THE USE OF CLUSTER ANALYSIS TO ASSESS THE DEMOGRAPHIC POTENTIAL OF RUSSIAN REGIONS

Oksana Shubat
Anna Bagirova
Irina Shmarova
Ural Federal University
620002, Ekaterinburg, Russia
Email: o.m.shubat@urfu.ru
Email: a.p.bagirova@urfu.ru
E-mail: i.v.shmarova@urfu.ru

KEYWORDS
Cluster analysis, Demographic potential, Demographic policy, Russian regions.

ABSTRACT

In recent years, Russia has been grappling with a serious economic crisis. The slowing pace of economic development is accompanied by adverse demographic trends. The purpose of our study is to assess the demographic potential of Russian regions and identifying groups that require the implementation of specific measures aimed at its development. We used hierarchical cluster analysis to model Russia's demographic space and segment regions with comparable problems related to forming demographic potential. Clustering was based on the indicators describing the demographic potential at the macro-level (regional) and meso-level (family level). The analysis identified the groups of regions that have the best and the worst conditions for the development of demographic potential. We proposed a set of measures that would be most relevant to the needs of specific groups of regions and could directly drive the development and actualisation of demographic potential. The analysis showed the need to use multi-factor classification in the demographics of countries that have a high level of regional differentiation. Modelling the demographic space on the basis of cluster analysis can be seen as an element of the system of supporting administrative decision-making and the development of effective demographic policy.

INTRODUCTION

In recent years, Russia has been grappling with a serious economic crisis. This manifests through a depreciation of the national currency, a drop in people's real incomes, increased unemployment and so on. The slowing growth and even decline in Gross Domestic Product also points to the adverse economic situation. Such negative dynamics are particularly stark in the context of global trends (table 1).

Table 1: Real GDP growth (annual percent change)
(Real GDP growth 2016)

	2012	2013	2014	2015	2016
Russia	3.5	1.3	0.7	-3.7	-0.8
World	3.5	3.3	3.4	3.2	3.1
	2017	2018	2019	2020	
	forecast				
Russia	1.1	1.2	1.5	1.5	
World	3.4	3.6	3.7	3.7	

The slowing pace of economic development in Russia is accompanied by adverse demographic trends. Thus, according to official forecasts, the number of people below working age will fall by 0.7 million people by 2030 (Federal State Statistics Service 2016). A drop in the Total Fertility Rate (hereinafter TFR) is also forecast.

It should be noted that there is a marked difference in the demographic situations across Russian regions. Thus, for example, in 2015 maximum TFR was 3.39 (in the Tyva Republic), while the minimum was 1.29 in Leningrad Region. The highest life expectancy was 80.05 years in Ingushetia, compared to 64.16 years in Chukotka (Single inter-departmental information and statistical system 2016). Such high regional differentiation indicates that while there should be an overall uniformity of approach, specific demographic policy measures need to be tailored to the needs of different regions.

Undoubtedly, it is not objectively possible to account for the full spectrum of diversity of regional situations in demographic policy. However, it is possible to cluster together regions with similar problems in this area. We believe that the ability to identify such types of regions is linked to the use of statistical cluster analysis, which will enable modeling Russia's demographic space.

Cluster analysis is often used for regional segmentation outside Russia. For example, Kronthaler identified groups of German regions based on their

Proceedings 31st European Conference on Modelling and Simulation ©ECMS Zita Zoltay Paprika, Péter Horák, Kata Váradi, Péter Tamás Zwierczyk, Ágnes Vidovics-Dancs, János Péter Rádics (Editors)
ISBN: 978-0-9932440-4-9/ ISBN: 978-0-9932440-5-6 (CD)

economic potential (Kronthaler 2005). Laboutkova, Bednarova and Valentova used cluster analysis to study relationship between regional decentralisation and economic imbalances in Europe (Laboutkova, Bednarova and Valentova 2016). Simpach segmented municipalities in a part of the Czech Republic on the basis of demographic development (Simpach 2013). Mertlova and Prokop used a set of macroeconomic indicators for regional clustering (Mertlova and Prokop 2015). Vahalik and Stanichkova drew out groups of countries with analogous competitiveness characteristics (Vahalik and Stanichkova 2016). Koisova and Haviernikova segmented regions of Slovakia using socio-economic indicators (Koisova and Haviernikova 2016). Zhang and Li used cluster analysis to identify groups of Chinese provinces and examine qualitative characteristics of the respective populations (Zhang and Li 2014).

Russian scientists carry out research that draws on cluster analysis for the segmentation of regions based on certain factors, such as the level of development of human capital (Petrykina 2013), levels of business and demographic activity (Ilyshev and Shubat 2008), migration characteristics (Abylkalikov 2015) and so on. Unfortunately, the results of this research are not used as foundation for regional demographic policies. In our opinion, on the one hand, this is linked to a lack of relevant skills among the official bodies developing demographic policies, and on the other hand – to the country's leadership being unwilling to recognise the regional specifics and attain a more profound understanding of existing problems.

In the course of our study, we modelled the country's demographic space on the basis of the level of development of demographic potential. We believe that precisely the development of demographic potential, rather than growth in fertility (which is today the official target benchmark), should be the main objective of Russian demographic policy.

The very concept of demographic potential only emerged in science relatively recently. It is assessed through various indicators. Thus, for example, Goraj et al suggest use purely quantitative population measures (Goraj, Gwiazdzinska-Goraj and Cellmer 2016), while Dobrokhleb and Zvereva include life expectancy (an indicator of the quality of life) as one of the indicators of demographic potential (Dobrokhleb and Zvereva 2016). We believe that an assessment of regional demographic potential should include both quantitative and qualitative measures. This is linked to the fact that the demographic potential of a particular region and the conditions for its actualisation determine not only the quantity, but also the quality of the future population. The very conditions for the development of demographic potential are formed at different levels. The macro level comprises the set of objective conditions related to the economic and demographic conditions in the region. The meso level encompasses the immediate social environment, including family and significant social groups. The micro level is based on an individual's behavioural determinants. We suppose that the more balanced the conditions for the development of demographic potential at different levels, the better the conditions for its overall actualisation.

The purpose of our study is to assess the demographic potential of Russian regions and identifying groups that require the implementation of specific measures aimed at its development. On the basis of cluster analysis we propose a set of measures that would be most relevant to the needs of specific groups of regions and could directly drive the development and actualisation of demographic potential.

DATA AND METHODS

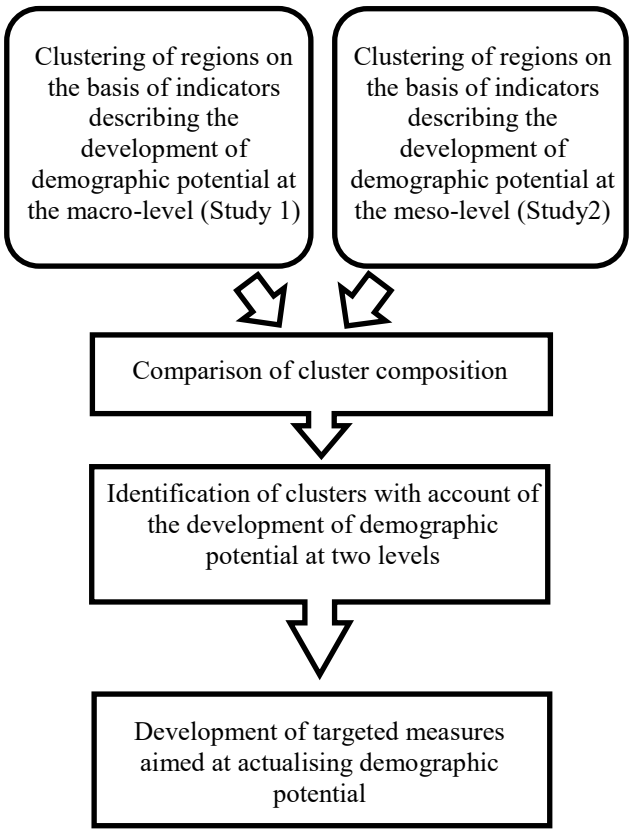
1. We used cluster analysis to model Russia's demographic space and segment Russian regions with comparable problems related to forming demographic potential. In carrying out this analysis, we undertook activities typical for this type of statistical work: the selection and transformation of input variables; the selection of distance measures and linkage rules; the selection of the clustering method; the selection of the number of clusters; the profiling of clusters and interpretation of the attained results.

2. We used hierarchical cluster analysis in our research. We used Euclidean distance as the distance measure and Ward's method to gauge distance between clusters. This decision was made on account of the analytical (discriminatory) abilities of these measures and their effectiveness (supported to the results of multiple studies). Moreover, these measures enabled making the clearest distinctions within the studied body of data, separating out uniform segments. To assess the robustness of the cluster solution, we performed several iterations of the clustering procedure, using different measures of distance between objects. Moreover, we applied partitioning methods of clustering (k-means procedure). For most Russian regions, their allocation into homogenous groups coincided. Some differences in cluster composition did not skew the profile of each identified group of regions. The shared characteristics and relationships identified in the course of the analysis did not change when different distance measures were used.

The decision on the number of identified groups of regions was taken on the basis of:

- Graphical representation of the clustering process (we examined a dendrogram);
- An evaluation of the between-group and within-group variability;
- Cluster size (we tracked the number of regions that form a single cluster to ensure that each group contained a sufficient number of regions).

3. The stages of our research are presented in Figure 1. We called this algorithm targeted clustering.



Figures 1: Research stages

The clustering of Russian regions on the basis of indicators describing the conditions for the formation of demographic potential at the macro-level – that is, the regional level (Study 1) – was based on the following variables:

- Birth rate;
- Perinatal mortality rate;
- Infant mortality rate;
- Under-five mortality rate;
- Pregnancy rate.

We presented the results of this clustering previously (Shubat et al 2016).

5. The clustering of Russian regions on the basis of indicators describing the conditions for the formation of demographic potential at the meso-level – that is, at the family level (Study 2) – was based on the following variables:

- Number of single mothers with children under the age of 18 (per 1,000 people);
- Number of single fathers with children under the age of 18 (per 1,000 people);
- Number of extramarital births (per 1,000 people);
- Coefficient of marriage instability (number of divorces per 1,000 marriages);
- Number of children without parental care (per 1,000 people).

6. We used both the clustering variables and other variables that describe the demographic situation in a region to interpret the clusters themselves. For this purpose we examined cluster centroids and executed tests of significance in difference between two means (or medians). We applied one-way analysis of variance (ANOVA) to evaluate differences between the means and Levene's test to assess the homogeneity of variances. To test the assumption of normality, we used the Shapiro-Wilks test. If we observed a violation of the assumptions of the one-way ANOVA, we used the non-parametric Kruskal Wallis Test.

7. Our samples included all Russian regions that had complete data for all input variables. Thus, Study 1 included 78 regions and Study 2 had 77 regions.

8. To carry out our research, we used data from our current demographic data, as well as data from the 2010 Census. The need to use data from different periods relates to the fact that not all necessary information is collected in Russia within the same period. We used open-source data provided by Russia's Federal Statistics Service (Federal State Statistics Service 2016).

RESULTS

1. In the course of our study, we identified a high degree of regional differentiation for all variables in both Study 1 and Study 2. There is a manifold difference in the maximum and minimum values of the variables that were used for the cluster analysis. Thus, the maximum value for "Marriage stability" is 5.5 times its minimum value. This ratio is even greater for "Number of children without parental care" - 17.5 times. The identified heterogeneity is evidently reason to use clustering methods for Russian regions.

2. The use of hierarchical cluster analysis for the data in Study 1 enabled identifying 3 clusters of regions:

- Cluster 1 – "Low fertility amid low economic activity";
- Cluster 2 – "Cautious" fertility amid high economic activity";
- Cluster 3 – "High fertility amid economic passivity".

We have previously demonstrated the profiling of these clusters and the assessment of the statistical significance of the obtained model of Russia's demographic space (Shubat et al 2016).

3. The application of hierarchical cluster analysis with respect to the data in Study 2 also enabled us to identify 3 regional clusters. The first cluster included 47 Russian regions; there were 10 regions in the second and 20 in the third. The evaluation of the cluster centroid confirmed the appropriateness of these three groups: the median values of the cluster variables differed significantly between the identified clusters (table 2). Results of the Kruskal Wallis Test are presented in Table 3. The dendrogram illustrates the arrangement of the clusters (Figure 2).

Table 2: Median values for cluster variables (Study 2)

Clustering variables	Cluster 1	Cluster 2	Cluster 3
Number of single mothers with children under the age of 18 (per 1,000 people)	38.4	31.3	34.9
Number of single fathers with children under the age of 18 (per 1,000 people)	3.5	5.3	3.5
Number of extramarital births (per 1,000 people)	4.0	2.5	2.7
Coefficient of marriage instability (number of divorces per 1,000 marriages)	575.6	516.0	558.8
Number of children without parental care (per 1,000 people)	0.8	0.3	0.4

Table 3: Kruskal Wallis Test (Study 2)

Clustering variables	Cluster	Mean Rank	Test Statisics		
			Chi-Square	df	Asymp.Sig
Number of single mothers with children under the age of 18 (per 1,000 people)	1	61.80	28280	2	0.000
	2	30.36			
	3	34.00			
Number of single fathers with children under the age of 18 (per 1,000 people)	1	35.40	16501	2	0.000
	2	65.80			
	3	34.83			
Number of extramarital births (per 1,000 people)	1	62.95	31154	2	0.000
	2	30.00			
	3	34.40			
Coefficient of marriage instability (number of divorces per 1,000 marriages)	1	49.35	8540	2	0.000
	2	24.70			
	3	37.64			
Number of children without parental care (per 1,000 people)	1	62.50	30455	2	0.000
	2	25.60			
	3	31.85			

4. The profiling of the groups of regions showed that they could be identified as clusters with different conditions for the development of demographic potential at the meso-level (family level).

Cluster 1 – "Worst conditions for the development of demographic potential at the meso-level"

This cluster includes around 20 Russian regions. The situation with the development of demographic potential here can be described as extremely poor. Indeed, this cluster has the poorest indicators as regards dissolved marriages, extramarital births, children left without parental care and single mother homes. Additional profiling through variables that were not used in the clustering revealed that this group of regions has the highest abortion indicators. Moreover, the figures for married couples without children are also high here.

Cluster 2 – "Best conditions for the development of demographic potential at the meso-level"

This cluster includes 10 Russian regions. The situation with the development of demographic potential can be described as most favourable. This group of regions had the most stable marriages, the lowest number of extramarital births and the least number of children without parental care. This cluster also had the lowest number of incomplete and single-mother families. Additional profiling through variables not used for clustering showed that compared to the rest of Russia, this group of regions has an above-average number of families with three or more children. One aspect of this group that is dissonant with the cluster's overall profile is the number of single fathers. On this, cluster 2 has the highest value out of all clusters.

Cluster 3 – "Satisfactory conditions for the development of demographic potential at the meso-level"

This cluster includes 20 Russian regions. It shows mid-way values across the analysed variables.

5. A comparison of the clusters obtained in the course of Study 1 and Study 2 allowed us to identify the groups of regions that have the best and the worst conditions for the development of demographic potential both at the macro- and the meso-levels.

We believe that it is possible to identify the most depressed groups of regions by finding overlapping entities in the following clusters: cluster 1 in Study 1 (Low fertility amid low economic activity) and cluster 1 in Study 2 (Worst conditions for the development of demographic potential at the meso-level). By comparing these two clusters, we identified a specific group that includes 11 Russian regions, that has the worst conditions for having and raising children at two levels. We labelled this group of regions the "depressed cluster", since the adverse processes happening at the family level coincide with a poor economic background (low levels of economic activity). This most problematic cluster evidently requires specific measures of demographic regulation.

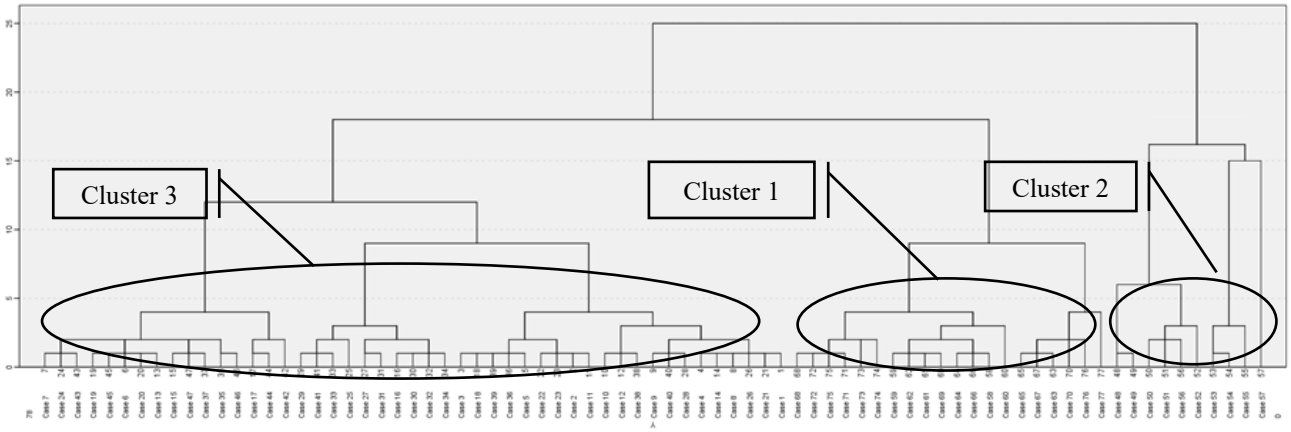


Figure 2: Dendrogram using Ward Linkage (Study 2)

We propose identifying the groups of regions with the most optimal conditions for the development of demographic potential by finding overlapping entities in the following clusters: cluster 3 in Study 1 (High fertility amid economic passivity) and cluster 2 in Study 2 (Best conditions for the development of demographic potential at the meso-level). Unfortunately, only three Russian regions fell within this intersecting zone. We labelled this group the "promising cluster".

DISCUSSIONS AND CONCLUSIONS

In our opinion that amid the adverse economic trends transpiring in Russia, the targeted development of measures for individual types of regions will help optimise budget spending to address such significant demographic policy challenges. The results of our analysis can be used to define priority aims and possible solutions that will reflect the unique needs of each identified regional cluster.

In our view, regions in the depressed cluster would benefit from the following set of measures (table 4).

The measures we propose place a strong emphasis on informational policies and social advertising for the development of demographic potential. We believe it should be segmented by different groups of young people.

Thus, university students can be targeted with social advertising through: 1) video clips on screens in university corridors; 2) advertising in student newspapers (essays, photographs, comic strips); 3) flashmobs; 4) advertising in social networks. The text of advertising materials should be profound and memorable, and convey the value of having and raising children. For example:

– "If I am born, I will bring you joy" (the text is accompanied by a photo of a baby);
– "Family is the meaning of life. Add it to your plans";
– "Plan your future. Plan a family";
– "A child is your future. Do it right"
– An Instagram photo competition around the topic of "This is the family I want".

Table 4: Overall aims and measures for the depressed cluster

Aim	Measures
Reduced perinatal and infant mortality	Improvements in the organisation of medical assistance to pregnant women and newborns, implementation of a "Health of future mothers and babies" programme
Increased pregnancy and birth rates	Informational policies aimed at strengthening young people's reproductive intentions, increasing the desire to have children, establishing a positive image of parenthood (through social networks, billboard advertising, mass media)
Greater marriage stability	Developing psychological support services for families (particularly for young families)
Fewer extramarital births and single-mother households	Informing young people about the legal consequences of unregistered marriages and psychological impact of raising children in incomplete families

Informational measures in regions in the depressed cluster should be supported by organisational and financial measures, aimed at enlarging the pool of state-funded education for children. Given the economic passivity of the region's population, we believe there is a need for a large number of extra-curricular children's activities funded by the local authorities.

Despite the seeming simplicity and obviousness of the described measures, they are hardly implemented in today's Russia. The results of our analysis suggest that the depressed group of regions could be a good 'pilot cluster' for such measures.

We would recommend that the relatively well-off three regions of the 'promising cluster' introduce a

reproduction-focused component into educational curricula. At the moment, these regions are experiencing favourable macro- and meso-conditions for the development of demographic potential. As such, there is scope to set strategic aims for these regions, which will not only address quantity objectives, but also seek to improve the quality of the population's future human capital. University curricula can include special modules in disciplines related to parental education, as regards its medical, legal, economic and psychological aspects. A university graduate that has undergone such training will not only be ready to fulfil professional duties, but also to carry out the functions of a parent. They will be more mindful of the parenting process and have a greater understanding of the upsides and challenges it brings. In our view, such serious preparation for parenthood will enable young people – future parents in these parts of the country – to form higher quality human capital in their children.

On the whole, the analysis we carried out showed the need to use multi-factor classification in the demographics of countries that have a high level of regional differentiation. Modelling the demographic space on the basis of cluster analysis can be seen as an element of the system of supporting administrative decision-making and the development of effective demographic policy at the regional and national levels. The results of cluster analysis can be used as a basis for forecasting demographic risks and threats, identifying points of demographic growth and recession zones.

We see scope to further our research by developing a methodology for analysing the conditions for forming demographic potential at the personal (micro-) level. This study did not take this factor into account given its highly subjective nature and also the difficulties in formalising the information required for such analysis. Moreover, a separate study requires research into indicators for single fathers: its variance across Russian regions and the determinants for this rather rare phenomenon for Russia.

ACKNOWLEDGMENT

The work was supported by Act 211 Government of the Russian Federation, contract № 02.A03.21.0006.

REFERENCES

Abylkalikov, S.I. 2015. "Typological analysis of Russian regions by migration characteristics", *Regional economics: theory and practice,* Vol 22(397), 21-30.

Dobrokhleb, V.G. and Zvereva, N.V. 2016. "The Potential of Modern Russian Generations", *Economic and Social Changes-Facts Trends Forecast,* Vol 44, Issue 2, 61-78.

Federal State Statistics Service. 2016. URL: http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/en/main/ (access date 01.02.2017)

Goraj, S., Gwiazdzinska-Goraj, M. and Cellmer, A. 2016. "Demographic Potential and Living Conditions in Rural Areas of North-Eastern Poland". *MSED-2016: 10th International Days of Statistics and Economics,* 482-493.

Ilyshev, A.M. and Shubat, O.M. 2008. "Multi-factor statistical analysis of business activity in regional micro-business". *Statistical issues,* Vol 4, 42-51.

Koisova, E. and Haviernikova, K. 2016. "Evaluation of selected regional development indicators by means of cluster analysis". *Actual Problems of Economics*, Vol 184, Issue 10, 434-443.

Kronthaler, F. 2005. "Economic Capability of East German Regions: Results of a Cluster Analysis". *Regional Studies*, Vol 39, Issue 6, 739-750.

Laboutková, S., Bednářová, P. and Valentová, V. 2016. "Economic Inequalities and the Level of Decentralization in European Countries: Cluster Analysis". *Comparative Economic Research*, Vol 19, Issue 4, 27-46.

Mertlova, L. and Prokop, M. 2015. "Cluster analysis as a method of regional analysis". *18th International Colloquium on Regional Sciences*, 56-63.

Petrykina, I. N. 2013. "Cluster analysis of regions of the Central federal district on the basis of the level of human capital development". *VSU gazette. "Economics and administration" series,* Vol 1, 72-80. URL: http://www.vestnik.vsu.ru/pdf/econ/2013/01/2013-01-11.pdf (access date 01.02.2017)

Real GDP growth. Annual percent. 2016. International Monetary Fund. URL: changehttp://www.imf.org/external/datamapper/NGDP_RPCH@WEO/OEMDC/ADVEC/WEOWORLD/RUS (access date 01.02.2017)

Shubat, O., Bagirova, A., Abilova, M. and Ivlev, A. 2016. "The Use of Cluster Analysis for Demographic Policy Development: Evidence From Russia". *ECMS-2016: 30th European Conference on Modelling and Simulation*, 159-165.

Single inter-departmental information and statistical system (SIDIS). 2016. Rosstat, Moscow. URL: https://fedstat.ru/ (access date 01.02.2017)

Simpach, O. 2013. "Application of Cluster Analysis on the Demographic Development of Municipalities in the Districts of Liberecky Region". *MSED-2013: 7th International Days of Statistics and Economics,* 1390-1399.

Vahalík, B. and Staníčková, M. 2016. "Key factors of foreign trade competitiveness: Comparison of the EU and BRICS by factor and cluster analysis". *Society and Economy*, Vol 38, Issue 3, 295-317.

Zhang, X. and Li, Z. 2014. "Application of cluster analysis to western China population quality assessment". *EEE-2014: International Conference on E-Commerce, E-Business and E-Service*, 239-242.

AUTHOR BIOGRAPHIES

OKSANA SHUBAT is an Associate Professor of Economics at Ural Federal University (Russia). She has received her PhD in Accounting and Statistics in 2009. Her research interests include demographic processes, demographic dynamics and its impact on human resources development and the development of human capital (especially at the household-level). Her email address is: o.m.shubat@urfu.ru and her Web-page can be found at http://urfu.ru/ru/about/personalpages/O.M.Shubat/

ANNA BAGIROVA is a professor of economics and sociology at Ural Federal University (Russia). Her research interests include demographical processes and their determinants. She also explores issues of labour economics and sociology of labour. She is a doctoral supervisor and a member of International Sociological Association. Her email address is: a.p.bagirova@urfu.ru and her Web-page can be found at http://urfu.ru/ru/about/personal-pages/a.p.bagirova/

IRINA SHMAROVA graduated from the Ural Federal University (Russia) in 2008. Now she is a senior lecturer of economics at Ural Federal University (Russia). Her research interests include demographic processes, the study of demographic policy implementation in the Russian regions and its impact on the state of labor resources. Her email address is: i.v.shmarova@urfu.ru and her Web-page can be found at http://urfu.ru/ru/about/personal-pages/i.v.shmarova/