

MICRO-ENTITIES AND SMALL ENTERPRISES SURVEYS IN BELARUS

NATALIA BOKUN

Belarus State Economic University

Minsk, BELARUS

e-mail: nataliabokun@rambler.ru

Abstract

The paper briefly describes the sampling methodology of micro-entities and small enterprises, problems of introduction of the micro-entities sample survey in practice of Belarusian official statistics. The sampling frame, sampling design and precision estimation are considered.

Keywords: micro-entities, sample survey, sampling frame, weighting, small enterprises

1 Introduction

In recent years, the growing number of small enterprises has motivated the development of specialized methodology and software for micro-entities and small enterprises sample surveys.

Since 2005 and until 2008 sample surveys of small enterprises spent quarterly. Survey objects were artificial persons of small business, i.e. SE. According to the legislation this was the organization with number of employees 100 persons and less. Sample frame was the file of SE. The territorial one-stage stratified sample was used. But in 2008 quarter survey was cancelled. Due to this reason only annual continuous small enterprises survey is conducted.

Nowadays, the National Statistical Committee of the Republic of Belarus together with Department of Statistics (BSEU) makes the preparatory work on implementation of the micro-entities and small enterprises sample surveys. In November 2014 a test sample survey was conducted; since 2015 Micro-entities Sample Survey (MS) is provided on a regular basis. The first results of Micro-entities Sample Survey indicated the appearance of significant organizational and methodological problems: non-responses, the need for localization of the sample, the presence of atypical units, using a combination of statistical weighting methods, samples in small domains.

This paper on small business sampling has the next parts:

- history of development of branch sample surveys;
- small enterprises sample survey;
- micro-entities sampling frames that incorporate two files of economic units: micro-entities and private farms;
- micro-entities sample design; territorial stratified univariate and multivariate (multidimensional) samples are used. The algorithm to receive optimal sample size for i -th kind of activity and j -th region is presented;

- statistical weighting that includes three methods: traditional Horvitz-Thomson estimator and calibration (GREG- and SYN-estimators).

2 History of development of branch sample surveys

In the conditions of command economy in the national statistics of Belarus, as well as in other countries of FSU Region, a priority it was given to methods of continuous survey with the exception of 3.5 thousand family budgets survey of workers, employees and collective farmers. Then the two-level stratified sample was used: at the first step the enterprises was selected within branches, than hired workers was selected. Such principle of selection ensured wages data representativeness.

In consequence of disintegration of the USSR and occurrence of market relations the economic situation has changed. Notably restrictions on individual labor activity have been removed, the structure of sources of revenue has changed, the number of small state and private enterprises has sharply increased in all economic branches. So, the total number of the small enterprises (SE) in republic has come to 28310 in 2000, 33094 in 2005, 111792 in 2014. From each of them was inexpedient to demand of statistic registration. Full coverage of population has become economically unjustified and almost unrealizable. As a result process stage-by-stage introduction of enterprises sampling in the practical statistics has begun:

1. 1997–2005 Theoretical workings out and pilot sample surveys (retail trade, services, small business);
2. Since 2006 until now. Theoretical workings out and selection of the enterprises on a regular basis (retail trade, small business, labor statistics).

At the first stage of introduction sampling in statistical practice (1997–2005) Statistics research institute provided with methodology and software of branch survey of the enterprises, based on using of group of methods of univariate selection: systematic sampling, random selection without allocation, simple random sample, stratified sample with proportional and optimal allocation. Pilot surveys of SE in retail trade were carried out in 1998–1999, survey of enterprises in services — in 2002, survey of small enterprises in economic branches — in 2003. In 2005–2006 problems of building of multivariate sample are investigated, the first version of the program is developed, trial multivariate samples of SE are spent.

At the second stage (since the end of 2006) researches of multivariate sampling and improvement data extrapolation are hold on. State statistics began to carry out quarter samples of SE in area of labor statistics on a regular basis. Since 2008 has added sample surveys in retail trade and in catering. Special quarter sample surveys of SE concerning employment and unemployment, and also personal subsidiary plots is predicted. Since 2015 micro-entities sample survey spent early.

Despite such advantages of sample, as enough low expenses, efficiency of and high reliability, statisticians was confronted with a number of problems:

Non-responses. The population of micro- and small enterprises is extremely dynamical: the creation of new entities, liquidations, changes in kinds of activity and size of the enterprises are taken place constantly. Sampling frame is based on the data of the previous year of complete survey, and not responded enterprises can be included into the sample (liquidated, changed the kind of activity or not presented the questionnaire).

Atypical units (outliers), i.e. presence in the frame of atypical units, inclusion (or not inclusion) of which in a sample strongly influences the estimates of parameters. Atypical units are the units, which have extreme values of variables, large sample weights, complex structure.

Samples in small domains. Construction samples of small enterprises by economic activity and regions, in some cases is connected with partition of survey population into the small groups and sample fractions become unacceptably high (50–60%). As a consequence, possibilities of control an admissible sampling error are problematic.

Problems of compromise between the accuracy requirement for various groups caused by stratification and restrictions on sample size.

Estimation. The problem of estimation still persists when the univariate stratified sample with admissible standard error and sampling fraction is built. Weights, raising factors allow to estimate precisely enough values of the parameter which was used for sample selection, but other estimates which number can reach 10–30 are of a low quality. In the case of multivariate sample, the error for some group of indicators will be in admissible limits (to 10%), but will be considerably above comparing with the case of univariate sample.

The problems of the software are caused by the complexity of mathematical apparatus of sample survey and the necessity of integration of the sample survey programs in the general system of collection and processing of statistical data.

Specific problems are met designing the *multivariate sample* (stratified by several variables): complexity of a choice of an optimal way of multivariate selection, complexity of a choice of a leading indicator (variable), technical difficulties of construction of multivariate general population (over 500 units), absence of the standard estimation methods.

The problems of non-responses and atypical units may be solved within traditional univariate sample; the solution is connected with the change of general population structure, allocation in separate files of atypical enterprises, use of weighting or replacement procedures. Multivariate sampling and different weighting schemes are used to handle remaining problems, it allows to receive the samples of small size, which are representative for many different parameters.

It is offered by the author to apply a combination of univariate and multivariate sampling methods in order to receive representative small business samples [1–3].

3 Small enterprises sample surveys

Survey objects are artificial persons of small business, i.e. small enterprises. According to the legislation this was the organization with number of the working from 100

persons in industry and transport branches till 25–30 persons in services, nowadays — 100 persons and less. Sampling carried out at each regions and Minsk by branches. The used sample design provides possibility of a choice to use a sampling method depending on population, number and character of survey variables (the program “Multivariate sampling”). It should be done several steps for searching optimal sample size for i -th branch and j -th region: allocation of observed variables, applying multivariate or univariate sample, selection is carried out by the cluster analysis.

Extrapolation of total value of variables on all population is carried out by traditional group raising factors (ratio of number of units in i -cluster of total population and corresponding cluster of sample) and simple errors.

Sampling frame is 20-30% from all number of small enterprises. As to branch sampling fraction is depending on number of SE and the degree of accuracy on a leading variable: a relative sampling error on regions less than 2%, on branches less than 5%, and on small branches less than 8–9%.

4 Micro-entities sampling frames and sample design

Sampling Frames are two files of economic units: 1) micro-entities, represented the state statistical reports on the financial results for basic years (report 1-MP (micro)); 2) set of the private farms. The first file is high — 80 thousands units, sample fraction depends on a character of the initial information, namely: the size of total population, kind of economic activity, region. The second array includes more than 2 thousands farms; it is observed completely (sample fraction is 100%). Predicted non-responses rate for republic is 12–13%, for regions — 6–12%, for Minsk is higher (18–20%). The combination of univariate and multivariate (multidimensional) sample is used.

To receive optimal sample size for i -th kind of activity and j -th region the author together with the colleagues-statisticians have developed the next algorithm:

1. The set of observed variables is allocated (for example, the wages fund, average number of employees, volume of production, revenues, profitability). Average, total values, variability of indicators are calculated.
2. Statistician chooses sampling method: univariate or multivariate. Univariate stratified samples with simple, proportional and optimal allocation are most often used.
3. It should be executed one of three conditions for applying multidimensional sampling: variation coefficient is more than 100%; survey objects are non-uniform on many variables; the small size of total population (top limit — 30–40 units). Otherwise univariate sampling should be used: random selection without allocation, simple random sample, proportional and optimal allocation.
4. It is expediently to use univariate stratified sample, total population is divided by rather homogenous groups. Then different variants of the sample size are ex-

ecuted (minimal is $0.05N$, maximal is $0.08N$). Predicted sample size is allocated by received groups. The choice of an optimal sample size and optimal kind of univariate stratified sample depends on a standard error. So, it (the minimal error) is a main criteria of the determination of sample size.

5. It is expediently to use multidimensional sample, selection is carried out by cluster analysis: total population is partitioned using cluster analysis (agglomerative hierarchical, iterative method of k -means) on homogenous groups to k -variables, i.e. clustering; in each received group the leading (basic) variable is determined and subsequent random selection of units is performed.

Optimal sample population is chosen for each cluster, where standard errors of k -variables are criteria of productivity. If the error exceeds admissible bounds, three methods of its reduction may be applied: a) increasing sample population in cluster; b) additional stratification of the enterprises in cluster to a leading variable; c) repetition of clustering, but with larger number of steps, or using an iterative method with the preliminary number of clusters $r > 1$.

Sample population is formed once in three-four years, i.e. fixed sample (yearly) is used.

5 Statistical weighting

To extrapolate sample data on the total population traditional group raising factors (weights) and standard errors have been used [2,3].

The methodology of weighting for univariate stratified sample is based on the assignment for each enterprise corresponding statistical weight (k_{ijl}):

$$k_{ijl} = N_{ijl}/n_{ijl}, \quad (1)$$

where k_{ijl} is individual weight for each enterprise of l -th group of i -th kind of activity (3 digit for NACE) in j -th region; N_{ijl} is the size of l -th group of i -th kind of activity in j -th region in total population; n_{ijl} is the size of sample group; l is the number of groups by observed variable value ($l = 1, \dots, m$).

Individual weights are equal within each group of micro-entities, calculated by region, kind of activity, observed indicator (output, employees or others). Individual weights, determined for multidimensional sample, are:

$$k_{ijr} = N_{ijr}/n_{ijr}, \quad k_{ijrh} = N_{ijrh}/n_{ijrh}, \quad (2)$$

where k_{ijr} is the weight of r -th cluster of enterprises; k_{ijrh} is the weight for h -th group of r -th cluster; r is the number of clusters in i -th branch of j -th region ($r = 1, \dots, \alpha$); h is the number of groups in r -th cluster ($h = 1, \dots, \gamma$).

To improve the representativeness by region weighting procedure can be complicated. It is possible to use GREG-estimators and calibration [2–4]. The results of trial calculations testing the first version of methodological and software sampling have shown that the main difficulties are associated with the use of different weighting

schemes, necessary estimation of the whole variables, splitting of the same population on the smaller groups, little subsamples. Sampling fraction — 10–15%. As to branch sampling fraction is depending on the number of enterprises and the degree of accuracy on a leading variable: a relative sampling error on regions less than 2–4%, on branches (kinds of activity) less than 5–6%, and on small branches less than 8–12%.

6 Concluding remarks

The use of combination of univariate and multidimensional samples, different weighting methods will provide very reliable information over larger number of variables: employment, wages fund, revenues and others. However, standard errors, calculated by separate indicators in the context of different kinds of activity at regional level are rather high. To improve the representativeness by region weighting procedure can be complicated by usage of auxiliary calibration estimators. Besides, it is important to take into account the necessity of annual sample updating. The creation of new entities, liquidations, changes in kinds of activity and size of enterprises are taken place constantly.

References

- [1] Bokun N., Chernyshova T. (1997). *Methods of sample surveys*. Minsk (in Russian).
- [2] Bokun N. (2010). Problems of multidimensional samples in retail trade. *Questions of statistics*. Vol. **3**, pp. 52–60 (in Russian).
- [3] Bokun N. (2014). Micro-entities sample survey: problems of design, formation and usage. *Workshop of Baltic-Nordic-Ukrainian network on Survey Statistics, Tallinn, Estonia, August 25-28, 2014*. pp. 25–30.
- [4] Sarndal C. E., Swensson B., Wretman J. (2003). *Model assisted survey sampling*. Springer-Verlag.