ALL-RUSSIAN SURVEY OF CONSUMER FINANCE (WAVE 5)

Methodology

2023

This methodology has been prepared by LLC Demoskop.

Cover photo: Shutterstock/FOTODOM

Address: 12 Neglinnaya Street, Moscow, 107016 12 Neglinnaya Street

Tel.: +7 495 771 91 00, +7 495 621 64 65 (fax)

Bank of Russia website: www.cbr.ru

Methodology for the All-Russian Survey of Consumer Finance (Wave 5)

1. Study design

This is a longitudinal study of household financial behaviour. The terms refers to large-scale, long-term surveys that assume that the bulk of the sample in each wave of the survey is a panel, i.e., that the units of observation in each part of the sample (households in this case) remain the same as those selected for the first, base wave of the survey.

The longitudinal nature of the survey brings it undeniable advantages over periodic surveys spaced at intervals in which each wave is conducted on a new independent representative sample. This is enabled by the ability to investigate time changes occurring at specific levels (household or individual), and it thereby significantly expands the range of issues that can be addressed. For example, in investigating the poverty problem, data obtained from regular surveys enable the measurement of time-dependent increases or decreases in the number of households with average per capita incomes below the subsistence level or of changes in the makeup of households in this group (aggregate data). However, they fail to help answer the question of how long an individual household has been in the poor group and measure the average time households remain in the group below the level of poverty. This is very important from the point of view of government programmes to support the economically disadvantaged. This is one of multiple challenges of this type.

The initial aim of this survey was to provide single-point estimates for each wave, in addition to the longitudinal estimates. This problem was solved by means of a special design for longitudinal surveys known as *split panel.* The split panel, a combination of single and panel samples in each survey wave, was proposed by Leslie Kish in 1987. The design is a series of single-point surveys in which the maximum possible share of units of the initial sample is retained in subsequent waves. Designs of this type are usually described as overlapping surveys and can be considered a version of a split panel. In this case, the study aims to obtain a sequence of single-point estimates while maintaining the possibility of longitudinal estimates for most of the original sample.

2. General overview of original sample of Russian households

To study household financial behaviour in our survey, we use **the design (model) of a stratified, multi-stage, probabilistic, and territorially targeted sample**.

The survey is based on a sample of households. A *household* is defined as people who live together at the same address and share income and expenses. Temporary residents (e.g. guests) who permanently reside elsewhere are not household members. Survey respondents were all members of a household aged 18 or older who stayed at the place of residence at the time of the survey.

Although the study involves the creation of a preliminary sample of households, the standard global practice in this case is the creation of a sample of dwellings (addresses) in which households live. This is explained by the fact that location-linked statistical information is available only about dwellings (more precisely, addresses), which enables a census and survey of households. Before the sample of households is created, numbered lists of dwellings are composed so that there is only one household per dwelling in almost all cases. Under this condition, the resulting sample of dwellings is essentially tantamount to a sample of households.

Since the All-Russian sample is grounded in the territorial principle, the first stage involves selecting the primary territorial sampling unit or primary territorial unit (PSU). As a basis for the PSU, we select the administrative-territorial units lying at the core of the administrative-territorial division of the Russian constituent entities. The administrative-territorial units (ATUs) are grouped into 2,029 converted administrative-territorial units based on territorial attributes, which are the primary sampling units. The PSUs are then grouped into 38 strata, mainly based on geographical factors and the level of urbanisation. When necessary, the ethnic component is used as the stratum-forming factor.

Similar to most nationwide sample surveys involving face-to-face interviews at respondents' places of residence, a number of remote and underpopulated areas of the Russian Federation are excluded from the sample for financial reasons.

Of the remaining areas, which account for 95.7% of the Russian population, mandatory conglomerates are selected. These are the three most heavily populated areas: Moscow, the Moscow Region, and Saint Petersburg. Because of their size, they each constitute a separate 'self-representing' (SR) stratum. The remainder – the converted administrative-territorial units – are grouped into 35 non-self-representing strata with approximately equal populations. This leads to a total of 38 strata. Then, by the *probability-proportional-to-size* (PPS) method, one area is selected from each non-self-representing stratum. This means that the probability that a certain area in a given stratum is chosen is directly proportional to the share of the area's population in the total population of the whole stratum.

Of the total target (planned) volume of the sample, 17.6% (of the total Russian population) are broken down into three self-representing strata. In accordance with the principles of PPS, the remaining households are evenly distributed among the converted administrative-territorial units, that is, the primary sampling units (PSUs), one in each of the 35 non-self-representing strata of approximately the same size. The selection of converted administrative-territorial units from the strata is governed by the *probability-proportional-to-size* principle.

Consistent with established practice, the absence of a consolidated list of households/dwellings for any of the 38 PSUs necessitates the introduction of an intermediate stage of selection. The population of each PSU is stratified into urban and rural substrata, and the volume of the target sample in a PSU is divided in proportion to the share of the population in each of these substrata. For example, if 40% of the population of a PSU live in rural areas, 40 out of 100 addresses (dwellings) are sampled from rural areas.

For both urban and rural households, the secondary sampling units (SSUs) are settlements. Urban settlements are stratified by population size, level of industrialisation, and remoteness from district centres. The volume of the sample is allocated in proportion to the population in each of these strata for urban settlements. Specific urban settlements are selected based on the *probability-proportional-to-size* (PPS) principle. Rural settlements are stratified by only two parameters: population and remoteness from district centres. Several administrative-territorial regions are additionally stratified by ethnic composition. The selection of rural settlements is governed by a procedure similar to that for urban settlements. The next sampling unit for rural settlements is the dwelling (address).

For cities and small towns, there is a third sampling level. The tertiary sampling units (TSUs) in urban settlements are constituencies. Constituencies with widely varying residents are sampled in proportion to population size. In the sampled constituencies, a list of dwellings is made through field surveys of their territories.

Household registers serve as lists of dwellings in rural settlements.

This is followed by the systematic selection of the required number of dwellings starting from a randomly selected address in the list. If the surveyor determines that more than one household resides at an address, the list will include as many dwellings as there are households living at the address.

In both urban and rural settlements, the surveyors are required to visit each dwelling at least three times to establish contact with the household and its individual respondents. No replacement of any type is allowed.

Throughout the waves, the survey covers the households of the original sample dwellings, regardless of whether these households have been polled in previous waves. If – in any wave – the household living at an address refuses to participate, the surveyor is obliged to make repeat attempts at contact in subsequent waves up to the point of definitive refusal. If a household has moved, the new household residing at its address is polled at the time of the survey. If that household is also unavailable or refuses to be interviewed, the original dwelling is replaced with another, selected by the same design, within the survey area or settlement that has recorded the loss. This approach is called *repeated dwelling sampling* and helps represent the general sample in every wave of the survey.

The gradual dropout of units in the original sample in longitudinal studies is a natural process known as attrition. As the volume of the original sample gradually declines, the households which exit need to be replaced to maintain the size of the target sample. Compared to the design of a fixed-panel longitudinal survey, this *split-panel* design enhances the longitudinal analysis by including households with shorter participation periods.

Since the attrition and aging of the original sample are not random, a shift is observed over time in the *age*, and the sample needs to be 'repaired' after a certain number of waves, that is, the researchers must replace part of the sample with new households which include household members of under-represented age groups in order to compensate for the effects of attrition and natural aging in the original sample. If this replacement does not take place, the share of pension-age respondents will tend to prevail.

The sample must be generated in a way that eliminates the need for reweighting.

3. Principles of selecting and forming primary sampling units (PSUs)

The majority of territorial sample models are grounded in their multi-step nature, since the object of a territorial sample is the population (or part of it) residing in the surveyed territory. In Russia, there are no lists of people, households, or dwellings based on administrative and territorial divisions. Moreover, the use of such lists to generate a sample for a large territorial entity such as the Russian Federation as a whole or one of its constituent entities would make no practical sense. Respondents selected from this list would be scattered throughout the surveyed territory and the costs of a survey of such a sample would be prohibitive, with little theoretical rationale. In such cases, **multi-stage sampling** is used**.** In this, the overall sample is broken down in a natural way into separate subsamples, or *clusters*, which serve as sampling units in the first step (stage) of sampling (primary sampling unit – PSU), with subsequent sampling of observation units taking place only in clusters that have been selected in the first step. Unlike the above-mentioned strata, the size of each cluster is relatively small, but the clusters themselves are numerous. The primary sampling units in a multi-stage probabilistic sample determine the first level of clustering of the observation units in the general sample. In sampling theory, the main requirement for such intermediate sampling units (clusters) is that they be as heterogeneous as possible in terms of the properties under study.

In practice, this underlying, theoretically substantiated, requirement for the PSUs is complemented with several other requirements related to the particularities of conducting mass sociological surveys. In the selection of the territorial sample, we are guided by the following requirements for selecting the PSUs:

1) The PSUs should have clearly defined geographical (territorial) boundaries. There must exist statistical materials appropriate for the creation of the sample.

2) There should be enough PSUs that the sampling error in the first stage is not too large.

3) The population of the PSUs should be large enough to enable a multi-year study based on the PSUs sampled. Sampling and, especially, the creation of an interview network are very costly if this condition is ignored in a study targeting the population of a sufficiently large territorial entity.

4) The distances in a PSU area should allow the interviewers to travel directly to the survey points.

What exactly does the second PSU requirement mean?

Let us consider the standard error in multi-stage sampling. Suppose we have a K-step sample. The population under study consists of N1 units of the first stage, each of which contains N2 units of the second stage, etc. Suppose also n1,n2,...nk were sampled, respectively, in each stage of sample generation. Then, if simple random sampling was used at each stage, the population mean is an unbiased estimate of the average value for the general population with variance:

$$V(y)=(1-f_1)*S_1**2/n_1 + (1-f_2)*S_2**2/(n_1*n_2) +..+ f_1*f_2*f_3*..((1-f_k)*S_k**2/(n_1*n_2*..n_k) \quad (3.1)$$

where S_i is the mean variance in the sampling unit of the i-th stage,

$f_i = n_i/N_i$ is the sample frequency at the i-th stage, and

$1-f_i$ - is the correction for the finiteness of the population at the i-th stage.

The unbiased estimate V(y) for the sample is:

$$v(y)=(1-f_1)*s_1**2/n_1+f_1*(1-f_2)*s_2**2/(n_1*n_2)+\ldots+f_1*f_2*f_3*..(1-fk)*s_k**2/(n_1*n_2*...n_k) \quad (3.2)$$

where s_i are sampling equivalents of S_i.

Formula (3.1) shows that if the sample size is fixed, each stage adds its share to the variance; that is, the fewer stages there are, the smaller the standard error. This in turn means that a two-or three-stage sample is the best from a theoretical point of view (a one-stage sample is impossible in the absence of a basis – a structured list of dwellings in Russia). The generally accepted value of the standard error is 10% of the mean-square deviation. Formula (3.2) shows that this condition is met when n1 > 100.

The Russian Federation is divided into 89 constituent entities. The constituent entities in turn are composed of 2,775 basic administrative-territorial units (including 1,868 districts [‘rayons’], 579 cities of republican, regional, or district subordination, and 328 intracity areas and urban districts).

There are too few constituent entities to meet item 2 of the PSU requirements, and they are very large in territory and fail to meet item 4 of the PSU requirements to serve as the PSUs. In contrast, the **administrative-territorial units are almost ideal primary sampling units for the creation of a representative sample of households for Russia as a whole.**

The definition of an administrative-territorial unit is marked by two aspects that necessitate the merger of a number of units before the sample is generated. First, there are cities of federal, republican, or regional subordination within the boundaries of certain districts. State statistics treat such cities as independent administrative-territorial units. Since there are many such independent cities in Russia, they are included in the districts where they are geographically located. This ensures greater heterogeneity of the PSUs and thus improves the quality of the sample.

Further, large Russian cities are divided into districts. In accordance with standard sampling principles, such cities are treated as separate units in the sample. Therefore, as a result of the internal redistribution of the initial areas, the final list of primary sampling units consists of 2,029 modified administrative-territorial units (ATUs).

Intentionally excluded territories. A significant fraction of Russia's territory is remote areas with very low population densities. For example, the Evenk Autonomous District's population density is a mere one person per 30 sq km, while the Kamchatka Region's is one person per 1 sq km. A portion of such territories are pre-emptively excluded from the sample. Consequently, territories amounting to about 4.3% of the Russian population are stripped out due to their low population densities, poor transport connections, and inappropriate surveying conditions.

Self-representing territories. Three constituent entities – Moscow, the Moscow Region, and Saint Petersburg – are included in the sample automatically. These highly populated territorial entities are 'self-representing' strata in the stratification stage.

Stratification. The accuracy of the estimates is improved by means of the stratification of administrative-territorial units (PSUs) that are not excluded from the sample and are not self-representing territories.

First, 10 modernised economic regions are generated (see Table 1). The regions are generated in such a way that each contains the whole number of strata of a given population. In this, changes to the boundaries of existing economic regions are kept to a minimum. The regions are then divided into strata according to population size in each modernised region to make sure that the sizes of the strata are approximately equal. For example, the Ural Region is divided into 6 strata, and the Volga-Vyatka Region into 3 strata. Table 1 shows the ten regions and the number of strata in each.

Table 1. Ten modernised economic regions of Russian Federation (net of 3 self-representing and excluded territories)

No.	Region	Number of strata
1	Northern Region and Kostroma Region	2
2	North-Western Region	1
3	Central Region excl. Kostroma Region	4
4	Volga-Vyatka Region	3
5	Central Black Earth Region excl. southern Voronezh Region	2
6	Volga Region excl. Astrakhan and Penza Regions and Kalmykia	4
7	North Caucasus Region, Astrakhan Region, southern Voronezh Region, Kalmykia	5
8	Ural Region	6
9	West Siberian Region	4
10	East Siberian and Far Eastern Regions	4
Total		**35**

The full description of all strata is presented in Tables 2 and 3. The first three strata are the self-representing territorial entities. Strata 4–38 are non-self-representing entities. Importantly, although strata 4–38 have approximately the same populations (in accordance with the sample design), the number of PSUs in the strata varies significantly (see the right column).

The corresponding number of strata is formed in each of the ten regions on the basis of the level of urbanisation. Geographical properties and the ethnic structure are also considered where they matter.

Table 2. Stratification of territory of Russian Federation: self-representing strata

No.	Self-representing strata
1	Saint Petersburg
2	Moscow
3	Moscow Region

Table 3. Stratification of territory of Russian Federation: non-self-representing strata

No.	**Non-self-representing strata**
	NORTHERN REGION AND KOSTROMA REGION
4	Urban population 87%
5	Urban population below 87%
	NORTH-WESTERN REGION
6	All districts of region
	CENTRAL REGION (excl. Kostroma Region)
	North: Vladimir, Ivanovo, Tver, Smolensk, Yaroslavl Regions
7	Urban population over 82%
8	Urban population below 82%
	South Bryansk, Kaluga, Oryol, Ryazan, Tula Regions
9	Urban population over 79%
10	Urban population below 79%
	VOLGA-VYATKA REGION and Penza Region
11	Regional centres and capitals of autonomies with populations over 300,000
12	Urban population over 55%
13	Urban population below 55%
	CENTRAL BLACK-EARTH REGION (excl. southern Voronezh Region)
14	Urban population over 75%
15	Urban population below 75%
	VOLGA REGION (excl. Astrakhan, Penza Regions, and Kalmykia)
16	Kazan, Tatarstan
17	Regional centres with more than 900,000 residents (Volgograd, Samara, and Saratov)
18	Urban population over 70%
19	Urban population below 70%
	NORTH CAUCASUS REGION,
20	Astrakhan Region, southern Voronezh Region, Kalmykia, North Caucasus autonomous republics excl. Adygea
21	Urban population over 95%
22	Urban population 58–95%
23	Urban population 36.5–58%
24	Urban population below 36.5%

No.	Non-self-representing strata
	URAL REGION
25	Regional centres and capitals of autonomies
26	(dual stratum)
27	Share of Russians below 45%
28	Urban population over 93%; Russians over 45%
29	Urban population 67.5–93%; Russians over 45%
30	Urban population below 67.5%; Russians over 45%
	WEST SIBERIAN REGION
31	Novosibirsk, Omsk, Tomsk
32	Urban population over 90%
33	Urban population 57.5%–90%
34	Urban population under 57.5%
	EAST SIBERIAN AND FAR EASTERN REGIONS
35	Eastern Siberia: Urban population over 89%
36	Far East: urban population over 84%
37	Urban population 64.7–89% (E.S.); 64.7–84% (F.E.)
38	Urban population below 64.7%

4. Principles for primary sampling units (PSUs)

Probability sampling assumes that at least one PSU is selected from each stratum, which is why one PSU is selected in each non-self-representing stratum by the *probability-proportional-to-size* method. This means that the greater the share of the population of the PSU is in the total population of a given stratum, the stronger the chance that the PSU will be selected.

5. Principles for selection and formation of secondary (SSUs) and tertiary sampling units (TSUs).

The sample may skip the third stage depending on the type of PSU.

A PSU consists of one city.	**The secondary sampling units (SSUs) are constituencies**. Specific constituencies are selected by simple mechanical sampling, with a fixed step, from the corresponding list of constituencies of the city. Participating households are identified in the sampled constituencies. To this end, the next step is to use the lists of dwellings available for the constituencies, which are checked by means of a field survey. Dwellings are sampled from the verified list (via simple mechanical sampling with a fixed step). If the field survey finds that that there is more than one household in a dwelling, the interviewer randomly selects one of them to be included in the sample of households.
PSUs are cities, towns, and urban-type and rural-type settlements.	**With this PSU structure, all three types of settlement are secondary sampling units (SSUs).** Initially, the population is divided by size into urban and rural strata. The sample of households is distributed in proportion to their shares. **Specific cities and urban-type settlements (SSUs) are selected from the list of cities and urban-type settlements** through mechanical proportional-to-size sampling. For each urban settlement selected, a third stage of sampling is run to select a **constituency, which becomes the tertiary sampling unit.** Specific constituencies are selected by simple mechanical sampling, with a fixed step, from the corresponding list of constituencies of the city. Participating households are identified in the sampled constituencies. To this end, the next step is to use the lists of dwellings available for the constituencies, which are checked by means of a field survey. Dwellings are sampled from

	the corrected list through a simple mechanical sample with fixed step. If the field survey finds that that there is more than one household in a dwelling, the interviewer randomly selects one of them to be included in the sample of households. **Specific rural-type settlements (SSUs) are selected from the list of rural-type settlements** through mechanical proportional-to-size sampling. **Rural settlements do not have TSUs since households are sampled according to the rural household register covering the entire rural-type settlement.** The list in the rural household register is checked by means of a field survey. Dwellings are sampled from the verified list (via simple mechanical sampling with a fixed step).

6. Methods of estimating survey parameters

Any indicators characterising the general population are called ‘parameters’ – as opposed to the indicators calculated for the sample population. First of all, researchers seek to calculate such indicators for the general population, or target groups of this population, as average aggregate amounts or as shares, their equivalents for nominal data. These are additive indicators. The procedure for estimating additive parameters from the values of the indicators in the sample population is summing the values of the variables included in these indicators for the sample population, multiplied by the selective weights of the observation units (which make up the whole sample population or the particular groups under scrutiny). There are two types of criteria for classifying observation units: quantitative criteria (such as a respondent’s level of income or per capita household income) and qualitative criteria (such as a respondent’s being a certain type, for example, the dichotomy of respondents with or without banks deposits).

The following formula (6.1) is used to estimate the totals for the populations under study based on the sample data:

$$Y_d = \sum_{k \in d} w_k y_k \tag{6.1}$$

Formula (6.1) sums the observation units in the target group (d) of the general population (this group may be households with more than two children);

Y_d is the estimated total indicator (for example, the total household income of the group under study);

w_k is the sample weight or distribution factor calculated based on the working design of the sample;

y_k is the value of the attribute of the k-th sampling unit.

Formula (6.2) estimates the average indicators of the study based on the sample data:

$$\overline{Y}_d = \frac{\sum_{k \in d} w_k y_k}{\sum_{k \in d} w_k} \qquad (6.2)$$

Formula (6.2) sums the observation units in the sample of the target group (d) of the general population (this group may be households with more than two children);

$\overline{Y}_d$ is the estimated average indicator (for example, the average household income in the target group under study);

w_k is the sample weight or distribution factor calculated based on the working design of the sample;

y_k is the value of the attribute of the k-th sampling unit.

The following formula estimates the share of the sample in the general population:

$$p_d = \frac{\sum_{k \in d} w_k I_k}{\sum_{k \in d} w_k} \qquad (6.3)$$

The summation in formula (6.2), similar to (6.1), is that of the sampling units belonging to the target group (d) of the general population (for example, pensioners);

pd is the estimated share of, for example, working pensioners;

I_k is an indicator that takes a value of 1 if the *k-th* sampling has the attribute of interest and 0 otherwise. This attribute may be, for example, whether the respondent is a working pensioner;

w_k is the sample weight or distribution factor calculated based on the working sample design;

The formulas discussed in the following section are used in the absence of dedicated software for calculating common target indicators and the relevant sampling.

7. Calculation of sampling errors adjusted for design effect

The level of representativeness of the sample is assessed by the quality of the input information, determined by how much the sample data can be extended to the general population and, to a certain extent, to the populations under scrutiny. The reliability criterion for the extended indicators is usually the variance coefficient for the estimates of a given indicator (the ratio of the standard error to the estimate itself).

Given that the sample under discussion is meant to be nation-wide, the reliability calculations should be conducted at the federal level. The estimates of the indicators calculated for the whole general population or for the target groups under the classification are then considered reliable if the variance coefficient is less than or equal to 2–3%. This means that the margin of the sampling error does not exceed 5% of the resulting estimate of the indicator under study.

To obtain the confidence interval of the estimated total, mean, or share, the standard error must be adjusted by the multiplier (quantile) corresponding to the selected level of reliability. Usually, 95% or 90% confidence probabilities are chosen in social and economic research. This means that the value of an indicator for the general population with a 95% or 90% confidence probability is within the interval of plus or minus 5% of the sample statistic. The corresponding adjustment ratio in this case is 1.96 or 1.64.

The formula to calculate the standard errors of the results obtained from a stratified, multi-stage, probabilistic, territorial, and targeted sample produced proportionally to the population size takes into account the design (plan) effect of this complex sampling plan.

The meaning of the effect design is that it shows how much the data obtained with the operating model, providing that the sample volume is the same, may differ from the data that could have been obtained from simple random sampling. For the standard mean error:

$$SE(\overline{y_d}) = \sqrt{\frac{\sigma_{y_0}^2}{n_d}\left(1 + iic\left(\frac{n_d}{a_d^*}\right)\right)} \qquad (7.1)$$

$SE(y_d)$ is the standard mean-square error Yd (or a share of p_d);

$\sigma_{y_0}^2$ is the variance of an attribute (y);

n_d is the sample volume in subgroup (d) of the general population;

icc is the assumed intra-class correlation, which cannot be calculated accurately based on the sample data.

Intra-class correlation measures the degree of difference between clusters (PSUs). If its value is close to zero, the clusters are approximately the same, which is a prerequisite for obtaining accurate estimates of the indicators;

a_d^* is the effective number of PSUs belonging to the target group (d) of the study.

The value on the right (7.1) in the parenthesis below the square root $\left(1 + iic\left(\frac{n_d}{a_d^*}\right)\right)$ is exactly the sample design effect (design effect). It is the ratio of the variance of the estimate of the indicator, calculated consistently with the sample plan, to the variance of the estimate of the same indicator calculated under the assumption of simple random sampling. Therefore, the less the effect design differs from 1, the more accurate the extended results of the study are.

It is quite difficult to calculate the design effect using formula (7.1). This is why it has not been a standard for conducting mass surveys in Russia, although multi-stage samples for mass surveys are in wide use. However, the absence of data on the size of the design effect significantly reduces confidence in the data provided by researchers on the assessment of the reliability of published indicators.

The design effect calculations of the RLMS survey, which is based on a similar sampling design, show that the design effects for almost all indicators used in this study are close to one (or insignificantly exceed 1). For this reason, simplified formulas for simple random sampling can be used to calculate the standard estimate error when this sample design is used.

To calculate the standard mean-square error ($\overline{Y_d}$):

$$SE(\overline{Y_d}) = \sqrt{\frac{s_y^2}{n_d}\left(1 + \frac{n_d}{N_d}\right)} \qquad (7.2)$$

s_y^2 is the corrected mean-square deviation of attribute (y) in the sample

To calculate the standard error of a proportion (p_d)

$$SE(p_d) = \sqrt{\frac{p_d * (1 - p_d)}{n_d - 1}\left(1 + \frac{n_d}{N_d}\right)} \qquad (7.3)$$

The design of the original sample ensures the representativeness of the general population of households and of respondents aged 18+ years for individual Russian population groups (if their number is sufficient in relation to the sample size) which are usually of interest to researchers, such as those living in a city, living in a village, or living in cities of different population sizes, and similar.

The sample is not representative:

– in relation to Russian regions since the administrative-territorial units are sampled to represent a stratum rather than the individual regions this stratum includes, wherein specific (ATUs) are selected;

– in relation to individuals and households in the upper income quintile, which is essentially unrepresented in the sample population of mass surveys. A special survey should be conducted to capture it and include it in the sample.