

## Methodological Aspects of Current Statistical Survey of Population Life Quality within the Unstable Regional Development Conditions

**Oksana A. Khokhlova**

*Doctor of Economics, Associate Professor, Plekhanov Russian University of Economics, Moscow, Russian Federation*

**Aleksandr V. Bezrukov**

*Ph.D. in Economics, Plekhanov Russian University of Economics, Moscow, Russian Federation*

**Elena V. Sibirskaia**

*Doctor of Economics, Professor, Plekhanov Russian University of Economics, Moscow, Russian Federation  
Email: e-sibirskaia@rambler.ru*

**Vladislav V. Borisov**

*Plekhanov Russian University of Economics, Moscow, Russian Federation*

Doi:10.5901/mjss.2015.v6n3s6p45

### Abstract

*The article covers the main approaches to current sample population survey on quality of life problematic in regional and municipal dissection. A block scheme is suggested for the system of indicators of current population survey on the aspects of: demographics and social sphere; environment and healthcare; labour resources; welfare and consumer prices; production and infrastructure. An algorithm of current population sample survey is developed in connection with operative monitoring of regional development core indicators. The statistical instrumentary of multivariate analysis of sample population survey results is suggested, with the listed parameter inputs for the methods applied. For the multivariate analysis the results of hierarchical cluster analysis are presented as to define the population classes within the sampling; at the next stage the k-means clusterization was applied to form the clusters, after which the most significant differentiating indicators are listed with the descriptive statistics for the clusters. There are then defined the opportunities of applying "Random forest" modeling methods for the purposes of classification of the population among the obtained clusters, and the test data were used to construct and probate the "random forest" model for further surveys. To reveal the main factors defining the life quality, a factor analysis was carried out using the maximum likelihood procedure.*

**Keywords:** *current population survey for life quality, multivariate analysis, regional analysis*

### 1. Introduction

Being one of the priority goals of the Government of Russia policies, the relevance of statistical study of population life quality is increasing within the conditions of unstable social and economic development of regions.

The development of operative regional monitoring in Russia, the focusing of attention to the stability and efficiency of separate regions of Russian Federation development, is defined by the necessity of all-side development and integrity of the country's economy within regional dissection. The territorial peculiarities of Russian Federation, evidently, define the need for concentrating on financial and economic state of each region, and its role in country's economy functioning in total. The effective planning of a region's activity thus requires a system of indicators for prognostication and early detection of macroeconomic indicators risks and the rise of a situation of financial stress within a Russian Federation territorial subject, with the account of regional peculiarities. A certain tendency for global economy decline in Russia, conditioned by the geopolitical situation, has revealed a set of problems in modeling the financial stability of the RF regions, reflecting their specifications.

Additionally, the tasks being presently solved by the Ministry of economic development concerning the details of macroeconomic indicators and their aggregation on municipal level, allow to give a more precise estimation of population

life quality in a region, reveal the more concrete reasons underlying the current situation.

Within these conditions the planning of economic activities in Russian Federation regions the need arises to forecast the possible risks and crisis situations and their development both by individual indicators and by overall characteristics. It is necessary to consider the analyzed indicators both in static and in dynamic view.

A specialized system of current population survey indicators aimed at problematic of life quality would allow to carry out the monitoring and necessary amendments of the unbalanced economic activities of the RF territorial subjects, to increase their stability and population well-being.

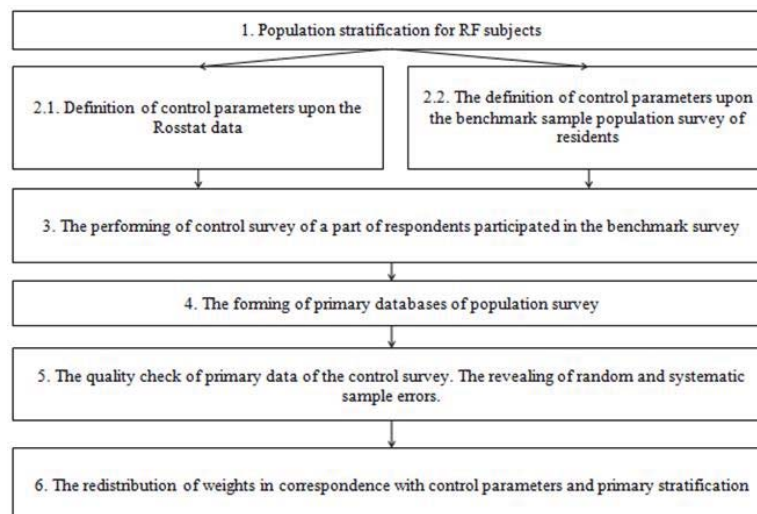
## 2. Methods and Materials

The methodological base of the research in theoretical and applied problems of current population survey is formed by the works of native and foreign authors in the area of current monitoring and sample survey (Zarova, 2013; Kish, 1965; Pearson, 1908; Kendall, 1978; Chetyrkin, 1977; Box, 1976).

The regulatory base for the research is the Constitution of Russian Federation, the legislative and regulating issues in Russian Federation, including: the Regulation of Russian Federation Government of Nov.27, 2010 №946 "On organizing in Russian Federation the system of federal statistical surveys on social and demographic problems and the monitoring of economical losses due to mortality, illnesses and invalidization of the population"; the List of indicators of real economy sector monitoring in financial, banking and social spheres of Russian Federation regions, Regulation of the RF Government №806-r, 2009; the Prognosis of social and economic development of Russian Federation for 2015 and the plan period of 2016-2017, Ministry of Economic development of RF, 2014.

The informational base of the research is composed of the results of sample survey of the population in Moscow, the Moscow district and the Ivanovo district, carried out within expeditions in May 2014 for the operative monitoring project of Plekhanov Russian University of Economics.

The population of the survey are the residents of Russian Federation above 16 years of age. The obtaining of operative data on the residents living standards and life quality assumes the organization and carrying out the sample survey of households and separate residents altogether. The form of sample survey is a stratified sampling with non-replicate random sample within the strata; the probability of a resident being sampled is thus proportional to the population structure. The method of survey is a questionnaire poll. The survey is quarterly periodical. Forming a panel sample is not stipulated. The algorithm of sample survey is presented on Figure 1.



**Figure 1.** The algorithm of current sample survey of regional residents

At the initial stage, the sample size is floating. The processing of preliminary survey of 500 of households in Moscow, the Moscow district and the Ivanovo district allowed to estimate a target sample size as not smaller than 5000 households in total for Russian Federation. However, a calibration of sample size is stipulated in proportion to costs of organizing and carrying out the survey in specific time and territory conditions. In addition, a new sample on the same population allows for using the optimal Neumann distribution principle (Neumann, 1951).

As the instrumentary of analysis there were used the methods of descriptive and multivariate analysis, the hierarchical and k-means cluster analysis, factor analysis, "random forest" modeling, groupings and classifications, analysis of variance (ANOVA).

### 3. Describing the Domain

The carrying out of the sample population survey for life quality problematic obtains a new most relevant meaning within the conditions of economic instability. The population is, undoubtedly, the main resource of any state, especially the labor force. The current monitoring of population well-being and administrative measures aimed at the increase of life quality will allow, in the opinion of authors, to significantly improve and stabilize the level of economic development of regions, to increase the quality of labor force, and, upon the current population monitoring, reveal the possible problems in economic development of a territory.

Thus, the development of a system of indicators and the recommendations on the use instruments of statistical analysis on sample survey data with close connection to the operative monitoring of regions is vital. Such surveying would allow to obtain the data on population life quality and broaden the informational base of regional monitoring, and, additionally, to obtain the data on municipal level, and adjust the data of official statistics.

Upon the study of economic literature, legal acts and methodological papers, among the most significant practices of population sample survey it is possible to point out the following, the recommendations and regulations of which were accounted in the research:

- Eurostat, 1998-2014. Labor Force Survey Regulation 0577;
- U.S. Bureau of Census for Bureau of Labor Statistics 2006-2013. Current Population Survey, technical recommendations 66;
- Rosstat, 2010. Population Survey on Problems of Employment, Statute N 452.

### 4. Results

The data on current sample regional population survey are simultaneous independent random variables, and, therefore, could be considered as multivariate statistical distribution, to which multivariate statistical analysis instruments and methods can be applied.

The following results have been obtained from the population survey of Moscow, the Moscow district and the Ivanovo district.

At the first stage, the cluster analysis methods were applied to classify the sample survey results and to obtain the initial assumption on the number of homogeneous classes of objects. The hierarchical analysis defined the supposed number of population clusters within the survey to be 4. For the amalgamation of objects the Ward's method was chosen, using the squared Euclidean distance metric. Let us remind that Ward's method allows for obtaining of homogeneous classes of objects with minimal within-group variation. The obtained icicle plot is displayed on Figure 2.

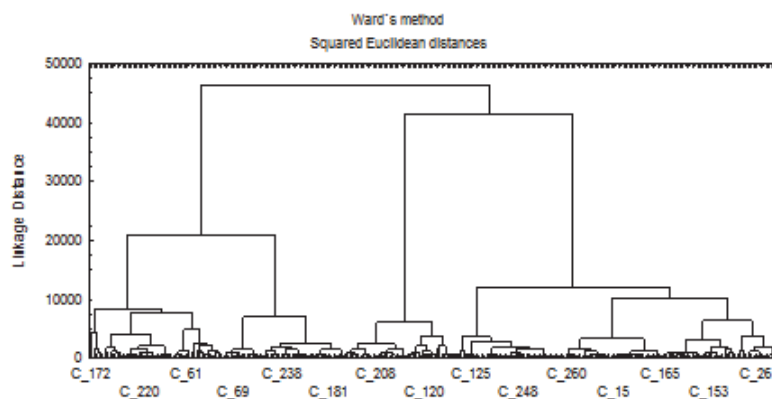


Figure 2. Clusterization of respondents in the sample survey.

At the next stage the clusterization of data using k-means method was performed, with the number of clusters k=4. The

analysis of variance of the results indicated that the obtained clusters are most significantly different by the parameters of: Actual income per household member, thousand rubles; Income level in monetary units per household member to provide for the necessary consumer goods and services, thousand rubles; Proportion of income used on housing bills payment, %; Proportion of income used to discharge current financial liabilities, %.

The average values and standard deviations for these indicators are presented in the table 1. The interval of mean plus-minus standard deviation represents the most typical (average) value of an indicator and the most common values of occurrences around it.

**Table 1.** Descriptive statistics of significant parameters within the clusters dissection of respondents: Mean±SD

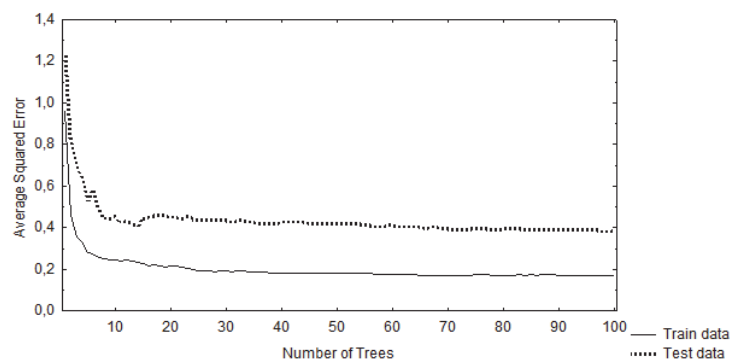
Indicator	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Actual income per household member, thousand rubles	28.0±8.4	32.6±4.3	15.7±6.0	12.7±6.4
Income level in monetary units per household member to provide for the necessary consumer goods and services, thousand rubles	30.7±6.4	32.9±4.6	23.4±8.3	22.0±8.5
Proportion of income used on housing bills payment, %	16.8±8.1	9.5±5.5	9.9±5.0	27.0±4.6
Proportion of income used to discharge current financial liabilities, %	26.8±7.4	2.9±5.5	3.2±5.5	7.8±10.1

SD = Standard (quadratic) deviation

Let us note that the obtained grouping enables to classify new observations into these clusters by using such methods as “Random Forest” or “Neural network modelling”. The use of “Random Forest” method is more preferable, for the reason of obtaining higher quality models compared to “Neural networks” (Hastie, T., Tibshirani R., Friedman J., 2009). Furthermore, the “Random Forest” models effectively process data with large number of variables, and equally efficiently interpret both discrete and continuous variables. “Random Forest” method also allows to construct models upon data with missing variables (which is a frequent occasion in sample population surveys as some inputs may be incorrect and thus considered missing and some questions may be left unanswered), thus avoiding the discard of these objects which contain useful information on other variables. Lastly, the “Random Forest” models allow estimation of the significance of every variable in the model.

To simulate the data with the purpose of classifying them into the initial clusters a “random forest” model was constructed. We can note that, upon the model, the main predictors were the indicators of material welfare – the income size, the current liabilities expenses and household billing and tariffs.

The graph of train and test data is presented on Figure 3.



**Figure 3.** The “Random Forest” model simulation: train and test data

To define the principal factors influencing the life quality of population, the factor analysis was carried out. Let us remind that the factor analysis is a collection of statistical methods aimed at revealing the common (latent) principal factors of large-scale events and processes. The main goal of factor analysis is to describe the many initial parameters of the studied object by a fewer number of primary substantive factors, transcendent to the initial variables, that cannot be directly measured but are revealed upon the study of initial variables correlations. The primary factors are an integral

characteristic of initial factors, that is, they are the variables of higher order that contain the information on the set of initial variables.

To reveal the principal factors the maximum likelihood method was used. This method analyzes the differences between sample correlation coefficients and hypothetical population correlation coefficients. While the maximum likelihood method is void of significant drawbacks, its advantages are the efficiency of factor solution, its robustness to variables scaling, and the ability to apply Chi-square significance test on the obtained model for a large set of data.

For the factor analysis, with the purpose of better interpretation of the principal factors and to estimate the principal factors for regional dissection, the factor analysis was performed on the data grouped by regions of Moscow, the Moscow district and the Ivanovo district. The minimal Eigenvalue was set at 2.000. The Varimax raw factor rotation method was then used for better interpretation of the solution. The results of principal factors influence on the population estimation of welfare are presented in the Table 3.

**Table 3.** Principal factors of life quality indication within regional dissection

Region	Number of principal factors	Interpretation of principal factors (in descending eigenvalue order)	Hierarchical analysis of secondary (oblique rotation) factors
Moscow	3	<ol style="list-style-type: none"> <li>1. Housing prices and tariffs growth (inverse);</li> <li>2. Consumer goods and services prices growth (inverse);</li> <li>3. Industrial production output growth</li> </ol>	Prices growth for separate housing services (water supply, heating)
Moscow district	4	<ol style="list-style-type: none"> <li>1. Housing prices and tariffs growth (inverse);</li> <li>2. Industrial production output growth;</li> <li>3. Residential market prices growth (inverse);</li> <li>4. Household size</li> </ol>	No secondary factors providing complimentary information to primary factors were revealed o
Ivanovo district	4	<ol style="list-style-type: none"> <li>1. Housing prices and tariffs growth (inverse);</li> <li>2. Commodity supply in retail trade (production goods and services market capacity);</li> <li>3. Residential market prices growth (inverse);</li> <li>4. Disposable population income</li> </ol>	<ol style="list-style-type: none"> <li>1. Primary residential market prices growth,</li> <li>2. Prices growth for separate housing services (water supply, heating)</li> <li>3. Foods market potential change;</li> <li>4. Nominal accrual wage change</li> </ol>
All regions	3	<ol style="list-style-type: none"> <li>1. Consumer goods and services prices growth (inverse);</li> <li>2. Disposable population income</li> <li>3. Housing prices and tariffs growth (inverse)</li> </ol>	No secondary factors providing complimentary information to primary factors were revealed o

It is evidently noticeable that the factor analysis results are similar for all regions, as household members in all regions point out the prices growth for consumer goods and services, disposable income change and housing prices growth as the defining factors. The secondary factors were pointed out to be the tariffs growth for separate housing services, water supply and heating in particular, and the change of nominal wage size.

## 5. Discussion

The conditions defining the population life quality may vary for some regions, sometimes significantly, and, as follows, the studied population may not be stated to be homogeneous. The reduction of shifts of the obtained sample statistics is enabled by stratification – the construction of qualitatively homogeneous non-intersecting subsets of the population. The base for the stratification is the administrative territorial subdivision of the Russian Federation.

On the other hand, the representativeness of the sample is provided by the process of population units sampling and is supported by the control of correspondence of the sample structure to the population structure using the so-called control parameters. Generally these control parameters are key social and demographic characteristics, such as gender, age, education etc. and the household parameters (the household size and composition). The informational base for forming these control parameters is suggested to be the data of Federal service of public statistics, and the data of specially organized benchmark survey.

The respondents in the benchmark survey may be offered a shorter version of the questionnaire, containing general data on the size and structure of household, its income and overall welfare, i.e. in this case the household is the population unit. It is suggested to survey only one member of the household. The tasks of the benchmark survey could then be formulated as such:

- the construction of a control parameters system, necessary to form a representative sampling;
- the gathering of overall information on living standards and life quality of a large proportion of households;
- the forming of contact data base to conduct a more detailed survey for separate residents.

Each resident passed the benchmark survey is offered to participate in the control survey. At this stage, the poll is offered to all household members older than 16 years of age, the questions may be answered individually or by one representative household member for all. As such, the unit of control survey is a single person, and the information collected may provide a complex evaluation of life quality on individual level.

At each stage of sample survey certain errors are possible, that may be of either random or systematic origin. Thus, a significant stage is the monitoring of the initial data quality and of the sample estimations of population parameters. The obtained estimations have to be, at the very least, consistent, that is, approach the probabilities of true population parameters. A more preferable property is the unbiasedness (an unbiased estimation is the one, the mathematical expectation of which is the same as the true population parameter). As this is the first time for the survey, the evaluation of characteristics efficiency may prove hard. Thus, the suggested efficiency indicators are the classic measures of dispersion: the standard (quadratic) deviation and the variation coefficient.

In addition, a more complex problem of the poll is the presence of questions for which a respondent has not provided an answer (the "unanswers"). As the result, the sampling loses in randomness and the obtained estimations will be biased. A detailed method is suggested by the authors of the research to resolve the "unanswers" problem:

1. If the proportion of unfilled polls is less than 0.1 it is possible to exclude these respondents from the survey;
2. if the proportion of unfilled polls is greater than 0.1 it is necessary to test a multivariate hypothesis upon the benchmark data survey:

$H_0$ : The value of a control parameter  $k_i$  among the respondents answered question  $i$  and not answered question  $i$  is statistically indifferent.

3. If the proportion of unfilled polls is greater than 0.3, it is necessary to test a multivariate hypothesis using the benchmark data:

$H_0$ : The value of all control parameters  $k_1, \dots, k_n$  among the answered question  $i$  and unanswered question  $i$  is statistically equal.

If  $H_0$  is accepted, excluding the respondents will not cause bias. If  $H_0$  cannot be accepted with given confidence probability, the missing data are filled with average values analogical to the distribution of control parameters.

After primary processing, the procedure of weights calibration (or secondary weighing) is applied to the data. As the result, the structure of sampling is brought in full correspondence to the structure of population – at the first stage in correspondence to control parameters, and at the next stage – in correspondence to the stratification. The procedure of secondary weighing enables to guarantee the representativeness of certain population categories that may have been represented in the sampling at a smaller or larger proportion than necessary.

## 6. Conclusion

Within the instability conditions of Russian Federation regional development, and with the consideration of economic sanctions leverage, the carrying of current population survey on the regional and municipal level is of the highest relevance as a means to obtain trustworthy indicators of population welfare in both social and economic aspects.

Thus, it is necessary to produce a more target-aimed system of indicators for sample population survey on problematic of life quality with the consideration of compatibility with the operative regional monitoring data.

In the opinion of authors, it is possible to allocate five categories of indicators, which could be considered and accounted for population survey. The statistical study of population living standards is thus based on the following system of indicators, forming an interrelated block scheme (Figure 4).

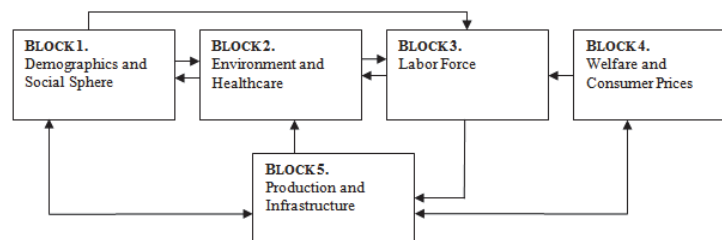


Figure 4. A conceptual block-scheme of system of indicators for sample population survey on life quality

Drawing certain conclusions, it can be stated that the modern changes in geopolitics and world economy have preconditioned to perform the current population monitoring based on sample survey for the indicators of life quality, with the purpose of complimenting the operative regional monitoring data both informatively and in detail related to living standards of the population.

The results of current population survey will allow to boost the quality of statistical information of the operative monitoring for social and economic development of the Russian Federation regions, and, with appropriate administrative measures, achieve the planned level of their sustainable development.

## References

- Berti, K., Salto, M., Lequien, M. (2012). An early-detection index of fiscal stress for EU countries. *European Economy, Economic Papers* 475, December, Brussels. pp: 25. DOI:10.2765/28160
- Bertrand Candelon, Elena-Ivona Dumitrescu, Christophe Hurlin. (2014). Currency crises early warning systems: why they should be Dynamic. *International Journal of Forecasting*, 30(4):1016-1029. DOI: 10.1016/j.ijforecast.2014.03.015
- Bragoli Daniela. (2010). Forecasting the Global Financial Crisis. Eurostat 6th Colloquium on Modern Tools for Business Cycle Analysis, Sep.26-29, Luxembourg, pp: 12. DOI: 10.2901/1977-3331.2011.009
- Buelens, C. (2012). Inflation forecasting and the crisis: assessing the impact on the performance of different forecasting models and methods. *European Economy. Economic Papers*. 451. March, Brussels. pp: 37. ISSN: 1725-3187
- Chetyrkin, E.M. (1977). Statisticheskie metody prognozirovaniya [Statistical methods of forecasting]. Statistika, Moscow. ISBN: n/a. pp: 199.
- Emilija Kairyte, William H. Meyers. (2010). Territorial Indicators For Rural Development: Targeting Lagging Areas In Lithuania. *Management theory and studies for rural business and infrastructure development* No. 20(1). ISSN 1822-6760.
- Eurostat. (2014). EU labour force survey – main features and legal basis. [http://epp.eurostat.ec.europa.eu/statistics\\_explained/index.php/EU\\_labour\\_force\\_survey\\_%E2%80%93\\_main\\_features\\_and\\_legal\\_basis](http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/EU_labour_force_survey_%E2%80%93_main_features_and_legal_basis) (Accessed on December 09, 2014)
- Frenkel', A.A. (1988). The application of regression analysis within the conditions of multicollinearity of economic indicators. A learner's guide. MESI, Moscow. ISBN: 978-5-282-02673-3, pp: 51.
- Gardo, S., Reiner, M. (2010). Economic and financial crisis on Central, Eastern and Southeastern Europe. A stock-taking exercise. *European Central Bank (ECB) Occasional paper series*, 114. ISSN 1725-6534
- Hastie, T., Tibshirani R., Friedman J. (2009). *Random Forests*. In: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed., Springer-Verlag, New York, NY, pp: 587-604. ISBN-13: 978-0387848570.
- Link, Michael W., Battaglia, Michael P., Frankel, Martin R., Osborn, Larry & Mokdad, Ali H. (2008). A Comparison of Address-Based Sampling (ABS) Versus Random-Digit Dialing (RDD) for General Population Surveys. *Public Opinion Q* 72, Feb.7, pp: 6-27. DOI: 10.1093/poq/nfn003
- Ministry of Economic development of Russian Federation. (2014). The prognosis of social and economic development of Russian Federation for 2015 and the plan period of 2016-2017 <http://economy.gov.ru/minec/activity/sections/macro/prognoz/201409261> (Accessed on December 14, 2014).
- Morrison, Donald F. (2004). *Multivariate Statistical Methods*. McGraw-Hill, New York. ISBN-13: 978-0534387785, pp.496.
- Richard Valliant, Alan H. Dorfman, and Richard M. Royall. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. Wiley, New York. ISBN: 978-0-471-29341-5, pp.536.
- Rosstat. (2010). Population Survey on Problems of Employment, Statute No. 452. [http://www.gks.ru/free\\_doc/new\\_site/population/trud/Untitled452-10.pdf](http://www.gks.ru/free_doc/new_site/population/trud/Untitled452-10.pdf) (Accessed on December 09, 2014).
- Rosstat. (2014). The methodological regulations on population survey for problems of employment. [http://www.gks.ru/bgd/free/b99\\_10/isswww.exe/stg/d030/i030110r.htm](http://www.gks.ru/bgd/free/b99_10/isswww.exe/stg/d030/i030110r.htm) (Accessed on December 09, 2014).
- Russian Federation Government. (2009). The list of indicators of real economy sector monitoring in financial, banking and social spheres. Regulation N 806-r <http://government.consultant.ru/page.aspx?1016390> (Accessed on December 09, 2014).
- Savin, I., Winker, P. (2012). Heuristic model selection for leading indicators in Russia and Germany. *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2(2012), pp. 67-89. DOI: 10.1787/19952899
- Sherbinin, A. de, Carr, D., Cassels, S. and Leuwen Jiang. (2009). Population and Environment. *Annual Review of Environment and Resources*, 32:345-373. DOI: 10.1146/annurev.energy.32.041306.100243
- Sibirskaya, E. V., Stroeva, O. A., Gubareva, L. I., & Mikheykina, L. A. (2014). The monitoring of the subject and object of the economic activity population in the innovative sector. *Life Science J.*, 11(8s), 292-296. DOI: 10.7537/j
- Stroeva, O. A., Sibirskaya, E. V., Khokhlova, O. A., & Oveshnikova, L. V. (2014). Regionalization of the innovation management process. *Life Science J.*, 11(8s), 297-301 DOI: 10.7537/j
- U.S. Bureau of Census for Bureau of Labor Statistics. (2006). Current Population Survey, technical recommendations 66. <http://www.census.gov/prod/2006pubs/tp-66.pdf> (Accessed on December 09, 2014).
- U.S. Bureau of Census. (2014). Office of management and budget. Standards and guidelines for statistical surveys. [http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards\\_stat\\_surveys.pdf](http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf) (Accessed on December 09, 2014)
- Von Neumann, J. (1951). Various techniques used in connection with random digits. *Monte Carlo methods*, Nat. Bureau Standards, 12, pp. 36–38. ISSN: 1049-4685

- Weisberg, Herbert F. (2005). *The Total Survey Error Approach*. University of Chicago Press, Chicago. ISBN: 9780226891279, pp.231.
- Zarova E. V., Prozhivin R.A. (2008). A balanced system of indicators of the region development: justification and economic modeling. *Questions of Statistics*, 8(2008), 59-66.