

Варианты измерения личных доходов на данных РМЭЗ: оценка смещений

Несогласованность или отсутствие ответов в опросах обычно воспринимаются в негативном ключе, могут трактоваться как ошибки. Как правило, такие анкеты или переменные с пропущенными значениями исключаются из анализа. Но ответы на одни вопросы можно использовать для «ремонта» пропусков или оценки надежности ответов на другие. К такому приему часто обращаются в анализе переменных личных и семейных доходов, если в анкету включена батарея вопросов о доходах из отдельных источников, а также об общей сумме доходов. Для российских экономистов и социологов базой данных, позволяющих осуществлять сравнение ответов и внутреннюю импутацию пропущенных ответов о доходах, является Российский мониторинг экономического положения и здоровья населения НИУ «Высшая школа экономики» (РМЭЗ НИУ ВШЭ). В вопросниках РМЭЗ о доходах спрашивают в «коротком» и «длинном» форматах [Duncan, Petersen 2001; Jansen et al., 2013]: в индивидуальной анкете есть вопросы о пяти видах личных доходов, полученных за последние 30 дней (выплаты на основном и дополнительном местах работы, приработки, пенсии и пособия по безработице), а также вопросы о средней заработной плате на основном месте работы за последний год или меньший фактически отработанный период (для работающих) и общей сумме личных доходов (для всех респондентов старше 15 лет).¹ Например, в одном из исследований мобильности заработных плат основной переменной была как раз среднемесячная заработная плата по основному месту работы за последние 12 месяцев, а для тех респондентов, у которых значения данной переменной отсутствовали, брались значения заработной платы за последние 30 дней [Лукиянова 2017, с. 297]. Но в таких примерах часто за текстом остается описание методической стороны работы с переменными доходов: каковы масштабы пропусков данных? какие возможности оценки смещений и коррекции значений переменных дает база данных? каковы последствия импутации значений переменных? как меняется распределение после заполнения пропусков? на каких отрезках переменной сконцентрированы пропущенные и импутируемые значения или они равномерно распределены в области низких, средних и высоких значений? какие именно поступления подразумевают респонденты, отвечая на вопросы о суммарных личных доходах?

В докладе будет представлено сравнение нескольких вариантов определения личных доходов на данных Российского мониторинга экономического положения и здоровья населения (2017 г.) с оценкой различий распределений этих переменных.

Определение смещений и ошибок в данных о доходах подразумевает наличие «эталона», и, как правило, для этого используют административные данные. На микроуровне ошибка фиксируется через сравнение значений, полученных в опросах от каждого респондента, с данными об этом же индивиде из различных документальных источников. При оценке на макроуровне сравнивают характеристики *совокупности* индивидов, определенные на основе данных из разных (опросных и административных) источников, например, средние значения, суммы и доли доходов из отдельных источников, показатели дифференциации и др. К коррекции показателей доходов населения на макроуровне прибегает, к примеру, Росстат. Расчеты распределения населения по уровню среднедушевых денежных доходов осуществляются с применением методов имитационного моделирования: эмпирическое распределение, полученное на данных

¹ Аналогичным образом спрашивают о доходах домохозяйства. Но для отдельных видов доходов просят назвать их величину, а для полученных в натуральном виде – дать денежную оценку, то суммарный доход домохозяйства подразумевает в анкете РМЭЗ только денежную оценку.

выборочных обследований, преобразуется в ряд распределения, соответствующий значению группировочного признака в генеральной совокупности (величина среднедушевого денежного дохода, полученная на основе баланса денежных доходов и расходов населения) [Методика расчета...]. Значение среднедушевого дохода на данных ОБДХ² ниже, чем получается при обобщении в балансе денежных доходов и расходов населения³ данных статистической отчетности различных субъектов и экспертной дооценки неучтенных доходов [Методологические положения...].

Как правило, административные данные, используемые для «внешней» оценки опросных данных, – это налоговые регистры, а также отчетность подразделений социального обеспечения. В целом все сравнения административных и опросных данных обнаруживают, во-первых, более низкие средние и медианные значения показателей доходов, полученные в опросах. Статистические эксперименты с разными измерениями доходов не меняют этого вывода (например, см. [Kreiner et al. 2015]). Во-вторых, в разных частях распределения искажения в большую и меньшую сторону наблюдаются с разной вероятностью: при низких доходах выше вероятность, что респонденты будут завышать свой доход (сообщаемый в опросе по сравнению с административными данными), с увеличением доходов, наоборот, выше вероятность занижить сообщаемый доход, что интегрально приводит к ошибке «скатывания в середину» (mean-reverting errors) [Angel et al. 2018; Kim, Tamborini 2014].

Конечно, сравнения опросных и административных данных не лишены ошибок и ограничений, не только из-за проблемы конфиденциальности. Одну из работ, посвященных сопоставлению опросных и административных данных о заработной плате, Джон М. Эбоуд и Марта Х. Стинсон начинают с тезиса, что «предположение о том, что некоторые данные содержат ошибки, в то время как другие данные – нет, является фундаментально ошибочным. Хотя процесс генерации ошибок может отличаться в двух источниках, ни один источник не может быть полностью безошибочным» [Abowd, Stinson 2013, p. 1451]. Они называют три причины, почему административные данные при таких сравнениях могут не являться истинными. Во-первых, измерения доходов в регистрах и в опросах могут различаться концептуально. Во-вторых, административные данные также могут быть искажены под действием «человеческого фактора». В-третьих, возможны ошибки в процессе сопоставления данных из нескольких источников, например, из-за ошибочно указанных идентификационных номеров [Abowd, Stinson 2013, 1460–1461].

В этом плане не стоит переоценивать административные данные о доходах, так как они фиксируют только официально полученные средства, поэтому данные обследований могут быть более точными для имеющих «неформальные» доходы. Тем не менее сопоставление информации о доходах из административных и опросных источников не только идентифицирует общие тенденции в смещениях опросных данных, но и привлекает внимание к методологии компонентного измерения доходов, не дает впасть в иллюзию существования идеальных данных о доходах.

В исследовании, которое планируется представить в докладе, осуществляется «внутренне сопоставление» ответов о личных доходах, содержащееся в базе данных одного обследования. При работе с РМЭЗ для идентификации экономического статуса респондентов через личный доход, как правило, оперируют ответами на вопрос «Попробуйте вспомнить, сколько всего денег в течение последних 30 дней Вы лично получили. Пожалуйста, посчитайте всё: зарплату, пенсии, премии, прибыли, пособия,

² Доходы, расходы и потребление домашних хозяйств / Федеральная служба государственной статистики. URL:

http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1140096812812.

Дата обращения: 12.10.2018.

³ Доходы, расходы и сбережения населения / Федеральная служба государственной статистики. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/level/#. Дата обращения: 12.10.2018.

материальную помощь, случайные заработки и другие денежные поступления, в том числе и в валюте, но валюту переведите в рубли» (j60). Наряду с ним каждого респондента спрашивают о выплатах на основном⁴ и дополнительном месте работы, о случайных приработках⁵, о величине пенсии⁶ (для пенсионеров) и пособия по безработице⁷ (для зарегистрированных в государственной службе занятости).

В исследовании все расчеты выполнены для ответов респондентов 18 лет и старше, N=10020. Полнота ответов на вопросы о личных доходах представлена в таблицах 1 и 2. В целом масштабы «ухода от ответов» о конкретных величинах доходов малы, хотя о величине заработной платы отказываются отвечать чаще, чем о пенсиях. Затруднились или отказались сказать о совокупном личном доходе 3,1%. Имеющиеся данные позволяют снизить это число до 2,5%, если определить совокупный доход через сумму отдельных доходов.

Таблица 1. Полнота ответов о различных видах личных доходов, РМЭЗ, 2017 г., в % по столбцам

Ответы о личных доходах	Виды личного дохода				
	выплаты на основном месте работы	выплаты на дополнительном месте работы	приработки	пенсия	пособие по безработице
нет такого вида дохода	47,5	98,0	92,7	57,6	99,7
такой доход есть, но в последние 30 дней выплат не было	2,2	0,2		0,5	0,0 (2 респондента)
названа величина дохода	48,6	1,8	7,1	41,2	0,3
затруднились назвать величину дохода	0,2	0,0 (2 респондента)	0,1	0,0 (3 респондента)	
отказались назвать величину дохода	1,3	0,0 (5 респондентов)	0,1	0,2	
нет ответа о величине дохода	0,1		0,0 (2 респондента)	0,5	
Итого	100,0	100,0	100,0	100,0	100,0

Таблица 2. Полнота ответов о суммарном личном доходе, РМЭЗ, 2017 г.

Ответы о суммарном личном доходе	%
личный доход назван, равен 0	6,3
названа сумма личного дохода, больше 0	90,8
затруднились назвать сумму личного дохода	0,7
отказались назвать сумму личного дохода	1,9
нет ответа о сумме личного дохода	0,3
Итого	100,0

⁴ «Сколько денег в течение последних 30 дней Вы получили по основному месту работы после вычета налогов и отчислений? Если всё или часть денег Вы получили в иностранной валюте, переведите валюту в рубли и назовите, пожалуйста, общую сумму» (j10).

⁵ «Скажите, пожалуйста, в течение последних 30 дней Вы занимались (ещё) какой-нибудь работой, за которую Вам заплатили или должны заплатить? Может быть, Вы сшили кому-то платье, подвезли кого-нибудь на машине, занимались репетиторством, помогли кому-то с ремонтом квартиры, машины, ухаживали за больными, продавали свои или купленные продукты или товары на рынке или на улице, челночили или делали что-то другое?» и «Сколько всего денег в течение последних 30 дней Вам заплатили за всю эту работу? Если это были разные виды работ, суммируйте. Если оплата производилась в неденежной форме, оцените, сколько это примерно будет в рублях?» (j57).

⁶ «Сколько денег Вы получили в качестве пенсии в течение последних 30 дней? Пожалуйста, учитывайте и основную пенсию, и доплаты к ней» (j363).

⁷ «Сколько денег Вы получили в качестве пособия по безработице в течение последних 30 дней?» (j89).

Расчетный совокупный личный доход определялся как сумма пяти указанных отдельных видов доходов. Чтобы минимизировать пропуски данных, в качестве расчетной величины личного дохода был принят ответ о суммарном личном доходе ($j60$) а) для учащихся и студентов, которые могут получать стипендию; б) для тех, кто получал один из частных видов личных доходов и не назвал его величину, но при этом указал суммарный личный доход. Сравнение суммарной сообщенной респондентом и расчетной величин личного дохода дает такие результаты (в % от общего количества респондентов):

сообщенная респондентом и расчетная величина личных доходов равны 0: 6,3%;

сообщенная респондентом и расчетная величина личных доходов выше 0 и примерно равны ($0,95 \cdot \text{расчетный личный доход} \leq j60 \leq 1,05 \cdot \text{расчетный личный доход}$): 76,2%;

сообщенная респондентом величина личных доходов выше 0 и на 5,1-25% превышает расчетный личный доход: 5,7%;

сообщенная респондентом величина личных доходов выше 0 и на 25,1% и более превышает расчетный личный доход: 5,5%;

сообщенная респондентом величина личных доходов выше 0, но расчетный личный доход равен 0: 3,3%;

респондент не сообщил суммарную величину личных доходов, сравнение с расчетной величиной невозможно: 2,9%.

Визуальное сравнение двух вариантов оценки личных доходов представлено на рис. 1.

Для 82,5% респондентов сообщенные ими величины личных доходов (примерно) совпадают с расчетными, при этом для 6,3% личный доход равен 0; для 2,9% респондентов мы не можем сделать корректное сопоставление этих переменных. 14,5% отвечавших указали суммарный личный доход выше, чем его расчетная величина; в их число входят 3,3% респондентов, чей расчетный личный доход равен 0, то есть у них не было доходов от труда, пенсий, пособий по безработице за 30 дней до опроса, но они включили в свой личный доход иные поступления. Ситуации, когда расчетный личный доход превышал сообщенный, не зафиксированы.

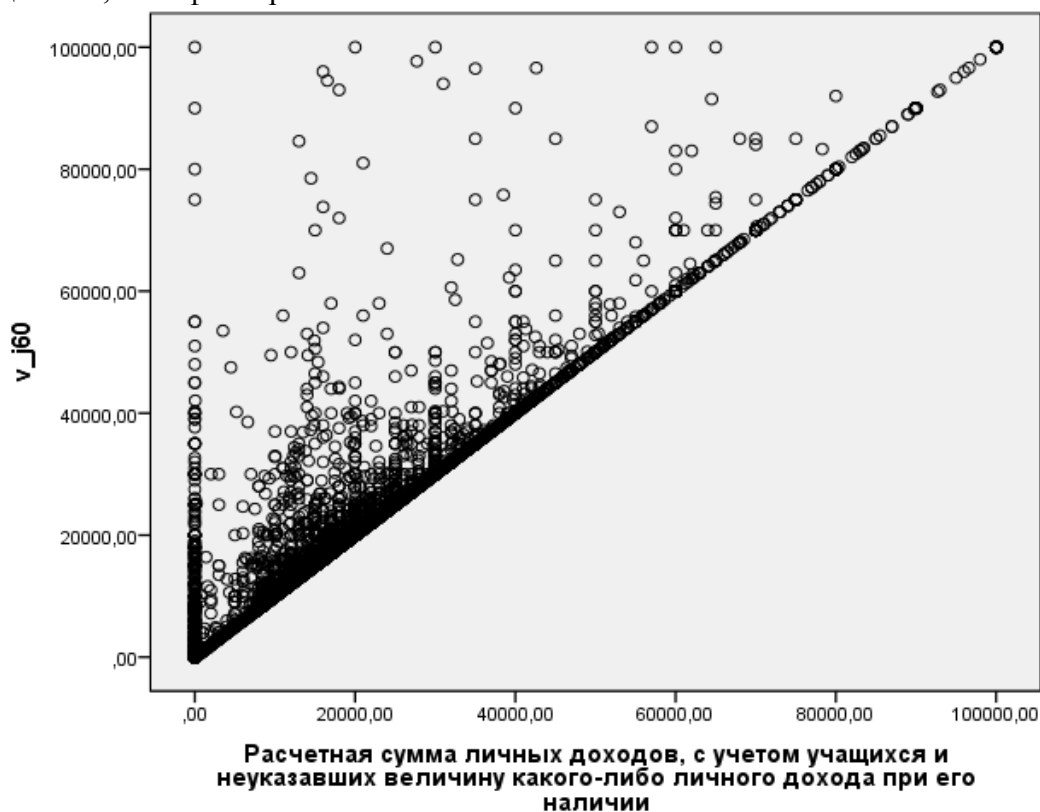


Рисунок 1. Визуализация соотношений расчетных и сообщенных респондентами суммарных величин личных доходов, РМЭЗ, 2017 г.

Можно предполагать, что отличия суммарных величин личных доходов, которые сообщают респонденты, от калькулируемой суммы отдельных видов доходов обусловлены, не столько неточностью суммирования и округления, сколько «концептуальными» особенностями понимания личных доходов. Так, исследовательский опыт автора показывает, что некоторые неработающие женщины могут включать в состав личных доходов детские пособия; часть неработающих респондентов любого пола могут считать личным доходом помощь родителей и т.п. При этом не приходится говорить об унификации мнений респондентов о составе личных доходов, и исследователь оказывается «в заложниках» концептуальных представлений респондентов.

В докладе будут представлены детальные количественные параметры распределений переменных суммарных личных доходов, в том числе оценки неравенства по доходам; определены возможные источники расхождений сообщенных респондентами и расчетных величин личных доходов.

Литература

Abowd, J. M., Stinson, M. H. (2013). Estimating measurement error in annual job earnings: A comparison of survey and administrative data. *Review of Economics and Statistics*, 95(5), 1451–1467.

Angel S., Heuberger R., Lamei N. Differences Between Household Income from Surveys and Registers and How These Affect the Poverty Headcount: Evidence from the Austrian SILC. *Social Indicators Research*. July 2018, Volume 138, Issue 2, pp 575–603.

Duncan G.J., Petersen E. The Long and Short of Asking Questions about Income, Wealth, and Labor Supply. *Social Science Research* 30, 248–263 (2001).

Kim C., Tamborini C. R. Response Error in Earnings: An Analysis of the Survey of Income and Program Participation Matched With Administrative Data. *Sociological Methods & Research*. 2014, Vol 43(1) 39-72.

Kreiner C. T., Lassen D. D., Leth-Petersen S. Measuring the Accuracy of Survey Responses Using Administrative Register Data: Evidence from Denmark / Chapter in NBER book: Improving the Measurement of Consumer Expenditures (2015), Christopher D. Carroll, Thomas F. Crossley, and John Sabelhaus, editors (p. 289-307). Conference held December 2-3, 2011. Published in May 2015 by University of Chicago Press. URL: <http://www.nber.org/chapters/c12663>. Дата обращения 31.07.2018.

Jansen W., Verhoeven W.-J., Robert P., Dessens J. The long and short of asking questions about income: a comparison using data from Hungary. *Quality & Quantity*. June 2013, Volume 47, Issue 4, pp. 1957–1969.

Лукьянова А. Л. Мобильность по заработной плате: до глобального кризиса и после // Мобильность и стабильность на российском рынке труда / Под общ. ред.: Р. И. Капелюшников, В. Е. Гимпельсон. М.: Издательский дом НИУ ВШЭ, 2017. Гл. 7. С. 292-334.

Методика расчета показателей распределения и дифференциации по уровню доходов населения, показателей, характеризующих уровень и распространение низких доходов / Федеральная служба государственной статистики. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/level/#. Дата обращения: 12.10.2018.

Методологические положения по расчету показателей денежных доходов и расходов населения / Федеральная служба государственной статистики. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/population/level/#. Дата обращения: 12.10.2018.