



ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Russian Journal of Economics 2 (2016) 56–85

Russian
Journal of
Economics

www.rujecon.org

Under-reported income of Russian households

Yaroslav Murashov, Tatiana Ratnikova*

National Research University Higher School of Economics, Moscow, Russia

Abstract

In the proposed paper, an attempt is made to estimate the proportion of unstated income for Russian households based on micro data. An overview of microeconomic approaches to estimating the scale of under-reported income is provided. These approaches are weakly represented in the national literature, so their strengths and weaknesses are also analyzed. A theoretical model of household consumer behavior is described that allows the size of under-reported income to be estimated. The structure of household incomes and expenditures is studied based on an RLMS sample for 2012. The model is estimated using household subsamples based on the type of household and household income. The estimation technique utilizes regression variables and random effects. The resulting subsample estimates were applied to the general population and compared with those obtained by other researchers using alternative methods and other data. A comparison is made to estimates of under-reported income developed for British households.

© 2016 Non-profit partnership “Voprosy Ekonomiki”. Hosting by Elsevier B.V. All rights reserved.

JEL classification: C21, C26, D11, D12.

Keywords: under-reported household income, household consumer behavior, under-reported income parameter, regression analyses, instruments, random effects.

1. Introduction

Under-reported income for households is one side of the shadow economy, whose definition includes those economic activities and the income derived from them that circumvent or otherwise avoid government regulation, taxation or observation (Del’Anno, 2003). Although these activities are unrelated to the production of criminal goods and services, they are conducted in a clandestine manner and remain outside of the social and pension security system. The shadow

* Corresponding author, *E-mail address:* taratnikova@yandex.ru.

Peer review under responsibility of Voprosy Ekonomiki.

economy includes covert production, income and expenses for end consumption and accumulation that are hidden from the statistics.

Typically, the size of the shadow economy is assessed in proportion to the gross domestic product. Conventional wisdom says that in developed countries it does not exceed 10% to 12% of GDP; in developing countries, it can reach 40% to 45%; and in transition economies, it accounts, for 22% to 25% of GDP (Burov, 2012).

In modern Russia—where, according to survey by the World Bank, 2012, half of the GDP is produced in the informal sector—the prevailing shadow economy is a particularly acute problem. In addition to adversely affecting the sociopolitical and economic life of the country, this phenomenon also undermines its international reputation.

The majority of Russian and foreign research papers on the shadow economy address the macroeconomic level, revealing that part of the economy—measured as a share of the GDP—that is hidden from taxes. The methods for evaluating the size of the shadow economy are based on certain macroeconomic indicators linked to the amount of goods and services produced within the country. Macroeconomists have a wide array of officially published statistical data at their disposal, and they utilize an extensive set of theoretical methods. However, the aggregate nature of macroeconomic data, while revealing the big picture, does not take into account the heterogeneity of contributions made by different social groups to the shadow economy, thus impeding the development of targeted measures that could pull household activities out from the shadows.

The alternative approach studies microeconomic data obtained from household budget surveys. It determines the size of the shadow economy based on the estimated share of income hidden by households. Unlike macroeconomics, the microeconomic approach provides access to the individual characteristics of surveyed subjects and analyzes the particular aspects of their consumer behavior. The advantage of this approach is that it is not linked to the direct accounting of tax revenues paid by households, thereby potentially enabling a more realistic estimate of the size of the shadow economy. Sample surveys to assess the shadow economy are used in many countries. The inconvenience of these methods is associated with the standard defects in all surveys. The results largely depend on the willingness of respondents to cooperate; however, the majority of respondents are reluctant to acknowledge fraudulent behavior, and thus, their responses are of questionable reliability. (On the other hand, the respondents' cooperation is not for issues related to their sources of income and spending on major purchases). Another disadvantage is that surveys cannot cover all clandestine activities and all income groups (respondents from high-income segments of the population who have the largest proportion of concealed revenue are not available for interviews). For these reasons, this approach always produces an underestimation.

As research for the OECD countries (Kirienko and Ivanov, 2013) shows, the highest estimates of the shadow economy are yielded by monetary methods (demand for cash, volume of transactions in the economy, and the ratio between cash and deposits). Additionally, there are the tax gap methods (measuring the discrepancy between declared income and income after inspection), with the lowest estimates obtained in the analysis of sample population surveys. Table 1 illustrates the gap in the estimated scale of the shadow economy in Russia, as obtained by different methods in different years.

Table 1

Estimated size of the shadow economy in Russia (% of GDP).

Year	Data source	Size of the shadow economy
1995	M. Lacko's calculation using the electricity cost method	39
1996	Report by Chernomyrdin to the Russian Ministry of Internal Affairs	50
1998	Estimate by the Goskomstat of Russia based on SNA data	22
1999	Schneider's calculation using the cash demand method	42
2000	Official data from the Main Economic Crimes Department of the Russian Ministry of Internal Affairs	40
2003	V. Soptaganov, secretary of the Security Council of the Russian Federation, V. Panskov, auditor of the Audit Chamber	25
2003	World Bank experts	40
2011	V. Zhukovsky, senior analyzt, Rikom-Trust investment company	35–40
2011	Rosstat	15

Sources: Kirienko and Ivanov (2013); Authors' compilation of data from official statistics, law enforcement agencies, scientific publications, and estimates by officials.

The estimates obtained from an analysis of microdata about concealed household income are scarce in the literature, as there are no sufficiently developed systematic approaches to adequately replicate these estimates for the entire population.

However, to make effective decisions on regulatory measures, it is important not only to know the estimated size of concealed income across the country as a whole but also to know in more detail which population groups are more inclined to use those practices. The micro-analysis enables a more objective approach to medium- and low-income segments of the population, the data on which reflect representative national household budget surveys.

The relevance of surveys of concealed household income in lower income groups may be associated with an analysis of the changes in prices for goods consumed by certain groups of households, which may serve as a basis for estimating the effects of price policies. For example, Ershov and Matitsin (2012) show that inflation is much higher for the poor than for the rich. Thus, it is possible to assume that escaping to the shadow is one of the few ways available to poor households to compensate for inflation costs in the short term and, therefore, is a method of survival. The modeling of concealed income for various socio-demographic groups can become a basis for understanding what groups, and to what extent, may be pushed into the shadow by increasing prices for vital commodities.

The study aims to contribute to the modeling of concealed income for low- and medium-income population groups within Russia. It provides an estimate of the share of concealed income for various Russian socio-demographic groups covered by RLMS statistics, using an approach that takes the consumer behavior of households into account. This analysis will be based on the consumption model of certain vital goods as a function of unobserved income, the indirect information about which can be obtained from declared income. The methodology for obtaining estimates of under-reported income parameters will be derived from a number of hypotheses about the distinctive features of consumer behavior for self-employed and hired workers. Using econometric analysis, we will obtain interval-based estimates of the under-reported income parameters for selected groups of households and, based on these estimates, calculate intervals for the proportion of concealed income for each group. To compare these esti-

mates with those of macro approaches, we will extrapolate the results by using the information provided by Rosstat on the distribution of income of the general population.

2. Review of the literature

Serious scientific studies on the shadow economy only started to appear during the second half of the 20th century. G. Becker, P. Gutmann, E. Sutherland, H. de Soto, E. Feige, and K. Hart, etc., can be considered the founders of this kind of analysis. They created the conceptual apparatus, investigated the nature of the shadow economy, and developed a methodology for its analysis (Burov and Samarukha, 2013). At the same time, in the USSR, G. Grossman, A. Kaliberda, T. Koryagina, D. Kaufmann and D. Schneider conducted studies to prove the existence of the shadow economy as part of the plan-based socialist economic system. A significant contribution to the developmental issues of the hidden economy was made by D. Makarov, A. Ponomarenko, B. Ispravnikov, and A. Oleinik, who explored the phenomenon of the shadow economy in Russia at the end of the 20th century and identified its various sectors (shadow, informal, concealed, illegal, black, etc.). The works by L. Kosals studied the origins of the shadow structures and institutions of the shadow economy. V. Tambovtsev, V. Radaev, S. Malakhov, and A. Oleinik analyzed the impact of transaction costs in the Russian economy on the proliferation of shadow practices. Radaev studied the prevalence of violence in the shadow economy. Ponomarenko (1995) analyzed the structure of the shadow economy and suggested the following main sectors: the production sector (illegal economy), the sector providing a real contribution to the GDP, and the redistribution sector of the shadow economy, which includes various crimes.

The objective of many studies was to identify *the causes behind the emergence of the shadow economy*. Cassel and Cichy (1987) conclude that, regardless of differences among national socioeconomic systems, the economic causes of the shadow economy are rooted in the economic role of the state, especially in regulating distribution relationships. In the West the primary role is played by increased tax pressure, while in the East it is driven by inflation.

Among the social causes of the shadow economy, one of the main recognized causes is the gap between the standard of living of the majority of the population and that of the middle class. Inequality in wealth and social status determine the number of potential participants in the shadow economy. The shadow economy is largely represented by the poor and marginalizes strata: youth, the unemployed, migrant workers, etc. In terms of raw numbers, they make up the greatest portion of participants in the shadow economy (Popov and Tarasov, 2005).

The imperfection of laws is the legal reason for the shadow sector of the economy to persist. The legislative framework fails to react quickly to the rapidly changing environment of a market economy, with gaps occurring in the legal field that create favorable conditions for the shadow economy (Burov and Samarukha, 2010).

The political reasons researchers refer to include the merger of power and major capital, which produces an oligarchy. By lobbying their oligarchic interests, large businesses place small companies in poor conditions and “push” them into the shadow.

Finally, there is the desire to receive more with less effort. This kind of rationale, in the absence or weakness of restrictions, encourages people to engage in shady activities (Burov and Samarukha, 2010).

The focus of a number of papers was to find answers to the question, *what are the dangers of the shadow economy?* There are multiple negative effects on society from the shadow economy (Timchenko, 2004):

- negative impacts on budgets at all levels and, as a result, limited financing to perform the state's functions, such as administration, defense, fundamental science, health care, and social programs;
- negative impact on the efficiency of macroeconomic policies, because the official statistics needed for macroeconomic decision-making become distorted and prevent timely informed decisions to encourage development in an economic sector, which results in the biased development of particular industries;
- negative impact on economic growth and development of the legal sector, because the ability to purchase goods and services in the shadow sector reduces the need for goods and services produced legally;
- lack of investment in the legal economy due to lower earnings and decreased competitiveness, as enterprises in the formal sector of the economy are initially at a disadvantage in economic terms compared with enterprises in the informal economy because of the high tax burden;
- negative impact on the conditions for reproducing labor in the legal economy, because failure to pay taxes in the informal sector enables enterprises in the shadow economy to allocate more funds to pay under-reported wages, thereby creating better financial opportunities for engaging skilled labor, without worrying about its reproduction;
- uncontrollable consequences for the environment due to the inability to formally record these harmful consequences and due to the avoidance of payments for using natural resources and fines for environmental pollution;
- negative effect on foreign investment, due to the unwillingness of foreign investors to work with inadequate economic and legal conditions in the criminal environment;
- deformation of the consumption structure due to allocating cash flow in the shadow economy predominantly in areas where one can earn high profits over a short period of time at minimal cost, due to the lack of governmental control and to legislative gaps;
- creation of conditions that encourage the growth of corruption in society, along with the criminal economy, organized crime, and terrorism.

However, there is an alternative view that argues that some kinds of shadow economies (especially the informal) actually support the development of the formal economy rather than hinder it. According to Cassel and Cichy (1987), there are three positive functions of the shadow economy in a market economy:

- “economic lubricant”, i.e., smoothing of fluctuations in the economic environment by redistributing resources between the legal and the shadow economy (during a crisis, the production resources of the legal economy are redistributed to the shadow economy and return after the crisis);
- “social shock absorber”, i.e., the mitigation of social inequality (informal employment improves financial conditions for the poor);

- “built-in stabilizer”, i.e., resources in the shadow economy that fuel the legal economy (concealed income is used to purchase goods and services in the legal sector, “laundered” illicit capital is taxed, etc.).

The answers to most questions of interest to researchers are closely related to *measuring the scale of the shadow economy*.

The effectiveness of such measurements is greatly limited by the very nature of the phenomenon, which is not covered by official statistics. The quantitative estimation methods for shadow economy parameters rely on data from reports by relevant ministries, agencies and organizations, and on sample surveys of enterprises and households. Depending on the type of information sources used, the approaches are divided into two groups: indirect (macroeconomic) and direct (microeconomic). Direct methods tend to underestimate the shadow economy; indirect methods usually overestimate it.

The majority of analytical tools are macroeconomic, among which are a number of monetary approaches associated with cash demand or associated with alternative calculations of the GDP (Barsukova, 2003). There are also popular methods for estimating the size of the shadow economy based on electricity consumption and the divergence between the official and real labor market (Italian method). The comparability of results is the main advantage of macroeconomic methods (Burov, 2012).

The number of microeconomic tools for analyzing the shadow economy is much smaller. Two primary areas of microeconomic estimates are being developed to determine the size of the shadow economy (Barsukova, 2003): the first explores the discrepancy between overall household income and expenditures, while the second examines the divergence between the consumption of certain specific goods by hired and self-employed households.

The first group of models analyze the divergence between income and expenditures, and the key conclusion is that the core of the shadow economy is self-employed households. This finding becomes the main hypothesis for the models in the second group, which compare the ratios of income and expenditures for certain categories of goods and services between self-employed and hired workers.

The second hypothesis of these models is that employees correctly report their wages, while the self-employed hide their income. The third premise of the models is that households correctly declare expenditures on current consumer goods.

The last two hypotheses can be considered a bottleneck of the model that compares the expenses of self-employed households and others. Suvorov (2008) notes that in certain sectors of the Russian economy there is an established practice of “shadow” wages. This is particularly true for the trade and finance sectors. In this case, wages are stated incorrectly, as with income. He also notes that there are significant differences in consumption estimates by the Russian population for a number of food products, according to trade statistics and household surveys. This implies that expenditures on food may be incorrectly stated by households.

Each of the aforementioned microeconomic areas uses household survey data. Both areas have similar defects. A selection shift (no samples from high-income households) prevents the full extent of the shadow economy from being determined based on sample data. There is the problem of comparing multi-temporal cash flows. Shadow incomes that do not appear in consumption are not captured (Barsukova, 2003), while the lack of reliable information on the total distribution of households by income and expenditures creates problems for replicating

estimates across the general population and prevents the comparison of results between macro and micro approaches. However, there is independent value in modeling concealed income based on household budget survey data, because such models can be built separately for different segments of the population and can analyze the determinants for escaping into the shadow in greater detail for different socio-demographic and socio-professional groups.

The first area was the basis for the paper by Gorodnichenko et al. (2009), who did not directly estimate the size of the informal economy but analyzed the impact of tax reforms in Russia (the transition to proportional taxation in 2001) on the amount of tax revenues and the incentives for households to exit the shadow economy. They also assessed the impact of the reforms on public welfare.

One of the key works mentioned in many empirical studies on microdata is Pissarides and Weber (1989). It slightly weakened the premises defined above and subjected them to criticism. The authors suggest that: 1) all households correctly state their expenses on certain types of goods and 2) some types of households correctly state their income. The term “type of household” is understood as belonging to one of the groups based on the source of income. According to the authors, the type of expenditures that are least likely to be concealed is food expenses (a hypothesis contrary to the contemporary Russian reality), and not all of the self-employed tend to hide their income, but only those who have a certain amount of income from “self-employment.” According to the authors, households whose main source of income is wages state their income correctly. The article offers a direct way to estimate the size of the shadow economy. The first stage estimates the expenditure function for certain categories of goods, designed on the basis of economic theory. The second stage considers this function using the method proposed by the authors, and real household income is reproduced based on reported expenditures.

Lyssiotou et al. (2004) developed the ideas of Pissarides and Weber and suggested two new approaches: the first involves assessing a single equation for food expenditures by using a non-linear method; the second estimates a system of demand for current consumer goods based on a breakdown of household expenses.

The basic idea of the predecessors’ model is to split demand for current consumer goods and durable goods. Considering the different categories of demand, the authors note that the portion of income a household spends on various categories depends not only on income but also on its source. As statistics show, self-employed households spend more on luxury goods while spending too little on food and other necessities, given the size of a household’s total income.

This method estimates the size of the shadow economy (hereinafter, the size of concealed household income) without studying the residual variations in household incomes. This is its advantage because there is no need for special calculations of the upper and lower limits of the shadow economy parameter. The disadvantage of this method is its lower statistical capacity compared to Pissarides and Weber’s approach.

These methods analyze concealed income drawing on the economic interests of individuals. However, a study of the incentives contingent upon social interactions among people is also noteworthy. The authors of many empirical studies suggest that an individual decides on his/her employment in the informal sector of the economy based on the interests of other people. Montmarquette and Gardes (2002) modeled this approach, taking into account the correlation between house-

hold decisions and the estimated size of the reference group (the group of people influencing the decisions of the individual).

Another empirical study by Gardes and Starzec (2009), which analyzed social interactions between different types of households, was based on an analysis of the shadow sector in Poland during the transition period (1990s). A common feature of transition economies is that the constraints on the consumer market are replaced by restrictions on the production market. Therefore, according to the surveys, most households indicate insufficient income as a reason for working in the shadow economy. For Poland in the 1990s, agriculture was the main sector of shadow employment, but after 2004, tax evasion became the main reason. The probability of participation in the shadow economy increases for the unemployed, young people, and people with higher education degrees.

Accounting for the social interaction of households and the allocation of expenses into purchases of goods in the formal and informal sectors are areas for future research, and there is currently no ability yet to implement them based on Russian statistical data. However, based on the RLMS database (Russia Longitudinal Monitoring Survey, see RLMS-HSE, 2012) used in the present paper, it is possible to apply a theoretical model and, with some substantial modifications, an empirical algorithm. These were used to estimate the size of concealed income in the (now classic) paper by Pissarides and Weber.

3. Theoretical model

As noted above, the theoretical model in this study is based on the model developed by Pissarides and Weber (1989), founders of the microeconomic approach to estimating the size of the shadow economy. Their approach is both simple and informative because, with the right expenditures category, the model obtains unbiased estimates of the share of concealed income for self-employed households with different characteristics.

The theoretical model presented by Pissarides and Weber (1989) describes the behavior of user households while taking into account hidden income on the one hand, and the relationship between income and consumption on the other.

The function of consumer spending i by households on a group of goods j reads as follows:

$$\ln C_{ij} = Z_i' \alpha_j + \beta_j \ln Y_i^p + \varepsilon_{ij} \quad (1)$$

where C_{ij} is the consumption of a group of goods j ; Z_i is the household i 's characteristics vector; Y_i^p is the "permanent income", i.e., the portion of income affecting consumer decisions; and β_j is the marginal propensity to consume the group of products j .

In turn, permanent income is related to actual (real) income Y_i through the following ratio:

$$Y_i = p_i Y_i^p \quad (2)$$

where p_i is the income variation due to unforeseen circumstances and is a random variable for households.

The authors suggest that the differences between permanent and actual income are attributable to external economic factors (e.g., adverse economic conditions across the country or in a separate branch) whose vector of influence is not dependent on the type of household by income source. Therefore, the average value of p_i is not dependent on the type of household. However, the variance of this parameter is related to the type of household: $V(p_i | i - ee) < V(p_i | i - se)$.

For self-employed households ($i - se$), the dispersion parameter p_i is greater than for hired households ($i - ee$) their income variance is higher (income for the self-employed is less stable). Hereinafter, the ee index refers to households whose main source of income is wages, and the se index refers to households whose primary source of income is derived from self-employment.

In accordance with the premises of the model, Y_i is not observable, because the income of self-employed households is stated incorrectly. If k_i is taken as the ratio indicating by how many times the real household income Y_i exceeds the stated income net of taxes Y_i^d , then the relationship between real and stated income will read as follows:

$$Y_i = k_i Y_i^d \quad (3)$$

According to the theory, employees state their income correctly; therefore, for households whose members are hired, $k_i = 1$. For the self-employed, the k_i parameter represents a random variable, and $k_i > 1$.

Thus, the unobserved permanent income, which is associated with the consumption function, can be expressed through current income and the model parameters as follows:

$$\ln Y_i^p = \ln Y_i^d - \ln p_i + \ln k_i \quad (4)$$

where $\ln p_i$ and $\ln k_i$ are two additional random regressors.

To verify the statistical hypothesis that households hide part of their income, we need assumptions about the distribution of the $\ln p_i$ and $\ln k_i$ parameters responsible for the discrepancy between observed income and permanent income. The authors suggest considering this distribution to be lognormal (because, according to the theory, such is the distribution of household income). In this case, the model parameters can be represented as the sum of their averages and deviations from the averages:

$$\ln p_i = \mu_p + u_i, \quad \ln k_i = \mu_k + v_i \quad (5)$$

Then, the consumption function will read as follows:

$$\ln C_{ij} = Z'_i \alpha_j + \beta_j \ln Y_i^d - \beta_j (\mu_p - \mu_k) - \beta_j (u_i - v_i) + \varepsilon_{ij} \quad (6)$$

The dependent variable is a certain category of household expenditures. Pissarides and Weber (1989) considered food expenditures.

The expected value of p_i is related to μ_p as follows:

$$\ln E(p_i) = \mu_p + \frac{1}{2} \sigma_u^2 \quad (7)$$

If $E(p_i)$ is not dependent on the type of household employment, $E(\ln p_i)$ does not depend on it either; therefore, it is possible to compare the mean logarithmic values for the hired the and the self-employed:

$$\mu_{pse} - \mu_{pee} = \frac{1}{2}(\sigma_{uee}^2 - \sigma_{use}^2) \leq 0 \quad (8)$$

If equation (6) is estimated separately for hired workers and the self-employed, the constant will vary within the model because the value of $\mu_{kse} - \mu_{kee}$ will be different for each group of households ($\mu_{kse} = 0$ for hired employees and $\mu_{kse} > 0$ for the self-employed). The information about the size of the shadow economy is conveyed by the value of this constant.

Rather than estimating the model based on two separate subsamples of hired employees and the self-employed, we can introduce a dummy variable to account for the type of household employment, and use a linear regression in the following form:

$$\ln C_{ij} = Z_i' \alpha_j + \beta_j \ln Y_i^d + \gamma_j D_{SE_i} + \zeta_{ij} \quad (9)$$

where $D_{SE_i} = 1$ if the household is self-employed, and 0 otherwise. This model can be estimated with ordinary OLS, adjusted for the heteroscedasticity of the error ζ_{ij} .

Estimating this equation, we can calculate the assessment of the under-reported income parameter k_i (indicating how many times the real income of a household exceeds the income stated).

If we assume that the marginal propensity to consume the j^{th} category of goods β_j coincides for self-employed and hired workers, the γ_j coefficient has the following meaning:

$$\gamma_j = E\{-\beta_j(\mu_p - \mu_k) | se\} - E\{-\beta_j(\mu_p - \mu_k) | ee\} \quad (10)$$

The above analysis and expression (9) show that

$$\gamma_j = \beta_j \left\{ \mu_k + \frac{1}{2}(\sigma_{use}^2 - \sigma_{uee}^2) \right\} \quad (11)$$

We can estimate the difference in the error variances of income between self-employed and hired workers based on an income decomposition, but this would require building auxiliary regressions of observed income for each of the two household categories.

Observable income can be represented as follows:

$$\ln Y_i^d = Z_i' \delta_1 + X_i' \delta_2 + \zeta_i \quad (12)$$

where Z_i is the vector of household characteristics; X_i is the set of instruments; $\zeta_i = u_i - v_i$ for self-employed and $\zeta_i = u_i$ for hired employees.

Then, $\sigma_{\zeta_{se}}^2 - \sigma_{\zeta_{ee}}^2 = \sigma_{use}^2 - 2\text{cov}(u, v)_{se} + \sigma_{vse}^2 - \sigma_{uee}^2$. Because (12) implies that

$$\mu_k = \frac{\gamma_j}{\beta_j} - \frac{1}{2}(\sigma_{use}^2 - \sigma_{uee}^2), \text{ then } \mu_k + \frac{1}{2}\sigma_{vse}^2 = \frac{\gamma_j}{\beta_j} + \frac{1}{2}(\sigma_{vse}^2 - \sigma_{use}^2 + \sigma_{uee}^2).$$

A point estimate of the expected value of the under-reported income parameter cannot be found because the value of $\text{cov}(u, v)_{se}$ is unknown, but it is possible to calculate an interval estimate for this parameter.

If we accept the assumption that $\text{cov}(u, v)_{se} = 0$, then the lower limit of the interval is determined subject to the condition that $\sigma_{vse}^2 = 0$, and the upper limit is obtained when making the assumption $\sigma_{use}^2 = \sigma_{uee}^2$ (self-employed income is at least as volatile as hired income).

In this case, the interval limits are expressed through residual income variances $\sigma_{\zeta_{se}}^2$ and $\sigma_{\zeta_{ee}}^2$:

$$\begin{aligned} (\ln k)_{low} &= \frac{\gamma_j}{\beta_j} - \frac{1}{2}(\sigma_{\zeta_{se}}^2 - \sigma_{\zeta_{ee}}^2) \\ (\ln k)_{up} &= \frac{\gamma_j}{\beta_j} + \frac{1}{2}(\sigma_{\zeta_{se}}^2 - \sigma_{\zeta_{ee}}^2) \end{aligned} \quad (13)$$

Based on the calculated limits of the under-reported income parameter, we came to a conclusion about the real income of households participating in the shadow economy (self-employed households according to the model).

Researchers are not only interested in the overall analytic result across the entire available sample of households but also in a more detailed study of the portion of concealed income across various socio-demographic groups.

4. Hypotheses of the study

Russia's shadow economy, according to Burov and Samarukha (2010), consists of four sectors—the informal economy, the illegal economy, the fictitious economy, and the concealed economy—and three types of relationships.

The subjects of the first type of relationship are large businesses that gained access to power, allocation, distribution, and assignment of material and financial resources. This segment of the population is not represented in the RLMS household budget survey data and, therefore, is not touched upon in this study.

The second type of relationship involves representatives of various socio-professional business groups whose goal is to obtain funds for business development and personal business income due to the paid satisfaction of the mass needs of the population (illegal sector and concealed component of activities). This segment of the population is poorly represented in the RLMS data.

The third type of relationship involves representatives of various socio-professional groups whose purpose is to make a living through paid satisfaction of mass needs of the population (informal sector). It is this segment of the population that is best represented in the RLMS data, and it is to this segment that our research hypotheses refer.

The shadow activities of individuals and households in the informal sector of the economy, according to Burov and Samarukha (2010), can be divided into four levels:

- the first level is associated with the basic physiological survival of a family;
- the second level is related to physiological and intellectual survival;
- the third level refers to striving for development and improvement;
- the fourth level is an attempt to lay a material foundation for future.

Depending on the level of household shadow activities, we will differentiate between reported income, expenditures for certain goods, and the size of under-reported income.

We will assume that the level of household shadow activities is largely determined by its level of monetary income and the opportunities that exist for households living in various types of settlements.

Based on monetary income, households can be divided into three groups:

- income below half the median across the sample (such households are classified as poor, and their shadow activities are supposedly driven by first- and second-level reasons);
- income between half of the median and one-and-a-half times the median (the rationale for such households presumably corresponds to second- and third-level reasons);
- income above one-and-a-half medians (affluent households with fourth-level reasons).

Hypothesis 1. The share of under-reported concealed income is higher in the lower and upper income groups than in the middle income group.

In support of this hypothesis, we can draw on the following arguments. In the lower income group, people may be unable to earn reportable income because of a lack of jobs in the legal sector as well as the low level of skill or its inconsistency with the structure of available jobs; therefore, almost all of the income will be concealed because of the need to ensure physical survival. The middle group includes households for whom a significant portion of income is derived from wages; these households are not as strongly motivated to conceal their earnings as members of the lower group. The upper income group contains representatives of the affluent segments of society who are not concerned with physical survival and for whom hired employment does not seem attractive. However, when promising business opportunities arise, the proceeds from them may be concealed to a great extent because of the desire not to attract attention from the criminal element of the shadow economy and the regulatory agencies, as well as the desire to optimize tax deductions.

Households can be grouped by settlement type based on two factors: size or administrative status. The hypotheses underlying the choice of these target groups are as follows.

Hypothesis 2. In small settlements and cities with million-plus populations, the share of under-reported income is higher than in medium-sized towns and villages.

This hypothesis seems valid because, in the former case, the number of formal jobs is very limited and people are forced to seek shadow income to support themselves, while in the latter case, on the contrary, there are a very large number of attractive opportunities for shadow earnings, especially in metropolitan areas.

Hypothesis 3. In rural areas and regional centers, the proportion of hidden income is higher than in urban settlements and secondary cities.

In a literature review on the shadow economy, Smith (1986) concluded that the informal economy is mostly run by the self-employed who own small family

businesses. Such enterprises are widely represented in rural areas, e.g., in agriculture, forestry, fisheries, construction, distribution, and repairs. In regional centers, where schools, medical facilities, shopping malls and markets are concentrated, there are additional opportunities such as tutoring, private medical services, private trucking, and various forms of freelancing.

5. Data description

This section describes sampling restrictions, classifies households, and examines in detail the characteristic features that differentiate self-employed households from others. Studying these features is necessary to determine the consistency of data and key hypotheses for the model and is necessary for the choice of control variables and variables for the instrumentation of income in the consumption equation.

The paper uses data from the 21st RLMS wave (2012), which represents a non-governmental longitudinal household survey. The RLMS covers a wide range of issues and produces an extensive base of socio-economic variables that can describe the structure of income and expenditure, the structure of food consumption, the level of material well-being of the population, education levels, investment, occupations, migration, health, etc. The wave includes individual and households data. The sample represents the current (2012) situation in households of the Russian Federation (more precisely, the income groups of Russian households available for the RLMS).

The following charts demonstrate the relevance of the RLMS sampling distribution by income in 2012 for the income distribution of the total Russian population in 2012 (according to the official Rosstat data).

Fig. 1 shows that the selected distribution of households by income is biased to the left: low-income groups in the population are much more widely represented than in the distribution for the general population. The right end of the sampling distribution is significantly thinner than the right end of the general population distribution, suggesting a small number of representatives from upper income groups in the RLMS sample. The global maximum for the sampling distribu-

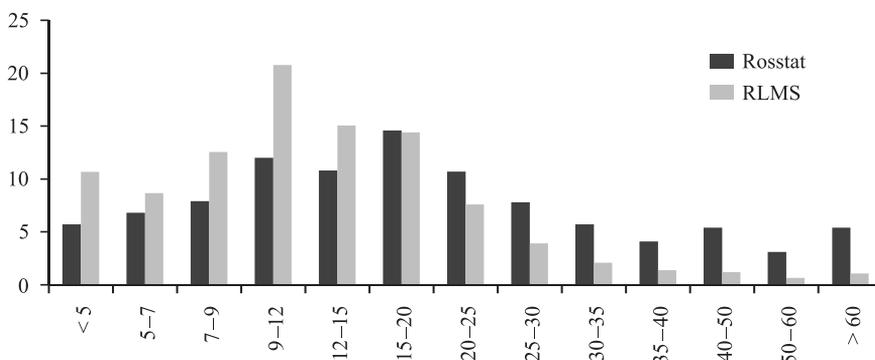


Fig. 1. Bar charts for the distribution of the Russian population by monthly income per capita, 2012 (RUB thousand).

Source: RLMS data for 2012, and the household budget survey data from Rosstat, adjusted for balance of income and expenditure indicators for the population during the same period.

tion falls within the interval from RUB 9,000 up to RUB 12,000 per month per person, whereas the global maximum for the general population distribution is within the interval from RUB 15,000 to RUB 20,000 per month per person.

A comparison of distributions shows that a direct replication of the findings from this study on the entire population would be impossible. We can only speak of modeling under-reported income for low- and medium-income segments of the population.

5.1. Income decomposition

To build the model, it is important to carefully examine the typology of households and household income distribution across the main sources. There are three such sources:

- 1) net wages;
- 2) “other income”, which includes pensions, scholarships, unemployment benefits, income from the sale of housing, income from housing rent, interest income from capital, dividend income from capital, alimony, debt payment subsidies and housing payment subsidies;
- 3) income from self-employment, defined as follows:

$$\text{income from self-employment} = \text{total income} - \text{wage} - \text{“other income”}$$

Formally, income from self-employment is part of “unearned income”:

$$\text{unearned income} = \text{total income} - \text{wage}$$

As we see from the descriptive statistics (Table 2), approximately 45% of households earn a positive income from self-employment. The average income from wages is the greatest among all mean values for various sources of income: households earn RUB 33,700 on average. The mean value of “other income” is RUB 13,478, which is 1.5 times higher than the average income from self-employment (RUB 9,405). At the same time, the mean values are shown for households with positive income.

An important part of the study is the analysis of variance of different income components. Income from self-employment has the greatest coefficient of variation, at 2.33 (the ratio of the standard deviation to the mean). On the other hand, the coefficient of variation for “other components” of income is the lowest (0.71)

Table 2
Descriptive statistics of income decomposition.

Statistics	Total income	Wages	Other income	Unearned income	Income from self-employment
$N (> 0)$	6359	4412	4475	5241	2879
Number of zero values	158	2105	2042	1276	3638
Average	37 126	33 700	13 478	16 676	9405
Standard error	33 929	29 838	9532	18 867	21 905
Standard error/mean	0.91	0.89	0.71	1.13	2.33
Minimum	130	990	85	50	0
Maximum	425 600	420 000	175 000	332 300	320 300
25% quartile	16 200	15 000	7800	8000	990
50% quartile	28 000	26 000	11 200	12 500	2724.5
75% quartile	46 408	44 000	17 800	20 200	9000

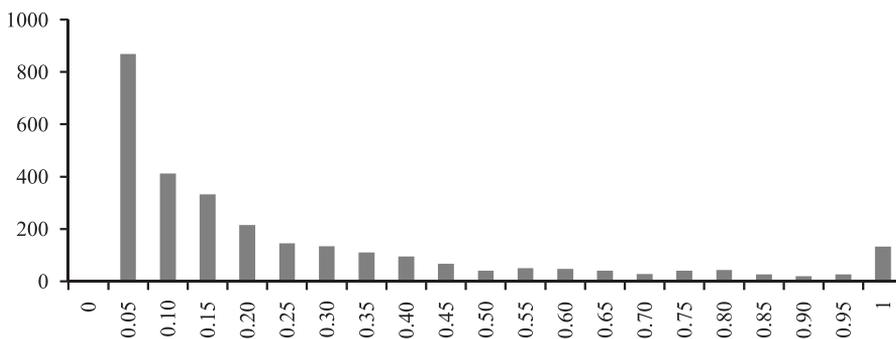


Fig. 2. Distribution of the proportion of household income from self-employment within the sample (among households with positive income from self-employment; $N = 2879$).

because the most important element within them is the least-variable part of income, i.e., pensions, scholarships, etc. The variation in wages is greater than for income from self-employment, but the relative variation, or the ratio of the standard deviation to the mean value, is higher for income from self-employment (2.33 compared with 0.89). This empirical fact means we do not have to reject the hypothesis that income from self-employment is more volatile, allowing us to carry out the decomposition by Pissarides and Weber (1989). This decomposition is based on the hypothesis that wages are stated correctly, while income from self-employment is under-reported. According to this model, the residual error in the income equation should be highest for the self-employed and can be used to estimate a parameter that would be used to adjust income from self-employment.

A separate task is the classification of households (division into self-employed and others) which can be solved by studying the proportion of income from self-employment out of total household income. Fig. 2 shows that the distribution of this proportion is asymmetrical: a large portion of households receive either zero or a small fraction of their total income from self-employment. The histogram identifies the share threshold value: it is 20% (Pissarides and Weber used 25%). All households deriving more than 20% of their income from self-employment will be categorized in the study as self-employed.

A noticeable spike in the distribution around a proportion of 1 indicates the presence of a large group of households for which income from self-employment is their primary income.

Table 3 explores the differences among households, indicating the presence or lack of income from self-employment. The officially self-employed have higher incomes, on average, which are also more volatile. We have seen above that income from self-employment has a higher coefficient of variation than income from wages. However, it is worth noting that inside the segment of officially self-employed households, income from wages has a higher coefficient of variation than income from self-employment. This fact, as well as an analysis of the chart, indicates the heterogeneity of self-employed groups and the need to divide households not into two groups (hired and self-employed), but into three: self-employed, people with a low share of income from self-employment, and people with zero income from self-employment.

Table 3

Descriptive statistics to decompose income for households with positive income from self-employment, grouped by the proportion of income from self-employment out of total income (RUB).

Type of income	Average	Standard error	Minimum value	Maximum value
<i>Statistics for households reporting low income from self-employment (< 20%) (N = 1947)</i>				
Total income	34 532.18	27 969.09	1000	276 500
Wages	20 335.45	26 283.81	0	243 000
Other income	10 769.81	10 052.66	0	99 000
Income from self-employment (estimate)	3426.92	8190.33	0	192 000
<i>Statistics for households reporting high income from self-employment (≥ 20%) (N = 932)</i>				
Total income	44 003.35	50 014.41	1500	425 300
Wages	14 866.87	24 496.69	0	220 000
Other income	7232.32	9178.87	0	120 000
Income from self-employment	21 904.16	33 352.10	600	320 300

The model's last important assumption required to estimate the shadow economy factor is the lognormality of household income. This assumption has been verified based on empirical data, using the Kolmogorov-Smirnov criterion. The hypothesis is not rejected at a significance level of 1% for total income or for income from self-employment (the estimated probability of a Type I error is 14.7% for total income and 1.4% for income from self-employment).

5.2. Descriptive statistics for the socio-demographic characteristics of households

The choice of control variables (household characteristics) is an important part of the technique for estimating the proportion of under-reported household income that will be implemented in this study. Appendix Table A1 describes the sampling framework for the characteristics of individuals classified according to type of employment: total; individuals not related to the self-employed; individuals whose income from self-employment is less than 20% of their total income; individuals whose income from self-employment is more than 20% of their income.

In this study, income from self-employment means all income from entrepreneurial activity, regardless of whether the employer is self-employed (working for himself) or an individual entrepreneur who hires employees. (This paper uses a slightly different principle for categorizing individuals as self-employed than that used in Lukyanova (2012), where an individual is assigned to this group if registered as an individual entrepreneur or the founder of a legal entity or if self-described as unofficially self-employed.)

People qualified as skilled labor include individuals within ISCO groups 1–5: lawyers or government employees, expert professionals of the highest category, skilled workers in agriculture, industry workers, or other skilled workers. Unskilled workers include respondents within ISCO groups 6–9: medium-skilled specialists, clerks, services and trade employees, and unskilled workers. The basic education group includes those with a completed or incomplete secondary education.

To describe the resulting patterns, we will, for the sake of brevity, regard the self-employed as individuals who earn more than 20% of their income from self-employment.

We can conclude from Appendix Table A1 that the self-employed are more often single than married. They are more often engaged in unskilled labor, which is true for heads of families as well as for regular members of the household. They have lower levels of education, whether higher (33.55% vs. 39.05%) or secondary vocational (37.74% vs. 42.49%), and they have a higher level of satisfaction with life (17.10% vs. 15.29%). Self-employed households have a significantly larger proportion of school students (37.6%) and a slightly larger proportion of college students (19.9%) than households with no income from self-employment (28.5% and 17.1%).

The self-employed want to get another job more often than hired employees (39.57% vs. 29.71%), but they do not state that they have a job at the moment (which perhaps means they are concealing the shadow portion of their income, or have no other paid sources of income). Financial opportunities and opportunities to improve living conditions, to take a vacation, or to pay for a child's education are no different for the self-employed from the average level of the full sample, although these opportunities exceed the capacity of those who have a small portion of income from self-employment.

We can conclude that a number of selected characteristics of household members should be used in a regression analysis because they correct some structural differences between households of the self-employed and of hired employees and may affect a household's intention to hide their income.

5.3. Descriptive statistics of the economic characteristics of households

Appendix Table A3 shows some of the categories used in the model of variables and variables that can be used as instrumental: settlement type, real estate ownership and the equipment in housing facilities, land ownership, the presence of crops (including those for sale), the ownership of vehicles and machinery for cultivating land, the possession of a computer and internet access, and the possession of other durable goods.

Based on descriptive statistics analysis, we can conclude that the self-employed live more often in rural areas (33.2% vs. 21.2% for the sub-sample of individuals with zero income from self-employment) than in towns (18.7% vs. 27.4% for the sub-sample of individuals with zero income from self-employment). In the regional centers, their number is roughly the same.

Regarding housing, the self-employed own a greater amount of total area and residential area, with a large number of rooms, than those having no income from self-employment or having a small income from it. They are less frequently equipped with central water supply, central plumbing, hot water supply, sewer, gas and telephone. They prefer satellite antennae to cable television. Less often than other types of households, they have cottages, lawnmowers, a foreign-made car with a GPS navigator, a washing machine and a microwave, but they more often own trucks, motorcycles and tractors, and they sell their crops more often (5.7% vs 2.2%, and 1.1% for the sub-sample of individuals with no income from self-employment). Desktop computers are less common with the self-employed,

while laptops are more common. Low-speed internet is more affordable for the self-employed than broadband.

The statistics show that among the self-employed with a high proportion of income from self-employment, there must be a substantial proportion of farmers (farms).

5.4. Cost decomposition

The final step of the preliminary data study is to decompose costs, on the basis of which we will build the variables for the equation model.

Households report information on the number of purchased products, their prices, and overall expenses over the past seven days. Monthly household expenditures on food represent the sum of expenditures for all products, normalized to a 30-day period.

According to the theory, household expenses can be divided into expenditures for the purchase of durable goods and current consumer goods. Expenditures for current consumer goods include food, clothing, and services. Spending on durable goods includes expenditures for purchasing household appliances.

Table 4 demonstrates that for households with positive income from self-employment, there is a high degree of heterogeneity in expenditures. All expenditures for the self-employed are higher for all categories than the average for the entire sample (excluding expenditures on food) and are higher than expenditures by households with a low proportion of income from self-employment. On the other hand, total income for the self-employed is also significantly higher (whereas the average household incomes from other parts of the sample are not

Table 4

Descriptive statistics for household expenditures, including the consumption of household-produced goods (RUB).

	Entire sample	No income from s/e	Share of revenue from s/e < 20%	Share of revenue from s/e ≥ 20%
Expenditures for meals at home, normalized to 30 days	9518	9462	9143	9502
(*) Consumption of home-produced goods	458	445	520	1403
Expenditures for meals at home, normalized to 30 days, adjusted for (*)	9976	9907	9663	10 905
Expenditures for eating out, normalized to 30 days	1550	1718	1283	1402
Estimated total expenditures for food	11 069	11 180	10 427	10 904
Estimated total expenditures for food, adjusted for (*)	11 519	11 625	10 946	12 307
Expenditures for clothes over the past 30 days	2436	2486	2182	2770
Expenditures for services over the past 30 days	7248	7499	6641	7529
Expenditures for health care over the past 30 days	2138	2127	2109	2247
Other expenditures over the past 30 days	2671	2765	2324	3024
Total current consumer expenditures on goods for 30 days	26 012	26 502	24 202	27 877
Expenditures for durable goods, normalized to 30 days	6930	6866	5925	9285
Total expenditures	32 942	33 368	30 127	37 162
Total income	36 082	35 020	34 514	43 500

significantly different). The question for the current research is whether the differences in expenditures are attributable to stated income or whether there is some portion unaccounted for that is part of the shadow economy.

Household consumption includes goods produced for a household's own consumption, and it is higher for self-employed households. This is more closely associated with the significant number of self-employed in farming households, for which this type of consumption is most characteristic, rather than concealed income. Cash expenditures on food for the self-employed households and the others do not vary greatly.

6. Estimating the econometric model

Our approach involves estimating an equation of expenditures for an individual good that depends on identifying household indicators, including income and employment type (self-employed and others), to obtain an unbiased estimate of the under-reported income parameter.

Food costs are often regarded in the literature as the most correctly presented in the statistics. The model uses a log-linear form for the dependency of expenses on income:

$$\ln C_h^F = Z_h' \alpha + \beta_F \ln Y_h + \gamma_F D_{SE_h} + u_h^F \quad (14)$$

where h is the number of households; C_h^F is food expenditures; Y_h is income; Z_h identifies the household's indicators; D_{SE_h} is a dummy variable that assumes the value of 1 for the self-employed and 0 otherwise; and u_h^F is random error, independently and identically (normally) distributed with zero mean.

During the evaluation of the model, the variable D_{SE_h} (which indicates the type of household) was found to be significant at the 1% level, but negatively (Table 5). This means that the self-employed spend less on food than other households, which does not correspond to the model's assumptions and makes it impossible to estimate the size of concealed income based on the food expenditure equation. This is quite consistent with Suvorov (2008), indicating the unreliability of information households report on their expenditures for food, and a strong overstatement of declared expenses on a number of the most demanded products.

However, it is possible to use the other household expenditure categories that are reported, perhaps more appropriately, but expenditures must depend on overall household income rather than income structure (a condition necessary to correctly assess the expenditure equation). The first reason is why expenditures for durable goods are unsuitable. The second reason is why we cannot use expendi-

Table 5

Comparison of estimates of expenditure models for certain categories of goods.

Coefficient	Model		
	Food expenditures (Pissarides–Weber)	Food expenditures (RLMS, 2012)	Spending on clothing (RLMS, 2012)
γ	0.0919***	-0.1815***	0.0921**
β	0.2695***	0.3263***	0.3604***

Note: The level of significance: * 10%, ** 5%, *** 1%.

tures that include the cost of transportation and are linked to the place of work. Clothing expenditures meet both conditions, so they can be used as an indicator of a household's well-being.

Thus, it is advisable to turn to the assessment of the equation where the dependent variable is the logarithm of spending on clothes for adults and children over the past 90 days (the scale of the variable is not important as long as the logarithm is used).

$$\ln C_h^{Cl} = Z_h' \alpha + \beta_{Cl} \ln Y_h + \gamma_{Cl} D_{SE_h} + u_h^{Cl} \quad (15)$$

The spending on clothing model produces a significant coefficient showing an excess of expenditures by the self-employed compared with other groups in the sample. Table 5 shows the correlation between the model estimates of expenditures on food and clothing in this study and those estimates obtained by Pissarides and Weber (1989) when modeling food expenditures. Differences in estimates of the β coefficient are due to both the choice of a different expenditures category, and to differences in the standard of living for the Russian and British households described in Pissarides and Weber. The negative estimate of the γ coefficient in the model of food expenditures for Russian households shows that the model's assumption of correctness for the declared food expenditures was incorrect for Russian households. However, the model of clothing expenditures allows a meaningful estimate of the γ parameter that is nearly identical to Pissarides and Weber's.

This model will be used as the basis for calculating the proportion of household income and researching the hidden dependencies of this value on the household's socio-economic and demographic characteristics.

It should be noted that the income value in equation (15) is correlated with an error because external shocks simultaneously affect household expenditures and income. Therefore, the income variable is endogenous in the cost model and needs instrumentation. (Another reason for the endogeneity of the income variable, as previously mentioned, is its measurement errors). As identifying instruments, X_h , we used the indicators of a household's possession of expensive durables and expensive foreign-made cars. The set of instruments turns out to be radically different from that used by Pissarides and Weber. The instruments are strong and exogenous and the details are listed in Appendix Table A2. (The correctness of the instruments was validated by the Hansen test; the relevance of the tools was validated by the Yogo test).

Another methodological feature of this research is the use of random effects in the expenditure model on a constant and a coefficient of γ :

$$\ln C_h^{Cl} = \alpha_0 + \alpha_{0i} + \sum Z_{jh} \alpha_j + \beta_{Cl} \ln Y_h + (\gamma_{Cl} + \gamma_{Cli}) D_{SE_h} + u_h^{Cl} \quad (16)$$

The i index means either the number of the income group to which a household belongs or the number of the group that corresponds to the size of the locality, or it means the number of the group corresponding to the administrative status of the locality where the household resides. Random effects for groups $\alpha_{0i} \sim iid(0, \sigma_{\alpha_0}^2)$ and $\gamma_{Cli} \sim iid(0, \sigma_{\gamma}^2)$ are assumed to be uncorrelated among themselves and between groups. The use of this set of instruments enables us to estimate the model across the entire sample, and it provides a number of observations sufficient to obtain

adequate estimates. This approach also improves the possibility of cross-group expenditure heterogeneity and the heterogeneity of the distribution of self-employed households between the groups, which ensures the consistency of estimates.

Table 6 contains estimates of the elasticity of expenditures on clothing by income β for certain household categories and estimates of the γ coefficient indicating, after multiplying by 100, the percentage by which self-employed expenditures exceed expenditures by other households, all other things being equal.

Table 6 shows that in terms of the settlement size, expenditures by the self-employed for clothing are higher than the expenditures of hired employees in metropolitan areas by approximately 15%, while in million-plus cities, small cities, and urban-type villages they are approximately 10% higher; and in villages, expenses for the self-employed are higher than those of hired employees by approximately 5%. In terms of the administrative status of settlements, expenditures on clothing by the self-employed in regional centers also exceed expenditures by hired workers by 15%, while for all other types of settlement statuses, the excess is approximately 5–6%, and there are no large differences in the excess. This greatly affects the specific style of clothing consumption and the price range available at different types of settlements. Regarding income groups, poor self-employed households spend 20% more on clothing than other poor households, which is approximately 4 times greater than the differences between the self-employed and other households from the middle group, and about twice as much as for the affluent households, all other things being equal.

These results suggest that differences in the consumption of clothing by the self-employed and other households do exist and vary considerably across income and settlement groups. This allows us to assess the proportion of concealed income within the selected subgroups.

The next step to estimate the proportion of concealed income requires an evaluation of residual income dispersions for the self-employed and other households

Table 6
Estimates of β and γ .

Parameter estimate	Size of the locality				
	Less than 100,000 people	From 100,001 to 400,000 people	From 400,001 to 1,000,000 people	From 1,000,000 to 10,000,000 people	Over 10,000,000 people
γ	0.0515	0.1198	0.1258	0.0981	0.1508
β	0.3553	0.3553	0.3553	0.3553	0.3553
Parameter estimate	A locality's administrative status				
	Village	Urban-type village	City	Regional center	
γ	0.0561	0.0621	0.0601	0.1425	
β	0.3478	0.3478	0.3478	0.3478	
Parameter estimate	Income group				
	< 0.5 of median income	0.5–1.5 of median income	> 1.5 of median income		
γ	0.1919	0.0409	0.1194		
β	0.3598	0.3598	0.3598		

on auxiliary regressions of the income logarithm against the household characteristics vector Z_i and a set of tools X_i and the subsequent calculation of lower and upper limits for the under-reported income parameter k (according to expressions (12) and (13)).

The estimate of the under-reported income share is based on interval estimates of k as follows:

$$Shadow_Share = \frac{\alpha I_{se}(k-1)}{\alpha I_{se} + (1-\alpha)I_{other}} \quad (17)$$

where k is the value of the under-reported income parameter (the higher the value, the greater the need to adjust the household income to meet the “true” level); α is the share of self-employed households in the sample relative to the total number of households (the higher the percentage, the more households escape to the shadow economy); I_{se} is the average income for the self-employed (the more earned by the self-employed, the greater the amount of adjusted income); I_{other} is the average income of other households (not self-employed). (Here, as in the past, the self-employed include those and only those households that obtain more than 20% of their total income from self-employment.)

From the above expression, it follows that, in the event of equal incomes between the self-employed and other households, the share of the concealed income will be proportional to the share of the self-employed within the sample. The higher the income of the self-employed compared to other household income, the higher the estimate.

Table 7 shows the results of estimating the under-reported income parameter and the proportion of income concealed across the analyzed sample of households. The median income for self-employed households exceeds the median income of the sample by 13%. Other households are below the sample average by 2.4%.

On average, the proportion of income concealed is 5% of common household income. Pissarides and Weber’s estimates for British households in 1989 were 5.5%, but the essential difference is that their sample was more related to the general population, and their analysis demonstrated that, on average, the share of the shadow economy accounts for 5.5% of GDP in the UK. We cannot extrapolate the sample estimates to the general population in this study, but we can try to extrapolate, subject to the interval distributions of per capita income based on Rosstat data and the RLMS sample for 2012, as described in Fig. 1 (details of the calculations are shown in Appendix Table A7).

Table 7

Interval estimate of concealed income share.

The sample size	4320
Number of self-employed	660
Percentage of self-employed	0.150
Average total income (RUB)	42 649
Average total income of the self-employed (RUB)	48 289
Average total income of others (RUB)	41 632
Lower limit of the under-reported income parameter	1.240
Upper limit of the under-reported income parameter	1.350
Lower limit of the proportion of concealed income	0.040
Upper limit of the proportion of concealed income	0.060

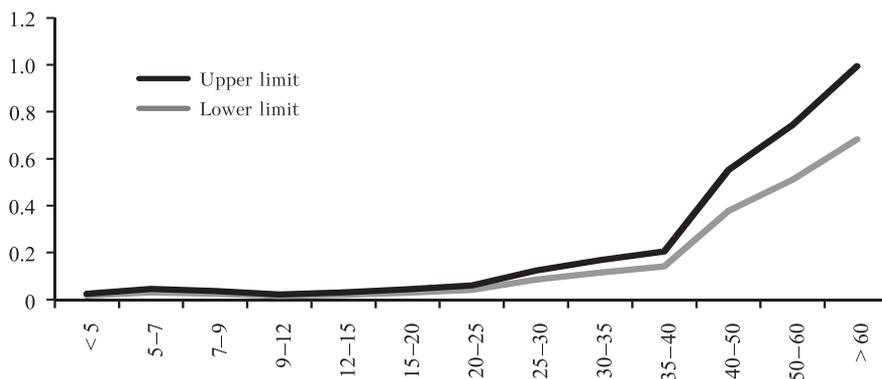


Fig. 3. Extrapolation of under-reported income share to the distribution of the general population by the income in the interval estimates based on per capita income and the distributions of per capita income (RUB thousand).

Source: RLMS data for 2012, and the household budget survey data from Rosstat, adjusted for balance of income and expenditure indicators for the population during the same period.

Fig. 3 demonstrates that in the last five intervals of per capita income the RLMS sample includes, on average, only 20% of the households from the corresponding income groups within the general population, so the extrapolation produces very rough estimates. Averaging across the extrapolated general population yields lower and upper limits of concealed income of 16% and 23%, respectively. These results fall within the range of estimates of the scale of the shadow economy in Russia in 2011 given by Rosstat based on SNA (15%) and the findings of individual experts (35–40%) for the same period (Table 1). However, they are half as high as the World Bank's estimates for 2011 (50%), shown in Table 1. The differences might be explained by the fact that this study does not take into account the contribution of the shadow economy within the criminal sector.

Fig. 4 demonstrates the interval estimates of the share of concealed income, grouped by sub-samples of the RLMS sample for 2012. More detailed information on the calculation results is provided in Appendix Tables A4–A6.

A study of the schedules shows that Hypothesis 3 (*in rural and regional centers, the share of concealed income is higher than in urban-type and secondary cities*) does not contradict the data of the study (Fig. 4a). Further analysis shows that the descriptive statistics for the sample proportion are essential to farms, which, if they are not required to receive subsidies for development, avoid disclosing their economic activity. This produces a spike on the left end of the graph. With regard to the proportion of concealed income, there is a surge on the right end in regional centers. Here, we need to be careful drawing conclusions, because regional centers are very heterogeneous in size and functions. This segment includes Moscow and St. Petersburg, and perhaps their presence in the subsample of households from regional centers leads to the high estimates of the share of concealed income.

If we look at the situation from a different angle, dividing the households by settlement size, it all looks slightly different (Fig. 4b). Therefore, Hypothesis 2 (*in small settlements and million-plus cities, the proportion of concealed income is higher than in medium-sized towns and villages*) is not completely consistent with the results of the estimation. With the growth of settlements up to 1 million people, the share of concealed income gradually grows to 10% of total household income.

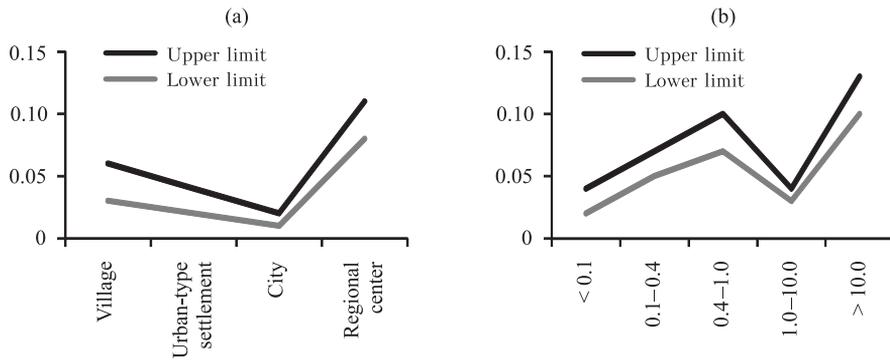


Fig. 4. Interval estimates of the proportion of concealed income based on (a) the administrative status of the locality where the household resides and (b) the size of the settlement (in millions).

However, in million-plus cities, it suddenly falls below 5%, while in metropolitan areas, it again rises and reaches its highest mark at 13%. Perhaps the failure of cities is associated with the bias and the small size of the household sample (42) for this category. If we do not take into account these technical considerations, we can assume that in such cities, unlike previous categories, opportunities do arise for formal employment. At the same time, such attractive conditions for self-employment like those in metropolitan areas do not necessarily arise, and active representatives of the population of these cities who strive for self-employment—not out of a need for physical survival, but with a view to achieving a high standard of living—migrate to metropolitan areas. This is supported by the smaller proportion of the self-employed (8%), which is half the size as in other categories.

Fig. 5 illustrates the estimates of the share of concealed income based on the various household income groups formed in relation to the median household income in the sample.

The results support Hypothesis 1 (*in the lower and upper income groups, the proportion of concealed income is higher than in the middle*). However, it turns out that the poorest households hide nearly 1.5 times more income than in the sub-sample of the most affluent households, whereas the percentage of self-employed households is the same for all three subsamples. The number of observations in the subsamples is sufficiently large (> 600), so it is unlikely that there are any technical reasons to consider the estimates inadequate. Rather, we are dealing here with essentially economic mechanisms. A lack of jobs and the need

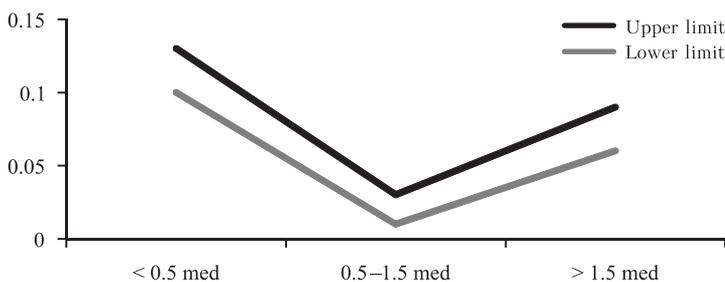


Fig. 5. Interval estimate of the proportion of concealed income per household income group relative to the median household income in the RLMS sample (med = RUB 27,160).

to compensate for the costs of inflation, which are much higher for poor households than for the wealthy (Ershov, Matitsin, 2012), force poor households to hide more of their income than other household categories, even the affluent.

7. Conclusion

In this study, an attempt was made to estimate the proportion of income concealed by Russian households based on RLMS data for 2012. As a theoretical framework for the study, we used a model proposed in Pissarides and Weber (1989). This model describes the behavior of consumer households accounting for concealed income on the one hand and the relationship between income and consumption on the other. The model assumptions that food consumption by self-employed households exceeds similar consumption by others and that food expenditures are reported correctly are inconsistent for Russian households in the sample. However, the cost of clothing is an expenditure category for current consumer goods that can help evaluate the proportion of concealed income.

Because revenues are measured with errors and because external shocks simultaneously affect income and expenses, income constitutes an endogenous regressor that requires instrumentation. Strong and exogenous instruments for income in the clothing expenditure equation are indicators of the presence of expensive durable goods and foreign-made cars in a given household. The resulting set of instruments is radically different from the set used in Pissarides and Weber (1989).

For economical and adequate accounting of the heterogeneity of expenditures in various income and settlement groups, we used a multilevel (hierarchical) simulation technique not previously used in similar studies. (We introduced random group effects into the model on the constant and a coefficient with a variable indicating whether the household is self-employed).

In their fundamental study, Pissarides and Weber (1989) were interested in a comparative analysis of the share of concealed income between white-collar and blue-collar households. In this study, it appeared more relevant to analyze the change in the share of concealed income by income and settlement groups of households, formed in relation to the median household income across the entire sample and based on the size and administrative status of the settlement. We formulated three research hypotheses regarding the behavior trends of the proportion of concealed income depending on income, the size of the locality, and its administrative status. The resulting estimates do not contradict the hypotheses that the proportion of income concealed in the extreme categories (“poor” and “rich” households; households in rural areas and regional centers) are higher than average. However, the assumption that revenues will more often be concealed in small settlements was not confirmed: as the size of settlements increases to 1 million people, the share of concealed income grows gradually to 10% of the total household income; in million-plus cities, it suddenly falls below 5%; and for metropolitan areas, it rises again and reaches a high point at 13%.

The estimated concealed incomes based on the RLMS sample, and extrapolated to the general population, were comparable to estimates by other experts and researchers. Averaging the results obtained by extrapolation produces estimates of the lower and upper limits of the share of concealed income at 16% and 23% respectively. These results fall within the interval of estimates of the scale

of the shadow economy in Russia conducted in 2011 by Rosstat based on SNA (15%), and the findings of individual experts (35–40%) for the same period; however, they are half as high as the World Bank's estimates for 2011 (50%). The differences may be explained by the fact that this study does not take into account the contribution of the criminal element to the shadow economy.

This paper has developed an approach to assessing the share of household income excluded from statistical observations, and it shows that, drawing on the results, we can get an idea of the size of the shadow economy in Russia, both in general and with respect to certain social groups. The estimate associated with certain groups can be especially significant, because when analyzing the effects of pricing and tax policies, it could prove useful to understanding which types of households are more likely to escape into the shadow. Such assessments can clarify the extent to which escaping to the shadow will mitigate the excesses of economic policies for different population groups and will thus serve as a “social shock absorber”. It can also illuminate the extent to which the shadow economy may also do harm to the pension system.

This approach may be more broadly used, as household income indicators are applied to estimate the well-being of different social strata within the population and the degree of stratification among them. Also, it is important to understand how accounting for income measurement errors can change the view of the situation and affect the mechanisms for providing targeted social support.

References

- Barsukova, S. (2005). Structure and institutes of informal economy. *Sotsiologicheskii Zhurnal*, 3, 118–134 (In Russian).
- Burov, V. Yu. (2012). Definition of scale of shadow economy. *Vestnik Ekonomist ZABGU* [On-line serial], 4 (In Russian).
- Burov, V. Yu., & Samaruha, V. I. (2010). *Shadow economy in the system of regional enterprise activity*. Irkutsk: BGUEP (In Russian).
- Cassel, D., & Cichy, U. (1987). The shadow economy and economic policy in East and West: A comparative system approach. In S. Alessandrini, & B. Dallago (Eds.), *The unofficial economy. Consequences and perspectives in different economic systems* (pp. 127–145). Aldershot: Gower.
- Del'Anno, R. (2003). Estimating the shadow economy in Italy: A structural equation approach. Discussion Paper, Department of Economics and Statistics, University of Salerno.
- Gardes, F., & Starzec, C. (2009). Polish households behavior in the regular and informal economies. *Revue économique*, 60 (5), 1181–1210.
- Gorodnichenko, Yu., Martinez-Vazquez, J., & Sabirianova Peter, K. (2009). Myth and reality of flat tax rate reform: Micro estimates of tax evasion response and welfare effects in Russia. *Journal of Political Economy*, 117 (3), 504–554.
- Kirienko, A. P., & Ivanov, Yu. B. (2013). Estimation of shadow economy on the base of indicators of level and quality of life. *Izvestiya IGA*, 4, 109–113 (In Russian).
- Lukyanova, A. L. (2012). *Pension insurance, informal employment and self-employment* (Preprint WP3/2012/02). Moscow: HSE Publ. (In Russian).
- Lyssiotou, P., Pashardes, P., & Stengos, T. (2004). Estimates of the black economy based on consumer demand approaches. *Economic Journal*, 114 (497), 622–640.
- Matitsyn, M. S., & Ershov, E. B. (2012). Research of Russian population differentiation on real incomes. *Ekonomicheskii zhurnal VShE*, 3, 318–340 (In Russian).
- Montmarquette, C., & Gardes, F. (2002). How large is your reference group. *CIRANO Working Papers*, 2002s-87.

- Pissarides, C., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39 (1), 17–32.
- Ponomarenko, A. (1995). What does the statistical term “shadow economy” mean and how it is represented in national accounts. *Voprosy Statistiki*, 6, 3–7 (In Russian).
- Popov, Yu. N., & Tarasov, M. E. (2005). *Shadow economy in the system of market economy*. Moscow: Delo (In Russian).
- RLMS-HSE (2012). *The Russia Longitudinal Monitoring Survey*. Moscow: National research university Higher school of economics (In Russian).
- Schneider, F., & Enste, D. (2000). Shadow economies: Size, causes and consequences. *Journal of economic literature*, 38 (1), 77–114.
- Suvorov, A. V. (2008). Problems of estimation of income differentiation in modern Russia. *Problemy prognozirovaniya*, 2, 3–18 (In Russian).
- Timchenko, V. A. (2004). Shadow economy: Definition, causes, socio-economic effects and scale. *Dengi*, 12 (In Russian).

Appendix A

Table A1

Descriptive statistics for the socio-demographic characteristics of household members (%).

Variable	Entire sample	No income from s/e	Share of revenue from s/e < 20%	Share of revenue from s/e ≥ 20%
Married members of the household	52.80	54.09	52.54	48.28
Household members engaged in skilled labor	45.80	49.79	42.12	38.17
Males heading households engaged in skilled labor	33.32	35.52	32.00	27.53
Household members with secondary vocational education	41.76	42.49	42.32	37.74
Members of households with higher education	36.69	39.05	33.80	33.55
Children in the household	89.87	90.05	90.70	87.42
Household members stating that they are fully satisfied with their life	15.51	15.29	15.15	17.10
Household members wishing to find another job	31.19	29.71	29.94	39.57
Any member of the household with an additional job	4.24	4.60	3.54	4.30
The household's ability to improve its living conditions	11.05	11.63	9.09	12.90
The household's ability to spend a holiday with all of the family members	20.57	22.68	17.72	18.28
Household ability to pay for the university education of its members	19.27	20.53	16.74	19.68
School children	30.4	28.5	30.6	37.6
Any member of the household receiving higher education	16.8	17.1	14.7	19.9
Household size (number of persons)—median	2	2	2	3
Household size (number of persons)—average	2.71	2.63	2.79	2.87
Number of children in the household—median	2	2	2	2
Number of children per household—average	1.70	1.64	1.77	1.76

Table A2

List of statistically significant explanatory and instrumental variables for regression models of expenditures (statistically significant at the 1% level of significance).

List of regressors (Z)	List of tools (X)
The number of members in the household, number of children	The presence of foreign cars, dishwasher, high speed internet, plasma TV, alternative housing,
Ability to pay for the education of the children	
The presence of a rented housing, Size of living space	
Crops harvested from the land	The opportunity to spend holidays with the whole family
Washing machine, cell phone, digital camera, DVD, computer, laptop, bicycle	
The size of the settlement (for models by type groups)	Head of household participation in the economically active population

Table A3

Descriptive statistics of household characteristics in general (%).

Variable	Entire sample	No income from s/e	Share of revenue from s/e < 20%	Share of revenue from s/e ≥ 20%
<i>Place of residence</i>				
Village	24.0	21.2	24.8	33.2
Urban-type settlement	5.6	5.9	5.2	5.3
City	26.3	27.4	27.7	18.7
Regional Center	44.1	45.4	42.3	42.8
<i>Possession of real estate and housing facilities</i>				
Residence in partial ownership	3.9	3.6	3.7	5.3
Housing is not in private ownership	6.0	5.4	7.3	5.7
Total area (sq.m.)	54.9	54.9	54.0	57.1
Living area (sq.m.)	36.7	36.4	35.8	39.5
Number of rooms	2.3	2.3	2.3	2.4
Central heating	71.7	74.6	70.4	62.9
Central water supply system	87.8	88.7	87.0	85.6
Hot water supply	66.2	68.5	65.4	58.7
Sewer	73.7	76.8	72.3	64.3
Phone	59.1	59.3	63.4	49.2
Gas supply	66.0	67.9	65.5	60.0
Possession of a country house	21.3	22.6	21.6	15.3
Possession of another apartment	8.5	8.6	7.9	9.7
<i>Possession of land</i>				
Ownership of land	50.1	49.1	52.2	49.4
Land lease	34.2	34.0	34.6	34.2
Crops harvested from the land	44.8	43.4	47.7	44.2
Sale of harvested crops	2.1	1.1	2.2	5.7
<i>Possession of technical equipment</i>				
Possession of a domestically-produced car	22.0	21.4	23.2	21.6
Possession of a foreign car	20.7	23.3	17.3	17.6
Possession of a truck	2.3	2.3	2.0	2.9
Possession of a motorcycle	3.1	2.9	3.2	3.7
Possession of a bicycle	21.2	20.1	23.2	20.9
Possession of a tractor	2.4	2.1	2.3	3.7
Possession of a lawn mower	7.5	8.0	7.5	5.2
<i>Possession of a computer and internet access</i>				
Possession of a computer	42.1	44.6	39.2	38.5
Possession of a laptop	32.3	33.9	28.1	34.7
Low-speed internet access	18.1	18.8	16.3	19.5
Broadband internet access	37.0	38.8	34.2	36.0
<i>Possession of other durable goods</i>				
Possession of electric oven	23.4	24.2	21.5	24.3
Possession of a refrigerator	49.5	51.3	46.8	48.0
Possession of a freezer	11.9	12.4	11.1	11.5
Possession of a washing machine	72.3	74.9	68.6	69.9
Possession of a microwave oven	61.2	63.0	58.9	58.5
Possession of a dishwasher	2.7	3.1	1.9	3.0
Possession of a color TV	78.5	77.9	81.2	75.7
Possession of a plasma TV	42.3	44.7	40.2	37.3
Possession of a player	24.7	26.3	23.2	21.9
Possession of a DVD player	45.5	45.1	44.3	49.0
Possession of a digital camera	36.9	38.8	33.6	36.5
Possession of camcorder	7.7	7.8	7.6	7.4
Possession of an MP3 player	10.4	11.1	8.8	11.0
Possession of a GPS navigator	7.2	8.1	6.1	6.5
Possession of an air conditioner	8.2	8.1	8.2	8.3
Possession of a satellite antenna	17.0	15.5	17.4	21.9
Possession of cable TV	30.5	32.4	29.5	25.1

Table A4

Interval estimate of the proportion of concealed income, depending on the size of the locality.

Model	Less than 100,000 people	From 100,001 to 400,000 people	From 400,001 to 1,000,000 people	From 1,000,000 to 10,000,000 people	More than 10 million people
The sample size	2038	653	700	498	431
Number of self-employed	338	93	118	42	69.0
Percentage of self-employed	0.17	0.14	0.17	0.08	0.16
Average total income (RUB)	35 865.24	38 868.37	42 991.25	44 772.72	77 445.13
Average total income of self-employed (RUB)	37 205.14	42 122.53	51 168.94	51 838.36	103 808.67
Average total income of others (RUB)	35 598.84	38 327.95	41 333.23	44 121.94	72 420.03
Grade k	1.16	1.40	1.42	1.32	1.53
Lower limit of estimate k	1.11	1.34	1.37	1.26	1.47
Upper limit of estimation k	1.21	1.46	1.49	1.37	1.59
Average proportion of concealed income	0.03	0.06	0.09	0.03	0.11
Lower limit of the proportion of concealed income	0.02	0.05	0.07	0.03	0.10
Upper limit of the proportion of concealed income	0.04	0.07	0.10	0.04	0.13

Table A5

Interval estimate of the proportion of concealed income, depending on the administrative status of the locality.

Model	Village	Urban-type settlement	City	Regional center
The sample size	995	238	1,146	1,941
Number of self-employed	215	34	121	290
Percentage of self-employed	0.22	0.14	0.11	0.15
Average total income (RUB)	33 036.45	42 973.50	37 879.59	50 352.90
Average total income of the self-employed (RUB)	38 772.06	42 367.32	33 578.28	62 176.63
Average total income of others (RUB)	31 455.48	43 074.53	38 387.35	48 276.05
Grade k	1.17	1.20	1.19	1.51
Lower limit of estimate k	1.11	1.13	1.12	1.42
Upper limit of estimation k	1.24	1.27	1.26	1.59
Average proportion of concealed income	0.04	0.03	0.02	0.09
Lower limit of the proportion of concealed income	0.03	0.02	0.01	0.08
Upper limit of the proportion of concealed income	0.06	0.04	0.02	0.11

Table A6

Interval estimate of the proportion of concealed income depending on the income group.

Model	Median income		
	< 0.5	0.5—1.5	> 1.5
The sample size	670	2 253	1 397
Number of self-employed	109	330	221
Percentage of self-employed	0.16	0.15	0.16
Average total income (RUB)	10 871.63	29 442.93	79 187.57
Average total income of the self-employed (RUB)	10 509.79	28 311.95	96 751.88
Average total income of others (RUB)	10 941.94	29 637.01	75 886.79
Grade k	1.70	1.12	1.39
Lower limit of estimate k	1.61	1.06	1.32
Upper limit of estimation k	1.80	1.19	1.47
Average proportion of concealed income	0.11	0.02	0.08
Lower limit of the proportion of concealed income	0.10	0.01	0.06
Upper limit of the proportion of concealed income	0.13	0.03	0.09

Table A7
Sample estimates of the limits of the proportion of concealed income and their extrapolation to the general population.

	Per capita income (RUB thousand)												
	< 5	5–7	7–9	9–12	12–15	15–20	20–25	25–30	30–35	35–40	40–50	50–60	< 60
Rosstat (% of general population) (1)	5.70	6.80	7.90	12.00	10.80	14.6	10.70	7.80	5.70	4.10	5.40	3.10	5.40
RLMS (% of sample) (2)	10.68	8.67	12.54	20.79	15.05	14.42	7.58	3.91	2.07	1.38	1.20	0.64	1.06
(1) / (2)	0.53	0.78	0.63	0.58	0.72	1.01	1.41	1.99	2.75	2.97	4.50	4.84	5.09
The share of the self-employed in the RLMS sample	0.14	0.16	0.15	0.11	0.12	0.14	0.13	0.18	0.17	0.22	0.35	0.31	0.46
Average income of self-employed households (RLMS) (RUB)	25 511.5	20 696.1	29 603.9	30 515.3	39 587.7	44 362.4	56 161.7	70 999.5	95 993.1	94 312.5	104 831.8	182 457.1	245 372.5
The median income of other households (RLMS) (RUB)	27 278.4	19 553.8	26 153.6	31 898.2	38 833.9	50 711.0	60 470.8	71 741.3	91 808.8	104 896.8	103 482.7	103 696.0	164 348.3
Average household income (RLMS) (RUB)	26 948.8	19 741.0	26 753.0	31 747.2	38 931.6	49 843.9	60 031.1	71 595.6	92 471.7	102 988.2	103 994.4	122 073.6	198 378.4
Estimate of the limit of the under-reported income parameter (RLMS)													
lower	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24	1.24
upper	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
Estimate of the limit of the proportion of concealed income (RLMS)													
lower	0.03	0.04	0.04	0.03	0.03	0.03	0.03	0.04	0.04	0.05	0.08	0.11	0.13
upper	0.05	0.06	0.06	0.04	0.04	0.04	0.04	0.06	0.06	0.07	0.12	0.15	0.20
Estimate of the limit of the proportion of concealed income (extrapolation)													
lower	0.02	0.03	0.03	0.02	0.02	0.03	0.04	0.08	0.12	0.14	0.38	0.51	0.68
upper	0.03	0.05	0.04	0.02	0.03	0.04	0.06	0.12	0.17	0.21	0.55	0.74	0.99