



Received: 13 February 2017  
Accepted: 13 March 2017  
Published: 05 April 2017

\*Corresponding author: Yaroslav Murashov, Higher School of Economics, National Research University, 26, Shabolovka Street, Moscow 119049, Russian Federation  
E-mail: [yaroslav.48532@rambler.ru](mailto:yaroslav.48532@rambler.ru)

Reviewing editor:  
Bernardo Spagnolo, Università di Palermo, Italy

Additional information is available at the end of the article

## ECONOMETRICS | RESEARCH ARTICLE

# Estimation of the Russian informal economy size on the household budget survey data

Yaroslav Murashov<sup>1\*</sup> and Tatiana Ratnikova<sup>1</sup>

**Abstract:** In the paper, we make an attempt to estimate the size of informal economy on the base of micro-data. Two main approaches are described and compared. They are implemented on the base of an RLMS sample for 2012. The first method, called single equation approach, is based on the specific category of household expenditures and the arbitrarily defined household type (self-employed or employee). The second method allows to obtain the results for income under-report for both wage income and self-employment income, and uses information on all the household current consumption categories. The single equation model is restricted to one expenditure category and two types of households, although it enables to perform the estimation on different subsamples of households with various socioeconomic characteristics. The comparison of the system of equations approach with single equation is made concerning the scale of informal economy and the role of wage-income under-report, which is possible to obtain through the system.

**Subjects:** Microeconomics; Employment & Unemployment; Econometrics; Mathematical Economics

**Keywords:** under-reported household income; household consumer behavior; under-reported income parameter; regression analyses

**JEL classifications:** C21; C26; D11; D12

### ABOUT THE AUTHORS

Yaroslav Murashov, postgraduate at NRU HSE, Topic of the thesis: estimation of the size of informal economy of Russian Federation on the base of household budget survey data. Education: bachelor degree, NRU HSE (2012); master degree, NRU HSE (2014); master degree, Paris1-Pantéon Sorbonne (2014); main field of study: quantitative economics. Career: economist at the Central Bank of Russian Federation (from 2015).

Tatiana Ratnikova is an associate professor of Economic Faculty (NRU HSE), senior Fellow in Laboratory of Russia Economic Macrostructure Modeling (NRU HSE). *Academic degree:* Candidate of Science - Keldysh Institute of Applied Mathematics of the Russian Academy of Sciences (1996). *Prior education:* NRU - Moscow Engineering Physics Institute (graduated in 1988, faculty: Theoretical and Experimental Physics, specialty: Theoretical Nuclear Physics). *Research interests:* Panel Data Analysis; FDI and Firm Productivity; Hedonistic Regression for Art Objects; Shadow Income of Household.

### PUBLIC INTEREST STATEMENT

The paper provides an introduction to the micro-based methods to estimate the size of informal economy, with the help of the reported information of the amount of goods consumed by a household and its income. This approach is alternative to the macro approaches to estimate the informal economy and in some cases it enables to provide comparative results, as far as the informal economy has monetary estimation (a share of concealed income). The edge of the method is the flexibility to take into account the socio-demographic characteristics of the household, as far as its income structure.

### 1. Methods to estimate the informal economy: Literature review

Black economy is by definition the activity concealed from law. So it will not be indicated in the official revenue documents. Although, if we are looking at the revenue surveys, they may include some information about the black economy activities, but to some extent, because people still have some incentives to conceal black economy activities. Therefore, researches may try to estimate the part of the economic activity which is treated as the black economy. One of the ways to estimate the black economy is microeconomic approach and is based on the data obtained from households budget surveys. The key idea is that the households indicate their expenditures more accurately than they indicate their income.

Actually, the main idea of the authors Pissarides and Weber (1989) is that (1) the reporting of expenditure on *some items* by *all groups* of population is accurate; (2) the reporting of income by *some groups* of the population is accurate. The income reporting is accurate not by some groups, but by some types of occupation of the population. The authors believe that the expenditure item which is recorded correctly is the expenditures on food (the less likely to conceal). The under-report of income comes from the people who are self-employed. Employees report their income correct. Although this may look like a strong assumption that only the self-employed are those who under-report their income, because for most of the people to indicate the wage expenditures is not always reasonable. They may obtain other wages which are not indicated in the responses to the questionnaire.

In the article written by Lyssiotou, Pashardes, and Stengos (2004) the authors actually use the method of Pissarides and Weber (1989) proposed in 1989 to estimate single food expenditure equation. But the authors go further in the estimation proposing the two other methods of black economy estimation.

The basic way of developing the model of Pissarides and Weber used by the authors is to include in the analyses the demand for durable goods. By dividing the demand into the demand for durable and the demand for non-durable goods one may notice that the share of income a household spends on durable goods depends not only on the level of income, but on the income source. Therefore, self-employment income may indicate not only income under-report, but the preference heterogeneity: self-employed people may tend to spend more on expensive goods (consumption of durable goods), therefore they are spending too little on food and other non-durable goods.

Another reason why the self-employment income is not a good proxy for the utility is that it tends to be more volatile, therefore, it influences savings. Households spend less and earn more than employees to meet the precautionary savings.

Estimation of the size of informal economy in Turkey estimated by Aktuna, Starzec, and Gardes (2011) is based on the model proposed by Lyssiotou et al. (2004). But now the authors assume that a household has three forms of income: wage income, self-employment and other income. The other income seems to be reported correctly, while the other income components may be hidden. Another enlargement is that the preference heterogeneity term depends on the share of each part of the income (the heterogeneity is viewed in the form of difference in savings between different occupation groups and consumption patterns).

The estimated system of equations is actually the same. The difference is that the system is estimated by generalized method of moments and the household production activities are included in the system in the form of household consumption and household production. So the initial model is enlarged and includes the consumption of household goods and the income from their production evaluated by the market price.

## 2. The basic models description

Let us describe the basic model as it is presented by Pissarides and Weber. The variables reported by the households: (1) consumption of individual items  $C_{ij}$  ( $i$  is the household index,  $j$  is the index of the index of the consumption item); (2) after-tax income  $Y_i^d$ ; (3) vector of household characteristics  $Z_i$ . According to our assumptions, there is under-report of the level of income for households in self-employment. Denote by  $k_i$  the coefficient by which a household  $i$  is underreporting their income. Then the connection between the actual and reported income is written in the following form:  $Y_i = k_i \cdot Y_i^d$ . So  $k_i = 1$  for employees and  $k_i > 1$  for the self-employed. The expenditure function of item  $j$  is written in the following form:  $\ln C_{ij} = Z_i' \alpha_j + \beta_j \ln Y_i^p + \varepsilon_{ij}$ . It is needed to note the definition of  $Y_i^p$ . We denote by this the measurement of the income influencing consumption decisions. The relation between the permanent and actual income is measured by  $Y_i = p_i \cdot Y_i^p$ . This parameter accounts for the variation of income due to the unforeseen circumstances. The mean of  $p_i$  does not depend on the type of the household and is the same for the employees and the self-employed. On the other hand, the variation of the parameter is different depending on the type of a household:  $\text{Var}(p_i|i - ee) < \text{Var}(p_i|i - se)$ . If the household is self-employed, then the variation of this parameter is higher, reflecting the higher variation of the self-employed income. That is why current consumption is a function of not a current income, but the permanent income.

Therefore, permanent income, which is directly related to the consumption function, can be decomposed into current income and the parameters of the model through the following way:

$$\ln Y_i^p = \ln Y_i^d - \ln p_i + \ln k_i$$

This implies existence of the two additional random regressors in the model if we put into the model observed income instead of the permanent income. To verify statistical hypothesis the authors must make assumptions concerning the statistical distribution of the coefficients responsible for the income mismatch. The coefficients have log-normal distribution. So now write them in the form of deviations from their means:  $\ln p_i = \mu_p + u_i$ ,  $\ln k_i = \mu_k + v_i$ . Now the trick performing the connection between the mean of  $p_i$  and the mean of its logarithm,  $E(\ln p_i) = \mu_p + \frac{1}{2}\sigma_u^2$ . As far as the mean of  $p_i$  does not depend on the type of the household, one may compare the mean of the log for the employees and the self-employed:  $\mu_{pse} - \mu_{pee} = \frac{1}{2}(\sigma_{uee}^2 - \sigma_{use}^2) \leq 0$ . How one can use this fact? Let us put the error terms decomposition into the expenditure function. Then we obtain the following decomposition of permanent income  $\ln Y_i^p = \ln Y_i^d - (\mu_p + u_i) + (\mu_k + v_i)$ .

This decomposition can be used for the expenditure function.

Then,  $\ln C_{ij} = Z_i' \alpha_j + \beta_j \ln Y_i^d - \beta_j (\mu_p - \mu_k) - \beta_j (u_i - v_i) + \varepsilon_{ij}$ . The dependent variable is the household expenditures on some particular item. In authors model food expenditures will be considered. The idea of this quite simple model is to obtain income differences for the self-employed. If this equation is estimated separately for employees and self-employed then the intercepts differ as far as  $\mu_{kse} - \mu_{kee}$  is not the same at each group. This constant term may give information on the size of black economy.

To estimate this relationship the following type of regression model was used:

$\ln C_{ij} = Z_i' \alpha_j + \beta_j \ln Y_i^d + \gamma_j D_{SE_i} + \xi_{ij}$ , where  $D_{SE_i}$  is a dummy variable taking value "1" if the individual is self-employed, and zero if he is employee. This model can be straightforward estimated by the method of least squares, accounting for the heteroscedasticity of error term  $\xi_{ij}$ . But how can we exactly compute the level of income under-reporting with the help of this model?

According with the theoretical income decomposition for the expenditure function the coefficient  $\gamma_j$  must be equal to the term  $E\{-\beta_j(\mu_p - \mu_k)|se\} - E\{-\beta_j(\mu_p - \mu_k)|ee\}$ . The estimated coefficient equals:  $\gamma_j = \beta_j \left\{ \mu_k + \frac{1}{2}(\sigma_{use}^2 - \sigma_{uee}^2) \right\}$ .

The income decomposition enables us to perform the reduced-form regressions for income. Therefore the observable income can be decomposed in the following form:  $\ln Y_i^d = Z_i' \delta_1 + X_i' \delta_2 + \zeta_i$ . Due to the income decomposition, we obtain that  $\sigma_{\zeta_{se}}^2 - \sigma_{\zeta_{ee}}^2 = \sigma_{use}^2 - 2\text{COV}(u, v)_{se} + \sigma_{vse}^2 - \sigma_{uee}^2$ . From the expression  $\gamma_j = \beta_j \left\{ \mu_k + \frac{1}{2} (\sigma_{use}^2 - \sigma_{uee}^2) \right\}$  we have  $\mu_k = \frac{\gamma_j}{\beta_j} - \frac{1}{2} (\sigma_{use}^2 - \sigma_{uee}^2)$ , then  $\mu_k + \frac{1}{2} \sigma_{vse}^2 = \frac{\gamma_j}{\beta_j} + \frac{1}{2} (\sigma_{vse}^2 - \sigma_{use}^2 + \sigma_{uee}^2)$ .

We get the lower bound for mean under-reporting if  $\sigma_{vse}^2 = 0$ ; the upper bound of mean under-reporting is obtained when  $\sigma_{use}^2$  is minimized. Having the income variance decomposition and assuming that the error terms  $u$  and  $v$  are not correlated we can determine the bounds for the level of income under-reporting. They are as follows:

$$(\ln k)_{\text{low}} = \frac{\gamma_j}{\beta_j} - \frac{1}{2} (\sigma_{\zeta_{se}}^2 - \sigma_{\zeta_{ee}}^2) \text{ and } (\ln k)_{\text{up}} = \frac{\gamma_j}{\beta_j} + \frac{1}{2} (\sigma_{\zeta_{se}}^2 - \sigma_{\zeta_{ee}}^2).$$

With the help of this decomposition the interval for the level of income under-reporting can be obtained, since we have estimated the model of expenditures and computed the income variations.

To perform the system of equations estimation it is needed to separate the consumption between durable and nondurable goods. The cost functions of consumption are defined:  $C(p; U) = F[c(p; U); d(r; U); U]$ , where  $c(\cdot)$ ,  $d(\cdot)$  denote the sub-cost functions. Using this function one can depict household expenditure on the  $i$ th good:

$$p_i q_i = p_i \frac{\partial F[\cdot]}{\partial c(\cdot)} \frac{\partial c(\cdot)}{\partial p_i} = p_i y \frac{\partial c(\cdot)}{c(\cdot) \partial p_i}.$$

The budget share of good  $i$  in the expenditures on nondurable goods,  $w_i = \frac{\partial \ln c(\cdot)}{\partial \ln p_i}$ .

The authors impose the unit cost of nondurables  $C(\cdot)$ . It has a quadratic logarithmic form:

$$\ln c(p; U) = a(p) + b(p) \frac{U}{1 - g(p)U}.$$

Using this function by differentiating one can obtain Hicksian shares of the demand ( $w_i$ ) denoted above. This can be parameterized with the following function

$$w_i = a_i(p) + b(p) \frac{U}{1 - g(p)U} + \lambda_i(p) \left[ \frac{U}{1 - g(p)U} \right]^2.$$

Given the cost function we can see that the expression  $\frac{U}{1 - g(p)U}$  is directly linked to the level of income of a household through the indirect utility function, so authors make linear decomposition of this expression as a function of  $\ln Y_h^*$  where  $Y_h^*$  is the true household income. As the result Hicksian demand shares can be written as:  $w_{ih} = a_i + \beta_i (\ln Y_h^*) + \lambda_i (\ln Y_h^*)^2$  where the parameters are a function of Hicksian demand shares parameters, the price index is dropped, since the prices are fixed at the level  $p_i = 1, r_i = 1$  for all  $i$ . The label  $h$  denotes a household, the label  $i$  denotes a good in the consumer bundle. This is the theoretical foundation for the Engel curve, since the consumption share depends on income as a quadratic function. The quadratic form of the Engel curve means that since the income of a household increases it tends to spend more expenditure share on luxury goods and less expenditure share on necessities.

The advantage of this model is that there is no need to arbitrary impose the type of a household. This can be seen through the following income decomposition:  $Y_{kh}^* = \theta_k Y_{kh}$ , where  $\theta_k$  -  $k$ th component of income.

The share of income earned by a household from a certain income source is computed, this is denoted by  $y_{kh} = \frac{Y_{kh}}{Y_h}$ . The total real income of a household equals  $Y_h^* = \sum_k \theta_k Y_{kh}$ . Dividing both part of the equation by the observable income:  $\frac{Y_h^*}{Y_h} = \frac{\sum_k \theta_k Y_{kh}}{Y_h} = \sum_k \theta_k y_{kh} \Rightarrow Y_h^* = Y_h \sum_k \theta_k y_{kh}$ . This helps to rewrite the Engel curve equation:

$$w_{ih} = \alpha_i + \beta_i \left[ \ln Y_h + \ln \left( \sum_k \theta_k Y_{kh} \right) \right] + \lambda_i \left[ \ln Y_h + \ln \left( \sum_k \theta_k Y_{kh} \right) \right]^2.$$

(1) The model is more complex than one considered by Pissarides and Weber (1989). But it can be linked to the model of Weber if one can impose a certain threshold in the share of household's self-employment income:

$$Y_{kh} \text{ is replaced by } \begin{cases} D_{kh} = 1 & \text{if } y_{kh} > \tilde{y}_{kh} \\ D_{kh} = 0 & \text{otherwise} \end{cases}, \lambda_i \text{ equals zero for all } i.$$

The equation for the commodity  $i$  according to Pissarides and Weber looks like:

$$w_{ih} = \alpha_i + \beta_i \ln Y + \sum_k \beta_i \ln \theta_k D_{kh}$$

This equation is estimated for the category of goods for which the income is reported the most correctly. This category is food expenditures.

Actually in the article the following model is estimated, almost similar to Pissarides and Weber:

$$\ln Y_{Fh} = \alpha_F + \delta_F S_h + \beta_F \ln Y_h + \lambda_F (\ln Y_h)^2 + \sum_j \alpha_{Fj} Z_{jh} + u_{Fh}.$$

The problem is that the expenditure on the left-hand side is actually varies along with not only the total income, but with the type of the household. Thus reflecting the preference heterogeneity:

$$\frac{\Delta w_{ih} / \Delta D_{kh}}{\Delta w_{jh} / \Delta D_{kh}} = \frac{\Delta w_{ih} / \Delta \ln Y_h}{\Delta w_{jh} / \Delta \ln Y_h}$$

The single equation approach does not distinguish these effects, that is why it is limited. The system approach can cope with this difficulty and is performed by the authors.

Estimating the size of informal economy in Russian Federation we are not computing the value of household production since we do not know the amount of time used on a certain activity. But there is the information on the household production of the agricultural goods. The money income from the goods sold is actually included in the household total income. The information of goods consumed is added to the goods consumption data to form full expenditures.

On the first stage it is possible to estimate a single food expenditure equation and to see whether an income decomposition and the black economy coefficient can be obtained. To say in advance, there are some problems concerning both the income decomposition and the estimation of this coefficient. Therefore, the more complicated econometrician estimation methods should be used.

### 3. RLMS database description

Russian Longitudinal Monitoring Survey is the only annual nongovernment monitoring of social-economic characteristics of the population of Russian Federation and the health conditions (NRU-HSE, 2012).

The key peculiarity of RLMS is the wide-spread base of socioeconomic variables. The variables include the income and expenditure structure, material welfare, investment, occupation, migration, health conditions, structure of food consumption, education. The cross-sectional data is combined with the panel.

Our attention is devoted to the study of households of cross-sectional analyses. The analyses is made on representative sample for the household in the 21st wave. The 21st wave of RLMS household budget survey has been conducted from October 2012 to December 2012.

**Table 1. Individual characteristics descriptive statistics (%)**

Variable	% total	% non_se	% 0–20_se	% 20–100_se
D_couple	52.80	54.09	52.54	48.28
D_white_collar	45.80	49.79	42.12	38.17
D_white_collar_male	33.32	35.52	32.00	27.53
D_educ_sec_special	41.76	42.49	42.32	37.74
D_educ_higher	36.69	39.05	33.80	33.55
D_children	89.87	90.05	90.70	87.42
D_life_satisfaction	15.51	15.29	15.15	17.10
D_wish_other_job	31.19	29.71	29.94	39.57
D_have_other_job	4.24	4.60	3.54	4.30
D_living_improvement	11.05	11.63	9.09	12.90
D_vacation_possible	20.57	22.68	17.72	18.28
D_child_study_pay	19.27	20.53	16.74	19.68

The following individual variables are used: family status, occupational status, level of education, level of life satisfaction, presence of some stomach diseases, desire to find another job, presence of children in the household, number of children in the household, ability of a household to improve the living conditions, ability of a household to have a vacation with all members of the family, ability of a household to pay for a child study in the University, presence of some other job at one member of the household.

The comparison of individual characteristics of the households with no income from self-employment and with high level of self-employment income can be seen in the Table 1.

Considering the family status self-employed are usually single than married. They are more involved in the non-qualified labor, both for the head and other members. They have lower level of education: both higher education (33.55 vs. 39.05%) and secondary special education (37.74 vs. 42.49%). They report a bit more higher level of life satisfaction (17.10 vs. 15.29%). The greater part of them does wish other job (39.57 vs. 29.71%). But they do not in general indicate that they have other job (maybe they conceal their shadow part of income, or simply do not have other job for which to be paid). The characteristics concerning the households are as follows: place of living, living conditions, some variables concerning land use, educational questions.

There can be made no conclusion that self-employed have a greater amount of durable goods and therefore that their wealth is higher. Maybe that is due to the fact that the ownership of these goods are not reported in the interview.

Despite this, some conclusions can be made. The self-employed tend to be more settle in the village (33.2 vs. 24.8% and 21.2% non\_se), and also have more share of people which sell their products grown (5.7 vs. 2.2% and 1.1% non\_se). Less part of self-employed have phone (49.2 vs. 63.4% and 59.3% non\_se) Therefore, maybe some part of their income may be the agricultural income from the selling of products. Also as we have mentioned in the literature review the agricultural goods consumption shall be included is household foods product consumption to provide unbiased value of food consumption. This is the point to correct our model.

#### 4. Income decomposition

An important part of our study is to see how income is decomposed between the different sources among the households. First of all, the total income of the household consists of three parts: (1) wage income free of taxes; (2) the part such as other income usually fixed income; (3) income from self-employment. The last one is actually computed as a difference between total income of a

household, the wage income and the income from all other sources. So the formula by which the income is built is looking as follows:

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

The other income of a household includes: pension; scholarship; unemployment benefit; income from equity sold; income from rent of equity; capital income in the form of the interest; capital income in the form of the dividend; insurance premium; aliments; money from the debt reimbursement; subsidies from the apartment payment. Actually the inclusion in this form the income from equity sold does not prove to be reasonable, because it is not a permanent source of income and therefore may be misleading.

Speaking about the income from capital (to the capital income we must denote the income from rent of equity, capital income in the form of interest and capital income in the form of dividend) we notice that there is very low rate of response (only 150 people have indicated that they have got income in some of the forms). This can be explain twofold: (1) most of respondents conceal the real amount of capital income because of the fact that they do not want to give any information about the total amount of capital owned; (2) the capital forms of income are not so popular in our country with comparison to the households in the western countries. We must say that other income computed using the information on capital income is under-reported. Therefore, the self-employment income as it is defined must include some forms of capital income.

We exclude income from equity sold from the items including in “other income”. The wage income, other income and total income provide the basis for the computing of self-employment income. The description of household income decomposition can be seen in Table 2.

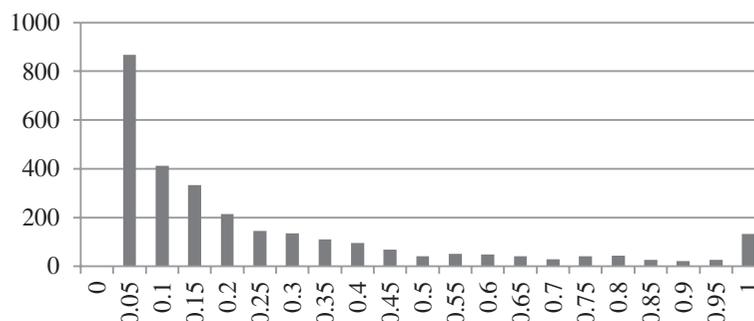
According to the descriptive statistics table, the mean value of wage income is the greatest among all the mean values of various income sources: households who get wage report that they earn on the average 33,700 rubles, whereas the mean for other income accounts for 13,478 rubles, self employment income is on average 9,405 rubles, and one must take into account the fact that this mean is only above those, who happen to have positive amount of income (near 45% of the households). So on average, the income from self-employment is much less than the wage income.

Coming to the variations of different components of household income we must notice that the assumptions proposed by Weber are fulfilled: self-employment part of income is much more *relative* variant. What does it mean? Although the variance of wage income is more than the variance of income from self-employment, the relative variance, or the variance related to the mean, is much greater for those who are self-employed (2.33 vs. 0.89). On the other hand, the relative variance of other income, which is defined above and consists of the sum of pensions, scholarships and so on, is the least (0.71).

**Table 2. Descriptive statistics of income decomposition**

Stats	Wage income	Total income	Other income	Self- employment income
N (>0, <0)	4,412	6,359	4,475	2,880
“zero” value	2,105	158	2,042	3,637
Mean	33,700	37,126	13478	9,405
SD	29,838	33,929	9,532	21,905
SD/mean	0.89	0.91	0.71	2.33
Min	990	130	85	0
Max	420,000	425,600	175,000	320,300

**Figure 1. The density of self-employment share of income distribution (self-employment income > 0).**



For those households who have self-employment income one can perform a histogram of the income share distribution (condition if the share > 0)—Figure 1.

We can see that the distribution of the share of the self-employment income in the total income of household is skewed towards 1. It means that even if a household has some self-employment income, the majority of the households is still having it as a minor source of income.

Now we must determine the criteria, by which the household is self-employed. Following Webber, the criteria is such that the share of income from self-employment is greater than some threshold value. In our model this value is defined with the help of histogram of income share distribution. Let us say that this value is 0.2 (in the paper authors take the value of 0.25).

Here there are the descriptive statistics of income decomposition for households with self-employment income (Table 3). Households with high self-employment income have mean higher total income and slightly less mean other income. We have seen earlier that self-employment income has higher coefficient of relative volatility than wage income. Surprisingly this is not true for our types of households with positive income from self-employment. Wage income tends to have the same volatility as income from self-employment, but even higher relative volatility for those who are self-employed. But total income for self-employed part of sample is anyway more volatile (in absolute and relative terms). To sum up, the hypothesis of income variance holds but only for the total income variance (except for the variance in wage income).

Concluding with the descriptive statistics analyses we must say that the income of Russian households does not necessarily come with the British households income behavior. Especially it can be seen that wages may be as much volatile as self-employment income. This fact says that maybe some new estimation methods should be implemented to estimate the size of black economy. The

**Table 3. Income decomposition descriptive statistics for the households with positive income from self-employment, grouped by the share of self-employment income in total income**

Variable	Obs	Mean	SD	Min	Max
<i>D_self_empl = 0</i>					
Total_Inc.	1,947	34,532.18	27,969.09	1,000	276,500
Wage_Inc.	1,947	20,335.45	26,283.81	0	243,000
Other_Inc.	1,947	10,769.81	10,052.66	0	99,000
Self_employment_Inc.	1,947	3,426.922	8,190.329	0	192,000
<i>D_self_empl = 1</i>					
Total_Inc.	932	44,003.35	50,014.41	1,500	425,300
Wage_Inc.	932	14,866.87	24,496.69	0	220,000
Other_Inc.	932	7,232.321	9,178.869	0	120,000
Self_employment_Inc.	932	21,904.16	3,3352.1	600	320,300

sample has been divided into three groups: non self-employed, people with low share of self employment income and people with high share of self-employment income (or self-employed).

### 5. Expenditure decomposition

As far as we know from the theory, the household expenditure can be decomposed into the expenditure on durable and nondurable goods. The items we are interested in are food expenditures, clothes expenditures, service expenditures (all of them form the expenditures on nondurable goods). The durable goods expenditures are also needed and include expenditures on household appliances (see Table 4).

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 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 is not significantly different). The question for the research is whether the differences in expenditures are attributable to stated income or is there some portion unaccounted for which 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. The one equation model estimation

At the first stage of econometric estimation we shall estimate the single equation of expenditures on a selected expenditure category (food), as far as the expenditures on food are reported the most correctly. The model is estimated in logarithms; the estimated specification is linear according to classical theoretical model. This model allows to apply the OLS estimation. The functional form of food equation is as follows:

**Table 4. Descriptive statistics for household expenses, including the consumption of household-produced goods**

	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	9,518	9,462	9,143	9,502
(*) Consumption of home-produced goods	458	445	520	1,403
Expenditures for meals at home, normalized to 30 days, adjusted for (*)	9,976	9,907	9,663	10,905
Expenditures for eating out, normalized to 30 days	1,550	1,718	1,283	1,402
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	2,436	2,486	2,182	2,770
Expenditures for services over the past 30 days	7,248	7,499	6,641	7,529
Expenditures for health care over the past 30 days	2,138	2,127	2,109	2,247
Other expenditures over the past 30 days	2,671	2,765	2,324	3,024
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	6,930	6,866	5,925	9,285
Total expenditures	32,942	33,368	30,127	37,162
Total income	36,082	35,020	34,514	43,500

$$\ln C_{Fh} = \alpha_F + \sum_j \alpha_{Fj} Z_{jh} + \gamma_F D_{SE_h} + \beta_F \ln Y_h + u_{Fh}$$

As the result of food equation estimation the variable corresponding to the type of a household proved to be not significant at 10% level. The possible explanation for this is that the differences in product consumption are fully attributed to the differences in the household characteristics, the impact of income is weak as far as the income elasticity is very low—the household tends to increase consumption up to a certain degree as far as the income increases.

This means that it is impossible to estimate the size of informal economy on the base of food equation. However, it is possible to implement the single equation approach with different categories of household expenditures. The idea of the choice of the expenditure category is the same: the expenditures should be reported correctly, and be the function of total income of a household, but not of the structure of income (thus we can estimate the single equation correctly). That is why, the service expenditures cannot be used, which include transport and are related to the workplace; on the other hand the clothes expenditures meet both restrictions, and that is why they can indicate the level of household welfare.

For the basis sample to estimate clothes expenditure model the sample including all the households is used.

The general model (model of all households with positive expenditures on clothes—full model) includes all households from 21st wave of RLMS. The purpose of the research is to define how the estimate of shadow economy share varies dependent on the socioeconomic characteristics of an individual (the occupation of household members, family status and the classification of a household both on occupation and family status). According to this approach the following subsamples are obtained:

On the base of occupation status: (2) households, some member of which is qualified; (3) households none of the members is a qualified worker, but has a job, therefore is a nonqualified worker;

On the base of family status: (4) households, living in a couple; (5) lonely people;

On the base of both qualification and family status: (6) lonely qualified; (7) lonely and nonqualified; (8) couple and some of the couple members is a qualified worker; (9) couple and both of the couple members are nonqualified; (10) couple and the head of the household is qualified; (11) couple, but only woman is qualified; (12) couple and the head of household is nonqualified;

The question is what sample should be used to compute the size of informal economy. The first basic sample represents all the households, while the second one is the sample of households with positive income from self-employment. The second sample is important since it focuses on households with certain part of self-employment income and therefore for whom the status “self-employed” can be given. For each of the two basic samples the size of shadow economy shall be computed (which is according to the hypothesis of Pissarides & Weber, 1989).

In Table 5 you can see the estimates of the size of informal economy for both samples.

The mean estimates of concealed income share grow when the size of informal economy is computed for the “short” sample of households with only positive income from self-employment. The amount of shadow economy estimated grows for the model as far as the share of self-employed people is greater. The mean share of income concealed depending on the base sample varies from 2.9 to 6.4%. For the lonely nonqualified the estimated value is close to 20%, which is more realistic estimate of the size of shadow economy.

**Table 5. The informal economy size estimation for two samples—first (full) and second (people with positive income from self-employment)**

N	Model	Difference in clothes expenditures for self-employed (%)	Income under-report coefficient bounds	Size of informal economy estimate (%)	
				Sample 1	Sample 2
(1)	General	8.39**	1.09–1.26	1.46–4.43	3.19–9.67
(2)	White_collar	11.49**	1.15–1.39	2.21–5.77	5.21–13.62
(5)	Non_couple	15.37**	1.21–1.43	3.99–8.13	8.60–17.50
(7)	Non_couple_blue_collar	16.40**	1.32–1.49	8.25–12.73	13.32–20.57
(8)	Couple_white_collar	11.36*	1.14–1.38	2.15–5.69	4.92–13.03
(11)	Couple_white_collar_female	33.73**	1.36–1.82	5.69–12.79	13.12–29.51

\*Significant difference at 10% level.

\*\*Significant difference at 5% level.

### 7. The system of equations estimation

The obtained results of the estimation of single equation have four main drawbacks: (1) do not account for the nonlinearity; (2) do not account for the preference heterogeneity; (3) do not account for the wage under-report of a household; (4) impose the fixed type of a household—either “employee” or “self-employed”. The system of equations of consumer budget shares approach allows to estimate the more complicated model, which allows to take into account the effects of the substitution of goods in dependence on the household income and type of employment. However, in the paper the simplified mode of Engel budget shares has been estimated, with no preference heterogeneity term and linear specification. This helped to obtain efficient estimates of the coefficients.

The equation of the expenditures on the category “i” is written in the following form:

$$w_{ih} = \alpha_i + \sum_j \alpha_{ij} z_{hj} + \beta_i [\ln Y_h + (\theta_0 y_h^w + \theta_1 y_h^s + y_h^o)] + v_{ih}$$

The system is estimated on the basis of current consumption, as well as it is reported more correctly then the consumption of durable goods. The following expenditure categories are used: food, clothes, service, healthcare and other expenditures. The system is estimated with the help of

**Table 6. The estimation of the informal economy size on the base of the system of equations**

	Sample 1	Sample 2
k wage_Inc.	1,224	1,224
k self_empl_Inc.	1,412	1,412
Sample size	6,508	2,877
Mean wage income	22,720	18,511
Mean self employment income	4,105	9,286
Mean other income	9,258	9,627
Mean total income	36,082	37,423
Mean wage income (*k wage)	27,813	22,662
Mean income of self employed (*k se)	5,795	13,108
Mean other income	9,258	9,627
Mean total income new	42,866	45,397
Informal economy estimate	18.80%	21.31%

nonlinear least squares though the system of seemingly unrelated equations. The model is estimated in the full sample without the division on the types of the households due to social characteristics.

In the Table 6 you can see the results of the system estimation.

According to the results of model estimation, both wage and the income from self-employment are concealed. The amount of estimated wage income concealed is 22%, while the amount of income from self-employment concealed is 41%. Total amount of shadow economy is 18.8%. Therefore, the estimation is made totally on the income decomposition, in contradiction to the single equation approach of Pissarides and Webber, as the result we have obtained the higher estimates of the size of shadow economy. The estimates are considerably higher than those obtained by Lyssiotou, Pashardes and Stengos for UK (2004) – 10,6%; by Fortin, Lacroix, Pinard (2009) for Canada – from 4,6 to 5,7%.

## 8. Conclusion

To conclude, we must say that methods implemented by Pissarides to the British economy and the ones implemented by Gardes to the economy of Turkey have a limited application to the households of Russian Federation. Although the consumer budget share estimation seems to give the results closer to the reality due to the fact that both wage income and self-employment income tend to be under reported. The results of this system estimation are very similar to the estimates of informal economy in Turkey. But while estimating the size of informal economy it is necessary to mention the specific feature of Russian households statistics because of which these estimation methods have limited implementation to the database. A few of households in the sample tend to have positive income from self-employment. Thus, the methods using the share of self-employment income are linked to this narrow part of database. The methods of Webber treat black economy only through the self-employment income, while the estimation of the system of equations helps to obtain separate estimates of wage income and self-employment income underreports.

### Funding

The authors received no direct funding for this research.

### Author details

Yaroslav Murashov<sup>1</sup>

E-mail: [yaroslav.48532@rambler.ru](mailto:yaroslav.48532@rambler.ru)

ORCID ID: <http://orcid.org/0000-0002-6782-9492>

Tatiana Ratnikova<sup>1</sup>

E-mail: [taratnikova@yandex.ru](mailto:taratnikova@yandex.ru)

<sup>1</sup> Higher School of Economics, National Research University, 26, Shabolovka Street, Moscow 119049, Russian Federation.

### Citation information

Cite this article as: Estimation of the Russian informal economy size on the household budget survey data, Yaroslav Murashov & Tatiana Ratnikova, *Cogent Economics & Finance* (2017), 5: 1307642.

### References

- Aktuna, G. A., Starzec, C., & Gardes, F. (2011). *A new estimation of the size of informal economy using monetary and full expenditures in a complete demand system*. Paris: Paris School of Economics University, Paris I Panthéon-Sorbonne.
- Fortin, B., Lacroix, G., & Pinard, D. (2009). Évaluation de l'économie souterraine au Québec Une approche micro-économétrique. *Revue économique*, 60, 1257–1274. <http://dx.doi.org/10.3917/reco.605.1257>
- Lyssiotou, P., Pashardes, P., & Stengos, T. (2004). Estimates of the black economy based on consumer demand approaches. *The Economic Journal*, 114, 622–640. <http://dx.doi.org/10.1111/j.1468-0297.2004.00234.x>
- NRU-HSE. (2012). *Russia longitudinal monitoring survey*. Moscow: RLMS-HSE.
- Pissarides, C., & Weber, G. (1989). An expenditure-based estimate of Britain's black economy. *Journal of Public Economics*, 39, 17–32. [http://dx.doi.org/10.1016/0047-2727\(89\)90052-2](http://dx.doi.org/10.1016/0047-2727(89)90052-2)



© 2017 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



***Cogent Economics & Finance* (ISSN: 2332-2039) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

