography: the unexplored relationship]. Moscow, M-studiya Publ., 2009. (In Russ.).
8. Shevyakov A.Yu., Kiruta A.Ya. *Ekonomicheskoye neravenstvo, uroven' zhizni i bednost' naseleniya Rossii i yeye regionov v protsesse reform; metody izmereniya i analiz prichinnykh zavisimostey. Zaklyuchitel'nyy otchet po grantu RPEI* [Economic inequality, poverty and the standard of living of the population of Russia and its regions amid reforms; methods of measurement and analysis of causal relationships. The final report for the RFEI grant]. Moscow, EPIKON Publ., 1999. (In Russ.).
9. Banerjee A., Yakovenko V.M., Di Matteo T. A study of the personal income distribution in Australia. *Physica A: statistical mechanics and its applications*, 2006, vol. 370.
10. Bilkova D. Lognormal distribution and using L-moment method for estimating its parameters. *International journal of mathematical models and methods in applied sciences*, 2012, issue 1, vol. 6.
11. Clementi F., Gallegati M. Power law tails in the Italian personal income distribution. *Physica A: statistical mechanics and its applications*, 2005, vol. 350.
12. Dagum C. A systemic approach to the generation of income distribution models. *Journal of Income Distribution*, 1997, vol. 6, no 1.
13. Deville J.-C. Generalized calibration and application to waiting for non-response. COMPSTAT: Proceedings in Computational Statistics 14th Symposium held in Utrecht, The Netherlands, 2000. Heidelberg: Physica-Verl., 2000.
14. Vanderhoeft C. Generalised Calibration at Statistics Belgium: SPSS Module g-CALIB-S and Current Practices. *Statistics Belgium Working Paper*, 2001, no 3.
15. Vanderhoeft C., Museux J.-M., Waeytens E. g-DESIGN and g-CALIB-S: SPSS modules for Generalized Calibration. *The Survey Statistician*, IASS Newsletter, 2000, no. 43.

## РОССТАТ ПРЕДСТАВИЛ ИТОГИ ПЕРЕПИСИ НАСЕЛЕНИЯ КРЫМСКОГО ФЕДЕРАЛЬНОГО ОКРУГА

Во исполнение распоряжения Правительства Российской Федерации от 17 июля 2014 г. № 1330-р Федеральная служба государственной статистики с 14 по 25 октября 2014 г. провела федеральное статистическое наблюдение «Перепись населения в Крымском федеральном округе» со 100%-ным охватом населения. Росстатом подведены окончательные итоги переписи населения о численности и размещении населения, его демографических, социально-экономических, этно-лингвистических характеристиках, жилищных условиях, числе и составе домохозяйств и семейных ячеек. Доклад об основных итогах переписи населения в Крымском федеральном округе представлен в Правительство Российской Федерации и опубликован в свободном доступе на сайте Росстата.

**3 и 4 июня 2015 г. в г. Ялте состоялась конференция, на которой были представлены итоги переписи населения Крымского федерального округа и обсуждены перспективы их использования при реализации государственной политики Российской Федерации в Крымском федеральном округе.**

В работе конференции приняли участие представители органов власти, федеральных министерств и ведомств, научных организаций.

Перепись населения в Крымском федеральном округе проводилась в полном соответствии с методологическими и организационными положениями Всероссийской переписи населения 2010 г. и рекомендациями ЕЭК ООН по проведению переписей населения и жилищного фонда 2010 г.

Подробные итоги переписи населения в Крымском федеральном округе до конца 2015 г. будут опубликованы на официальном сайте Росстата [www.gks.ru](http://www.gks.ru) и сайтах территориальных органов Росстата по Республике Крым [crimea.gks.ru](http://crimea.gks.ru) и г. Севастополю [sevastopol.gks.ru](http://sevastopol.gks.ru).