А. Ю. Тимофеева

К проблеме эндогенности потребительских расходов при оценивании системы спроса домашних хозяйств[1]

Работа посвящена анализу одного из случаев проявления проблемы эндогенности, а именно наличию погрешностей измерения потребительских расходов при построении системы спроса по данным выборочных обследований домашних хозяйств. Предложен и апробирован оригинальный подход на основе факторного анализа, позволяющий ослабить смещение оценок эластичности спроса и не требующий дополнительной информации (инструментальных переменных, повторных наблюдений). Его преимуществом также является возможность оценки параметров распределения истинных потребительских расходов, дающих представление о степени искажения информации о расходах населения.

Ключевые слова: эндогенный регрессор; погрешность измерения; потребительские расходы; эластичность спроса по доходу; факторный анализ; метод инструментальных переменных.

JEL classification: C21; C51; D12; D31.

1. Введение и постановка проблемы

Исследование структуры и закономерностей спроса можно считать одним из первых и ключевых этапов в эмпирическом изучении экономических явлений и процессов. Толчком к использованию больших массивов статистической информации и соответствующего инструментария здесь послужили базы данных, сформированные в ходе массовых бюджетных обследований (Сычева, 1998). Традиционно, в силу исторических реалий бюджетные обследования связываются с анализом проблем бедности. Современные задачи анализа спроса намного шире, они включают, например, разработку шкал эквивалентности, построение моделей спроса на специфические товары (табачные изделия, алкоголь).

Несмотря на достаточно продолжительную историю бюджетных исследований спроса, все еще остаются проблемы методологического характера. Одной из самых существенных можно назвать трудность определения индикатора дохода, выступающего основным фактором-детерминантом спроса в моделях пространственных данных (одномоментный срез) при предположении о фиксированных ценах товаров и услуг. С одной стороны, ввиду ряда недостатков статистики доходов населения, многими экономистами предлагается использовать в качестве показателя дохода сумму расходов по всем затратным статьям домохозяйств (Айвазян, 2012; Бирман, 2000). Сторонники такого подхода считают, что именно расходы являются «реальной экономической величиной, которую следует рассматривать в потреби-

[1] Работа выполнена при финансовой поддержке Министерства образования и науки РФ (государственное задание № 2014/138, проект № 1689).

тельском бюджете в данном промежутке времени» (Интрилигатор, 1975). Кроме того, в пользу включения суммарных потребительских расходов в качестве объясняющей переменной в систему спроса (в денежном выражении) говорит и более высокая корреляция с расходами по каждой отдельной статье.

Однако, с другой стороны, наличие расходов как в правой, так и в левой частях уравнений спроса приводит к так называемой проблеме эндогенности, поскольку нарушается одно из стандартных предположений регрессионного анализа о независимости и детерминированности входных факторов (Hausman, 2001). Проанализируем эту проблему при оценивании модели расходов с постоянной эластичностью (двойной логарифмической):

$$Y_j = a_j X^{b_j},\ \forall j = \overline{1,k},$$

где Y_j — расходы по j-й статье; b_j — эластичность спроса по доходу X; a_j — параметр масштаба; k — число статей бюджета. Такого рода модели являются наиболее простыми и часто используемыми при обработке данных домохозяйств (Brown, Deaton, 1972, p. 1176), в частности, для исследования структуры потребления по российской статистике (Шабанов, 2012).

Очевидно, что в ходе обследования значения расходов домашних хозяйств фиксируются с некоторыми случайными погрешностями ε_j. При этом наблюдаемый в практике анализа данных бюджетных обследований рост вариации в расходах домашних хозяйств с увеличением их благосостояния приводит к мультипликативному виду уравнений спроса:

$$y_j = Y_j \varepsilon_j = a_j X^{b_j} \varepsilon_j,\ \forall j = \overline{1,k}. \quad (1)$$

Если в качестве объясняющей переменной X использовать сумму всех наблюдаемых расходов $x = \sum_{j=1}^{k} Y_j \varepsilon_j$, то такой регрессор будет коррелировать со случайными ошибками в каждом уравнении. При этом между ошибками каждого уравнения системы будет взаимосвязь.

Сведем систему уравнений (1) к линейной путем логарифмирования:

$$\tilde{y}_j = \tilde{a}_j + b_j \tilde{X} + \tilde{\varepsilon}_j,\ \forall j = \overline{1,k}, \quad (2)$$

где $\tilde{y}_j = \ln y_j$, $\tilde{X} = \ln X$, $\tilde{a}_j = \ln a_j$, и относительно ошибок $\tilde{\varepsilon}_j = \ln \varepsilon_j$ предполагается выполнение стандартных предположений регрессионного анализа: они постулируются независимыми одинаково распределенными случайными величинами с нулевым математическим ожиданием и постоянной дисперсией.

Использование вместо $\tilde{X}$ эндогенного регрессора $\tilde{x} = \ln x$ при идентификации уравнений спроса (2) на основе метода наименьших квадратов (LS) приводит к оценкам параметров

$$b_j^{LS} = \frac{\widehat{\mathrm{cov}}(\tilde{x}, \tilde{y}_j)}{\hat{\sigma}_{\tilde{x}}^2},\ \tilde{a}_j^{LS} = \hat{\mu}_{\tilde{y}j} - b_j^{LS} \hat{\mu}_{\tilde{x}}, \quad (3)$$

где $\widehat{\mathrm{cov}}(\tilde{x}, \tilde{y}_j)$ — оценка ковариации $\tilde{x}$ и $\tilde{y}_j$; $\hat{\sigma}_{\tilde{x}}^2$ — оценка дисперсии $\tilde{x}$; $\hat{\mu}_{\tilde{y}j}$, $\hat{\mu}_{\tilde{x}}$ — оценки среднего $\tilde{y}_j$, $\tilde{x}$. Оценки (3) являются смещенными и несостоятельными (Айвазян и др., 1985). Эта проблема особенно сильно сказывается на оценках эластичностей спроса по до-

ходу, восстановленных по микроданным. Заметим, что в ходе анализа агрегированных данных временных рядов такие проблемы проявляются в меньшей степени, поэтому в рамках статьи рассматривается именно анализ индивидуальных данных бюджетных обследований в фиксированный момент времени.

2. Современные подходы к идентификации систем спроса

Основным подходом к анализу систем спроса, учитывающим эндогенный характер потребительских расходов, является метод инструментальных переменных (instrumental variables, IV). Он основан на использовании инструмента Z, тесно взаимосвязанного с входным фактором X и не коррелированного с ошибкой ε_j. В случае эндогенного регрессора $\tilde{x}$ оценки параметров системы (2) по методу инструментальных переменных определяются как

$$b_j^{IV} = \frac{\widehat{\mathrm{cov}}(Z, \tilde{y}_j)}{\widehat{\mathrm{cov}}(Z, \tilde{x})},\ \tilde{a}_j^{IV} = \hat{\mu}_{\tilde{y}j} - b_j^{IV}\hat{\mu}_{\tilde{x}}. \quad (4)$$

Простота вычисления и хорошие свойства оценок при выборе подходящего индикатора являются основными достоинствами метода IV. Однако на практике выбор инструмента представляет существенную проблему, которая недостаточно формализована и решается в зависимости от особенностей исходных данных и теоретических соображений. Так, например, в качестве инструмента может быть использован общий заработок главы семьи до вычета взносов и налогов (Blundell et al., 2007).

Использование метода IV нередко сопряжено с проблемой «слабых» инструментов (Эббес, 2007), включение которых в модель не только не приводит к улучшению свойств оценок, но, наоборот, значительно их ухудшает. Хотя разработан ряд тестов на пригодность инструментальных переменных, выбор инструментов можно считать скорее искусством, чем формально регламентированной процедурой.

Другой подход к идентификации систем спроса связан с использованием повторных наблюдений, например, дохода семьи за несколько периодов (Hausman et al., 1995). Такой подход имеет теоретическим основанием идею перманентного дохода: в краткосрочном периоде доход домашнего хозяйства случайно колеблется вокруг истинного значения X. Однако для идентификации системы спроса в этом случае требуется наличие сбалансированных панельных данных, которые, однако, не всегда доступны.

В настоящее время активно развиваются «экономные» альтернативные подходы к оцениванию эконометрических уравнений при наличии эндогенных регрессоров. Один из таких подходов предложен Ebbes et al. (2009) и предполагает инструментальную переменную ненаблюдаемой (латентной). При этом исходят из того, что Z является дискретной величиной (качественной переменной) с заданным числом групп (уровней). Такой подход в целом имеет схожие черты с методами группировки (Wald, 1940), но не требует наличия наблюдаемой переменной группировки, которая должна быть не взаимосвязана с ошибками и упорядочена в соответствии с уровнем благосостояния домашних хозяйств. Латентная инструментальная переменная в свою очередь оценивается одновременно с параметрами уравнений спроса, что является безусловным достоинством, поскольку на практике подобрать переменную группировки, в полной мере отражающую социально-экономическое положение семей, достаточно трудно.

Латентная инструментальная переменная отражает скрытую структуру доходов домашних хозяйств и оценивается методом максимального правдоподобия путем разделения смеси нормальных распределений. Следовательно, подход непригоден в случае относительной однородности доходов населения, когда нельзя выявить их гетерогенную структуру, в частности, при нормальном распределении эндогенного регрессора $\tilde{x}$.

В данной работе рассматривается как раз такой случай более или менее однородной по доходу совокупности населения, для примера взяты домохозяйства пенсионеров. Задача состоит в оценивании неизвестных параметров уравнений (2) в условиях латентного характера истинных (неискаженных случайными погрешностями) потребительских расходов X, которые рассматриваются как случайная величина, не взаимосвязанная со случайными ошибками. Такая постановка отвечает идеям факторного анализа (FA) и не требует привлечения дополнительной информации о значениях дисперсий ошибок, а также инструментальных переменных. Отметим, что в теории потребления представление системы спроса в виде модели FA используется для определения ранга такой системы (Bai, 2003). Здесь же на основе FA разработан подход к оцениванию системы спроса (1).

3. Подход на основе факторного анализа

Предлагаемый подход реализуется в несколько этапов, на каждом из которых находятся оценки определенной группы неизвестных параметров.

Этап 1. Устойчивое оценивание дисперсий ошибок на основе метода факторизации.

Этап 2. Оценивание параметров распределения объясняющей переменной.

Этап 3. Оценивание параметров сдвига и наклона уравнений спроса (2).

Предлагаемый подход накладывает некоторые ограничения на форму распределения переменных. В частности, для дальнейшего анализа существенны следующие предположения:

- ошибки в разных уравнениях не коррелируют, что требуется для применения метода факторизации;
- случайные величины $\tilde{X}$ и $\tilde{\varepsilon}_j$ независимы и нормально распределены, что необходимо для восстановления параметров распределения истинных потребительских расходов.

Что касается первого предположения, то отсутствие взаимосвязи между погрешностями измерения расходов по разным статьям бюджета должно достигаться за счет независимого их определения. В частности, опросный лист домохозяйства должен содержать необходимую детализацию расходов, не допускающую какой-либо подтасовки данных со стороны респондента (перекидывания расходов между статьями). Что касается второго предположения, то использование логнормального распределения дохода часто критикуется (Кирута, Шевяков, 1995; Панкратьев, 1997), поскольку оно недостаточно хорошо согласуется с наблюдаемым в действительности распределением. Однако, как отмечает Айвазян (2012), при исследовании достаточно однородных совокупностей домохозяйств (по источникам формирования доходов, территориальному и социально-профессиональным признакам) правомерно предполагать распределение населения по среднедушевому доходу логнормальным.

Следует отметить, что существуют также некоторые ограничения, накладываемые на число статей бюджета k. Так, система спроса, составленная только из двух благ, например, продовольственных и всех остальных (Шабанов, 2012), не предоставляет достаточно информации для корректного применения FA. Поэтому, с одной стороны, излишнее агрегирование

товарных групп приводит к возможной потере существенной информации о потребительском поведении. Однако, с другой стороны, чрезмерная детализация расходов снижает надежность результатов из-за более хаотичного изменения спроса и проблем с нулевым потреблением, что затрудняет восстановление закономерностей покупательского выбора (Deaton, 1986). Следовательно, на практике нужно выбрать некоторый компромиссный вариант.

Далее более подробно опишем этапы оценивания.

3.1. Проблема оценки дисперсий ошибок

Присутствие в исследуемой системе уравнений (2) ненаблюдаемой переменной $\tilde{X}$ делает перспективным ее оценивание с помощью факторного анализа (Иберла, 1980). В этом случае $\tilde{X}$ — единственный латентный фактор, который полностью описывает вариацию логарифмов расходов на отдельные группы товаров, не связанную с влиянием случайных ошибок. Тогда ковариационная матрица $V = \left\{V_{ij}\right\}$ признаков $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_k$ имеет следующий вид:

$$V = \begin{bmatrix} b_1^2\sigma_{\tilde{X}}^2 + \sigma_{\tilde{\varepsilon}1}^2 & & & \\ b_1 b_2 \sigma_{\tilde{X}}^2 & b_2^2\sigma_{\tilde{X}}^2 + \sigma_{\tilde{\varepsilon}2}^2 & & \\ \vdots & & \ddots & \\ b_1 b_k \sigma_{\tilde{X}}^2 & b_2 b_k \sigma_{\tilde{X}}^2 & \cdots & b_k^2\sigma_{\tilde{X}}^2 + \sigma_{\tilde{\varepsilon}k}^2 \end{bmatrix},$$

где $\sigma_{\tilde{X}}^2$, $\sigma_{\tilde{\varepsilon}j}^2$ — дисперсии $\tilde{X}$ и $\tilde{\varepsilon}_j$ соответственно.

В силу того, что дисперсии ошибок, с которыми фиксируются объясняемые переменные, неодинаковы, то наряду с коэффициентами $b_1, \ldots, b_k$ необходимо оценить и так называемые характерности (Иберла, 1980), в нашем случае полностью обусловленные ошибкой, т. е. матрицу $\Psi = \mathrm{diag}\left\{\sigma_{\tilde{\varepsilon}j}^2\right\}$. Тогда имеет место следующее разложение ковариационной матрицы признаков:

$$V = w^T w + \Psi = U + \Psi,$$

где U — редуцированная ковариационная матрица признаков (Иберла, 1980), w — вектор факторных нагрузок, в нашем случае представляемый в виде

$$w = \left(w_1, w_2, \ldots, w_k\right) = \left(b_1\sigma_{\tilde{X}}, b_2\sigma_{\tilde{X}}, \ldots, b_k\sigma_{\tilde{X}}\right).$$

Для решения проблемы оценки общностей предлагается использовать известный итерационный алгоритм (Иберла, 1980, с. 167). В качестве исходных данных, в отличие от обычного FA, будут выступать выборочные оценки ковариационной матрицы $\hat{V}$ признаков $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_k$.

Полученные таким образом диагональные элементы матрицы Ψ представляют собой оценки дисперсий ошибок, с которыми наблюдаются расходы на отдельные группы товаров. Тогда неизвестные параметры определяются следующим образом:

$$b_j = \frac{w_j}{\sigma_{\tilde{X}}},\ \tilde{a}_j = \mu_{\tilde{y}j} - b_j \mu_{\tilde{X}}, \tag{5}$$

где $\mu_{\tilde{y}j}$, $\mu_{\tilde{X}}$ — математические ожидания $\tilde{y}_j$ и $\tilde{X}$. Вместо значений w_j подставляются компоненты вектора факторных нагрузок $\hat{w}$, оцененные методом факторизации. Оценка $\mu_{\tilde{y}j}$ может быть получена как среднее арифметическое значений $\tilde{y}_j$. Следовательно, для нахождения оценок параметров нужно дополнительно определить моменты распределения $\tilde{X}$.

3.2. Проблема устойчивости

На практике при анализе данных бюджетных обследований приходится сталкиваться с большим разбросом значений признаков и нередко с аномальными наблюдениями, вызванными как значительным завышением или занижением опрашиваемыми членами семьи своих расходов, так и нестандартным экономическим поведением домохозяйства. Независимо от природы таких искажений, их наличие отрицательно сказывается на результатах использования описанного выше подхода вследствие высокой чувствительности выборочных оценок ковариации.

Для того чтобы обнаружить аномальные наблюдения в многомерном пространстве исследуемых признаков, предлагается использовать метод минимума определителя ковариационной матрицы (Rousseeuw, Leroy, 1987), реализация которого осуществляется в несколько шагов (Fauconnier, Haesbroeck, 2009).

Шаг 1. Из всего множества исходных наблюдений *N* выделяется некоторое подмножество заданного объема $h = [\alpha N],\ 0.5 < \alpha < 1$, обеспечивающее наименьшее значение определителя ковариационной матрицы, для чего используется алгоритм FAST-MCD (Rousseeuw, van Driessen, 1999). Обозначим полученную в результате выборку как

$$\left(\tilde{y}^0_{11}, \tilde{y}^0_{21}, \ldots, \tilde{y}^0_{h1}\right), \ldots, \left(\tilde{y}^0_{1k}, \tilde{y}^0_{2k}, \ldots, \tilde{y}^0_{hk}\right).$$

Шаг 2. Рассчитывается грубое приближение оценок моментов по соотношениям:

$$\hat{\mu}^0_{\tilde{y}j} = \frac{1}{h}\sum_{i=1}^{h} \tilde{y}^0_{ij},\ \hat{V}^0_{sj} = \frac{c(\alpha)}{h}\sum_{i=1}^{h}\left(\tilde{y}^0_{is} - \hat{\mu}^0_{\tilde{y}s}\right)\left(\tilde{y}^0_{ij} - \hat{\mu}^0_{\tilde{y}j}\right), \tag{6}$$

где $c(\alpha)$ — коэффициент, обеспечивающий состоятельность оценки ковариационной матрицы при предположении о совместном нормальном распределении исследуемых признаков, задаваемый на основе χ^2-распределения.

Шаг 3. Определяются точки исходных данных, расположенные в пределах эллипса рассеяния, построенного на основе оценок (6). Для этого вычисляется расстояние Махаланобиса d^0_i от каждой точки данных до центра $\left(\hat{\mu}^0_{\tilde{y}1}, \hat{\mu}^0_{\tilde{y}2}, \ldots, \hat{\mu}^0_{\tilde{y}k}\right)$. Индикатор неаномальности *i*-го наблюдения задается как

$$\omega_i = \begin{cases} 1, & \text{если } d^0_i \le \chi^2_{k,0.975}, \\ 0, & \text{иначе,} \end{cases}$$

где $\chi^2_{k,0.975}$ — 97.5%-ный квантиль χ^2-распределения с *k* степенями свободы. В результате очищенная выборка содержит $n = \sum_{i=1}^{N} \omega_i$ наблюдений.

Шаг 4. Вычисляются окончательные значения

$$\hat{\mu}_{\tilde{y}j} = \frac{1}{n}\sum_{i=1}^{N}\omega_i \tilde{y}_{ij}\ ,\ \hat{V}_{sj} = c\left(\frac{n}{N}\right)\frac{1}{n}\sum_{i=1}^{N}\omega_i\left(\tilde{y}_{is} - \hat{\mu}_{\tilde{y}s}\right)\left(\tilde{y}_{ij} - \hat{\mu}_{\tilde{y}j}\right),$$

выступающие в качестве оценок $\mu_{\tilde{y}j}$ и V_{sj} соответственно.

Размер h выборочного подмножества задается исходя из предполагаемой доли выбросов. При этом обычная практика использования $h = 0.5(N + k + 1)$ не всегда себя оправдывает (Fauconnier, Haesbroeck, 2009). Поэтому здесь предлагается применять методы размножения выборки, в частности бутстрап (Эфрон, 1988), для определения характеристик оценок искомых параметров и выбора такого объема h подмножества, который обеспечивает наиболее приемлемую величину этих характеристик (в первую очередь, наименьшие значения стандартных отклонений оценок параметров). Кроме того, использование бутстрапа позволит проанализировать статистические свойства оценок путем восстановления их эмпирического распределения по многократно генерируемым выборкам. Число повторений должно быть достаточным для получения стабильных результатов.

3.3. Проблема оценки латентного фактора

В соответствии со свойством аддитивности, истинные потребительские расходы равны сумме всех не искаженных погрешностями расходов по разным статьям семейного бюджета. Тогда случайная величина X представляет собой сумму логнормальных случайных величин Y_j, и ее распределение может быть аппроксимировано логнормальным (Vanduffel et al., 2005). Его моменты определяются исходя из параметров распределений $\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_k$, средние значения которых, в силу нулевого математического ожидания ошибок, равны $\mu_{\tilde{y}1}, \mu_{\tilde{y}2}, \ldots, \mu_{\tilde{y}k}$, а ковариационная матрица соответствует введенной ранее редуцированной ковариационной матрице $U = \left\{U_{sj},\ s, j = 1,...,k\right\}$. Опираясь на результаты Vanduffel et al. (2005), представим моменты распределения X следующим образом:

$$\mu_X = \sum_{j=1}^{k}\exp\left(\mu_{\tilde{y}j} + \frac{1}{2}U_{jj}\right),\ \sigma_X^2 = \sum_{j=1}^{k}\sum_{s=1}^{k}\left(\exp(U_{js}) - 1\right)\exp\left(\mu_{\tilde{y}j} + \mu_{\tilde{y}s} + \frac{1}{2}\left(U_{jj} + U_{ss}\right)\right), \quad (7)$$

где μ_X, σ_X^2 — математическое ожидание и дисперсия X соответственно. Тогда параметры логнормального распределения X определяются как

$$\mu_{\tilde{X}} = \ln\mu_X - \frac{1}{2}\sigma_{\tilde{X}}^2\ ,\ \sigma_{\tilde{X}}^2 = \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right). \quad (8)$$

Использование в соотношениях (7)–(8) найденных с помощью метода факторизации оценок элементов матрицы U и выборочных средних логарифмов расходов на каждую группу товаров позволяет восстановить распределение потребительских расходов, не искаженных случайными ошибками, и вычислить параметры уравнений (2) на основе выражений (5).

Далее на основе вычислительных экспериментов, проведенных в среде статистической обработки R[2], исследованы свойства оценок, полученных с использованием предложенного подхода, и произведено сравнение с результатами оценивания классическим LS и его устойчивой модификацией — методом усеченных наименьших квадратов (least trimmed squares, LTS).

4. Результаты вычислительных экспериментов

Проводилось два варианта вычислительных экспериментов.

1. Исследование свойств оценок параметров системы спроса при справедливости введенных предположений (в частности, об известной и фиксированной функциональной форме модели).

2. Исследование точности оценивания уравнений спроса при неверной спецификации модели.

В обоих случаях рассматривались системы спроса из трех товаров с мультипликативными логнормальными ошибками. Предполагался разный характер изменения спроса на товары в зависимости от дохода. В первом варианте экспериментов первый товар имел эластичность спроса меньше единицы, а два других — больше единицы. Во втором варианте первый и третий товары выступали предметами роскоши (эластичность превышала единицу), а второй товар — необходимым благом.

Для оценивания методами LS и LTS, игнорирующими проблему эндогенности, в качестве регрессора использовались суммарные расходы по всем статьям семейного бюджета $x = \sum_{j=1}^{3} y_j$. Переменная $\tilde{X}$ во всех экспериментах моделировалась нормально распределенной со средним значением 2.91 и стандартным отклонением 0.365.

В первом случае моделирование данных производилось на основе системы (1) с истинными значениями параметров, равными

$$a_1 = 0.725,\ a_2 = 0.013,\ a_3 = 0.332;\ b_1 = 0.902,\ b_2 = 1.5,\ b_3 = 1.062.$$

Значения параметров подобраны так, чтобы бюджетное ограничение (условие аддитивности) выполнялось с наименьшими погрешностями. Распределение логарифма ошибок задавалось нормальной смесью с функцией распределения вида

$$F(x) = (1-\lambda)\Phi\left(x;0,\sigma^2\right) + \lambda\Phi\left(x;0,k\sigma^2\right),$$

где $\Phi\left(x;0,\sigma^2\right)$ — функция нормального распределения с нулевым математическим ожиданием и дисперсией σ^2. Величина $k > 1$ определяет степень неоднородности данных, связанную с наличием выбросов; $k = 16$, $\sigma^2 = 0.3$ для всех уравнений. Параметр $\lambda \in [0,1]$ характеризует степень засорения данных аномальными наблюдениями и задан равным нулю для базового варианта экспериментов без выбросов (схема 0) и 0.05 для варианта с засорением (схема 1).

[2] http://www.R-project.org/.

Объемы выборок варьировались от 100 до 1000 с шагом 100. Результаты оценивания параметров, осредненные по 500 повторениям, при $N = 500$ представлены в табл. 1. В скобках указаны среднеквадратические отклонения полученных оценок. Результаты экспериментов базовой схемы (без выбросов) при разных объемах выборок только для оценок коэффициента эластичности спроса на третий товар $\hat{b}_3$ отображены графически на рис. 1. Точками отмечены средние значения оценок, отрезками — среднеквадратические отклонения, горизонтальная линия показывает истинное значение параметра.

Таблица 1. Результаты оценивания параметров в первом варианте экспериментов, $N = 500$

Схема-метод	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$	$\hat{b}_1$	$\hat{b}_2$	$\hat{b}_3$
0-FA	0.757 (0.163)	0.015 (0.009)	0.351 (0.091)	0.895 (0.072)	1.516 (0.227)	1.055 (0.089)
0-LS	0.491 (0.046)	0.097 (0.016)	0.435 (0.052)	1.010 (0.032)	0.794 (0.056)	0.947 (0.041)
1-FA	0.756 (0.178)	0.015 (0.009)	0.357 (0.099)	0.896 (0.078)	1.515 (0.235)	1.050 (0.096)
1-LTS	0.425 (0.039)	0.147 (0.039)	0.373 (0.065)	1.070 (0.030)	0.660 (0.084)	1.016 (0.058)

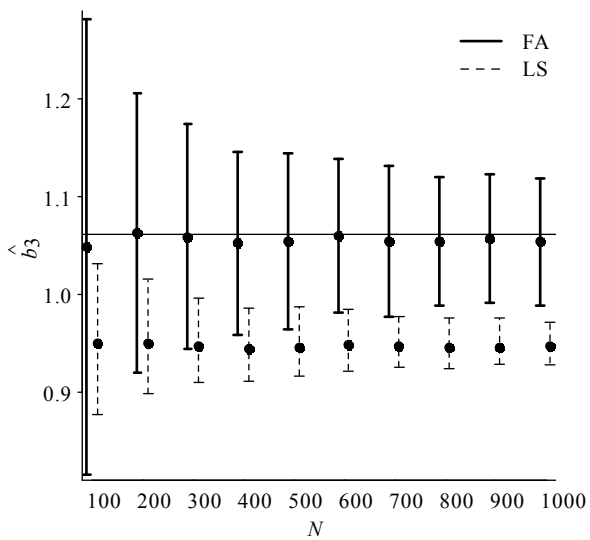


Рис. 1. Характеристики оценки параметра наклона в зависимости от объема выборки

Из рисунка 1 очевидно, что оценки, полученные по LS, характеризуются систематическим смещением, не уменьшающимся с ростом числа наблюдений. В то же время смещение оценок, полученных с помощью FA, практически нулевое. Однако оценки FA обладают гораздо большей дисперсией, особенно существенной на малых выборках. С ростом объема анализируемых данных разница между дисперсиями оценок LS и FA лишь немного сокращается. При $N = 1000$ стандартные отклонения оценок, полученных с помощью FA, примерно в 2.3, 3.6 и 2.3 раза больше, чем по LS, для первого, второго и третьего уравнения соответственно.

Из таблицы 1 видно, что отмеченные особенности характерны для оценок всех трех уравнений. Ситуация принципиально не меняется и в схеме 1 при наличии аномальных наблюдений. При этом использование метода минимума определителя ковариационной матрицы позволяет и при наличии и при отсутствии выбросов получить примерно одинаковые средние оценки параметров. Тогда как оценки LTS более существенно отличаются от LS-оценок.

Стоит отметить, что проверка согласия эмпирического распределения оценок параметров, полученных с помощью FA, с нормальным распределением указывает на то, что только в редких случаях нулевая гипотеза (о согласии) не отвергается. Для сравнения критерий Шапиро–Уилка подтвердил нормальность распределения LS-оценок на 95%-ном доверительном уровне во всех экспериментах. Следовательно, гипотеза о нормальности предложенных FA-оценок не обоснована, и для исследования их свойств предлагается использовать методы размножения выборки.

Как и оценки параметров уравнений спроса, оценки параметров распределения $\tilde{X}$ незначительно смещены. Для выборки объема 500 оценка среднего составила 2.911 ± 0.025, оценка стандартного отклонения 0.366 ± 0.03. Следовательно, при предположении о функциональной форме модели предложенный подход позволяет получить практически несмещенные оценки эластичности спроса и достаточно точно восстановить параметры распределения потребительских расходов.

Однако модель с постоянной эластичностью обладает рядом недостатков, в частности, в силу нелинейности по доходу не выполняется свойство аддитивности. В то же время многие теоретические системы спроса являются нелинейными, и этот недостаток часто игнорируется (Brown, Deaton, 1972). Вдобавок к этому существует немало эмпирических доказательств того, что эластичность спроса по доходу не остается постоянной, а убывает с ростом дохода (Brown, Deaton, 1972). Поэтому дополнительно проанализированы последствия выбора неверной спецификации модели.

В качестве истинной модели во втором варианте экспериментов принята линейно-логарифмическая модель спроса

$$Y_j = a_j X\left(\ln X - b_j\right), \ j = 1,2,3. \tag{9}$$

Это аналогично представлению долей расходов как линейной функции от логарифма дохода, так называемой модели Уоркинга–Лизера, используемой, в частности, в (Hausman et al., 1995). Истинные значения параметров задавались следующим образом:

$$a_1 = 0.1, \ b_1 = 1.27; \ a_2 = -0.23, \ b_2 = 5.58; \ a_3 = 0.13, \ b_3 = 1.203077.$$

Значения параметров подобраны так, чтобы выполнялись свойства аддитивности, неотрицательности функций спроса, убывающей и положительной эластичности спроса по доходу. Подробные условия, накладываемые на параметры и область определения линейно-логарифмических кривых Энгеля для обеспечения перечисленных свойств, можно найти в (Тимофеев, Колесникова, 2009). В генерируемой системе спроса эластичность спроса по доходу на первый и второй товары варьируется в широком диапазоне примерно от 1.3 до 2.6. Второй товар обладает эластичностью спроса по доходу от 0.4 до 0.7. Следовательно, второе уравнение характеризуется наименьшей степенью отклонения от модели (1) с постоянной эластичностью.

Дисперсии логарифмов ошибок, мультипликативно включаемых в уравнения (9), задавались одинаковыми и равными 0.38, засорения данных аномальными наблюдениями не производилось. Качество оценивания модели определялась с помощью средней ошибки аппроксимации:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{Y_i - \hat{y}_i}{Y_i}\right|,$$

где Y_i, $\hat{y}_i$ — истинное и прогнозное значение расходов i-го домохозяйства. При LS-оценивании предполагалось, что истинная модель (9) известна, в то время как при использовании подхода на основе FA постулировалась модель (1). Результаты, осредненные по 500 повторениям, представлены в табл. 2. В скобках приведены стандартные отклонения.

Таблица 2. Показатель точности оценивания модели во втором варианте экспериментов

Уравнение	Метод	$N = 100$	$N = 300$	$N = 500$	$N = 700$	$N = 1000$
1	FA	0.177 (0.159)	0.091 (0.059)	0.075 (0.038)	0.065 (0.033)	0.058 (0.030)
	LS	0.292 (0.126)	0.267 (0.115)	0.278 (0.079)	0.271 (0.079)	0.275 (0.069)
2	FA	0.069 (0.044)	0.036 (0.021)	0.031 (0.016)	0.025 (0.012)	0.023 (0.011)
	LS	0.123 (0.043)	0.119 (0.033)	0.121 (0.028)	0.120 (0.025)	0.121 (0.022)
3	FA	0.147 (0.139)	0.079 (0.040)	0.065 (0.030)	0.056 (0.028)	0.051 (0.023)
	LS	0.244 (0.112)	0.218 (0.084)	0.215 (0.077)	0.210 (0.065)	0.211 (0.06)

Из таблицы 2 видно, что даже при неверной спецификации предложенный подход обеспечивает лучшее качество оценивания, чем LS при известной функциональной форме модели. При этом с ростом числа наблюдений точность оценивания модели с помощью FA значительно возрастает — средняя ошибка аппроксимации сокращается практически в три раза, если сравнивать результаты при минимальном и максимальном объеме выборки. Точность LS-оценивания существенно не зависит от объема выборки. Как и ожидалось, лучше всего оценено второе уравнение, наиболее близкое к модели с постоянной эластичностью.

Отметим также, что оценка дисперсии латентного признака во втором варианте экспериментов уже несколько смещена. Так, для выборки объемом 500 элементов оценка среднего $\tilde{X}$ составила 2.914 ± 0.026, оценка стандартного отклонения — 0.388 ± 0.029. При всех исследованных объемах выборок оценка дисперсии смещена вправо относительно истинного значения, смещение сокращается с ростом N. Следовательно, из-за завышенной оценки параметра $\sigma^2_{\tilde{X}}$ в случае неверной спецификации модели оценка модального значения неискаженных потребительских расходов будет занижена, а оценка среднего окажется завышенной.

Перейдем к решению практической задачи с использованием предложенного подхода. Представленный далее анализ эмпирических данных производился в среде статистической обработки R.

5. Оценка системы спроса домашних хозяйств пенсионеров

В более ранних работах автора исследовались проблемы анализа и моделирования потребления продуктов питания малообеспеченным населением Новосибирской области, в частности пенсионерами (Колесникова и др., 2009; Тимофеев, Колесникова, 2010). При решении этих задач важную роль играет определение эластичности продовольственного спроса по доходу. Применим для этой цели предложенный подход к идентификации уравнений спроса от дохода.

Для анализа использованы индивидуальные данные о домохозяйствах, полученные по итогам выборочных обследований, проводимых Росстатом, в 1–4 кварталах 2009 года[3]. Для формирования однородной совокупности отбор наблюдений осуществлялся по следующим признакам:

- территория — Новосибирская область;
- тип населенного пункта — городской;
- число наличных лиц в домохозяйстве равно числу лиц пенсионного возраста.

В результате объем выборки N составил 609 домохозяйств. В качестве статей расходов рассматривались укрупненные группы: продукты питания (PROD), непродовольственные товары (NEPROD), услуги (USLUG). Расходы на покупку алкогольных напитков и на питание вне дома исключены из анализа, поскольку отражают скорее индивидуальные особенности членов семьи, чем закономерности потребления. Поэтому объясняющей переменной выступали суммарные расходы по трем рассмотренным категориям, т. е. потребительские расходы без учета затрат на алкоголь и общественное питание. Далее для краткости будем называть их просто потребительскими расходами.

По представленным в логарифмической шкале значениям расходов построены корреляционные поля, изображенные на рис. 2–4. Из рисунков видно, что в исходных данных присутствуют аномальные наблюдения, доля которых по приблизительным оценкам составляет 5%. Визуальный анализ не позволяет сделать однозначных выводов о тесноте зависимости, поэтому проведено оценивание формы корреляционного поля путем построения контурных эллипсов (эллипсов рассеяния) (Тимофеев и др., 2013).

Для того чтобы продемонстрировать, насколько сильно искажается характер взаимосвязи между изучаемыми переменными из-за наличия грубых ошибок наблюдения, построены два варианта эллипсов рассеяния:

- классический (classical) при игнорировании проблемы выбросов;
- устойчивый (robust) на основе оценок моментов эмпирического распределения признаков, полученных методом минимума определителя ковариационной матрицы.

При построении эллипсов доверительная вероятность задавалась равной 0.975. Начальный размер выборочного подмножества определялся как $h = [0.95N] = 578$. Из рис. 2–4 видно, что использование робастных оценок ковариационной матрицы позволило выявить более тесную связь между признаками, о чем можно судить по меньшей площади устойчивого эллипса по сравнению с классическим. Кроме того, после исключения из выборки аномальных наблюдений использование критерия Шапиро–Уилка показало, что нулевая гипотеза о нормальности распределения каждого из признаков не отвергается с доверительной вероятностью 0.95.

[3] См. Микроданные выборочных обследований бюджетов домашних хозяйств, http://obdx.gks.ru/.

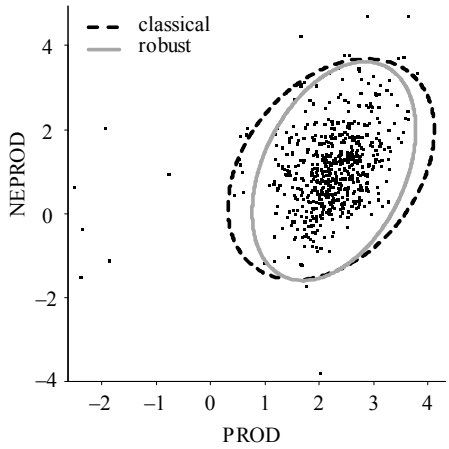


Рис. 2. Корреляционное поле продовольственных и непродовольственных расходов

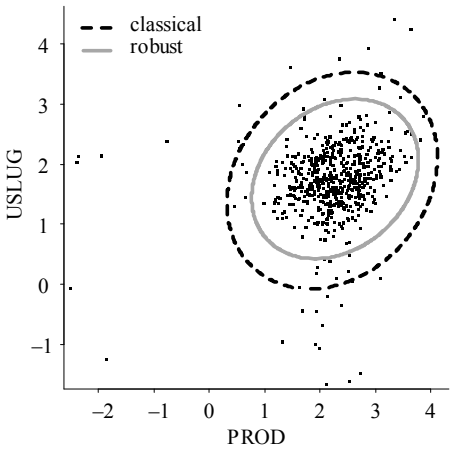


Рис. 3. Корреляционное поле продовольственных расходов и расходов на оплату услуг

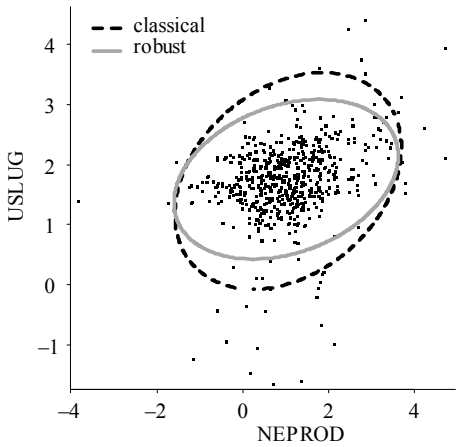


Рис. 4. Корреляционное поле непродовольственных расходов и расходов на оплату услуг

С помощью методов размножения выборки (бутстрапа) проведены дополнительные исследования, направленные на выбор оптимального значения размера оценочного подмножества h для использования метода минимума ковариационной матрицы. С этой целью было сгенерировано 1000 псевдовыборок, для каждой из которых производилось оценивание эластичности спроса с помощью предложенного подхода при заданных исходных значениях доли α наблюдений, включенных в оценочное подмножество (см. шаг 1 алгоритма устойчивого оценивания ковариационной матрицы). Далее по всем повторениям вычислялись средние значения оценок параметров, представленные в табл. 3. В скобках указаны их среднеквадратические отклонения.

Таблица 3. Результаты оценивания параметров в зависимости от объема подмножества

	h				
	304 ($\alpha = 0.5$)	548 ($\alpha = 0.9$)	578 ($\alpha = 0.95$)	593 ($\alpha = 0.975$)	609 ($\alpha = 1$)
$\hat{b}_1$	1.03 (0.18)	0.92 (0.09)	0.89 (0.09)	0.80 (0.09)	0.84 (0.12)
$\hat{b}_2$	1.73 (0.42)	1.86 (0.20)	1.89 (0.18)	1.90 (0.19)	1.76 (0.25)
$\hat{b}_3$	0.44 (0.09)	0.56 (0.07)	0.60 (0.07)	0.61 (0.08)	0.70 (0.12)

Очевидно, что наименьшие отклонения (наилучшая точность оценивания) достигаются при размере оценочного подмножества $0.95N$, которое и было задано для получения итоговых результатов, приведенных в табл. 4 (средние значения и стандартные отклонения по 2000 псевдовыборкам).

Таблица 4. Результаты оценивания на основе FA, усредненные по псевдовыборкам

Расходы	$\hat{a}$	$\hat{b}$	$\hat{\Psi}$	$\hat{w}^2$	$\hat{\Psi}/\hat{w}^2$
PROD	0.725 (0.188)	0.902 (0.087)	0.209 (0.024)	0.104 (0.026)	2.188 (0.766)
NEPROD	0.013 (0.008)	1.884 (0.181)	0.480 (0.091)	0.451 (0.106)	1.161 (0.473)
USLUG	1.037 (0.210)	0.598 (0.071)	0.194 (0.017)	0.045 (0.011)	4.574 (1.412)

По величине $\hat{\Psi}/\hat{w}^2$ можно судить о том, насколько существенны погрешности измерения расходов домохозяйств. Значение этого показателя интерпретируется как отношение дисперсии ошибки признака к дисперсии его истинного значения. Наибольшее расхождение между ними наблюдается для расходов на оплату услуг, где дисперсия погрешности в 4.6 раза превышает дисперсию истинных расходов. Для продовольственных товаров — в 2.2 раза, для непродовольственных — в 1.2 раза. Такая существенная величина дисперсии случайных ошибок согласуется с выводами, сделанными ранее в (Hausman et al., 1995). В условиях такого сильного искажения данных бюджетных обследований имеет место проблема малой величины коэффициента детерминации восстанавливаемых уравнений зависимости (Hausman, 2001). Это служит еще одним веским доводом в пользу необходимости корректного учета погрешностей в объясняющей переменной.

С использованием сгенерированных псевдовыборок осуществлено сравнение результатов оценивания параметров уравнений спроса с помощью различных методов: FA, LS и IV. Оценки параметров по LS получены из соотношений (3), по методу IV — исходя из выражений (4). В качестве инструмента выбран дециль по среднедушевым располагаемым ресурсам. На рисунке 5 координатам пересечения отрезков соответствуют средние значения оценок параметров уравнений (1), а половина длины отрезка равна величине стандартного отклонения.

Из рисунка 5 видно, что применение LS обеспечивает наименьшую дисперсию оценок параметров, что соответствует теоретическим представлениям и экспериментальным исследованиям, проведенным, например, в (Тимофеев и др., 2013). Однако оценки (3) сильно смещены относительно оценок, построенных другими методами. Причем в силу корреляции между ошибками разных уравнений, возникающей при использовании эндогенного регрессора, наблюдается не только известный эффект смещения оценок эластичности спроса к нулю (Hausman et al., 1995), но и положительное смещение, а именно переоценка эластичности спроса на продовольственные товары и услуги. Оценки эластичностей, полученные по методу IV и с помощью FA, почти сопоставимы по вариации и близки по средней величине. При этом средние оценки IV отклоняются относительно средних оценок FA в том же направлении, что и LS.

Благодаря использованию методов размножения выборки удается получить информацию об эмпирических распределениях оценок эластичности. В таблице 5 значения оценок эластичностей, построенных предложенным методом, для удобства восприятия и интерпретации представлены в сгруппированном виде. По полученным значениям можно судить о характере изменения расходов домохозяйств с ростом дохода. Значения, не превышающие единицы, свидетельствуют о первостепенной необходимости определенной категории товаров. К таким здесь, безусловно, отнесены услуги. Отсюда становится понятно, что речь идет преимущественно о таких базовых расходах, как оплата жилья. Существенная часть расходов малообеспеченных пенсионеров приходится именно на эту статью: 69–73% от расходов на оплату услуг для первых четырех децильных групп.

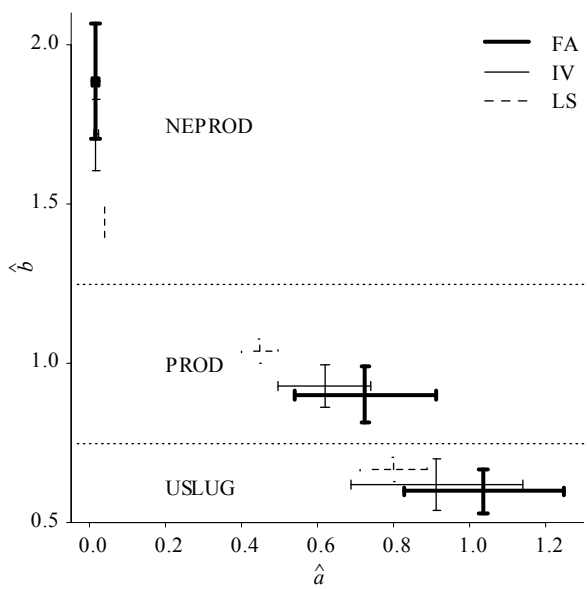


Рис. 5. Результаты оценивания параметров разными методами

Таблица 5. Распределения оценок эластичности, построенных на основе FA, %

Расходы	Оценка эластичности		
	<1	[1; 2]	>2
PROD	87	13	0
NEPROD	0	74.65	25.35
USLUG	100	0	0

Расходы на питание в большинстве случаев (табл. 5) также рассматриваются как необходимые, но в 13% случаев они растут быстрее, чем общие потребительские расходы. Это согласуется с выводом, полученным ранее (Денисов и др., 2013) по результатам непараметрической идентификации уравнения продовольственных расходов: для некоторых децильных групп населения эластичность спроса на продукты питания превышает единицу.

Наконец, непродовольственные товары выступают предметами роскоши, но характер изменения расходов на эту категорию также различен. Так, примерно в 75% случаев доля непродовольственных товаров в бюджете растет более медленными темпами, чем общие потребительские расходы, тогда как в остальных случаях рост доли происходит более быстрыми темпами. Это указывает на то, что показное (демонстративное) потребление, не обусловленное рациональными нуждами, наиболее сильно проявляется именно в группе непродовольственных расходов.

Преимуществом предложенного подхода к оцениванию уравнений спроса является также возможность восстановления параметров распределения истинного (неискаженного) дохода. В таблице 6 проведено сравнение наблюдаемого (искаженного) и восстановленного (истинного) распределения потребительских расходов по значениям квантилей. Осреднение произведено по 2000 псевдовыборок. В скобках указаны стандартные отклонения.

Таблица 6. Характеристики наблюдаемого (искаженного) и восстановленного (истинного) распределения потребительских расходов

Потребительские расходы	Первый квартиль	Мода	Медиана	Среднее значение	Третий квартиль
Наблюдаемые	14.255 (0.274)	15.444 (0.717)	18.990 (0.405)	22.984 (0.658)	25.376 (0.651)
Восстановленные	14.500 (0.300)	16.228 (0.345)	18.414 (0.373)	19.618 (0.520)	23.393 (0.789)

Наиболее близкими можно считать значения 25%-ного квантиля и медианы. Это говорит о том, что представители малообеспеченных домохозяйств в меньшей степени склонны искажать свои расходы. В то же время в среднем (по медиане) случайные ошибки компенсируются. Остальные характеристики распределений довольно сильно различаются. При этом оценка математического ожидания, полученная по восстановленным параметрам распределения истинных потребительских расходов, в среднем меньше наблюдаемой почти на 3.4 тыс. руб. Что касается оценки моды, то, наоборот, восстановленные значения в среднем превышают наблюдаемые, но не столь значительно. Это свидетельствует о том, что основное искажение данных о потребительских расходах приходится на домохозяйства пенсионеров с высоким доходом.

6. Заключение

Предложенный подход направлен на решение проблемы искажения оценок эластичности спроса по доходу в условиях эндогенности потребительских расходов, а также дополняет разработанную автором методику оценки экономического положения населения. Проанализированной постановке задачи присущи некоторые ограничения. Наиболее существенны из них следующие.

1. Фиксация функциональной формы зависимости между расходами на отдельные категории товаров и потребительскими расходами. Несмотря на традиционность использования двойной логарифмической модели, ей присущи определенные недостатки. В частности, жесткое фиксирование постоянной эластичности спроса по доходу для всех слоев населения не позволяет обнаружить закономерности изменения эластичности с ростом дохода.

2. Необходимость выделения гомогенной группы потребителей может приводить к существенному сокращению объема выборки, что уменьшает надежность оценок эластичности спроса.

3. Предположение о логнормальном распределении потребительских расходов часто критикуется, в то время как использование других подходящих классов распределений (McDonald, 1984) позволило бы лучше описать расслоение населения по материальному благосостоянию.

Преодоление такой ограниченности возможно в рамках развиваемых автором подходов к непараметрическому и полупараметрическому оцениванию моделей с ошибками в объясняющих переменных (Денисов и др., 2013; Timofeeva, 2014). Эти подходы обладают большей гибкостью, но предполагают известной дисперсию ошибок входного фактора (или ее отношение к дисперсии ошибки отклика), в то время как подход на основе факторного анализа свободен от такого предположения и позволяет найти оценки дисперсии ошибок. Комбинирование этих подходов открывает более широкие возможности для развития методов оценивания эластичностей спроса по данным выборочных бюджетных обследований домохозяйств в направлении повышения их адекватности и применимости на практике.

Список литературы

Айвазян С. А. (2012). *Анализ качества и образа жизни населения: эконометрический подход*. М.: Наука.

Айвазян С. А., Енюков И. С., Мешалкин Л. Д. (1985). *Прикладная статистика: Исследование зависимостей*. М.: Финансы и статистика.

Бирман И. (2000). Уровень жизни: проблемы измерения. *Экономическая наука современной России*, 2, 35–52.

Денисов В. И., Тимофеева А. Ю., Хайленко Е. А., Бузмакова О. И. (2013). Устойчивое оценивание нелинейных структурных зависимостей. *Сибирский журнал индустриальной математики*, 4, 47–60.

Иберла К. (1980). *Факторный анализ*. М.: Статистика.

Интрилигатор М. (1975). *Математические методы оптимизации и экономическая теория*. М.: Прогресс.

Кирута А., Шевяков А. (1995). Дифференцированный баланс доходов и потребления населения: новые аспекты теории и практических применений. *Вопросы статистики*, 7, 3–13.

Колесникова А. Ю., Скосырский В. А., Тимофеев В. С., Храмцова О. В., Чистяков В. М. (2009). Оценка доступности социально-значимых товаров для малообеспеченного населения. *Сибирская финансовая школа*, 2, 24–30.

Панкратьев Н. (1997). Статистические методы анализа экономического расслоения населения. *Вопросы статистики*, 1, 22–28.

Сычева В. С. (1998). Исторический очерк бюджетных исследований в Западной Европе и США. *Социологические исследования*, 4, 48–56.

Тимофеев В. С., Колесникова А. Ю. (2009). Идентификация моделей зависимости спроса от дохода в рамках неоклассической теории. *Доклады Академии наук высшей школы РФ*, 2, 51–65.

Тимофеев В. С., Колесникова А. Ю. (2010). Исследование устойчивости в задачах оптимизации потребительского выбора. *Доклады Академии наук высшей школы РФ*, 1, 47–61.

Тимофеев В. С., Щеколдин В. Ю., Тимофеева А. Ю. (2013). Идентификация зависимостей признаков стохастической природы на основе регрессии Деминга. *Информатика и ее применения*, 7 (2), 60–68.

Шабанов В. Л. (2012). Уровень и уклад жизни сельского населения: оценка трансформации на основе анализа структуры потребления. *Вопросы статистики*, 7, 72–77.

Эббес П. (2007). Инструментальные переменные и эндогенность: нетехнический обзор. *Квантиль*, 2, 3–20.

Эфрон Б. (1988). *Нетрадиционные методы многомерного статистического анализа*. М.: Финансы и статистика.

Bai J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71 (1), 135–171.

Blundell R., Chen X., Kristensen D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, 75 (6), 1613–1669.

Brown J. A. C., Deaton A. S. (1972). Surveys in applied economics: Models of consumer behaviour. *Economic Journal*, 82, 1145–1236.

Deaton A. S. (1986). Demand analysis. In: *Handbook of Econometrics*, vol. 3. New-Holland: Amsterdam and New York, 1767–1839.

Ebbes P., Wedel M., Boeckenholt U. (2009). Frugal IV alternatives to identify the parameter for an endogenous regressor. *Journal of Applied Econometrics*, 24 (3), 446–468.

Fauconnier C., Haesbroeck G. (2009). Outliers detection with the minimum covariance determinant estimator in practice. *Statistical Methodology*, 6, 363–379.

Hausman J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspective*, 15 (4), 57–67.

Hausman J., Newey W. K., Powell J. L. (1995). Nonlinear errors in variables: Estimation of some Engel curves. *Journal of Econometrics*, 65 (1), 205–233.

McDonald J. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52 (3), 647–663.

Rousseeuw P. J., Leroy A. M. (1987). *Robust regression and outlier detection*. New York: Wiley-Interscience.

Rousseeuw P. J., van Driessen K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics,* 41 (3), 212–223.

Timofeeva A. Yu. (2014). Orthogonal regression for nonparametric estimation of errors-in-variables models. *International Journal of Mathematical, Computational, Physical and Quantum Engine*ering, 8 (8), 1051–1055.

Vanduffel S., Hoedemakers T., Dhaene J. (2005). Comparing approximations for risk measures of sums of nonindependent lognormal random variables. *North American Actuarial Journal*, 9 (4), 71–82.

Wald A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11 (3), 284–300.

References

Ajvazjan S. A. (2012). *Analiz kachestva i obraza zhizni naselenija: jekonometricheskij podhod*. M.: Nauka.

Ajvazjan S. A., Enjukov I. S., Meshalkin L. D. (1985). *Prikladnaja statistika: Issledovanie zavisimostej*. M.: Finansy i statistika.

Birman I. (2000). The level of life: The measuring problems. *Economics of Contemporary Russia*, 2, 35–52.

Denisov V. I., Timofeeva A. Ju., Hajlenko E. A., Buzmakova O. I. (2013). Robust estimation of nonlinear structural models. *Journal of Applied and Industrial Mathematics*, 8 (1), 28–39.

Iberla K. (1980). *Faktornyj analiz*. M.: Statistika.

Intriligator M. (1975). *Matematicheskie metody optimizacii i jekonomicheskaja teorija*. M.: Progress.

Kiruta A., Shevjakov A. (1995). Differencirovannyj balans dohodov i potreblenija nasele-nija: novye aspekty teorii i prakticheskih primenenij. *Voprosy statistiki*, 7, 3–13.

Kolesnikova A., Skosyrsky V., Timofeev V., Khramtsova O., Chistyakov V. (2009). Assessment of essential goods availability for population with poor financial security. *Siberian Financial School*, 2, 24–30.

Pankrat'ev N. (1997). Statisticheskie metody analiza jekonomicheskogo rassloenija naselenija. *Voprosy statistiki*, 1, 22–28.

Sycheva B. C. (1998). Istoricheskij ocherk bjudzhetnyh issledovanij v Zapadnoj Evrope i SShA. *Sociologicheskie issledovanija*, 4, 48–56.

Timofeev V. S., Kolesnikova A. Ju. (2009). Identification of demand models of income with a view to neoclassical economics. *Proceedings of the Russian Higher School Academy of Sciences*, 2, 51–65.

Timofeev V. S., Kolesnikova A. Ju. (2010). The sensitivity analysis for optimization problem of consumer choice. *Proceedings of the Russian Higher School Academy of Sciences*, 1, 47–61.

Timofeev V. S., Schekoldin V. Yu., Timofeeva A. Yu. (2013). The error-in-variables model identification on the basis of Deming's approach. *Informatics and Applications*, 7 (2), 60–68.

Shabanov V. L. (2012). Living standard and way of life of the rural population: Transformation estimation on the basis of consumption structure analysis. *Voprosy Statistiki*, 7, 72–77.

Ebbes P. (2007). A non-technical guide to instrumental variables and regressor-error dependencies. *Quantile*, 2, 3–20.

Jefron B. (1988). *Netradicionnye metody mnogomernogo statisticheskogo analiza*. M.: Finansy i statistika.

Bai J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71 (1), 135–171.

Blundell R., Chen X., Kristensen D. (2007). Semi-nonparametric IV estimation of shape-invariant Engel curves. *Econometrica*, 75 (6), 1613–1669.

Brown J. A. C., Deaton A. S. (1972). Surveys in applied economics: Models of consumer behaviour. *Economic Journal*, 82, 1145–1236.

Deaton A. S. (1986). Demand analysis. In: *Handbook of Econometrics*, vol. 3. New-Holland: Amsterdam and New York, 1767–1839.

Ebbes P., Wedel M., Boeckenholt U. (2009). Frugal IV alternatives to identify the parameter for an endogenous regressor. *Journal of Applied Econometrics*, 24 (3), 446–468.

Fauconnier C., Haesbroeck G. (2009). Outliers detection with the minimum covariance determinant estimator in practice. *Statistical Methodology*, 6, 363–379.

Hausman J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. *Journal of Economic Perspective*, 15 (4), 57–67.

Hausman J., Newey W. K., Powell J. L. (1995). Nonlinear errors in variables: Estimation of some Engel curves. *Journal of Econometrics*, 65 (1), 205–233.

McDonald J. (1984). Some generalized functions for the size distribution of income. *Econometrica*, 52 (3), 647–663.

Rousseeuw P. J., Leroy A. M. (1987). *Robust regression and outlier detection*. New York: Wiley-Interscience.

Rousseeuw P. J., van Driessen K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics,* 41 (3), 212–223.

Timofeeva A. Yu. (2014). Orthogonal regression for nonparametric estimation of errors-in-variables models. *International Journal of Mathematical, Computational, Physical and Quantum Engine*ering, 8 (8), 1051–1055.

Vanduffel S., Hoedemakers T., Dhaene J. (2005). Comparing approximations for risk measures of sums of nonindependent lognormal random variables. *North American Actuarial Journal*, 9 (4), 71–82.

Wald A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11 (3), 284–300.