

**ECONOMICS EDUCATION AND RESEARCH CONSORTIUM RUSSIA**

---

**РОССИЙСКАЯ ПРОГРАММА ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЙ**

С.А. Айвазян, С.О. Колеников

**УРОВЕНЬ БЕДНОСТИ И ДИФФЕРЕНЦИАЦИЯ  
ПО РАСХОДАМ НАСЕЛЕНИЯ РОССИИ**

ИТОГОВЫЙ ОТЧЕТ

Май 2000 г.

## СОДЕРЖАНИЕ

1. Аннотация .....	3
2. Введение .....	3
3. Обзор литературы .....	9
4. Модельная спецификация и результаты оценивания .....	14
4.1. Верификация базовых рабочих гипотез исследования .....	14
4.2. Основные переменные, используемые в исследовании, и его информационное обеспечение .....	18
4.3. Описание модели и содержательная интерпретация ее параметров.....	20
4.4. Методология эконометрического анализа модели.....	22
4.5. Результаты эконометрического анализа модели .....	29
5. Выводы .....	36
6. Библиография .....	38
<i>Приложение 1.</i> Результаты непараметрического анализа распределения населения регионов России и страны в целом по величине совокупных среднедушевых расходов .....	43
<i>Приложение 2.</i> Результаты статистического анализа моделей смесей распределений в рамках статистически наблюдаемого диапазона значений совокупных среднедушевых расходов домашних хозяйств .....	50
<i>Приложение 3.</i> Вероятность уклонения домашнего хозяйства от обследования как функция некоторых его характеристик (результаты анализа) .....	59

## УРОВЕНЬ БЕДНОСТИ И ДИФФЕРЕНЦИАЦИЯ ПО РАСХОДАМ НАСЕЛЕНИЯ РОССИИ

### 1. Аннотация

Задача измерения уровня бедности и экономического неравенства в современном российском обществе рассматривается в контексте общей проблемы снижения социальной напряженности путем рациональной организации социальной помощи малоимущим слоям населения. Соответствующие индикаторы уровня бедности (типа индекса Фостера-Гриира-Торбекка) предлагается строить на основании распределения населения *по расходам* (а не по доходам, как это обычно делается), что продиктовано спецификой российского переходного периода. Эта же специфика учитывается в предложенной эконометрической модели распределения населения по среднедушевым расходам, включающей в себя специальные методы *дооценки* (калибровки) распределения, построенного на базе официальной статистики бюджетов домашних хозяйств. Приводятся результаты экспериментальной апробации предложенной методологии эконометрического анализа, основанные на исходных статистических данных RLMS (5-й ~ 8-й раунды) и выборочных бюджетных обследований домашних хозяйств Республики Коми, а также Волгоградской и Омской областей (2-й квартал 1998 г.).

### 2. Введение

Различные измерители уровня бедности населения и показатели дифференциации населения по доходам и расходам относятся к *ключевым индикаторам качества проводимой социальной политики*, и в частности, используются при построении адресной социальной помощи малоимущим слоям населения, нацеленной на максимальное (в рамках имеющихся средств) снижение социальной напряженности в обществе.

Применяемые сегодня российскими официальными службами ([1], [2], [3]) и предлагаемые другими известными нам исследователями ([4], [5], [6]) показатели и способы их оценивания по выборочной бюджетной статистике домашних хозяйств (даже с учетом используемых шкал эквивалентности и калибровки исходных данных, ориентированной на

макропоказатели баланса доходов и расходов населения) имеют явные недостатки, приводящие к серьезному искажению реальных значений этих характеристик<sup>1)</sup>.

Мы видим несколько основных причин, объясняющих подобную ситуацию. Среди них:

- (i) неработоспособность используемой статистическими службами (Госком-статом) двухпараметрической модели логнормального распределения населения региона и всей России по среднему душевому доходу; при этом, главные искажения эта модель обнаруживает именно на «хвостах» распределения, по которым, собственно, и строятся оценки упомянутых характеристик;
- (ii) ограниченность используемой калибровки модели, при которой «подтягивание» анализируемого распределения производится в направлении известной социально-демографической структуры домашних хозяйств и заданной (из балансов доходов и расходов) величины среднего значения среднедушевых семейных доходов [3]; при этом считается, что общий вид модели плотности распределения и ее мода остаются неизменными;
- (iii) предложенные другими исследователями (напр. [4], [5]) методы аппроксимации анализируемого распределения и способы «взвешивания» (калибровки) исходных наблюдений также приводят к существенным искажениям реальной ситуации, т.к. не позволяют оценить удельный вес и структуру *ненаблюдаемого* спектра «богатых» и «очень богатых» домохозяйств (т.к. «взвешивание» придает веса уже имеющимся наблюдениям, но *не генерирует наблюдений из скрытой части диапазона*);
- (iv) в качестве индикатора уровня бедности обычно рассматривается только доля домашних хозяйств со среднедушевым доходом, не превосходящем уровня прожиточного минимума ([7] ~ [9]); однако *выбор индикатора уровня бедности (или критериев, по которому домашнее хозяйство следует относить к категории бедных) следует производить в зависимости от конечной прикладной цели экономического анализа*; в частности, для построения адресной политики социальной помощи — это одни критерии, для измерения уровня благосостояния — другие (см. [10], [11], [12]);

---

<sup>1)</sup> В ряде исследований (напр., [4], [6], [16]) показано, что такой показатель дифференциации доходов, как «коэффициент фондов», занижается при этом не менее чем в 2 раза, а оценки «доли домашних хозяйств со среднедушевым доходом, не превосходящем прожиточного минимума», полученные описанными в [1]~[6] и [9] способами, могут различаться между собой в 1,5~2 раза (подтверждение последнего факта явилось одним из результатов и данного исследования, см. ниже, п. 4.5).

- (v) в специфических условиях современной российской экономики определение индикаторов уровня бедности и критериев, по которым домашнее хозяйство следует относить к категории бедных, должно базироваться на величине среднедушевого **расхода** (*а не дохода*, как это принято в большинстве других исследований); в обоснование этого тезиса отметим, что при рассмотрении расходов вместо доходов:
- а) снимается проблема «учета-неучета» несвоевременно выплаченной части заработной платы членам домашнего хозяйства;
  - б) становятся несущественными вопросы, связанные с намеренно или непреднамеренно скрытой частью доходов, включая доходы, полученные из теневого сектора экономики;
  - в) правомерно расширяется спектр факторов, определяющих уровень благосостояния домашнего хозяйства, в первую очередь, за счет включения в этот спектр *личного подсобного хозяйства и имущественных компонентов* (недвижимости, личного транспорта, ювелирных изделий и т.п.), аренда или продажа которых может существенно поддерживать уровень благосостояния.
- (vi) в современной российской экономической теории и практике до сих пор не решалась (и даже не ставилась) *задача оптимального распределения денежных средств*, выделенных на социальную поддержку малоимущих слоев населения, где оптимальность понималась бы в смысле минимизации анализируемого индикатора уровня бедности (см. выше п. (iv)).

**Цели данного исследования** определяются стремлением участников проекта к преодолению недостатков, отмеченных в пп. (i)~(vi). В частности, речь идет о разработке методологии эконометрического анализа (на базе данных о семейных бюджетах) распределения населения России по величине среднедушевых расходов, о построении и статистической оценке на базе этого распределения основных характеристик уровня бедности и имущественной дифференциации населения (различных индикаторов уровня бедности, коэффициента фондов и т.п) и о постановке и теоретическом решении задачи оптимального распределения денежных средств, выделенных на адресную социальную поддержку беднейших слоев населения.

**Общие постановки решаемых в проекте задач** обусловлены сформулированными выше целями исследования. В агрегированном виде решаемые в проекте вопросы могут быть сформулированы в форме следующих двух задач.

**Задача 1.** Построить, обосновать теоретически и апробировать экспериментально интерпретируемую в содержательных терминах эконометрическую модель распределения населения региона России по величине среднедушевых совокупных денежных расходов, включая разработку методологии ее идентификации, основанной на официальных данных выборочных бюджетных обследований домашних хозяйств (БОДХ) и некоторых макропоказателях баланса доходов и расходов населения.

Решение этой задачи необходимо строить с учетом специфических условий современной российской экономики. В частности, в современных российских условиях существенно гиперболизируется (по сравнению с другими временами и странами) роль *трех основных источников искажений* закона распределения населения по среднедушевым доходам (расходам) при оценке этого закона по результатам выборочных бюджетных обследований домашних хозяйств (ДХ). Речь идет о следующих источниках:

( $\alpha$ ) сознательное уклонение ДХ от обследования («*truncation*»), нарушающее намеренный план выборочного обследования, а следовательно искажающее представительность выборки;

( $\beta$ ) намеренное искажение ответов респондентов («*misreporting*») с целью, например, сокрытия неофициальных источников получения дополнительных доходов;

( $\gamma$ ) принципиальная невозможность попадания в выборку представителей той категории домашних хозяйств, совокупные душевые доходы (расходы) которой превосходят некоторый заданный (достаточно высокий) уровень  $x_0$  («*censoring*»).

Конечно, построить такую эконометрическую модель распределения населения региона по величине среднедушевых доходов (или расходов), которая максимально бы элиминировала искажающее действие трех упомянутых факторов, без формулировки и, возможности, обоснования и статистической проверки ряда дополнительных рабочих гипотез невозможно. В нашем исследовании к таким гипотезам относятся:

- гипотеза  $H_1$  об общем виде исследуемого закона распределения вероятностей (з.р.в.);
- гипотеза  $H_2$  о форме зависимости вероятности «выпадения» домашнего хозяйства (ДХ) из сети официально статистически обследуемых ДХ от ряда его социально-экономических и территориальных характеристик, и в частности, от совокупных душевых расходов семьи;

- гипотеза  $H_3$  о постоянстве коэффициента вариации среднедушевых расходов (по отношению к изменению социально-экономической страты населения, для которой он подсчитывается);
- гипотеза  $H_4$  об общем виде з.р.в. среднедушевых расходов внутри *ненаблюдаемой* («супер-богатой») страты населения.

*Гипотеза  $H_1$*  базируется на характере трансформаций социально-экономической структуры российского общества (см. ниже п. 4.1), а ее статистическая проверка и использование позволят построить относительно лаконичную и интерпретируемую в содержательных терминах модель распределения населения региона России по величине среднедушевых совокупных расходов. Статистическая проверка и использование *гипотезы  $H_2$*  позволит снизить смещение оценки з.р.в. по доходам (расходам), обусловленное действием фактора ( $\alpha$ )-«truncation». *Гипотезы  $H_3$  и  $H_4$*  выполняют чисто техническую роль. Их использование позволит учесть в определенной мере искажающий эффект фактора ( $\gamma$ )-«censoring».

*Задача 2. Рассмотреть достаточно широкий класс индикаторов уровня бедности, основанных на распределении населения по величине **совокупных среднедушевых расходов**, и сформулировать проблему наилучшего распределения суммы  $S$ , выделенной на социальную поддержку малоимущих слоев населения, в виде специальной оптимизационной задачи, в которой в качестве минимизируемых критериев социальной напряженности рассматриваются индикаторы уровня бедности из упомянутого выше класса.*

В качестве класса индикаторов бедности рассмотрим семейство

$$I(w(x), f(x)) = \int_0^{z_0} w(x) f(x) dx, \quad (1)$$

где  $f(x)$  — функция плотности распределения населения по величине совокупных среднедушевых расходов,  $z_0$  — так называемая «черта бедности» (величина прожиточного минимума), а весовая функция  $w(x)$  — непрерывная, дифференцируемая, убывающая и выпуклая вниз на интервале  $[0, z_0)$  функция (эти ее свойства определяются естественным предположением, что при передаче любой суммы денег от бедных к менее бедным значение индикатора бедности (1) возрастет). Очевидно, семейство (1) включает в себя (при подходящем выборе весовой функции  $w(x)$ ) такие распространенные показатели как «де-

фицит бедности» («poverty gap»), «индекс Фостера-Гриира-Торбекка» («Foster-Greer-Thorbeck index»), «индикаторы класса Дальтона» («Dalton class indicators»), наконец, так называемые «меры с разрывной чертой бедности» («Poverty-Line-Discontinuous mesures» или «PLD mesures»), см. [13], [14], [15].

Пусть  $S$  (сумма, выделенная на адресную социальную поддержку бедного населения) меньше того количества денег, которое необходимо для полной ликвидации бедности. И пусть  $\varphi(x|S)$  — функция, задающая правило распределения суммы  $S$  среди населения с душевыми расходами  $x < z_0$  (например, это может быть функция плотности распределения суммы  $S$  среди бедного населения), а  $\tilde{f}(x|\varphi, S)$  — плотность распределения населения по совокупным среднедушевым расходам, получающаяся *после* реализации социальной помощи в соответствии с правилом  $\varphi(x|S)$ . С учетом этого изменится и значение индикатора уровня бедности вида (1), а именно:

$$I(w(x), \tilde{f}(x|\varphi; S)) = \int_0^{z_0} w(x) \tilde{f}(x|\varphi; S) dx. \quad (1')$$

Тогда задача 2 сводится к определению такой функции  $\varphi_0(x|S)$ , при которой индикатор уровня бедности (1') достигает своего минимума (при заданных  $w(x)$  и  $S$ ), т.е.:

$$\varphi_0(x|S) = \arg \min_{\varphi} \int_0^{z_0} w(x) \tilde{f}(x|\varphi; S) dx. \quad (2)$$

Необходимо подчеркнуть, что задача 2 рассматривается в данной работе **в контексте конкретного проекта борьбы с относительно длительной бедностью** (см. [8], [9]). Этим обусловлены следующие два обстоятельства. В о - п е р в ы х, в рамках категории *длительно* («перманентно») бедных не работает тезис об относительно высокой мобильности доходных групп (см. Богомолова Т.Ю., Тапилина В.С., Ростовцев П.С. Рост мобильности по доходам в изменении неравенства в распределении доходов. — Новый проект, РПЭИ, декабрь 1999). В о - в т о р ы х, главными инструментами ослабления *такой* бедности (alleviation of poverty) являются *различные формы прямых выплат дотаций нуждающимся*, а не меры стимулирования, побуждения (incentive) социальной и трудовой активности нуждающихся, хотя последние, безусловно, эффективнее применительно к контингенту *временно* бедных, например, из числа не востребовавшей части потенциального среднего класса.



### 3. Обзор литературы

Вначале обсудим известные нам работы, в которых ставились цели, наиболее близкие к целям задачи 1 (см. предыдущий пункт). Предлагаемая в данном проекте **модель распределения российского населения по величине совокупных среднедушевых расходов** является, по существу, развитием и модификацией модели распределения населения по *доходам*, впервые описанной в [16]. *Модификация* заключается во введении и статистической оценке «вероятности уклонения» домашнего хозяйства от обследования (см. выше гипотезу  $H_2$ ), в замене доходов *расходами* и в формулировке основанного на гипотезах  $H_2$ ,  $H_3$  и на знании усредненного значения макрохарактеристики среднедушевых расходов алгоритма калибровки имеющихся и генерирования (с помощью Монте-Карло-моделирования в соответствии с гипотезой  $H_4$ ) дополнительных наблюдений из *ненаблюдаемого* спектра душевых расходов.

Работы [4], [5], [6], [16] содержат различные доводы, подтверждающие справедливость наших критических замечаний (i) ~ (iv) (п. «Введение»). В работе [2] описан подход, также основанный на модели смеси логнормальных распределений; однако он не оснащен необходимым инструментарием, позволяющим проводить грамотный эконометрический анализ этой смеси, и не предлагает никаких способов учета скрытых от прямого наблюдения данных. Описанный в [5] подход, основанный на полиномиальной аппроксимации плотности анализируемого закона распределения, слишком формален и не позволяет построить феноменологическую модель изучаемого явления, придать содержательную интерпретацию параметрам модели, учесть ненаблюдаемый спектр расходов.

Основная слабость подхода, описанного в [6], — в явной неадекватности базового предположения о логнормальном виде закона распределения населения по доходу, хотя авторы и рассматривают *трехпараметрическую* модель (в отличие от *двухпараметрической* «госкомстатовской»). Однако, анализ доходов вместо расходов, неубедительность базового допущения о «правильности» определения *модального* уровня дохода по результатам выборочного БОДХ (которое по многочисленным доказательствам специалистов, в том числе — и «госкомстатовских», является *существенно смещенным*), а также формально-аппроксимационный метод подбора неизвестных параметров предлагаемой модели не позволяет сколько-нибудь серьезно отнестись к *модельной части* этой работы. В то же время предпосланный модельной части *экономический анализ* ситуации с доходами в России 90-х годов, по существу помогающий понять механизм формирования ненаблюдаемо-

го в БОДХ «правого хвоста» исследуемого распределения (как результата, в основном, широких возможностей определенных узких слоев населения в области продажи и эксплуатации элементов национального богатства страны), заслуживает, с нашей точки зрения, самого серьезного внимания.

*Несколько подробнее остановимся на критическом анализе подхода А.Ю. Шевякова и А.Я. Кируты (см [4]) и на его принципиальных отличиях от подхода, предлагаемого в нашем проекте. Это объясняется тем, что в этой работе предпринята наиболее серьезная, с нашей точки зрения, попытка описать реальное распределение населения региона по среднедушевому доходу, основываясь на данных БОДХ и «Баланса доходов и расходов населения». Эта попытка основана на непараметрическом подходе к оцениванию и включает в себя, в частности, некий прием устранения смещения, присущего данным БОДХ, а также описание процедуры агрегирования региональных данных с учетом региональных дефляторов и шкал эквивалентности. Отметим наиболее существенные, с нашей точки зрения, недостатки данного подхода.*

а) *Используемый в [4] способ взвешивания («калибровки») имеющихся наблюдений БОДХ, по существу, игнорирует все население, расположенное правее максимума наблюдаемых значений. Т.е. полностью игнорируется «правый хвост» распределения, и, следовательно, не учитывается фактор ( $\gamma$ ) цензурированности исходных статистических данных. В рамках нашей модели этот «хвост» восстанавливается с опорой на гипотезу  $H_4$ .*

б) Как прямое следствие упомянутого выше недостатка — принципиально ошибочный вывод авторов [4] о том, что «избыточное экономическое неравенство полностью обусловлено избыточной бедностью». При игнорировании правого хвоста распределения авторы и не могли прийти к другому выводу.

в) Привлекательная, на первый взгляд, «непараметричность» подхода, в действительности имеет два существенных недостатка. Во-первых, получаемая при этом оценка закона распределения населения по среднедушевому доходу является *чисто формальной аппроксимацией* анализируемого неизвестного закона и *не поддается содержательной интерпретации*. Во-вторых, полученная таким способом модель не пригодна для решения *прогнозных задач*.

г) В оценках уровня бедности, имущественной дифференциации и других показателей благосостояния в российских условиях переходного периода лучше апеллировать в

расходам, а не к доходам населения. Это снимает проблемы несвоевременных выплат заработной платы, скрытых доходов и т.п. (см. п. (v) во «Введении»).

Теперь обратимся к работам, наиболее тесно связанным с решением задачи 2. В этой связи, в первую очередь, следует отметить реализованные по инициативе и при финансовой поддержке Всемирного банка проект [8] и пилотные программы [9]. В них делается правомерная попытка измерения уровня бедности на базе *дооценки реальных душевых доходов* домашнего хозяйства (которые в [8] называются *потенциальными потребительскими расходами*). Однако, к сожалению, каждый из предложенных в этих работах конкретных способов такой дооценки имеет существенные недостатки (они подробно проанализированы С.А. Айвазяном в [17]). Кроме того, в упомянутых работах единственным критерием бедности остается доля бедного населения (т.е. критерий (1) при  $w(x) \equiv 1$ ) и не ставится задача определения *оптимального* способа распределения суммы, выделенной на социальную поддержку малоимущих семей (т.е. не решается задача (2)).

Весьма полный обзор индикаторов уровня бедности дается в [18]. В этой работе обсуждается, в частности, и один из частных случаев критерия (1) — так называемый индекс Фостера-Гриира-Торбекка, приведены результаты расчетов этого индекса на примере данных квартальной бюджетной статистики домашних хозяйств за 1996 г. Однако и в [18] этот индекс вычисляется на базе распределений по *доходу*, а главное, он не связывается с задачей оптимизации адресной социальной поддержки малоимущих слоев населения.

Таким образом, в *российской* экономической теории и практике задача 2, насколько нам известно, не ставилась и не решалась. Тем не менее, различные аспекты этой проблемы исследованы в ряде работ зарубежных авторов, которые, правда, тоже опираются в своих исследованиях на индикаторы бедности, вычисленные на базе распределений по **доходу** (см. [15], [19] ~ [24]). В частности, в [23] доказано, что в «доходных» вариантах индикаторов бедности (1) с весовой функцией  $w(x)$  вида

$$w(x) = \left( \frac{z_0 - x}{z_0} \right)^\alpha, \quad 0 \leq x < z_0, \quad \alpha > 1 \quad (3)$$

(что соответствует индексу Фостера-Гриира-Торбекка) оптимальной в смысле (2) является так называемая «*чистая стратегия подтягивания самых бедных к пороговому значению*  $\bar{z}_0 < z_0$ », где пороговое значение  $\bar{z}_0$  определяется из условия

$$N \left( \int_0^{\bar{z}_0} f(x) dx \right) \left( \int_0^{\bar{z}_0} (\bar{z}_0 - x) f(x) dx \right) = S \quad (4)$$

(здесь  $N$  — общая численность населения). Эта стратегия («allocation of p-type» в терминологии [15] и [23]) состоит в том, что после определения порогового значения  $\bar{z}_0$  каждый член общества со среднедушевым доходом  $x < \bar{z}_0$  получает пособие в размере  $\bar{z}_0 - x$ . В [23] доказано, что «смешанная стратегия» («allocation of mixed-type»), при которой часть суммы  $S$  идет на «подтягивание» к уровню  $\bar{z}_0$  самых бедных, а оставшаяся часть  $S$  — на «подтягивание» к уровню  $z_0$  самых богатых из бедных (при этом, конечно, в правой части соотношения (4), с помощью которого определяется величина  $\bar{z}_0$ , стоит некоторая сумма  $S_1 < S$ ), может быть оптимальной только при условии  $w(z_0) = \delta > 0$  (что соответствует в терминологии [15] индикаторам бедности с «разрывной чертой бедности» — «Poverty-Line-Discontinuous measures», см. выше).

Отметим, что фактор «цензурирования» исходных статистических данных, т.е. выпадение страты «супер-богатых» из общего эконометрического анализа распределения населения России по величине среднедушевых совокупных расходов (доходов), на первый взгляд, почти не повлияет на индикаторы бедности населения, представляющие главный интерес в решении проблемы адресной социальной поддержки малоимущих семей. Ведь в расчете индикаторов бедности типа (1) используется *только «левый хвост»* анализируемого распределения, в то время как эксплуатация гипотезы  $H_4$  может уточнить поведение этого распределения *на своем «правом хвосте»*.

Однако, учет страты супер-богатых приводит к существенному увеличению значений различных индикаторов имущественной дифференциации населения — коэффициента Джини, коэффициента фондов (т.е. отношения суммарных доходов 10% богатейшего населения к суммарным доходам 10% беднейшего населения) и т.п.<sup>2</sup> В свою очередь, характеристики дифференциации и *поляризации* населения по расходам (доходам) являются индикаторами уровня социальной напряженности в обществе, а потому от них существенно зависит и структура весовой функции  $w(x)$  в индикаторах бедности типа (1): ведь эти индикаторы в задаче снижения уровня длительной («постоянной») бедности (1)—(1')—(2) интерпретируются, прежде всего, как индикаторы *именно социальной напряженности в обществе*. Так, например, вычисление введенной в 1994 г. Жоан—Марией Эстебан и Дебреем Рэйем *меры поляризации* населения по доходу (см. [25]) существенно опирается на знание именно «хвостовых» элементов соответствующего распределения. В то же время в

ряде работ эта мера эффективно эксплуатируется (наряду, например, с коэффициентом Джини) в качестве *индикатора социальной напряженности* в обществе и фактора, находящегося в причинной зависимости с уровнем преступности (см., например, [26]). Именно этим объясняется то внимание, которое уделяется в данном проекте показателям дифференциации населения по расходам.

## 4. Модельная спецификация и результаты оценивания

### 4.1. Верификация базовых рабочих гипотез исследования

Решение сформулированной выше задачи 1 базируется на теоретическом обосновании и/или экспериментальной статистической проверке ряда рабочих гипотез.

- **Гипотеза  $H_1$**  о том, что распределение российских домашних хозяйств по величине среднедушевых совокупных денежных расходов на самом деле может быть адекватно описано смесью логарифмически-нормальных законов. Эта гипотеза поддается статистической проверке с помощью одного из критериев согласия (ее статистическая верификация на данных 1996 г. проведена в [16]).

В плане теоретическом эта гипотеза базируется на трех постулатах:

- распределение населения по величине среднедушевых расходов  $\xi$  **внутри однородной по структуре источников дохода, территориальному и социально-профессионально-демографическим признакам страте** подчинено логарифмически-нормальному закону с параметрами  $a = E(\ln \xi(a))$  и  $\sigma^2(a) = D(\ln \xi(a))$ .
- если представить все общество **состоящим из непрерывного (по средней величине логарифмов расходов  $a$ ) спектра таких страт, то при некотором естественном виде смешивающей (весовой) функции  $q(a)$**  распределение *всего населения* по величине среднедушевых расходов снова будет подчиняться логарифмически-нормальному закону;
- при нарушении непрерывности спектра составляющих общество различных страт (т.е. при существенном «вымывании» или элиминировании отдельных страт, например, так называемого «среднего класса») или при нарушении монотонного убывания смешивающей функции  $q(a)$  по мере удаления аргумента  $a$  от общего среднего из логариф-

---

<sup>2</sup> Расчеты, проведенные в [16] по данным 1995—96 гг., показали, что «доучет» частично уклонившихся от обследований «богатых» и полностью уклонившихся от обследований «супер-богатых» повышает коэффициент Джини с 0,376 до 0,531, а коэффициент фондов с 12,9 до 22,8.

мов расходов  $a_0$  упомянутое в п. (b) общее логрифмически-нормальное распределение трансформируется в смесь логнормальных законов.

Обсудим подробнее каждый из этих постулатов.

*Постулат (a)* является весьма распространенным в исследованиях, посвященных моделированию распределительных отношений в обществе, и основан на мультипликативном характере воздействия на расход (доход, заработную плату) факторов в рамках населения однородной социально-экономической страты. Механизм генерирования логарифмически-нормального распределения в подобных ситуациях подробно описан в литературе (применительно, правда, к заработной плате, он был нами описан еще в 1967 году, см. [27]).

*Постулат (b)* является простым следствием того, что при нормальном законе распределения различных страт по величине  $a = \mathbf{E}(\ln \xi)$  (т.е. при условии, что смешивающая функция  $q(a)$  описывается  $(a_0; \Delta^2)$  — нормальной плотностью) смесь нормальных распределений логарифмов доходов ( $\zeta = \ln \xi$ ) различных страт вида

$$\varphi(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \sigma(a)} e^{-\frac{(z-a)^2}{2\sigma^2(a)}} q(a) da$$

представляет собой (при  $\sigma^2(a) = \sigma^2 = const$ ) композицию нормальных законов и, следовательно, снова будет нормальным распределением с параметрами  $a_0 = \mathbf{E}(\ln \xi)$  и  $\sigma_0^2 = \sigma^2 + \Delta^2$ . Этот факт был впервые подмечен и объяснен также в [27].

*Постулат (c)* очевидно справедлив в вырожденной ситуации, когда смешивающая функция  $q(a)$  определена лишь в конечном наборе дискретных точек  $a_1, a_2, \dots, a_k$ . Реальная ситуация в современной российской экономике, конечно, сложнее. Однако, она бесспорно характеризуется существенной трансформацией смешивающей функции  $q(a)$ . Специфика переходного периода, не отменяя описанной в пп. (a) и (b) общей конструкции модели формирования распределения населения по среднемудушевым расходам, внесла свои коррективы в природу функции  $q(a)$ .

- **Гипотеза  $H_2$**  о том, что вероятность уклонения домашнего хозяйства от официальных бюджетных обследований является определенного вида функцией от ряда его социально-экономических и территориальных характеристик. Эта гипотеза также поддается статистической проверке (необходимые для этого статистические данные предпо-

лагается извлечь из результатов RLMS, см. [28], и некоторой дополнительной информации Госкомстата РФ). Гипотеза кажется естественной и подсказана в ходе обсуждения проблемы Е.Б. Фроловой, возглавляющей Управление статистики уровня жизни Госкомстата РФ. Очевидно, она основана на опыте работы статистиков-регистраторов БОДХ и RLMS. Конкретизация гипотезы  $H_2$  и способ ее эконометрического анализа описаны ниже в п. 4.4.

- **Гипотеза  $H_3$**  о постоянстве коэффициента вариации среднедушевых семейных расходов, т.е. о его независимости от номера социально-экономической страты, в рамках которой он подсчитывается. Эта гипотеза также поддается статистической проверке с помощью одного из критериев однородности дисперсий. В рамках теоретически и эмпирически доказанного положения о том, что население  $j$ -той однородной по географическим, социально-демографическим и профессиональным признакам страты распределено по среднедушевым доходным и расходным характеристикам  $\xi(j)$  в соответствии с логнормальным законом с параметрами  $a(j) = \mathbf{E}(\ln \xi(j))$  и  $\sigma^2(j) = \mathbf{D}(\ln \xi(j))$  (см., например, [29]), гипотеза  $H_3$  эквивалентна утверждению

$$H'_3: \mathbf{D}(\ln \xi(j)) = \sigma^2 = const$$

Эквивалентность гипотез  $H_3$  и  $H'_3$  следует из известного соотношения, справедливого в рамках логнормальной генеральной совокупности:

$$\frac{[\mathbf{D}(\xi(j))]^{\frac{1}{2}}}{\mathbf{E}\xi(j)} = \left( e^{\sigma^2} - 1 \right)^{\frac{1}{2}}.$$

- **Гипотеза  $H_4$**  о том, что распределение населения по величине среднедушевых совокупных денежных расходов  $x$  в статистически ненаблюдаемом диапазоне значений этих расходов (т.е. при  $x > \max_{1 \leq i \leq n} \{x_i\}$ , где  $x_i$  — величина среднедушевых совокупных денежных расходов в  $i$ -м статистически обследованном домашнем хозяйстве, а  $n$  — общее число статистически обследованных хозяйств) может быть аппроксимировано *трехпараметрическим логнормальным законом* с параметром сдвига, равным  $x_{(n)} = \max_{1 \leq i \leq n} \{x_i\}$ , и параметром  $\mathbf{D}(\ln \xi(k)) = \sigma^2$ , не зависящим от номера страты  $k$  и оцененным по наблюдениям, относящимся к статистически обследованным стратам населения (см. выше гипотезу  $H_3$ ).

Эта гипотеза не является статистической, т.к. она не может быть прямо (непосредственно) статистически верифицирована с помощью того или иного статистического критерия (необходимые для этого исходные статистические данные для нас недоступны). Поэтому гипотезу  $H_4$  следует считать некоторым *исходным модельным допущением*, а *априорная* справедливость которого может быть обоснована лишь с помощью подходящих теоретических соображений, а *апостериорная* — сопоставлением *реальных* значений основных «выходных» характеристик с соответствующими значениями, полученными на базе предлагаемой модели. В качестве упомянутых выше теоретических соображений можно привести здесь следующие факты и оценки специалистов.

К одному из значительных последствий стремительного развала СССР и его общественно-экономической системы следует отнести формирование (из рядов высшей партийно-бюрократической и хозяйственной элиты, дополненных наиболее «продвинутыми» представителями организованной преступности) того узкого слоя «избранных», который, используя специфические методы приватизации, получил возможность явной и тайной реализации (на мировом и внутреннем рынках) элементов присвоенного им национального богатства. Различные расчеты специалистов (см., например, [6]) показывают, что **«товарная интервенция» на рынках национального богатства в размере 0,2~0,3% его физического объема (в год) эквивалентна дополнительному увеличению валовых доходов населения на 10~20 %**. Очевидно, что подавляющая доля этого прироста валового дохода приходится как раз на упомянутый слой **«избранных», который в силу однородности социальных позиций и уровня властных полномочий может быть квалифицирован как специфическая социально-экономическая страта**. Поэтому закон распределения населения по расходам, о котором идет речь в гипотезе  $H_4$ , относится к населению именно этой страты.

Заметим, что в обычной ситуации для описания распределения той части населения по доходам (расходам), доходы (расходы) которой превышают некоторый фиксированный уровень  $x_0$ , используется *закон Парето*. Однако, это правомерно лишь в ситуациях, когда функция плотности распределения **всего** населения *монотонно убывает для всех  $x \geq x_0$*  (что, как правило, и имеет место при нормально функционирующей экономике). В нашем же случае описанная выше специфика формирования страты «супер-богатых» вполне допускает существование локального максимума функции плотности и правее точки  $x_0$ .



Использование гипотез  $H_1 \sim H_4$  позволит построить неформальную (т.е. интерпретируемую в содержательных терминах) модель распределения населения России по величине среднедушевых совокупных денежных расходов и разработать на базе этой модели методологию статистической оценки показателей уровня бедности и дифференциации, основанную на данных бюджетных обследований домашних хозяйств и некоторых макропоказателях баланса доходов и расходов населения.

#### 4.2. Основные переменные, используемые в исследовании, и его информационное обеспечение

1) *Совокупные (общие) среднедушевые денежные расходы ( $\xi$ ) в случайно «извлеченном» из генеральной совокупности домашних хозяйств или в  $i$ -м статистически обследованном домашнем хозяйстве ( $x_i$ ).*

Следуя [7], мы определяем (в привязке к выбранному такту времени — *кварталу*) совокупный денежный расход ДХ как сумму следующих слагаемых :

- $\xi^{(1)}$  — *потребительские квартальные расходы*, складывающиеся из расходов на продукты питания, алкоголь, непродовольственные товары для личного потребления членов ДХ и на получение услуг личного характера;
- $\xi^{(2)}$  — *расходы на промежуточное потребление* (затраты ДХ на ведение личного подсобного хозяйства );
- $\xi^{(3)}$  — *среднеквартальные валовые накопления основного капитала ДХ* (покупка земельных участков и недвижимости, изделий из драгоценных металлов, расходы на строительство и ремонт жилья) ;
- $\xi^{(4)}$  — *квартальная сумма всех оплаченных налогов и других обязательных платежей* (включая алименты, долги, клубные и общественные взносы) ;
- $\xi^{(5)}$  — *сумма остатков денег на руках и прироста организованных сбережений* (включая покупку валюты и ценных бумаг, банковские вклады и т.п.);
- $\xi^{(6)}$  — *стоимостная оценка потребленных за квартал продуктов, произведенных в личном подсобном хозяйстве.*

Таким образом,

$$\xi = \sum_{l=1}^6 \xi^{(l)},$$

где величины  $\xi^{(l)}$  для  $l = 1, 2, \dots, 6$  определены выше.

Наблюдаемые значения  $x_i, x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(6)}$  случайных величин, соответственно,  $\xi, \xi^{(1)}, \xi^{(2)}, \dots, \xi^{(6)}$  — это результаты статистического обследования  $i$ -того домашнего хозяйства, полученные с помощью стандартных БОДХ Госкомстата, (см. [7]) или из RLMS (см. [28]).

2) *Макрохарактеристики совокупных среднедушевых расходов населения региона* ( $\mu^{\text{макро}}$ ) полученная из ежеквартальных «Балансов доходов и расходов населения» Госкомстата РФ (см. [30]).

Величина  $\mu^{\text{макро}}$  имеет тот же смысл и ту же структуру, что и переменная  $\xi$ . Однако значения  $\mu^{\text{макро}}$  и ее составляющих подсчитываются не из выборочной статистики БОДХ, а из макро-региональных данных *торговой статистики, налоговых служб, информации банков и рынка ценных бумаг.*

3) *Относительная частота (доля)  $p(x)$  домашних хозяйств со среднедушевыми и валовыми расходами  $x$ , уклонившихся (или отказавшихся) от статистического обследования в течение наблюдаемого периода.* Источники информации : Госкомстат РФ и RLMS (раунды 5–8).

4) *Параметры социально-демографической структуры семей региона* (в среднем по региону : количество членов семьи, доля детей, доля пенсионеров и т.п.).

Остановимся несколько подробнее на данных RLMS и БОДХ, которые предполагается использовать в качестве информационного обеспечения нашего исследования.

(1) Данные RLMS, 5-й, 6-й, 7-й и 8-й раунды (см. [28]).

(2) Выборочные бюджетные обследования Госкомстата домашних хозяйств во 2-ом квартале 1998г. трех регионов РФ в ходе совместного проекта рабочей группы ЦЭМИ и Госкомстата. Согласно методике Госкомстата, выборка формируется по принципу представительности типов домашних хозяйств на основе микро-переписи 1994г. Квартальное обследование бюджета состоит из заполнения домохозяйством дважды в течение квартала двухнедельных дневниковых записей расходов и промежуточного месячного опроса. На основе собранных первичных данных о расходах Госкомстат выводит следующие агрегированные показатели, которые и будут использоваться в работе: денежные расходы ( $denras$ ; сумма фактических затрат, произведенных членами д/х в течение учетного периода; включают в себя потребительские расходы и расходы, не свя-

занные с потреблением); потребительские расходы (potras; часть денежных доходов, направленных на приобретение потребительских товаров и услуг); расходы на конечное потребление домохозяйств (konpot; потребительские расходы за вычетом стоимости продуктов питания, отданных д/х, в сумме с натуральными доходами д/х, т.е. сумма безналичных и натуральных поступлений продуктов питания, дотаций и льгот); располагаемые ресурсы д/х (rasres; сумма денежных средств, denres, и натуральных поступлений, natдох, которые находились в распоряжении д/х в течение учетного периода, т.е. денежные расходы, отложенные на конец учетного периода сбережения и натуральные доходы д/х). Бюджетные обследования были также дополнены специальным опросным листом, направленным на исследование качества жизни (см. [11]).

### 4.3. Описание модели и содержательная интерпретация ее параметров

Обозначим  $\xi$  (тыс.руб.) — среднегодовой расход случайно извлеченного представителя населения России и  $\xi_j$  (тыс.руб.) — среднедушевой расход индивидуума, случайно извлеченного из населения  $j$ -ой однородной социально-экономической страты. Тогда, в соответствии с гипотезами  $H_1$  и  $H_4$  плотность распределения случайной величины  $\xi$  будет описываться моделью смеси логнормальных законов вида

$$f(x|\Theta) = \sum_{j=1}^k q_j \frac{1}{\sqrt{2\pi} \sigma_j x} e^{-\frac{(\ln x - a_j)^2}{2\sigma_j^2}} + q_{k+1} \frac{1}{\sqrt{2\pi} \sigma_{k+1} \cdot (x - x_0)} e^{-\frac{(\ln(x-x_0) - a_{k+1})^2}{2\sigma_{k+1}^2}}, \quad (5)$$

где  $\Theta = (k; q_1, \dots, q_{k+1}; a_1, \dots, a_{k+1}; x_0; \sigma_1^2, \dots, \sigma_{k+1}^2)$  — параметры модели, имеющие следующую содержательную интерпретацию:

$k+1$  — число компонентов смеси (каждый из компонентов интерпретируется как однородная по социально-экономическим характеристикам страта населения);

$q_j$  ( $j = 1, 2, \dots, k+1$ ) — априорная вероятность появления наблюдений, представляющих  $j$ -ый компонент смеси (удельный вес  $j$ -ой однородной страты во всем населении региона);

$x_0$  — уровень цензурирования исходных данных, т.е. пороговое значение среднедушевых расходов, отделяющее статистически доступный диапазон изменения расходов ( $x \leq x_0$ ) от статистически недоступного диапазона ( $x > x_0$ );

$a_j = \mathbf{E}(\ln \xi_j)$  ( $j = 1, 2, \dots, k + 1$ ) — теоретические средние значения логарифмов среднедушевых расходов (усреднение произведено по всему населению  $j$ -ой страты);

$\sigma_j^2 = \mathbf{D}(\ln \xi_j)$  ( $j = 1, 2, \dots, k + 1$ ) — дисперсии логарифмов среднедушевых расходов, подсчитанные по населению  $j$ -ой страты;

Предполагается, что все население ( $k + 1$ )-ой (самой богатой) страты имеет среднедушевые расходы, превосходящие пороговое значение  $x_0$ , и *полностью уклоняется от выборочных бюджетных обследований домашних хозяйств*. Остальные домашние хозяйства доступны для статистического обследования, хотя и могут уклоняться (отказываться) от него с вероятностью  $p(x)$ , где  $p(x)$  — монотонно возрастающая по величине среднедушевого расхода  $x$  функция (см. выше гипотезу ( $H_2$ )).

Эконометрический анализ с целью идентификации модели (5) подразумевает, в частности, оценку параметров  $\Theta$  по данным БОДХ и БДРН.

#### 4.4. Методология эконометрического анализа модели

##### 4.4.1. Вероятность $p$ уклонения домашнего хозяйства от обследования как функция некоторых его характеристик

В качестве переменных, от которых зависит «вероятность уклонения»  $p$ , в данной работе рассматриваются три характеристики:

$z^{(1)} = \ln \xi$  — логарифм (натуральный) совокупных душевых расходов ДХ;

$z^{(2)}$  — характеристика места проживания ДХ (с градациями «город», «столичные регионы», «сельская местность» и «поселки городского типа»);

$z^{(3)}$  — уровень образования главы семьи (с градациями «ниже среднего», «среднее», «промышленно-технические училища», «техникумы» и «высшее»).

Зависимость «вероятности уклонения»  $p$  от  $Z = (1, z^{(1)}, z^{(2)}, z^{(3)})^T$  анализировалась в рамках *логит-модели* вида

$$p(Z) = P\{\eta_i = 0 | Z\} = \frac{1}{1 + e^{\beta^T Z}}, \quad (6)$$

где

$$\eta_i = \begin{cases} 0, & \text{если } i\text{-е ДХ уклонилось от обследования,} \\ 1 & \text{в противном случае,} \end{cases}$$

а  $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$  – вектор-столбец искомых (подлежащих статистической оценке) параметров модели (6). При этом, характеристики  $z^{(2)}$  и  $z^{(3)}$  вводятся в модель в форме *фиктивных переменных* (при гипотезе независимости значения параметра  $\beta_1$  от разных сочетаний градаций переменных  $z^{(2)}$  и  $z^{(3)}$ ), так что переменные  $z^{(2)}$  и  $z^{(3)}$  можно рассматривать как «сопутствующие», а полученная в результате эконометрического анализа модели (6) функция  $p(Z) = p(z^{(1)}, z^{(2)}, z^{(3)})$ , действительности, дает нам *целый набор моделей*, описывающих связь между «вероятностью уклонения от обследования»  $p$  и логарифмом совокупных душевых расходов  $z = z^{(1)}$  при различных сочетаниях градаций переменных  $z^{(2)}$  и  $z^{(3)}$ . Нам будет удобно обозначить элементы этого набора функций с помощью

$$p_{kl}(z) = p(z | z_k^{(2)}, z_l^{(3)}) = P\{\eta_i = 0 | z^{(1)} = z, z^{(2)} = z_k^{(2)}, z^{(3)} = z_l^{(3)}\}, \quad (6')$$

$$k = 1, 2, 3, 4; \quad l = 1, 2, 3, 4, 5$$

(очевидно, общее число таких функций составит 20).

В параграфе 4.5 приведены результаты оценивания параметров  $\beta$  функции (6) по данным RLMS (раунды 5~8). Расчеты подтвердили статистически значимую монотонно возрастающую зависимость вероятности  $p$  от  $z$  при любых сочетаниях градаций сопутствующих переменных  $z^{(2)}$  и  $z^{(3)}$ .

Там же приведены результаты эконометрического анализа упрощенного (пáрного) варианта модели (6), в котором исследуется зависимость вероятности  $p$  только от  $z = z^{(1)} = \ln \xi$ :

$$p(z) = P\{\eta_i = 0 | z^{(1)} = z\} = \frac{1}{1 + e^{\beta z}}. \quad (6'')$$

#### 4.4.2. Калибровка (взвешивание) имеющихся наблюдений

Анализ зависимостей (6) и (6'') может представлять автономный интерес. Однако в нашем исследовании функции (6') и (6'') используются в дальнейшем для калибровки имеющихся наблюдений и, соответственно, оцениваемого по бюджетным данным распре-

деления региона (страны) по величине совокупных душевых расходов: если исходные статистические данные содержат по каждому ( $i$ -му) обследованному домашнему хозяйству помимо значения его совокупных душевых расходов  $x_i$  еще и «значения» сопутствующих переменных  $z_{k_i}^{(2)}, z_{l_i}^{(3)}$ , то для калибровки используются функции (6') (здесь  $k_i$  и  $l_i$  — номера градаций, зарегистрированных, соответственно, по переменным  $z^{(2)}$  и  $z^{(3)}$  в  $i$ -м наблюдении); если же мы располагаем *только значением совокупных душевых расходов*, то приходится ограничиваться так называемой «грубой калибровкой», используя для этого функцию (6''). С учетом этого замечания в дальнейшем в целях упрощения обозначений мы будем обозначать функции (6') и (6'') с помощью  $p(z)$ , если речь идет о калибровке *прологарифмированных* наблюдаемых значений расходов, и с помощью  $p(x)$ , если речь идет о калибровке *исходных* наблюдений (измеряемых в тыс. руб.).

Пусть  $f(x)$  — функция плотности распределения населения региона по среднедушевым расходам. Тогда, если  $n$  — общее число статистически обследованных жителей региона и  $x^*$  — некоторое заданное значение среднедушевых расходов, то число наблюдений  $v(x^*)$ , попавших в малую (ширины  $\Delta$ ) окрестность точки  $x^*$  при условии, что никто не уклоняется от обследования, будет

$$v(x^*) \approx nf(x^*)\Delta . \quad (7)$$

*Реальное* (наблюдённое в выборке объема  $n$ ) число наблюдений  $\tilde{v}(x^*)$ , подсчитанное с учетом известных вероятностей уклонения от обследования  $p(x)$ , будет

$$\tilde{v}(x^*) \approx nf(x^*)(1 - p(x^*))\Delta . \quad (8)$$

Из (7) и (8) следует, что

$$v(x^*) = \tilde{v}(x^*) \cdot \frac{1}{1 - p(x^*)} . \quad (9)$$

В частности, выбирая в качестве точек  $x^*$  наблюдаемые в выборке значения среднедушевых расходов  $x_i$  ( $i = 1, 2, \dots, n$ ) и беря достаточно малые значения  $\Delta$ , мы будем иметь

$$\tilde{v}(x_i) = 1$$

$$v(x_i) = \frac{1}{1 - p(x_i)} .$$

А это значит, что если по имеющейся выборке

Наблюдённые значения $x$	$x_1$	$x_2$	$\dots$	$x_n$
Веса наблюдённых значений	$\frac{1}{n}$	$\frac{1}{n}$		$\frac{1}{n}$

(10)

мы хотим оценить *истинную* функцию плотности  $f(x)$ , то мы должны перейти к взвешенной (откалиброванной) выборке

Наблюдённые значения $x$	$x_1$	$x_2$	$\dots$	$x_n$
Веса наблюдённых значений	$\omega_1$	$\omega_2$	$\dots$	$\omega_n$

(11)

где веса  $\omega_i$  определяются по формуле

$$\omega_i = \frac{1}{\sum_{j=1}^n \left( \frac{1}{1-p(x_j)} \right)} \quad (11')$$

Заметим, что веса  $\omega_i$  тем больше, чем больше вероятность уклонения от обследования  $p(x_i)$ , причем  $\sum_{i=1}^n \omega_i = 1$ .

#### 4.4.3. Оценка параметров статистически наблюдаемых компонентов смеси

На данном этапе решается задача оценки параметров  $k, \tilde{q}, \dots, \tilde{q}_k, a_1, \dots, a_k, \sigma_1^2, \dots, \sigma_k^2$  в смеси логнормальных распределений вида

$$\tilde{f}(x) = \sum_{j=1}^k \tilde{q}_j \frac{1}{\sqrt{2\pi} \sigma_j x} e^{-\frac{(\ln x - a_j)^2}{2\sigma_j^2}} \quad (12)$$

по выборке (11).

Задача сводится к оценке тех же параметров в смеси *нормальных* распределений вида

$$\tilde{\varphi}(z) = \sum_{j=1}^k \tilde{q}_j \frac{1}{\sqrt{2\pi} \sigma_j} e^{-\frac{(z-a_j)^2}{2\sigma_j^2}} \quad (13)$$

по выборке

Наблюдённые значения $z$	$z_1$	$z_2$	$\dots$	$z_n$
Веса наблюдённых значений	$\omega_1$	$\omega_2$	$\dots$	$\omega_n$

(8')

где  $z_i = \ln x_i$  ( $i = 1, 2, \dots, n$ ).

Результаты решения данной задачи на исходных статистических данных 8-го раунда RLMS, а также — на данных бюджетных обследований домашних хозяйств Республики Коми, Волгоградской и Омской областей (за 2-й квартал 1998 г.) описаны в следующем параграфе.

Задача решена методами, описанными в [31]–[35], с помощью программного обеспечения, реализованного в пакетах «КЛАССМАСТЕР» и STATA (краткое описание использованных методов и алгоритмов см. в Приложении 3).

#### 4.4.4. Оценка ненаблюдаемого компонента смеси и всего распределения в целом

Пусть удельный вес ненаблюдаемого ( $\hat{k} + 1$ -го) компонента смеси равен  $q_{\hat{k}+1}$ , а среднее значение логарифмов среднедушевых расходов этой страты равно  $a_{\hat{k}+1}$ . Тогда общее среднее значение среднедушевых расходов  $\mu$  для *всего* населения региона, вычисленное с помощью модели (5) с учетом полученных на предыдущей стадии оценок  $\hat{k}; \hat{q}_1, \dots, \hat{q}_{\hat{k}}; \hat{a}_1, \dots, \hat{a}_{\hat{k}}; \hat{\sigma}_1^2, \dots, \hat{\sigma}_{\hat{k}}^2$ , определится формулой

$$\mu = \int_0^{\infty} x \left( \sum_{j=1}^{\hat{k}} \hat{q}_j \frac{1}{\sqrt{2\pi} \hat{\sigma}_j x} e^{-\frac{(\ln x - \hat{a}_j)^2}{2\hat{\sigma}_j^2}} + q_{\hat{k}+1} \frac{1}{\sqrt{2\pi} \sigma_{\hat{k}+1} (x - x_0)} e^{-\frac{(\ln(x-x_0) - a_{\hat{k}+1})^2}{2\sigma_{\hat{k}+1}^2}} \right) dx, \quad (14)$$

где



$$\hat{q}_j = \hat{\tilde{q}}_j(1 - q_{\hat{k}+1}), \quad j = 1, 2, \dots, \hat{k}. \quad (15)$$

Принимая во внимание свойства логнормального распределения, имеем :

$$\mu = \sum_{j=1}^{\hat{k}} \hat{q}_j e^{\frac{1}{2}\hat{\sigma}_j^2 + \hat{a}_j} + q_{\hat{k}+1} \left( x_0 + e^{\frac{1}{2}\hat{\sigma}_{\hat{k}+1}^2 + \hat{a}_{\hat{k}+1}} \right). \quad (14')$$

Величина  $\mu$ , определенная формулой (14'), зависит от неизвестных значений  $q_{\hat{k}+1}, a_{\hat{k}+1}$ , а также от  $x_0$  и  $\hat{\sigma}_{\hat{k}+1}^2$ . По построению величина  $x_0$  принимается равной максимальному наблюдаемому значению среднедушевых расходов, т.е.

$$x_0 = \max_{1 \leq i \leq n} \{x_i\} \quad (16)$$

В условиях справедливости гипотезы  $H'_3$  (см. выше п. 4.1) вычисляется общая оценка  $\hat{\sigma}^2$  величины  $\sigma^2$  по формуле :

$$\hat{\sigma}^2 = \sum_{j=1}^{\hat{k}} \hat{q}_j \hat{\sigma}_j^2 \quad (17)$$

и величина  $\hat{\sigma}_{\hat{k}+1}^2$  полагается равной  $\hat{\sigma}^2$ .

Затем в плоскости координат  $(q_{\hat{k}+1}, a_{\hat{k}+1})$  вычисляется линия уровня из условия

$$\mu(q_{\hat{k}+1}, a_{\hat{k}+1}) = \mu^{\text{макро}}, \quad (18)$$

где  $\mu(q_{\hat{k}+1}, a_{\hat{k}+1})$  подсчитывается по формуле (14') при  $x_0 = \max_{1 \leq i \leq n} \{x_i\}$  и  $\hat{\sigma}_{\hat{k}+1}^2 = \hat{\sigma}^2$ , а  $\mu^{\text{макро}}$  — среднее значение среднедушевых расходов, полученное из «Балансов доходов и расходов населения» для анализируемого региона и соответствующего такта времени.

Конкретный выбор точки  $(\hat{q}_{\hat{k}+1}, \hat{a}_{\hat{k}+1})$  на линии (18) требует дополнительных условий или экспертной информации.

При практическом построении линии уровня (18) полезно учесть следующие соображения:

а) из общих соображений очевидно, что

$$q_{\hat{k}+1} \ll \min_{1 \leq j \leq \hat{k}} \{q_j\}$$

(знак  $\ll$  переводится как «много меньше»; в нашем случае речь следует вести о том, что  $q_{\hat{k}+1}$  приблизительно на порядок меньше величины  $\min_{1 \leq j \leq \hat{k}} \{q_j\}$ ).

б) линия уровня (18) строится в виде таблицы, в которой «входом» является сетка значений  $q_{\hat{k}+1}$ , а «выходом» соответствующие значения  $a_{\hat{k}+1}$ , подсчитанные из условия (18) с учетом (14'). Дробность сетки значений  $q_{\hat{k}+1}$  целесообразно выбрать, например, следующим образом (при  $\min_{1 \leq j \leq k} \{q_j\} = m \cdot 10^{-2}$ ,  $1 \leq m \leq 9$ ):

$$q_{\hat{k}+1} = \begin{cases} v \cdot 10^{-2}, & v = m-1, m-2, \dots, 1; \\ v \cdot 10^{-3}, & v = 9, 8, \dots, 1; \\ v \cdot 10^{-4}, & v = 9, 8, \dots, 1. \end{cases}$$

в) отправляясь от (14'), можно получить следующую оценку сверху для величины удельного веса  $q_5$  ненаблюдаемой страты:

$$q_5 < \frac{1}{x_0} \left( \mu^{\text{макро}} - \sum_{j=1}^{\hat{k}} \tilde{q}_j e^{\hat{a}_j + \frac{\sigma_j^2}{2}} \right). \quad (19)$$

#### 4.4.5. Индикаторы уровня бедности в задаче адресной социальной поддержки малоимущих семей

Если ограничить класс весовых функций  $w(x)$ , участвующих в выражениях индикаторов бедности (1), функциями вида (3), воспользоваться результатами из [23] о виде оптимального распределения финансовой поддержки малоимущих семей в этом случае, а также результатами оценивания функции  $f(x)$  плотности распределения по совокупным среднедушевым расходам, то можно сформулировать следующее оптимальное правило адресной социальной поддержки малоимущих семей:

- (i) по заданным («входным») параметрам задачи — общей численности населения  $N$  анализируемого региона, черте бедности  $z_0$ , сумме  $S$ , выделенной на адресную социальную поддержку населения, функции плотности  $f(x)$ , описывающей распределение населения данного региона по совокупным душевым расходам, и значению  $\alpha > 1$ , конкретизирующему используемый в задаче индекс Фостера-Гриира-Торбекка

$$I_{\alpha}^{(z_0)}(f) = \int_0^{z_0} \left( \frac{z_0 - x}{z_0} \right)^{\alpha} f(x) dx, \text{ определяется пороговое значение } \bar{z}_0 \text{ из уравне-}$$

ния (4), т.е. из соотношения

$$N \cdot I_0^{(\bar{z}_0)}(f) \cdot \bar{z}_0 \cdot I_1^{(\bar{z}_0)}(f) = S; \quad (4')$$

- (ii) каждому жителю региона с совокупными среднедушевыми расходами  $x < \bar{z}_0$  предоставляется пособие в размере  $\bar{z}_0 - x$ .

Очевидно, изменение вида весовых функций  $w(x)$  может привести и к соответствующей модификации формулировки оптимального правила распределения адресной социальной помощи.

В данной работе подсчитаны и проанализированы доли бедных (т.е. индексы Фостера-Гриира-Торбека  $I_\alpha^{(z_0)}(f)$  при  $\alpha = 0$ ) и индикаторы социальной напряженности (т.е. индексы Фостера-Гриира-Торбека  $I_\alpha^{(z_0)}(f)$  для  $\alpha = 2$ ) для каждого из упомянутых выше регионов России тремя различными способами: 1) с помощью прямых непараметрических оценок; 2) с помощью оценок, основанных на логарифмически-нормальной модели распределения населения по совокупным душевым расходам; 3) с помощью оценок, основанных на модели смеси лог-нормальных распределений. Результаты этих подсчетов и основанного на них анализа приведены в параграфе 4.5.

#### 4.5. Результаты эконометрического анализа модели

Сформулированные во «Введении» задачи 1 и 2 и описанная в предыдущем параграфе методология данного исследования обусловили следующую **общую логическую схему эконометрического анализа**:

1) проводится *непараметрический анализ распределений* населения по величине совокупных среднедушевых расходов на базе данных Госкомстата России за II-й квартал 1998 г. по Республике Коми, Волгоградской и Омской областям (по каждому из регионов в отдельности), а также на базе данных RLMS (8 раунд, данные за ноябрь 1998 г.) по России в целом; на «выходе» этого этапа исследований — гистограммы и основные числовые характеристики соответствующих распределений (см. Приложение 1);

2) на базе панельных данных RLMS (5~8 раунды) об отказах домашних хозяйств от участия в обследовании<sup>3</sup> в рамках многомерной и парной логит-моделей оценены функции, соответственно,  $p_{kl}(x)$  (см. формулы (6) и (6')) и  $p(x)$  (см. формулу (6'')), описывающие *зависимость вероятности  $p$  уклонения домашнего хозяйства от обследования от величины  $x$  его совокупных среднедушевых расходов* при различных сочетаниях града-

<sup>3</sup> Эти данные получены от сотрудников Института социологии РАН П. Козыревой и Е. Артомоновой, за что авторы выражают им свою признательность.

ций переменных, определяющих характер местности проживания этого ДХ и уровень образования его главы семьи;

3) в соответствии с методологией, описанной в п. 4.4.2, производится *калибровка исходных статистических данных*: **грубая** (т.е. с использованием парной логит-модели (6'')) для региональных данных, и **более тонкая** (т.е. с использованием многомерной логит-модели (6)) — для общероссийских данных RLMS, 8-й раунд; цель — устранение смещения в оценке анализируемых з.р.в., вызываемого эффектом «*truncation*» (см. выше, «Введение»).

4) повторяется п. 1), т.е. снова проводится *непараметрический анализ исследуемых распределений*, но уже **на базе откалиброванных данных**; результаты сравниваются с результатами пункта 1);

5) в соответствии с методологией, описанной в п. 4.4.3, производится *оценка параметров в моделях смеси распределений* (трех региональных и общероссийской) по откалиброванным данным в рамках статистически наблюдаемого диапазона значений совокупных среднедушевых расходов домашних хозяйств;

6) в соответствии с методологией, описанной в п. 4.4.4, производится *оценка ненаблюдаемого компонента смеси распределений* в каждой из четырех анализируемых моделей (в трех региональных и в общероссийской); цель — устранение смещения в оценке анализируемых з.р.в., вызываемого эффектом «*цензурирования*» исходных данных («*censuring*», см. «Введение»);

7) на базе распределений, полученных в пп. 4) и 6), вычисляются и анализируются *индикаторы уровня бедности и социальной напряженности* для трех рассматриваемых регионов (по состоянию на II-й квартал 1998 г.) и для России в целом (по состоянию на ноябрь 1998 г.).

#### **4.5.1. Статистический анализ и калибровка распределений совокупных среднедушевых расходов населения**

Представленные на рис. П.1~П.4 и в таблицах П.1~П.8 Приложений 1 и 2 результаты этого статистического анализа свидетельствуют о следующем.

1) *Распределение населения по величине среднедушевых расходов ни внутри региона, ни в масштабах всей страны не может быть описано ни логарифмически-нормальным законом, ни двухкомпонентной смесью логарифмически-нормальных законов.*

Действительно, из содержания 5-х и 6-х столбцов таблиц П.5~П.8 следует, что гипотеза о логарифмически-нормальном з.р.в. должна быть отклонена во всех случаях с ошибкой 1-го рода, имеющей не менее трех нулей перед первой значащей цифрой, а гипотеза подчиненности анализирующих распределений з.р.в., описываемому *двухкомпонентной* смесью лог-нормальных законов, отклоняется с уровнем значимости, не превосходящем (за исключением Омской области) 0,085. Видимая непротиворечивость этой гипотезы данным Омской области, однако, не может служить достаточным обоснованием для ее принятия: все критерии качества идентификации модели (значение логарифмической функции правдоподобия, критерии Акаике и Боздогана, наконец, значение статистики  $\chi^2$  соответствующего критерия согласия) оказываются существенно лучшими для модели *трехкомпонентной* смеси и в случае Омской области.

2) *Запрограммированные в пакетах «Классмастер» и STATA алгоритмы статистического оценивания параметров в модели смеси нормальных распределений как правило выводят на оценки числа компонентов смеси  $\hat{k} = 3$  или  $\hat{k} = 4$ , т.е. сигнализируют о том, что население региона и страны в целом может быть представлено (в статистически обследованном диапазоне значений душевых расходов) как смеси трех или четырех социально-экономических страт, хотя это не обязательно выражается в виде, соответственно, трех или четырех локальных максимумов соответствующей функции плотности распределения.*

Волгоградская область оказалась единственным исключением из этого правила. В то время как по всем остальным случаям (т.е. по России в целом, по Республике Коми и по Омской области) увеличение числа компонентов смеси, начиная с  $k = 5$ , приводит к резкому ухудшению значений всех рассматриваемых критериев качества идентификации модели, анализ данных по Волгоградской области дал оценочное значение числа компонентов смеси  $\hat{k} = 5$  как наилучшее.

3) *В сравнении с аналогичной картиной 1996 года (см. [16]) расслоение населения на страты по расходам в 1998 году гораздо менее четко выражено, что соответствует упомянутой выше тенденции к постепенному возвращению анализируемого распределения к лог-нормальному виду по мере выхода экономики из полосы переходного периода.*

Графики из [16], описывающие з.р.в. населения России по душевым доходам по состоянию на осень 1996 г., демонстрируют наличие локальных максимумов плотности распределения. Основанные на этой модели процедуры классификации населения позволили

провести определенную его стратификацию и социально-экономический анализ каждой из выявленных страт в отдельности. Выявление страт и проведение аналогичного социально-экономического анализа *по данным 1998 года* оказалось гораздо более проблематичным, требует привлечения дополнительной информации, и потому мы были вынуждены вынести эту работу за рамки данного проекта.

4) В совокупности использованные в работе метод *калибровки* наблюдений (как средство нивелирования эффекта «truncation») и «нестатистический» прием «*досчета*» *ненаблюдаемого компонента смеси* (как средство нивелирования эффекта «censoring») позволяют объяснить существенную часть приблизительно сорокапроцентной разницы, существующей между *реальными* и *официальными* (зарегистрированными государственными органами статистики) доходами (расходами) населения.

Оценка ненаблюдаемого (латентного) компонента смеси строится с точностью до «*линейного уровня*» параметров  $q_{\hat{k}+1}$  и  $a_{\hat{k}+1}$  при определенных ограничениях на диапазоны изменения этих параметров (см. формулы (14)~(18) в п. 4.4.4). Пример такой линии уровня (подстроенной по данным RLMS, 8-й раунд) приведен на рис. 1.

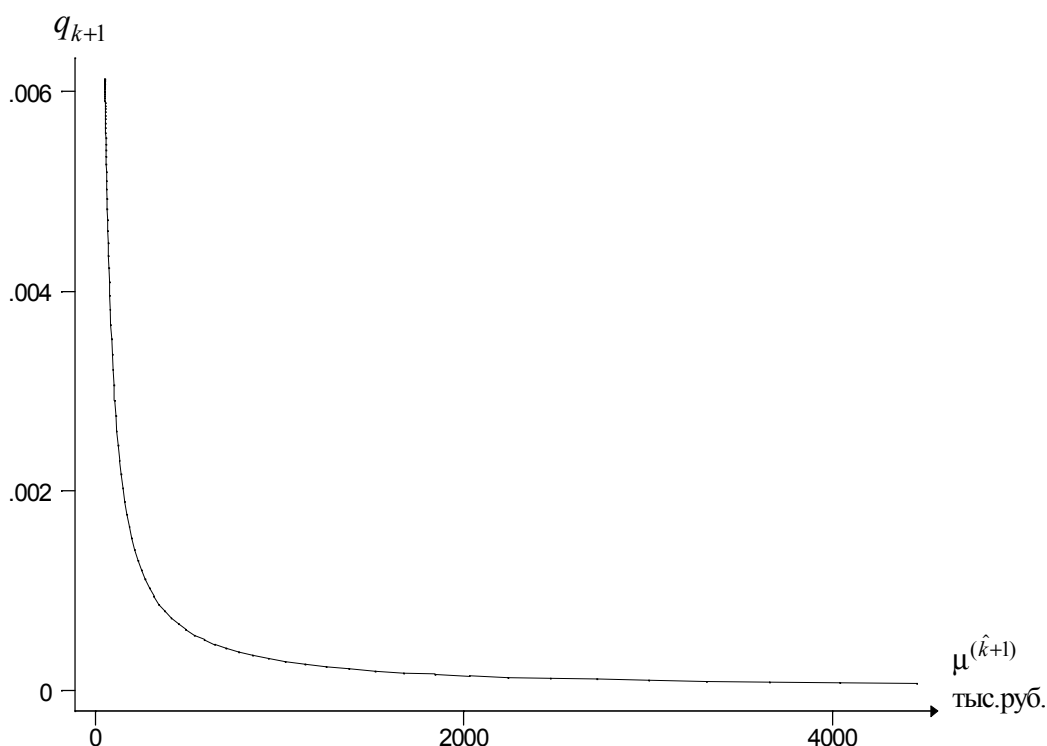


Рис. 1. Связь между удельным весом и средней величиной душевых расходов, используемая при определении параметров в латентной страты населения.

Оказывается, однако, что с точки зрения интересующих как характеристик распределения (индикаторов бедности, среднего значения, характеристик дифференциации и поляризации населения по совокупным душевым расходам) конкретизация выбора определенной точки  $(\hat{\mu}^{(\hat{k}+1)}, \hat{q}_{\hat{k}+1})$  на этой линии уровня практически не имеет значения. Индикаторы бедности зависят, по существу, только от «левого хвоста» распределения, а характеристики дифференциации и поляризации, конечно, существенно возрастут при включении в модель латентной страты, однако их значения практически не зависят от того, в каком именно месте мы выберем точку  $(\hat{\mu}^{(\hat{k}+1)}, \hat{q}_{\hat{k}+1})$  на линии уровня. Заметим лишь, что по оценкам различных исследований доля скрытых доходов (расходов) составляет от 25 до 40% (см., например, [16]). В предложенной в данной работе методологии анализа калибровка наблюдений повышает официальные расходы на 2~3%, а введение в модель латентной страты — еще на 20~30% (по данным RLMS, 8-й раунд это приращение составило  $(1211-913)/913=0,326=32,6\%$ ).

#### 4.5.2. Оценки индикаторов бедности и социальной напряженности

В табл. 1 приведены значения индикаторов бедности и социальной напряженности. В терминах семейства  $I_{\alpha}^{(z_0)}(f)$  индексов Фостера-Гриира-Торбека (см. пп. «Введение» п.4.4.2) это индексы  $I_0^{(z_0)}(f)$  — доля бедного населения и  $I_z^{(z_0)}(f)$  — индикатор социального напряжения, обусловленного наличием слоя бедного населения. В таблицу включены официальные данные Госкомстата РФ (столбец 4), данные, заимствованные из пилотных программ [9] (столбец 5, только для двух регионов, включенных в эти пилотные программы), прямые непараметрические оценки этих показателей, подсчитанные по откалиброванной выше исходной информации RLMS и трех анализируемых регионов (столбцы 8 и 9), а также оценки, основанные на логарифмически-нормальной модели (столбцы 6 и 10), и на модели смеси логарифмически-нормальных распределений (столбцы 7 и 11).

Таблица 1\*. Индикаторы бедности и социальной напряженности (Россия — по состоянию на ноябрь 1998г., регионы – по состоянию на II-й квартал 1998г.)

№ региона	Название региона	Черта бедности (тыс.руб.)	Доля бедных					Индексы бедности (социальной напряженности) Фостера-Гриира-Горбека		
			Официальные данные Госкомстата	Оценка пилотных программ	Оценка по лог-норм. модели	Оценка, основ. на модели смеси	Прямая непарам. оценка	Прямая непарам. оценка	Оценка по лог-норм. модели	Оценка, основ. на модели смеси
1	Россия	0,636	28,4	—	52,5	52,8	53,9	0,137	0,139	0,130
2	Республика Коми	0,466	20,6	26,7	53,8	56,2	56,7	0,127	0,130	0,128
3	Волгоградская область	0,368	31,5	49,2	62,0	62,7	63,0	0,177	0,177	0,177
4	Омская область	0,372	25,2	—	42,6	43,2	44,2	0,082	0,130	0,081

\*) Данные столбцов 3 и 4 заимствованы из [37]~[39].



Анализ данных табл. 1 приводит к следующим выводам.

1) Имеет место существенный разброс в значениях анализируемых показателей как между регионами, так и (для каждого фиксированного региона) между способами оценивания. С нашей точки зрения наиболее точным способом вычисления значений этих характеристик является метод прямой непараметрической оценки, основанной на откалиброванных данных (столбцы 8 и 9). И хотя он практически не изменяет рейтинги регионов по этим показателям, но дает существенно более высокие значения показателей бедности (по Республике Коми и Волгоградской области — в два раза!), чем официальная статистика. В то же время оценки, полученные на основании моделей смеси, дают почти во всех случаях результаты, существенно более близкие к прямым непараметрическим оценкам, чем официальная статистика.

2) Хотя удельный вес включенной в модель «ненаблюдаемой» страты супербогатых относительно невелик (в большинстве случаев он измеряется десятками и даже сотыми долями процента), влияние этой страты на значения ключевых характеристик дифференциации и поляризации населения весьма существенно. Так, например, коэффициент Джини по официальной статистике составлял по стране в ноябре 1998г. Значение равное 0,372, по прямой непараметрической оценке откалиброванных данных RLMS (8-й раунд) он был равен 0,488, а при учете латентной страты его значение возрастает до 0,610! Примерно в такой же пропорции возрастает коэффициент фондов.

## 5. Выводы

1) Специфика переходного периода в России, не отменяя общей логарифмически-нормальной конструкции модели формирования распределения населения по среднедушевым доходам (расходам), внесла свои коррективы в природу «смешивающей функции»  $q(a)$ . Феномен появления *дискретной* смеси лог-нормальных распределений (вместо действующей в стационарном режиме стабильно развивающейся экономики специального вида *непрерывной* смеси, которая снова дает лог-нормальное распределение) объясняется тем, что в переходный период происходят изменения структуры спроса на труд и человеческий капитал. В силу этих изменений работники невостребованных более профессий вынуждены переключаться на другие, как правило, менее доходные источники заработка. В частности, подобные изменения коснулись «советского среднего класса» квалифицированных работников. В сочетании с низкой мобильностью, характерной для российского рынка труда, эти структурные перестройки обусловили «вымывание» отдельных групп ра-

ботников. В то же время, присвоение значительных потоков рент обусловило появление совершенно новых «очень богатых» групп населения. Таким образом, в обществе выделился ряд достаточно четко обозначенных (на доходной шкале) групп, что привело к дискретности смеси распределений (лог-нормальных внутри отдельных групп). Отсюда — естественность попытки аппроксимации искомого распределения с помощью дискретной смеси лог-нормальных законов. Заметим, что в ходе приближения временного переходного периода российской экономики к нормальному стационарному состоянию будет возвращаться к нормальному виду и функция  $q(a)$ , а, следовательно, и искомое распределение населения по расходам будет становиться все больше похожим на обычное двухпараметрическое логнормальное распределение. Проведенные в рамках данного проекта предварительные расчеты и сравнение результатов, относящихся к 1996 г. и к 1998 г., подтверждают эту тенденцию.

2) Статистический анализ предложенной в проекте модели предусматривает: а) оценку параметров  $\Theta = (k; q_1, q_2, \dots, q_k; a_1, a_2, \dots, a_k; \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2)$  конечной дискретной смеси распределений в наблюдаемом диапазоне изменения среднедушевых расходов с помощью соответствующих статистических процедур (см. [31] ~ [35]); б) специальную калибровку полученного распределения с учетом введенной в модель вероятности отказа от обследования как функции от величины среднедушевых расходов; в) дооценку (восстановление) ненаблюдаемого  $(\hat{k} + 1)$ -го компонента анализируемой смеси распределений с помощью повторной калибровки модели, построенной на «подтягивании» модельного среднего значения к экзогенно заданной (из макростатистики) величине средних душевых расходов и основанной на частично верифицируемых рабочих гипотезах  $H_3$  и  $H_4$ .

3) Предложенная структура модели распределения населения по совокупным среднедушевым расходам, включающая процедуру двукратной калибровки, которая позволяет учитывать уклонившиеся от обследования домашние хозяйства и настраивать модель на соответствующие макропоказатели совокупных расходов, обеспечивает более точную оценку основных характеристик бедности и имущественной дифференциации населения по сравнению с методами, используемыми в экономической практике ([1] ~ [3], [7], [28]) и другими исследователями ([4] ~ [6]). В частности, она может быть использована при расчетах, связанных с организацией адресной социальной поддержки малоимущим слоям населения.

4) Основанные на предлагаемой модели методики следует реализовывать на *региональном уровне*. Лишь после «приведения регионов к общему знаменателю» с помощью соответствующих дефляторов и коэффициентов, учитывающих региональные различия в покупательной способности рубля, в составе основной потребительской корзины, в расчете прожиточного минимума и т.п., результаты по регионам можно агрегировать с целью сведения их к общероссийскому итогу.

## 6. Библиография

- [1] К методике оценки распределения генеральной совокупности населения по величине среднедушевого дохода на основании эмпирических данных. — Материалы Госкомстата РФ (тезисы к докладу на Ученом совете в ЦУЖе), 1999.
- [2] *Великанова Т., Колмаков И., Фролова Е.* Совершенствование методики и моделей распределения населения по среднедушевому доходу. — «*Вопросы статистики*», 1996, № 5.
- [3] *Великанова Т.Б., Фролова Е.Б.* Методология прямой оценки величины уровня бедности, основанная на распространении результатов выборочного обследования на генеральную совокупность. — Материалы Управления статистики уровня жизни населения Госкомстата РФ, 1999.
- [4] *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. — Финальный отчет по проекту EERC, июль 1999.
- [5] *Ершов Э.Б., Майер В.Ф.* Методологические и методические проблемы определения уровня, объема и дифференциации доходов населения. — Материалы к заседанию ВЦУЖ 28 декабря 1998 г.
- [6] *Суворов А.В., Ульянова Е.А.* Денежные доходы населения России: 1992-1996 гг. — «Проблемы прогнозирования», 1997 г.
- [7] Определение основных показателей агрегирования данных обследования бюджетов домашних хозяйств. — Материалы Госкомстата РФ, 1999.
- [8] *Брейтуэйт Дж.* Адресность и относительно длительная бедность в России. Февраль 1999. Доклад на семинаре Всемирного банка 19 апреля 1999 г.

- [9] Пилотные программы по введению адресной социальной поддержки малоимущих семей в Республике Коми, Воронежской и Волгоградской областях. Предварительные итоги. Министерство труда и социального развития МФ. — М.: 1999. — 104 с.
- [10] *Айвазян С.А.* Интегральные индикаторы качества жизни населения: их построение и использование в социально-экономическом управлении и межрегиональных сопоставлениях. — Москва: ЦЭМИ РАН, 2000. — 117 с.
- [11] *Айвазян С.А., Герасимова И.А.* Социальная структура и социальное расслоение населения Российской Федерации (по материалам выборочного обследования населения трех регионов РФ). — М.: ЦЭМИ РАН, 1998.
- [12] *Колеников С.О.* Методы анализа качества жизни. Серия РЭШ «Лучшие студенческие работы», 1999 (в печати).
- [13] *Hagenaars A.* A Class of Poverty Indices. — «International Economic Review», 28, 1987, pp. 583—607.
- [14] *Foster J., Greer J., Thorbeck E.* A Class of Decomposable Poverty Measures. — «Econometrica», 52(3), 1984, pp. 761—766.
- [15] *Bourguignon F., Fields G.* Discontinuous loss from poverty, generalized  $P_\alpha$  measures, and optimal transfers to the poor. — XI-th World Congress of the International Economic Association. Tunis, December 1995.
- [16] *Айвазян С.А.* Модель формирования распределения населения России по величине среднедушевого дохода. — «Экономика и математические методы», том 33 (1997), № 4, с. 74-86.
- [17] Национальная оценка и распространение информации. Предварительный отчет в рамках проекта Всемирного банка по теме «Структурная перестройка системы социальной защиты населения» №SPIL-2.2.2/11. ГУ-ВШЭ, Москва, 1999.
- [18] *Корчагина И., Овчарова Л., Турунцев Е.* Система индикаторов уровня бедности в переходный период в России. — Научный доклад №98/04 по секции «Микроэкономика-2 (поведение домохозяйств)» Российской программы Консорциума экономических исследований. РПЭИ/Фонд Евразия, апрель 1999.
- [19] *Sen A.K.* A Sociological Approach to Measurement of Poverty. — Oxford Economic Papers, 37, pp. 669—667.
- [20] *Atkinson A.B.* On the Measurement of Poverty. — Econometrica, vol.55 (1987), N4, pp.749-764.

- [21] *Kanbur S.M.R.* Measurement and Alleviation of Poverty. — «IMF Staff Papers», 34, 1987, pp. 60—85.
- [22] *Foster J.E., Shorrocks A.F.* Poverty Orderings. — «Econometrica», 56(1), 1988, pp. 173—177.
- [23] *Bourguignon F., Fields G.S.* Poverty Measures and Anti-Poverty Policy. — «Recherches Economiques de Louvain», 56(3-4), 1990, pp. 409—427.
- [24] *Ravallion M.* Poverty Comparisons. — Chur, Switzerland, Harwood Academic Publishers, 1994.
- [25] *Esteban J.-M., Ray D.* On the Measurement of Polarization. — *Econometrica*, 62, No.4, pp. 819—851.
- [26] *Fajnzulber P., Lederman D., Loayza N.* Inequality and Violent Crime. — The research project «Crime in Latin America» of the World Bank, 1999.
- [27] *Айвазян С.А., Рабкина Н.Е., Римашиевская Н.М.* Методика расчета ожидаемого распределения рабочих и служащих по размерам заработной платы. — М.: НИИ труда ГКМСССР по вопросам труда и заработной платы, 1967.
- [28] *Mroz T., Popkin B., Mancini D., Glinskaya T., Lokshin V.* Monitoring Economic Conditions in the Russian Federation: The Russia Longitudinal Monitoring Survey 1992-1996. — Report submitted to the U.S. Agency for International Development. Carolina population Center. University of North Caroline at Chapel Hill. February, 1997.
- [29] *Aivazian S.A.* Probabilistic-Statistical Modelling of the Distributary Relations in Society. — In: «Private and Enlarged Consumption», North-Holland Publ. Comp., 1976.
- [30] Методологические положения по статистике. Выпуск 1. — М., Госкомстат РФ, 1996.
- [31] *Day N.E.* Estimating the Components of a Mixture of normal distributions. — *Biometrika*, vol. 56 (1969), N3, pp. 463—474.
- [32] *Dempster A., Laird G., Rubin J.* Maximum Likelihood from Incomplete Data via the EM-algorithm. — *J.R. Statist. Soc.*, B.39 (1977), pp. 1—38.
- [33] *Aivazian S.A.* Mixture-Model Cluster Analysis Using the Projection Pursuit Method. — In: «Computational Learning and Probabilistic Reasoning», John Wiley and Sons Ltd, 1996, pp. 278-286.
- [34] *Rudzkis R., Radavicius M.* Statistical Estimation of a Mixture of Gaussian Distributions. — In: «Acta applicandae Mathematicae», vol. 38, N1, 1995.

- [35] *Jakimauskas G., Sushinkas J.* Computational aspects of statistical analysis of gaussian mixture combining EM algorithm with non-parametric estimation (one-dimensional case). — Preprintas N96-6, Matematikos ir Informaticos Institutas, Vilnius, Lietuva, 1996.
- [36] The Russia Longitudinal Monitoring Survey: «Family questionnaire» and «Sample of Russian Federation». Rounds V and VI. Technical Report. August-October 1996.
- [37] Социально-экономическое положение России. — Государственный комитет Российской Федерации по статистике, 1998, I—XII.
- [38] Краткосрочные социально-экономические показатели. Госкомстат РФ, апрель 1999.
- [39] Социальное положение и уровень жизни населения России. Официальное издание. — М.: Госкомстат России, 1999.
- [40] Богомолова Т.Ю., Тапилина В.С., Ростовцев П.С. Рост мобильности по доходам в изменении неравенства в распределении доходов. Новый проект, РПЭИ, декабрь 1999.
- [41] Deaton, Angus. Understanding consumption, Clarendon Lectures in Economics. Oxford University Press, Clarendon Press, 1992.
- [42] Stata Reference Manual. Release 6. Stata Press, 1999.
- [43] Little, R.J.A and D.B. Rubin. Statistical Analysis with Missing Data. Wiley, 1987.

**Результаты непараметрического анализа распределения населения регионов России и страны в целом по величине совокупных среднедушевых расходов**

**1. Российская Федерация (данные RLMS, 8-й раунд, ноябрь 1998 г., общее число наблюдений  $n = 3618$ )**

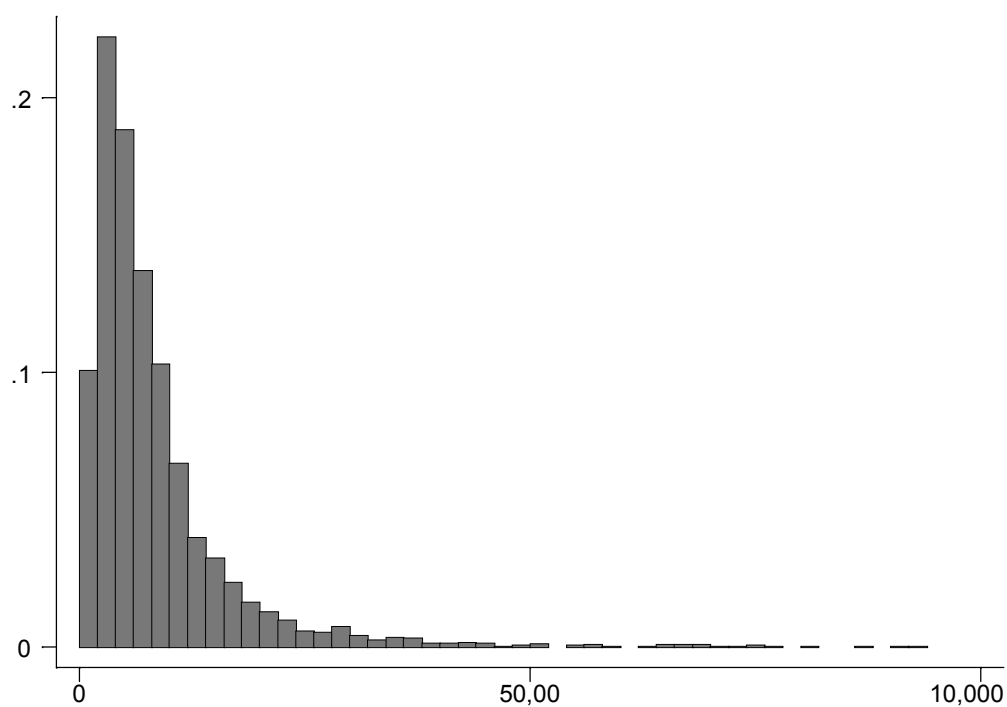


Рис. П.1<sup>а</sup> Гистограмма распределения населения России по совокупным душевым расходам, построенная на базе и с х о д н ы х (некалиброванных) данных.

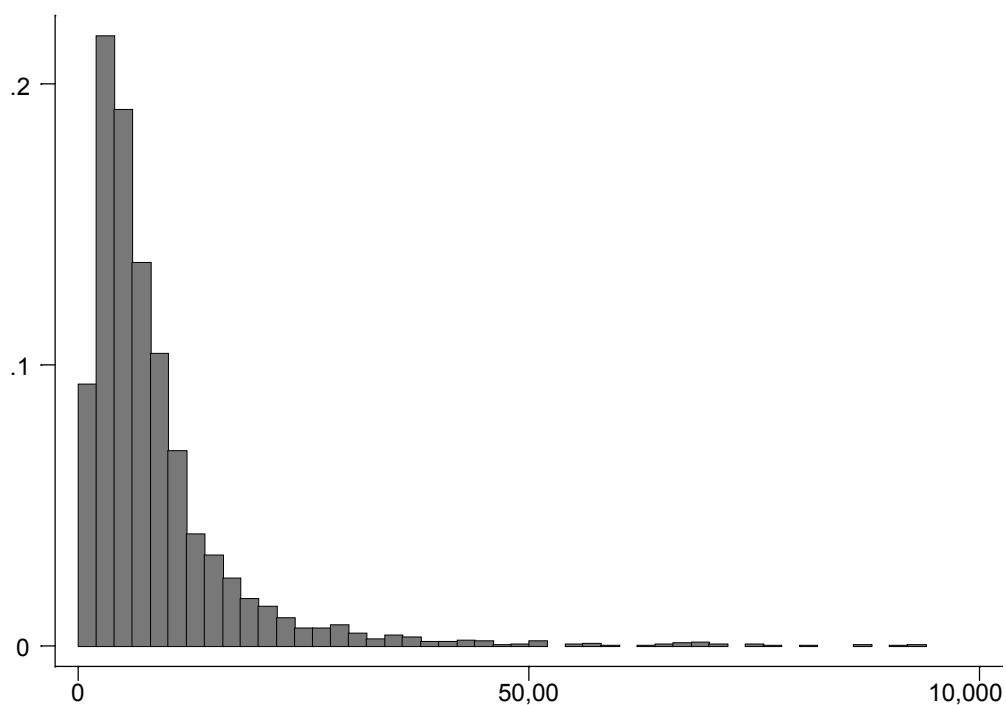


Рис. П.1<sup>б</sup> Гистограмма распределения населения России по совокупным душевым расходам, построенная на базе откалиброванных данных.

Таб. П.1 Выборочные значения основных количественных характеристик распределения населения России по совокупным душевым расходам.

№№ пп.	Название характеристики (ед. измерения - тыс.руб.)	Значения выборочных характеристик	
		По исходным данным	По откалиброванным дан- ным
1	Среднее значение ( $\hat{\mu}$ ) совокупных душевых расходов	0,914	0,932
2	Среднеквадратическое отклонение ( $S$ )	1,765	1,682
3	Минимальное значение в выборке ( $x_{\min}$ )	0,008	0,008
4	Максимальное значение в выборке ( $x_{\max}$ )	49,344	49,344
5	Нижняя дециль ( $\hat{x}_{0,1}$ )	0,200	0,209
6	Верхняя дециль ( $\hat{x}_{0,9}$ )	1,699	1,763



2. Республика Коми (данные выборочного обследования домашних хозяйств, II-й квартал 1998 г., общее число наблюдений  $n = 1089$ )

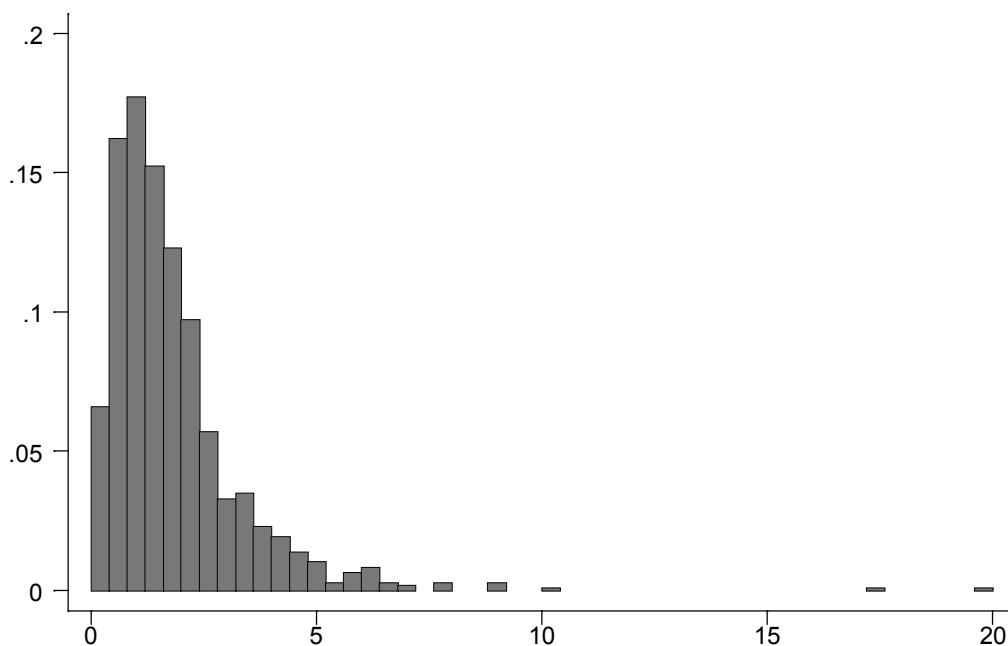


Рис. П.2<sup>а</sup> Гистограмма распределения населения Республики Коми по совокупным душевым расходам, построенная на базе и с х о д н ы х (некалиброванных) данных.

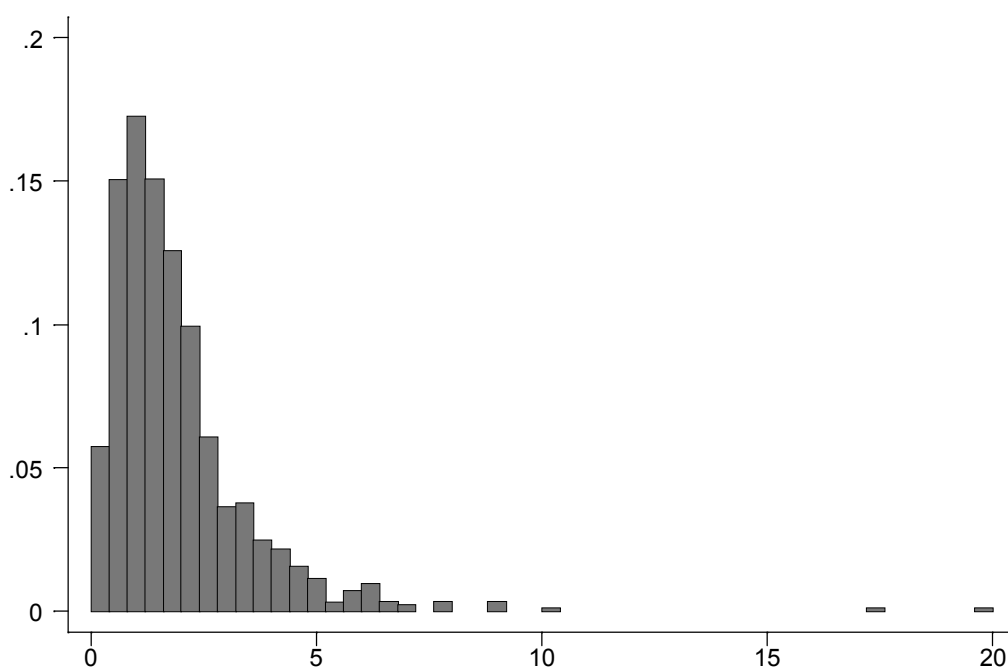


Рис. П.2<sup>б</sup> Гистограмма распределения населения Республики Коми по совокупным душевым расходам, построенная на базе о т к а л и б р о в а н н ы х данных.

Таб. П.2 Выборочные значения основных количественных характеристик распределения населения Республики Коми по совокупным душевым расходам.

№№ пп.	Название характеристики (ед. измерения - тыс.руб.)	Значения выборочных характеристик	
		По исходным данным	По откалиброванным данным
1	Среднее значение ( $\hat{\mu}$ ) совокупных душевых расходов	0,633	0,686
2	Среднеквадратическое отклонение ( $S$ )	1,087	1,249
3	Минимальное значение в выборке ( $x_{\min}$ )	0,054	0,054
4	Максимальное значение в выборке ( $x_{\max}$ )	24,797	24,797
5	Нижняя дециль ( $\hat{x}_{0,1}$ )	0,154	0,163
6	Верхняя дециль ( $\hat{x}_{0,9}$ )	1,208	1,302

**3. Волгоградская область (данные выборочного обследования домашних хозяйств, II-й квартал 1998 г., общее число наблюдений  $n = 1263$ )**

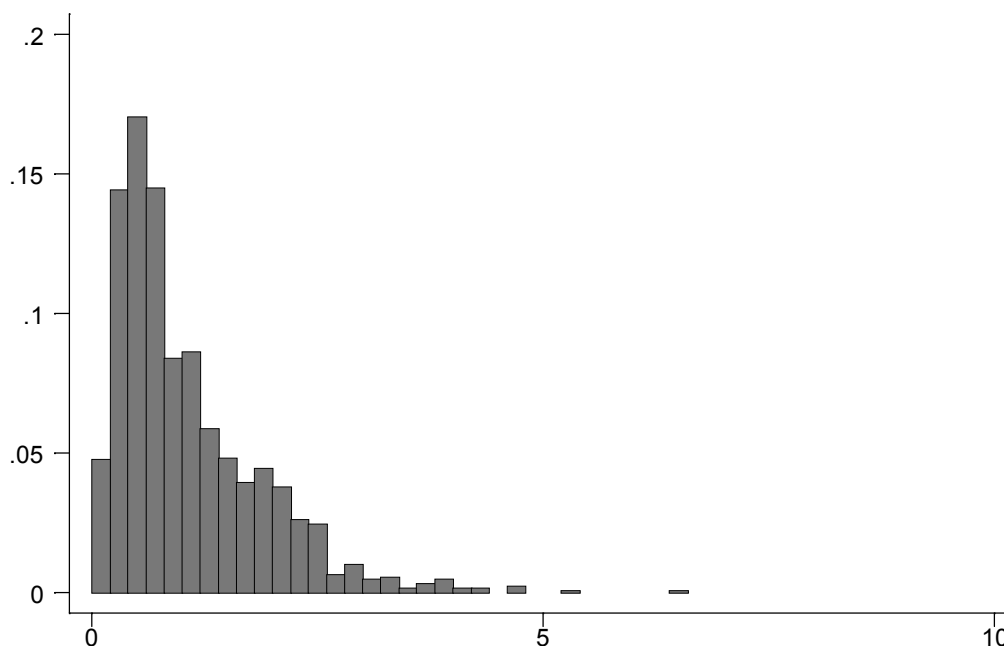


Рис. П.3<sup>a</sup> Гистограмма распределения населения Волгоградской области по совокупным душевым расходам, построенная на базе и с х о д н ы х (некалиброванных) данных.

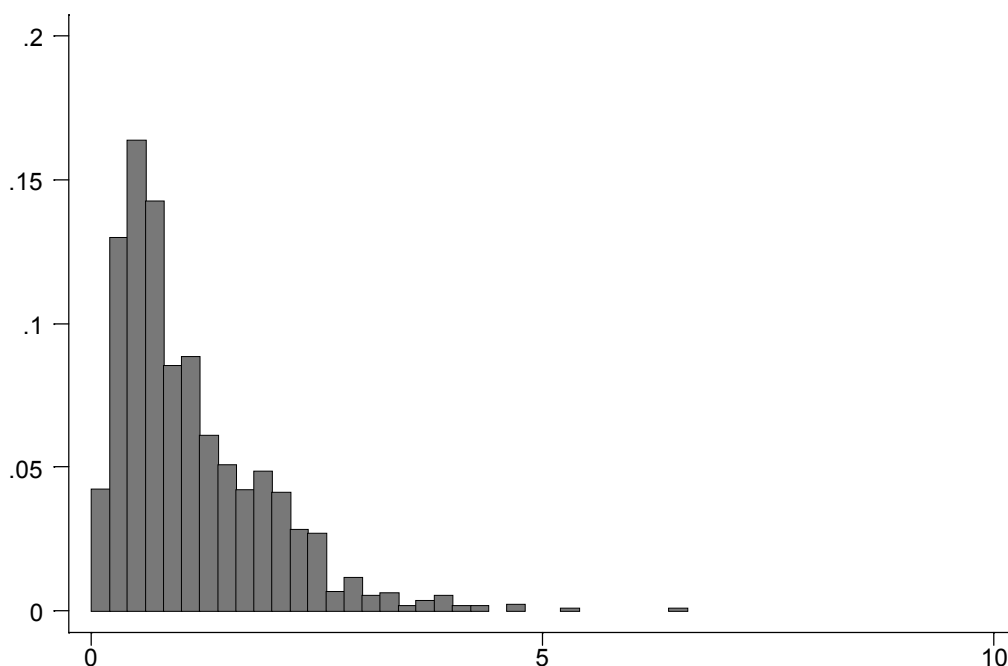


Рис. П.3<sup>б</sup> Гистограмма распределения населения Волгоградской области по совокупным душевым расходам, построенная на базе откалиброванных данных.

Таб. П.3 Выборочные значения основных количественных характеристик распределения населения Волгоградской области по совокупным душевым расходам.

№№ пп.	Название характеристики (ед. измерения - тыс.руб.)	Значения выборочных характеристик	
		По исходным данным	По откалиброванным дан- ным
1	Среднее значение ( $\hat{\mu}$ ) совокупных душевых расходов	0,412	0,433
2	Среднеквадратическое отклонение ( $S$ )	0,458	0,479
3	Минимальное значение в выборке ( $x_{\min}$ )	0,017	0,017
4	Максимальное значение в выборке ( $x_{\max}$ )	6,101	6,101
5	Нижняя дециль ( $\hat{x}_{0,1}$ )	0,101	0,110
6	Верхняя дециль ( $\hat{x}_{0,9}$ )	0,766	0,794

4. Омская область (данные выборочного обследования домашних хозяйств, II-й квартал 1998 г., общее число наблюдений  $n = 1244$ )

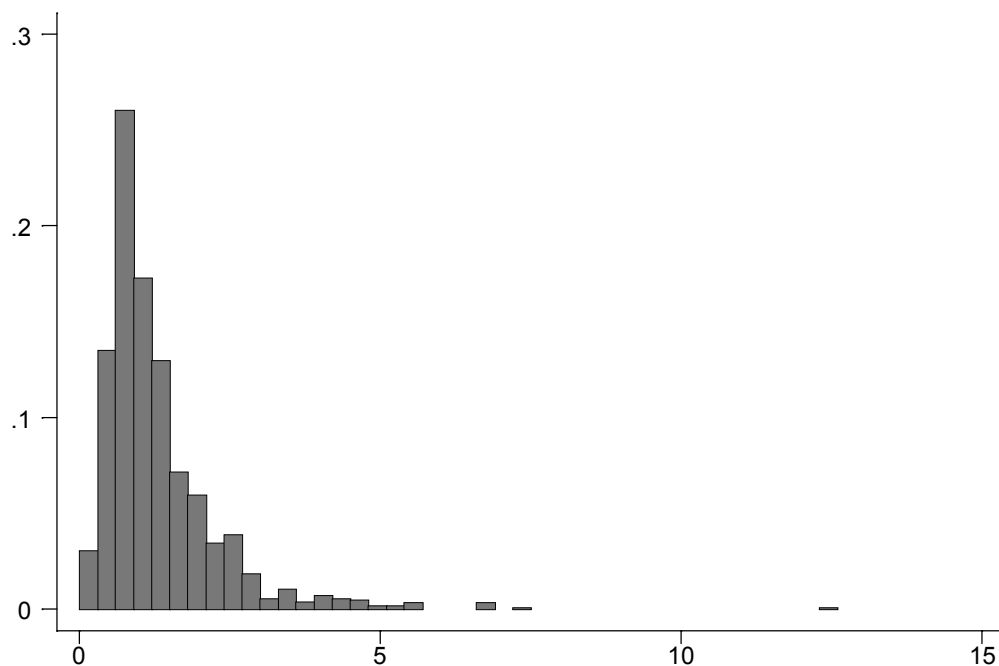


Рис. П.4<sup>а</sup> Гистограмма распределения населения Омской области по совокупным душевым расходам, построенная на базе и с х о д н ы х (некалиброванных) данных.

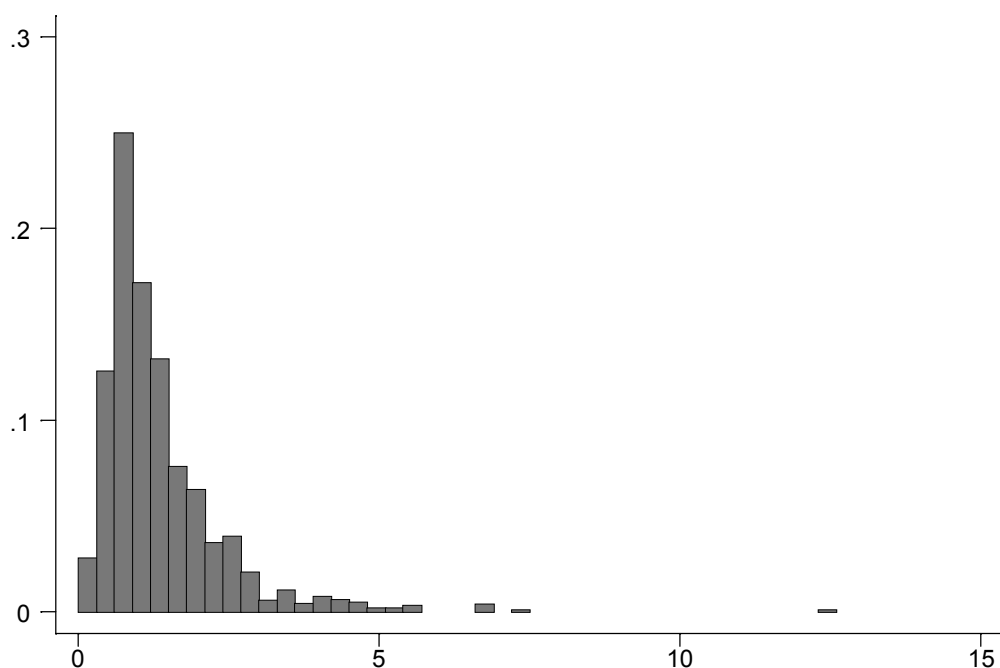


Рис. П.4<sup>б</sup> Гистограмма распределения населения Омской области по совокупным душевым расходам, построенная на базе откалиброванных данных.

Таб. П.4 Выборочные значения основных количественных характеристик распределения населения Омской области по совокупным душевым расходам.

№№ пп.	Название характеристики (ед. измерения - тыс.руб.)	Значения выборочных характеристик	
		По исходным данным	По откалиброванным дан- ным
1	Среднее значение ( $\hat{\mu}$ ) совокупных душевых расходов	0,611	0,641
2	Среднеквадратическое отклонение ( $S$ )	0,708	0,761
3	Минимальное значение в выборке ( $x_{\min}$ )	0,034	0,034
4	Максимальное значение в выборке ( $x_{\max}$ )	11,809	11,809
5	Нижняя дециль ( $\hat{x}_{0,1}$ )	0,160	0,163
6	Верхняя дециль ( $\hat{x}_{0,9}$ )	1,211	1,238

**Результаты статистического анализа моделей смесей распределений в рамках статистически наблюдаемого диапазона значений совокупных среднедушевых расходов домашних хозяйств**

**1. О методологии статистического оценивания параметров модели смеси распределений (EM-алгоритмы и их модификации)**

Опишем кратко процедуру статистического оценивания параметров

$$\Theta(k) = (\tilde{q}_1, \dots, \tilde{q}_k; a_1, \dots, a_k; \sigma_1^2, \dots, \sigma_k^2) \quad (\text{П.1})$$

функции плотности вероятности

$$\tilde{\varphi}_k(z|\Theta) = \sum_{j=1}^k \tilde{q}_j \varphi(z|a_j; \sigma_j^2) \quad (\text{П.2})$$

по случайной выборке (8') с помощью метода максимального правдоподобия при заданном значении числа компонентов смеси  $k$  (в соотношении (П.2) под  $\varphi(z|a_j; \sigma_j^2)$  подразумевается функция плотности нормального распределения со средним значением  $a_j$  и дисперсией  $\sigma_j^2$ ).

Задача состоит в нахождении таких значений

$$\hat{\Theta}(k) = (\hat{q}_1, \dots, \hat{q}_k; \hat{a}_1, \dots, \hat{a}_k; \hat{\sigma}_1^2, \dots, \hat{\sigma}_k^2), \quad (\text{П.3})$$

при которых логарифмическая функция правдоподобия

$$l_k(\Theta(k)) = \sum_{i=1}^n \omega_i \left[ \ln \sum_{j=1}^k \tilde{q}_j \varphi(z_i|a_j; \sigma_j^2) \right] \quad (\text{П.4})$$

достигает своего максимума, т.е.

$$\hat{\Theta}(k) = \arg \max_{\Theta(k)} l_k(\Theta(k)) \quad (\text{П.5})$$

(в соотношении (П.4)  $z_i$  — элементы выборки (11),  $\omega_i$  — веса наблюдений, определенные формулой (11'), а  $n$  — объем имеющейся выборки).

Итерационный *EM-алгоритм* (алгоритм «*Expectation-Maximization*»), с помощью которого решается задача (П.5), основан на следующей логической схеме (см. [31], [32]):

- (i) логарифмическая функция правдоподобия (П.4) представляется в виде

$$l_k(\Theta(k)) = \sum_{i=1}^n \omega_i \sum_{j=1}^k g_{ij} \ln \tilde{q}_j + \sum_{i=1}^n \omega_i \sum_{j=1}^k g_{ij} \ln \varphi(z_i | a_j; \sigma_j^2) - \sum_{i=1}^n \omega_i \sum_{j=1}^k g_{ij}, \quad (\text{П.6})$$

где величины

$$g_{ij} = \frac{\tilde{q}_j \varphi(z_i | a_j; \sigma_j^2)}{\tilde{\varphi}_k(z_i | \Theta(k))} \quad (\text{П.7})$$

в соответствии с правилом вычисления условных вероятностей определяют вероятность наблюдать класс  $j$  при условии, что в качестве  $i$ -го элемента выборки мы располагаем наблюдением  $z_i$  (так называемые *апостериорные вероятности* наблюдения класса  $j$ );

(ii) этап «*Expectation*»: пусть на  $t$ -м шаге итерационной процедуры получено значение

$$\hat{\Theta}^{(t)}(k) = \left( \hat{q}_1^{(t)}, \dots, \hat{q}_k^{(t)}; \hat{a}_1^{(t)}, \dots, \hat{a}_k^{(t)}; (\hat{\sigma}_1^2)^{(t)}, \dots, (\hat{\sigma}_k^2)^{(t)} \right) \quad (\text{П.8})$$

оценки параметра  $\Theta(k)$ ; подставляя в формулу (П.7) эти значения, получаем величины  $g_{ij}^{(t)}$ , которые вставляем затем в правую часть (П.6) вместо значений  $g_{ij}$ ;

(iii) этап «*Maximization*»: для получения следующей итерации максимизируем по  $\hat{\Theta}^{(t)}(k)$  выражение

$$l_k(\hat{\Theta}^{(t)}(k)) = \sum_{i=1}^n \omega_i \sum_{j=1}^k g_{ij}^{(t)} \ln \hat{q}_j^{(t)} + \sum_{i=1}^n \omega_i \sum_{j=1}^k g_{ij}^{(t)} \ln \varphi(z_i | \hat{a}_j^{(t)}; (\hat{\sigma}_j^2)^{(t)}) - \sum_{i=1}^n \omega_i \sum_{j=1}^k g_{ij}^{(t)} \quad (\text{П.9})$$

при *фиксированных* значениях  $g_{ij}^{(t)}$ ; в результате получаем решения:

$$\begin{aligned} \hat{q}_j^{(t+1)} &= \sum_{i=1}^n \omega_i g_{ij}^{(t)}, \\ \hat{a}_j^{(t+1)} &= \frac{1}{\hat{q}_j^{(t+1)}} \sum_{i=1}^n \omega_i g_{ij}^{(t)} z_i, \\ (\hat{\sigma}_j^2)^{(t+1)} &= \frac{1}{\hat{q}_j^{(t+1)}} \sum_{i=1}^n \omega_i g_{ij}^{(t)} (z_i - \hat{a}_j^{(t+1)})^2, \\ & j = 1, 2, \dots, k. \end{aligned}$$

После этого возвращаются к этапу «*Expectation*», т.е., используя значения  $\hat{q}_j^{(t+1)}$ ,  $\hat{a}_j^{(t+1)}$  и  $(\hat{\sigma}_j^2)^{(t+1)}$  ( $j = 1, 2, \dots, k$ ), подсчитывают по формуле (П.7) величины  $g_{ij}^{(t+1)}$ , вставляют их в правую часть (П.9), переходят к этапу «*Maximization*» и т.д. В [32] и других более поздних

работах<sup>4</sup> доказаны (при достаточно общих предположениях, наиболее жестким из которых является *требование ограниченности логарифмической функции правдоподобия*) полезные свойства EM-алгоритмов и, в частности, их сходимость (по вероятности) к искомому решению (П.5).

Модификации EM-алгоритма, использованные в наших расчетах, носят технический характер. Они состоят в приписывании весов  $\omega_i$  наблюдениям  $z_i$ , а также — в использовании в начальной стадии алгоритма вспомогательного («нулевого») так называемого фонового («background») компонента, который в отличие от остальных (нормальных) компонентов полагается *равномерно* распределенным, грубо говоря, на всем диапазоне изменений выборочных данных. Подробное описание версии EM-алгоритма, реализованной в пакете «Классмастер», можно найти в [35].

Выше была описана схема определения оценок максимального правдоподобия  $\hat{\Theta}(k)$  параметров  $\Theta(k)$  (см. (П.1)) при *известном* значении числа компонентов смеси  $k$ . Теперь кратко опишем **процедуру оценивания числа компонентов смеси  $k$** , т.е. тех компонентов, *которые можно статистически выявить в рамках наблюдаемого диапазона значений совокупных среднедушевых семейных расходов*.

Эта процедура заключалась в последовательной проверке простых гипотез вида

$$H_0: k = j$$

при альтернативе

$$H_1: k = j + 1, \quad j = 1, 2, \dots, \dots$$

с использованием критической статистики

$$\gamma(j) = -2 \ln \frac{l_j(\hat{\Theta}(j))}{l_{j+1}(\hat{\Theta}(j+1))}.$$

Первое значение  $j = \hat{k}$ , при котором гипотеза  $H_0$  оказалась не отвергнутой, принималось за оценку числа компонентов в смеси (П.2). Эта процедура дополнялась описанным в [33] приемом приблизительного определения числа кластеров, основанным на методе целенаправленного проецирования (Projection Pursuit Method), а также — содержательным анализом получаемых при этом классов.

---

<sup>4</sup> В действительности общая схема алгоритмов, позднее названных EM-алгоритмами, была, по-видимому, впервые предложена в работе Шлезингер М.И. О самопроизвольном различении образов. — «Читающие автоматы», Киев, Наукова думка, 1965, с. 38—45. Там же были исследованы и основные свойства этих алгоритмов. Однако, эта работа труднодоступна и малоизвестна среди зарубежных специалистов.



Параллельно с описанным выше модифицированным EM-алгоритмом (программно реализованным в пакете «Классмастер») для решения той же задачи эксплуатировалась также созданная С.О. Колениковым программа, использующая средства пакета STATA (в качестве — его внутренний «максимизатор», см. *W. Gould, W. Scribney. Maximum Likelihood estimation with STATA. — Stata Press, 1999*). Этот алгоритм максимизации, реализованный в STATA, можно вкратце описать следующим образом:

1. находятся начальные значения: если пользователь не предложил некоторых начальных значений, то случайным образом;
2. в окрестности этих начальных значений производится случайных поиск лучших значений;
3. производится одномерная оптимизация по каждому из параметров модели;
4. запускается алгоритм многомерной оптимизации.
  - 4.1. Определяется, выпукла ли функция правдоподобия в данной точке (производные первого и второго порядка вычисляются численно).
  - 4.2. Если функция выпукла, то итерация осуществляется с помощью метода Ньютона—Рафсона;
  - 4.3. если нет, то с помощью градиентного метода наискорейшего сауска.
5. Завершение работы определяется как сочетание
  - 5.1. стабилизации функции правдоподобия (по умолчанию,  $10^{-6}$ ; в случае необходимости можно изменить);
  - 5.2. стабилизации значений коэффициентов (по умолчанию, отн. Изменение меньше  $10^{-7}$ );
  - 5.3. градиент функции правдоподобия мал по величине (устанавливается как параметр; использовано значение  $10^{-3}$ );
  - 5.4. сделано слишком много итераций (по умолчанию, 16000);
  - 5.5. невозможность вычислить производные (плато функции правдоподобия) вызывает аварийный останов.

Если максимизация завершилась успешно, то Stata выводит таблицу оценок коэффициентов модели со стандартными отклонениями и доверительными интервалами (см. ниже). Помимо этого, автор дополняет выводимые результаты значениями различных критериев качества оценивания; информационными статистиками (AIC, ICOMP) и результатами теста  $\chi^2$  на согласие с модельным распределением. Число интервалов группирования  $N$

(степеней свободы хи-квадрат), на которые разбивается весь диапазон значений анализируемого признака, выбирается с учетом рекомендации  $N \approx \log_2 n$ , где  $n$  — общее число наблюдений. В соответствии с гипотезой  $H_3$  (в форме  $H'_3$ , см. выше п. 4.1) описанные процедуры реализуются при упрощающем условии  $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$ , где  $\sigma_j^2 = D(\ln \xi)$ .

Результаты применения этих процедур к данным RLMS, 8 раунд (по России), а также к каждому из региональных массивов данных (по Республике Коми, Волгоградской и Омской областям) представлены в табл. П.5~П.8.

Табл. П.5. Результаты оценивания параметров в модели смеси нормальных законов, описывающей распределение населения России по величине логарифма совокупных душевых расходов

Число компонент смеси ( $k$ )	Оптимальное значение логарифмич. Функции правдоподобия	Критерий Акаике (AIC)	Критерий Боздогана (ICOMP)	Статистика $\chi^2(N)$ критерия согласия	Уровень значимости критерия согласия	$\hat{\sigma}^2$	$\hat{a}_1$	$\hat{q}_1$ %	$\hat{a}_2$	$\hat{q}_2$ %	$\hat{a}_3$	$\hat{q}_3$ %	$\hat{a}_4$	$\hat{q}_4$ %	$\hat{\mu}_{\text{мод.}}$ (тыс.руб.)
1	-4244,8	8489,5	8490,0	53,23	$< 5 \cdot 10^{-5}$	0,790	6,400	100	—	—	—	—	—	—	0,893
2	-4222,6	8453,1	8461,7	34,35	$3 \cdot 10^{-4}$	0,722	6,430	98,71	4,118	1,29	—	—	—	—	0,879
3	-4192,4	8384,8	8398,0	13,72	0,248	0,600	6,412	95,60	4,397	2,57	8,578	1,83	—	—	0,920
<b>4</b>	<b>-4189,1</b>	<b>8378,0</b>	<b>8393,0</b>	<b>12,12</b>	<b>0,354</b>	<b>0,538</b>	<b>6,376</b>	<b>91,09</b>	<b>4,469</b>	<b>3,19</b>	<b>7,759</b>	<b>5,44</b>	<b>9,816</b>	<b>0,27</b>	<b>0,937</b>

Выбрана модель с **четырьмя** компонентами

Табл. П.6. Результаты оценивания параметров в модели смеси нормальных законов, описывающей распределение населения Республики Коми по величине логарифма совокупных душевых расходов

Число компонент смеси ( $k$ )	Оптимальное значение логарифмич. функции правдоподобия	Критерий Акаике (AIC)	Критерий Боздогана (ICOMP)	Статистика $\chi^2(N)$ критерия согласия	Уровень значимости критерия согласия	$\hat{\sigma}^2$	$\hat{a}_1$	$\hat{q}_1$ %	$\hat{a}_2$	$\hat{q}_2$ %	$\hat{a}_3$	$\hat{q}_3$ %	$\hat{a}_4$	$\hat{q}_4$ %	$\hat{a}_5$	$\hat{q}_5$ %	$\hat{\mu}_{\text{мод.}}$ (тыс. руб.)
1	-1313,74	2627,48	2627,78	32,08	0,000	0,654	-0,842	—	—	—	—	—	—	—	—	—	0,598
2	-1302,37	2604,75	2615,06	16,54	0,085	0,610	2,114	0,50	-0,857	99,50	—	—	—	—	—	—	0,630
3	-1300,69	2601,38	2613,91	23,67	0,009	0,475	-0,162	22,88	-1,058	76,79	2,527	0,33	—	—	—	—	0,636
<b>4</b>	<b>-1299,34</b>	<b>2598,69</b>	<b>2615,72</b>	<b>9,77</b>	<b>0,461</b>	<b>0,285</b>	<b>-1,976</b>	<b>8,99</b>	<b>0,010</b>	<b>25,41</b>	<b>-1,038</b>	<b>65,17</b>	<b>2,338</b>	<b>0,43</b>	—	—	<b>0,628</b>
5	-1294,49	2588,97	2609,11	5,90	0,824	0,168	0,890	3,07	-1,082	55,02	2,756	0,27	-0,115	29,47	-2,029	12,17	0,633

Выбрана модель с **четырьмя** компонентами

Табл. П.7. Результаты оценивания параметров в модели смеси нормальных законов, описывающей распределение населения Волгоградской области по величине логарифма совокупных душевых расходов

Число компонент смеси ( $k$ )	Оптимальное значение логарифмич. функции правдоподобия	Критерий Акаике (AIC)	Критерий Боздогана (ICOMP)	Статистика $\chi^2(N)$ критерия согласия	Уровень значимости критерия согласия	$\hat{\sigma}^2$	$\hat{a}_1$	$\hat{q}_1$ %	$\hat{a}_2$	$\hat{q}_2$ %	$\hat{a}_3$	$\hat{q}_3$ %	$\hat{a}_4$	$\hat{q}_4$ %	$\hat{a}_5$	$\hat{q}_5$ %	$\hat{\mu}_{\text{мод.}}$ (тыс. р. уб.)
1	-1587,56	3175,11	3175,36	43,63	0,000	0,723	-1,259	100	—	—	—	—	—	—	—	—	0,408
2	-1586,26	3172,52	3185,81	42,78	0,000	0,673	0,116	2,58	-1,295	97,42	—	—	—	—	—	—	0,414
2	-1587,49	3174,98	3191,68	42,72	0,000	0,716	-1,254	99,77	-3,044	0,23	—	—	—	—	—	—	0,407
3	-1585,25	3170,50	3191,55	37,53	0,000	0,577	-2,456	4,10	-1,280	90,57	0,021	5,34	—	—	—	—	0,414
4	-1573,11	3146,23	3157,29	32,694	0,000	0,180	0,613	4,10	-2,833	7,00	-0,647	38,56	-1,661	50,34	—	—	0,413
<b>5</b>	<b>-1568,44</b>	<b>3136,88</b>	<b>3149,21</b>	<b>12,03</b>	<b>0,283</b>	<b>0,099</b>	<b>-2,943</b>	<b>6,16</b>	<b>-1,927</b>	<b>28,58</b>	<b>-4,81</b>	<b>27,89</b>	<b>0,650</b>	<b>4,21</b>	<b>-1,266</b>	<b>33,16</b>	<b>0,411</b>

Выбрана модель с **пятью** компонентами

Табл. П.8. Результаты оценивания параметров в модели смеси нормальных законов, описывающей распределение населения Омской области по величине логарифма совокупных душевых расходов

Число компонент смеси ( $k$ )	Оптимальное значение логарифмич. Функции правдоподобия	Критерий Акаике (AIC)	Критерий Боздогана (ICOMP)	Статистика $\chi^2(N)$ критерия согласия	Уровень значимости критерия согласия	$\hat{\sigma}^2$	$\hat{a}_1$	$\hat{q}_1$ %	$\hat{a}_2$	$\hat{q}_2$ %	$\hat{a}_3$	$\hat{q}_3$ %	$\hat{a}_4$	$\hat{q}_4$ %	$\hat{a}_5$	$\hat{q}_5$ %	$\hat{\mu}_{\text{мод.}}$ (тыс. руб.)
1	-1503,17	3006,34	3006,63	23,83	0,008	0,656	-0,838	—	—	—	—	—	—	—	—	—	0,600
2	-1499,93	2999,85	3013,09	13,75	0,184	0,602	0,671	2,34	-0,875	97,66	—	—	—	—	—	—	0,612
2	-1497,50	2995,00	3002,55	25,63	0,004	0,591	-2,915	1,48	-0,807	98,52	—	—	—	—	—	—	0,592
<b>3</b>	<b>-1482,10</b>	<b>2964,20</b>	<b>2972,02</b>	<b>13,47</b>	<b>0,198</b>	<b>0,382</b>	<b>-0,960</b>	<b>84,72</b>	<b>-2,911</b>	<b>2,20</b>	<b>0,294</b>	<b>13,09</b>	—	—	—	—	<b>0,607</b>
4	-1479,72	2959,44	2979,77	23,30	0,010	0,351	2,192	0,18	0,165	16,95	-2,908	2,28	-0,998	80,59	—	—	0,613
4	-1481,45	2962,90	2977,45	14,77	0,141	0,278	-3,003	2,06	-1,260	46,09	0,548	8,67	-0,563	43,18	—	—	0,606
5	-1476,16	2952,33	2970,11	18,42	0,048	0,211	-3,047	2,01	0,436	11,87	-1,433	34,27	-0,662	51,68	2,302	0,18	0,612

Выбрана модель с **тремя** компонентами

**Вероятность уклонения домашнего хозяйства от обследования как функция некоторых его характеристик (результаты анализа)**

Модель, в рамках которой исследуется вероятность ( $p$ ) уклонения домашнего хозяйства от обследования как функция от логарифма его совокупных душевых расходов ( $z^{(1)}$ ), характеристики места проживания этого ДХ ( $z^{(2)}$ ) и уровня образования главы семьи ( $z^{(3)}$ ), описана в п. 4.4.1 данной работы.

Перед тем как описать результаты статистического анализа этой модели, *охарактеризуем природу и структуру исходных статистических данных, использовавшихся* в качестве информационной базы этого анализа.

В каждом раунде RLMS второй волны (т.е. начиная с 1994 г.) интервьюеры обходят все домохозяйства, входящие в первичную выборку (4718 домохозяйств) и записывают, был ли проведен опрос в данном домохозяйстве или нет, и если нет, то по какой причине. Регистрируемые коды ответов приведены в таблице П.9. Таблица П.10 представляет частоту отказов в отношении к общему числу интервьюируемых.

**Таблица П.10. Частота отказов от участия в обследовании**

	Раунд 5	Раунд 6	Раунд 7	Раунд 8
Опрос не проведен	743	963	1118	1254
Число отказов	410	539	489	701
Отказ из-за нежелания предоставлять информацию о благосостоянии			17	19
Опрос проведен	3973	3781	3750	3831

Конечной целью анализа является ответ на вопрос: «Зависит ли вероятность уклонения от социологического обследования от благосостояния семьи?». На основе вышеупомянутых данных об отказах в сочетании с данным о доходных и расходных характеристиках домохозяйств, приводимых в основных файлах RLMS, можно сформировать эконометрические модели с бинарной зависимой переменной (отказ/участие) и расходами домохозяйств (или какой-то более удачной мерой благосостояния) в качестве объясняющей переменной.

Табл. П.9. Кодификатор результатов посещения

01	Опрос проведен	27	Действия против интервьюера
<b>Объективные причины</b>		28	Другое
02	Нежилое помещение	<u>Мотивы отказа</u>	
03	В квартире (доме) в данный момент никто не живет	41	Немотивированный отказ
04	Квартира недоступна	42	Ссылка на занятость
05	Квартиру снимают иностранцы	43	«Очень некогда»
06	Никого нет дома (3 посещения)	44	«Никому не открываю»
07	Не открывают дверь, не вступая в разговор	45	«Эти вопросы ничего не дают, ничего не меняют»
08	Опрос невозможен из-за болезни	46	«Не хочу никому рассказывать о своей жизни»
09	Опрос невозможен из-за инвалидности	47	«Имею права не отвечать»
10	Дома нет никого из взрослых	48	«Хочу отдохнуть»
11	Открывший дверь пьян	49	«Не хочу быть в компьютере»
14	Семья отсутствует в течение всего периода опроса	50	«Недавно участвовали в другом социологическом опросе»
15	Семья бывает дома только поздно вечером	51	«Мы здесь люди временные»
16	Семья фактически живет в другом месте	52	Семейные обстоятельства
18	Другое	53	Не устраивает (не интересует) тема опроса
<b>Отказы</b>		54	Надоела политика
30	Отказ от участия	55	Отказ из чувства протеста
<u>Ситуация общения</u>		56	Опасения за последствия интервью: нежелание предоставлять информацию о своих политических взглядах
21	Отказ через закрытую дверь	57	<i>Опасения за последствия интервью: нежелание предоставлять информацию о благосостоянии семьи</i>
22	Отказ открывшего дверь	58	Не верят интервьюеру
23	Отказ респондента	59	Другое
24	Отказ другого члена семьи		
25	Отказ во время проведения интервью		
26	Отказ-обман		

Очевидно, если домохозяйство отказалось от участия в опросе в данном раунде, то данные о его расходах не могут быть получены. Однако, поскольку данные являются *панельными*, т.е. в разных раундах участвует одно и то же домохозяйство (по крайней мере,



потенциально участвует, и другие домохозяйства попасть в выборку не могут), информацию об уровне благосостояния семьи можно получить из других раундов, считая, что благосостояние семьи в разные раунды примерно постоянно. Данное предположение может быть подвергнуто критике на том основании, что мобильность по доходам в России представляется достаточно высокой (см., например, [40]). По нашим сведениям, мобильность не является критически высокой: дисперсия логарифмов потребительских расходов по выборке RLMS находится в пределах от 0.018 (т.е. отклонения от среднего составляют менее 2%) до 1.32 (т.е. потребительские расходы меняются в 3.7 раза), составляя в среднем 0.25 с медианой 0.21 (т.е. потребительские расходы колеблются вокруг «постоянного» уровня примерно на 25%). Используемый нами алгоритм расчета меры благосостояния позволяет получить величины типа «постоянного дохода» (или, в нашем случае, «постоянного потребления»), рассматриваемые в контексте гипотезы Фридмана о доходах в течение жизненного цикла [41]. В качестве меры благосостояния домохозяйства используются усредненные (по доступным периодам) расходы. Использование других мер благосостояния (медианы расходов за доступные периоды, временных (imputed) расходов<sup>4</sup>, первой главной компоненты расходов) приводит к содержательно похожим результатам, при этом эластичность вероятности отказа по доходу (в соответствующих единицах) меняется в пределах нескольких процентов. В дальнейшем будет использоваться именно среднее расходов как наиболее понятная величина.

Базовыми переменными, которые используются в анализе вероятности уклонения от обследования в зависимости от дохода, служат потребительские расходы домохозяйств, дефлированные к единому начальному периоду (переменные *totexpr\** за разные годы; в базе данных RLMS используется дефлятор «Обзора российской экономики», конструируемый Российско-Европейским центром экономической политики), а также уровень урбанизации местности, в которой проживает домохозяйство ( $z^{(2)}$ ), и уровень образования его главы (точнее, члена домохозяйства с наибольшим доходом  $z^{(3)}$ ). Используемой мерой благосостояния домохозяйства служит среднее значение логарифмов дефлированных рас-

---

<sup>4</sup> Используемое программное обеспечение Stata позволяет восстановить (impute) пропущенные значения расходных характеристик домохозяйства на основе линейной регрессионной модели. Для каждой структуры (pattern) пропусков строится своя регрессионная модель, на основе которой восстанавливаются пропущенные значения [42], [43]. Иными словами, для каждого наблюдения с пропущенным значением восстанавливаемой переменной формируется набор регрессоров, значения которых *не* пропущены в данном наблюдении; оценивается линейная регрессия; и строится прогнозное значение искомой переменной. При дальнейшем использовании восстановленных подобным образом данных в качестве регрессоров оценки соответствующих коэффициентов, вероятно, будут смещены вследствие ошибки прогноза, причем, как показывает практика, чаще всего в сторону нуля.

ходов за доступные периоды (максимум — четыре периода, раунды V—VIII). Использование среднего за несколько периодов позволяет приблизиться к значению «постоянного потребления» за указанные периоды. В качестве зависимой переменной  $\eta$  фигурирует наблюдаемый факт, отказывалось ли домохозяйство от участия хотя бы в одном из четырех раундов RLMS. Проводился также анализ с использованием категории отказа «Не хочу сообщать сведения о доходах», однако эта категория немногочисленна (порядка 0.5%), тогда как логит-модель (как и подавляющее большинство методов анализа бинарных зависимых переменных) хорошо работает при доле успехов (в нашем случае, отказов по указанной причине) в пределах 10—90%.

Оценки нескольких спецификаций логит-модели, основанные на описанных выше данных, приводятся в табл. П.11. Зависимая переменная модели — индикатор модели — индикатор того, уклонялось ли д/х от обследования по произвольной причине (795 из 4239 наблюдений; последняя цифра больше всех частот участия в обследовании из табл. П.10 в силу того, что в данную панель входят все домохозяйства, хотя бы раз участвовавшие в обследовании, т.е. надмножество домохозяйств для каждого периода). Переменные, характеризующие уровень благосостояния домохозяйства — медиана или среднее за четыре раунда логарифмов реальных (дефлированных к 1992 г.) расходов. В качестве базовой категории региональных переменных при введении соответствующих фиктивных переменных используется градация «город» (U; буквы в скобках относятся к графику, приведенному ниже). Образовательные категории построены по накопительной, а не индикаторной системе, следующим образом: базовая категория — образование ниже среднего (L); далее идет среднее образование (коэффициент при этой категории показывает отличие от категории образование ниже среднего); далее ПТУ, техникум, высшее — дополнительно к среднему (т.е. измеряют отличия указанных категорий от лиц со средним образованием; среднее образование не исключает наличия технического или высшего, поэтому индикаторы S не исключают индикаторов P, T или H). Используемые в дальнейшем веса рассчитываются по модели с наивысшим отношением правдоподобия на одну степень свободы, т.е. по последней.

На рис. П.5 приводятся графики прогнозных значений вероятности отказа от участия в обследовании для нескольких категорий домохозяйств (шкала совокупных душевых расходов дана в логарифмическом масштабе). Поскольку модель насчитывает четыре географических и пять образовательных категорий, общее количество частных логистических

кривых на графике должно составлять 20. На рис. П.5 показаны некоторые наиболее «населенные» и представительные из этих кривых.

Таблица П.11. Результаты анализа многофакторной модели для вероятности отказа от участия в обследовании

	(1)	(2)	(3)	(4)
Медиана расходов	0.396 (0.084)**	0.355 (0.075)**		
Средние расходы			0.429 (0.089)**	0.399 (0.079)**
Столичные регионы (M)		1.052 (0.206)**		1.043 (0.203)**
Сельская местность (R)		-1.583 (0.292)**		-1.576 (0.291)**
ПГТ (P)		-0.876 (0.310)**		-0.878 (0.308)**
Среднее образование (S)		-0.862 (0.156)**		-0.868 (0.156)**
ПТУ (P)		-1.826 (0.184)**		-1.825 (0.182)**
Техникум (T)		-1.268 (0.212)**		-1.277 (0.213)**
Высшее (H)		-0.857 (0.142)**		-0.880 (0.142)**
Константа	-4.532 (0.653)**	-3.140 (0.588)**	-4.788 (0.691)**	-3.464 (0.632)**
Кол-во наблюдений	4239	4239	4239	4239
Тест Вальда	Wald(1)= 22.05	Wald(8)= 317.86	Wald(1)= 23.39	Wald(8)= 334.78
Эмпирический уровень значимости	0.00	0.00	0.00	0.00

В скобках указаны стандартные отклонения коэффициентов, рассчитанные с поправкой на кластеризацию наблюдений (стратификацию выборки).

\* значимость на уровне 5%; \*\* значимость на уровне 1%.

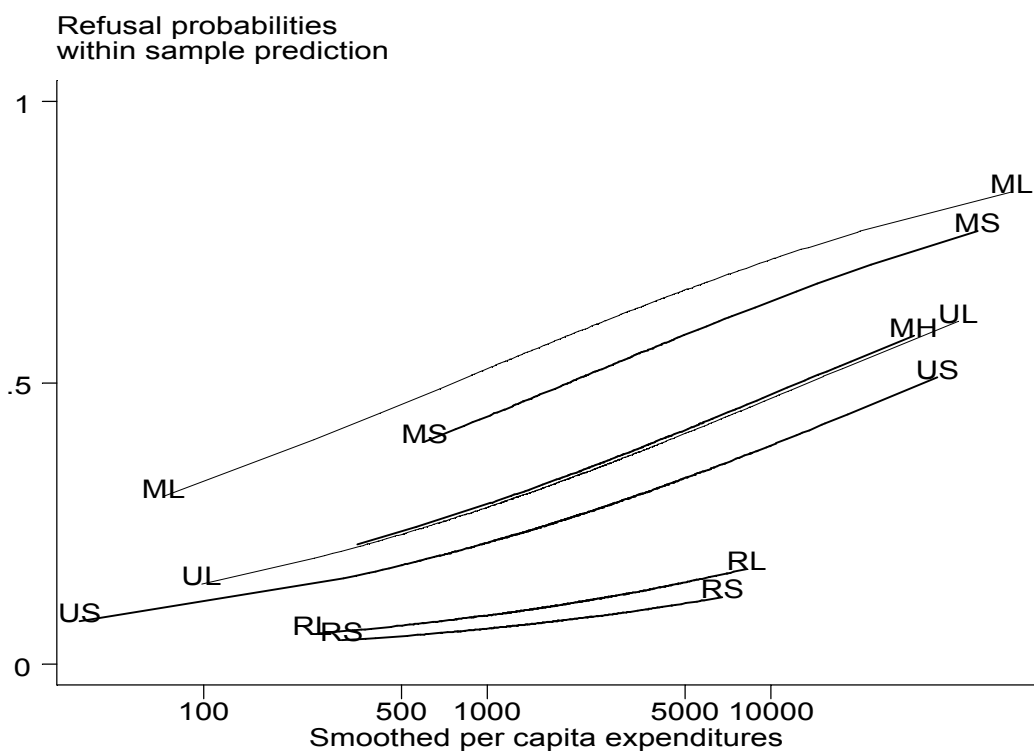


Рис. П.5. Семейство кривых, описывающих зависимость вероятности  $p$  от расхода  $x$ .

Полученные результаты, хотя и интересны сами по себе, служат исключительно для уточнения выборочных весов домохозяйств, т.е. вероятности участия конкретного домохозяйства в выборке обследования. В промежуточном отчете по проекту использовалась однофакторная модель с использованием среднедушевых расходов в качестве единственной объясняющей переменной. Как видно, использование многофакторной модели позволяет уточнить результаты, что, безусловно, должно позитивно сказаться на дальнейших результатах. Поскольку эластичность вероятности отказа от участия в обследовании статистически одинакова в спецификациях без дополнительных факторов и в расширенных спецификациях, можно считать, что общий результат о зависимости вероятности участия от расходов домохозяйства надежно верифицирован.