FUZZY LOGIC MODELLING OF THE RUSSIAN DEMOGRAPHIC SPACE

Anna Bagirova
Oksana Shubat
Alexander Akishev
Ural Federal University
620002, Ekaterinburg, Russia
Email: a.p.bagirova@urfu.ru
Email: o.m.shubat@urfu.ru
Email: alexander.akishev@urfu.me

KEYWORDS
Fuzzy logic, clustering, demographic potential, Russian regions

ABSTRACT

Demographic processes are extremely difficult to manage and require the different methods of their research. Our studies aimed at modelling the demographic space of Russian regions by using fuzzy clustering. Our analysis is based on the indicators of the regions' demographic potential. We used our own original methodology combining the statistical procedures of fuzzy clustering and expert survey data. We considered indicators characterizing the reproduction potential and variables characterizing the potential quality of the future population. As a result of fuzzy clustering, five clusters were formed. Our experts evaluated the reproduction potential and the quality of the future population for each cluster. The data for each region were used to calculate their reproduction potential and the quality of their future population. In comparison to hard clustering, fuzzy clustering enhances the flexibility of evaluation: our assesments of each region do not depend exclusively on the potential of the nearest cluster, as we also take into consideration the region's possible similarities with other neighbouring clusters with different potential. Such modelling allows us to identify those Russian regions that could be considered as 'growth points' in the implementation of demographic policy.

INTRODUCTION

Demographic processes are extremely difficult to manage for a number of reasons. Firstly, they are influenced by a range of external and internal political, economic, cultural, and religious factors. Secondly, demographic processes are inertial: it will take time even for the most efficient demographic policy to make a difference. In addition, the goals of the country's demographic policy might change over the course of time, which will result in a further slowdown of demographic processes. Thirdly, demographic processes and situations in some regions can differ considerably. Given these factors, studying demographic processes to seek out determinants and develop robust forecasts is highly topical. Fertility is of utmost importance, since it determines natural population growth. However, this demographic process provokes the most vigorous scientific and political debate in Russia – questions related to fertility determinants and ways to regulate it remain unresolved.

It should be noted that the latter situation can be found only in certain countries: for instance, some researchers point out differences in the demographic development between northern and southern Italy (Pastuszka 2017), provinces of China (Wang et al. 2017), and Russian regions (Shubat et al. 2017). In Russian regions, demographic indicator values vary greatly (see table 1).

Table 1: Minimax values of demographic indicators in Russian regions in 2016 (Russian Regions 2017)

Indicators	Minimum		Maximum		Ratio of maximum to minimum values
	Value, ‰	Region	Value, ‰	Region	
Crude birth rate	9.2	Leningrad	23.2	Tuva	2.5
Crude death rate	3.3	Ingushetia	17.9	Pskov	5.4
Crude marriage rate	3.8	Ingushetia	8.6	Kam-chatka	2.3
Crude divorce rate	0.9	Chechnya, Ingushetia	6.4	Ma-gadan	7.1
Migration rate	-103.2	Chukotka	140.7	Mos-cow	-

The uncertainty of the demographic situation in countries with highly imbalanced regional development can be reduced if we apply adequate mathematical and statistical tools for demographic analysis and modelling. One such tool is clustering. Traditional cluster analysis, however, does not always provide results reliable enough to develop a demographic policy. One of the reasons for this is that clustering is often based on the researcher's intuition, which also means subjectivity.

Proceedings 32nd European Conference on Modelling and Simulation ©ECMS Lars Nolle, Alexandra Burger, Christoph Tholen, Jens Werner, Jens Wellhausen (Editors)
ISBN: 978-0-9932440-6-3/ ISBN: 978-0-9932440-7-0 (CD)

To enhance objectivity in cluster analysis when modelling demographic processes, we can use fuzzy logic. The fuzzy set theory appeared in the mid-1960s, when Lotfi A. Zadeh proposed a new approach to describe objects and systems which are difficult to formalize (Zadeh 1965). Zadeh showed that for more realistic modelling, it is necessary to go beyond the traditionally accepted standard quantitative methods and introduce so-called linguistic variables into the analysis. The values of these variables can be words or sentences, which enables us to take into account the uncertainty or 'fuzziness' of human knowledge. One of Zadeh's key ideas was to extend classical set theory. In the case of fuzzy sets, the membership function can vary between 0 and 1 rather than just take the value of 0 or 1.

Fuzzy set methods make it possible for one and the same object to belong to several or even all clusters at the same time but with different membership degrees. In many cases, the results of fuzzy clustering seem to be more natural and visual than those of hard clustering. Thus, the application of fuzzy clustering enables us to formalize the uncertainty inherent in demographic processes.

It should be noted that fuzzy set theory is used comparatively rarely in demography. Generally, fuzzy logic methods are used for population forecasts (Sasu 2010; Abbasov and Mamedova 2003). In the Russian Science Citation Index, which is a national database that contains over 12 million scientific publications by Russian authors, only five such publications have been registered since 2005. Seletkov and Martsenyuk have used fuzzy sets to analyze population dynamics (Seletkov and Martsenyuk 2016); Shokin and Fedorov have investigated the relationship between dynamic macroeconomic and demographic indicators in Russian regions (Shokin and Fedorov 2012); and Kulikova and Nikishina have forecasted the population of a specific region (Kulikova and Nikishina 2015).

This research is aimed at modelling the demographic space of Russian regions by using fuzzy clustering. Our analysis is based on the indicators of the regions' demographic potential. Thus, such modelling allows us to identify those Russian regions that could be considered as 'growth points' in the implementation of demographic policy.

The paper is structured as follows: first, we justify the need to use methods with unclear logic to analyse the demographic space of Russian regions. Then we describe the methodology we developed to study demographic potential of Russian regions. Afterwards we outline the results of applying it to assess the potential of population reproduction and the quality potential of the future population of Russian regions. We also explain possible reasons for the imbalance between these two types of demographic potentials and describe possible points of debate in our analysis.

DATA AND METHODS

In this research, we used our own original methodology combining the statistical procedures of fuzzy clustering and expert survey data. The main stages in the application of this methodology are as follows:

1. We created a database on the demographic potential of Russian regions. It should be noted that demographic potential is a comparatively new concept in social and economic studies. There are different approaches to evaluating demographic potential: Goraj et al. measure regions' demographic potential with the help of quantitative population indicators (Goraj et al. 2016), while Dobrokhleb and Zvereva believe that qualitative indicators (in particular life expectancy) can also serve as indicators of demographic potential (Dobrokhleb and Zvereva 2016). We believe that to measure the demographic potential of regions we should include both quantitative and qualitative indicators.

Therefore, this research considers two sets of indicators: the first set deals with a region's reproduction potential and reflects the quantitative aspects of its demographic potential. These variables include the following:

- the rate of natural increase, which is the crude birth rate minus the crude death rate;
- the proportion of children (aged between 0 and 15) in the population;
- the rate of reproduction potential realization – the quotient of the total number of child births by the total number of conceptions (calculated as the sum of the number of child births and abortions).

The second set describes the qualitative aspect of demographic potential (the potential quality of the future population). These variables include the following:

- the proportion of healthy children in the overall population of school-age children;
- the disability rate in the child population (proportion of disabled children aged between 0 and 15);
- the juvenile crime rate (proportion of juvenile delinquents aged 14-17);
- the involvement of children in supplementary educational programs (proportion of children aged 5-18 provided supplementary educational services);
- the proportion of state (municipal) educational institutions meeting modern standards in the total number of such institutions.

The data were provided by the Federal State Statistics Service. Due to certain peculiarities of the national system for the collection of statistical data, the available information did not refer to one year but characterized the situation in the given regions in 2015 and 2016. The resulting database included the characteristics of the demographic potential of 79 Russian regions (six regions were excluded because the necessary data was lacking).

2. To evaluate the demographic potential of these regions, fuzzy clustering was applied on the basis of the fuzzy *c*-means (FCM) algorithm (Bezdek 1981). At the preliminary stage, the initial data were normalized.

It is convenient to consider the result of the clustering of M elements into c clusters by using the following characteristic function:

$$U = [\mu_{ki}], \mu_{ki} \in \{0,1\}, k \in \overline{1,M},\ i \in \overline{1,c}, \qquad (1)$$

where the kth row of matrix U shows the membership degree of the kth object in the ith cluster (0 designates that the object does not belong to the cluster while 1 means that it does). The membership function in the fuzzy set can take any values within the interval [0, 1].

The algorithm of fuzzy c-means has the following parameters: c is the number of clusters while m is the exponential weight $m \in (1, \infty)$. Exponential weight m affects the matrix of membership degrees. The higher m is, the fuzzier the final c-means matrix; at $m \to \infty$, the degrees of membership tend to $1/c$, which is a bad solution as all objects belong to all clusters with the same membership degree.

In this analysis, Euclidean distance was used as a measure of distance. This algorithm enabled us to solve the task of criterion minimization:

$$\sum_{i=\overline{1,c}} \sum_{k=\overline{1,M}} (\mu_{ki})^m \|V_i - X_k\|^2, \qquad (2)$$

where V_i are cluster centres; X_k are clustering elements; $m \in (1, \infty)$ is the exponential weight; and μ_{ki} is the membership degree of the kth element in the ith cluster. The values of cluster centres V_i are calculated as

$$V_i = \sum_{k=\overline{1,N}} (\mu_{ki})^m X_k / \sum_{k=\overline{1,N}} (\mu_{ki})^m,\ i = \overline{1,c} \qquad (3)$$

3. The analysis further included expert evaluation of the resulting cluster centroids. Five specialists in demography separately evaluated the reproduction potential X_i^1 and the potential quality of the future population X_i^2 in each cluster. Then their evaluations were converted into numerical values (a high level corresponded to 5 and a low level to 1).

4. Expert evaluations of the clusters and the membership degrees of the regions in the clusters were used to calculate the reproduction potential P_i^1 and the quality of the future population P_i^2 in each region:

$$P_i^1 = \sum_{j=1}^{c} \mu_{ij} X_i^1,\ P_i^2 = \sum_{j=1}^{c} \mu_{ij} X_i^2,\ i = \overline{1,M} \quad (4)$$

5. The calculated values of the regions' potential were then used to rank these regions. The resulting values R_i^1 and R_i^2 are the ordinal indicators of a region's potential compared to other regions (these indicators show how high or low the given region ranks according to its potential).

6. To develop a differentiated and focused demographic policy framework, it is necessary to consider not only the values of each region's demographic potential, but also how balanced this demographic potential is. Therefore, values R_i^1 and R_i^2 were used to calculate the indicator of balance Q_i:

$$Q_i = |R_i^1 - R_i^2|,\ i = \overline{1,M}. \qquad (5)$$

It is clear that the lower value Q_i is, the more balanced the demographic situation in the region is in terms of its reproduction potential and the quality of its future population.

RESULTS

1. As a result of fuzzy clustering, five clusters were formed (exponential weight $m = 1.4$). Table 2 shows the values of cluster centroids.

Table 2: Cluster centroids

Cluster	Indicators characterizing the reproduction potential		
	Rate of natural increase, %	Proportion of children in the population, %	Rate of realization of reproduction potential, %
1	13.4	29.8	0.87
2	1.2	20.7	0.62
3	-3.6	16.3	0.69
4	-1.4	18.2	0.63
5	1.1	18.6	0.74

Cluster	Indicators characterizing the potential quality of the future population				
	Proportion of healthy children, %	Disability rate in the child population, %	Juvenile delinquency rate, %	Involvement of children in the system of supplementary education, %	Proportion of modern educational institutions, %
1	78.2	4.30	0.15	40.9	68.2
2	83.6	2.03	1.62	54.5	79.4
3	80.6	1.98	0.81	70.2	82.7
4	78.4	1.90	1.38	71.7	81.1
5	82.0	2.02	0.67	55.7	82.3

Fuzzy clustering does not provide an exact list of regions for each cluster. Thus, we obtained models of demographic situations in different regions of the country. For instance, the first model (Cluster 1) is characterized by a high level of reproduction potential: this cluster has a high rate of natural increase, a large proportion of children, and a high rate of reproduction potential realization. The potential quality of the future population in this demographic model is extremely low, as this model combines the lowest proportion of healthy children with the highest disability rate, the lowest involvement of children in supplementary education, and the lowest proportion of modern education institutions.

2. Our experts evaluated the reproduction potential X_i^1 and the quality of the future population $X_i^2 (i = \overline{1, M})$ for each cluster (table 3). It can be observed that there is a lack of balance between the two kinds of potential in the majority of clusters.

Table 3: Expert evaluation of the potential of cluster centres

Cluster	Reproduction potential		Potential quality of the future population	
	Expert evaluation	Numeric value in expert evaluation	Expert evaluation	Numeric value in expert evaluation
1	high	5	low	1
2	medium	3	above medium	4
3	low	1	above medium	4
4	low	1	medium	3
5	above medium	4	above medium	4

3. In the next step, the data for each region were used to calculate their reproduction potential P_i^1 and the quality of their future population P_i^2. The results of these calculations were then used to rank regions R_i^1, R_i^2 and to estimate balance Q_i of the two types of potential. In other words, the regions were ranked according to the two types of potential and how balanced they are.

It should be noted that some regions did not belong exclusively to one cluster (based only on the membership degrees). However, expert evaluations and membership degrees P_i^1, P_i^2 made it possible to get a clear picture of these regions' reproduction potential and the potential quality of their population in the future.

Let us now focus on the case of Vologda region. The values of the membership degrees of this region in the clusters are 0.000113, 0.553551, 0.033691, 0.395077, and 0.017568, respectively. Therefore, this region cannot be described as representing only one demographic model since its membership degrees in the second and fourth clusters are quite high.

According to table 3, experts evaluated the reproduction potential of this region in the second and fourth clusters as 'medium' and 'low', respectively. The quality of the future population is evaluated as 'above medium' and 'medium', respectively. These data alone are insufficient to identify the demographic potential of Vologda region.

Further calculations and ranking provided us with a more detailed picture of this region's demographic potential. The region received the following expert evaluations: $P_1 = 2.16$, $P_2 = 3.6$. Accordingly, the region ranked $R_1 = 40, R_2 = 56$ in terms of its reproductive potential and the potential quality of its future population. Taking into consideration the overall number of the regions in this study (79), the ranking position of Vologda region demonstrates that this region's reproduction potential corresponds to the medium level, while the potential quality of its future population is below the medium level. As for the balance of these two types of potential, the region ranks twenty-first (that is, in the upper third of the ranking) and can be described as a region with comparatively balanced levels of these two types of potential.

Thus, fuzzy clustering enabled us to make an exact evaluation of the demographic potential of the regions whose characteristics were shared by several clusters. These regions' evaluations depended on the values of the demographic potential of the closest clusters proportional to the degrees of membership in these clusters. Such a result would be impossible to achieve if we applied only clear clustering algorithms, which would mean that the evaluation of the region's potential would be equal to the evaluation of the potential of the nearest cluster, regardless of the possible similarities with other clusters.

DISCUSSION

This research undoubtedly contains a number of debatable points, one of which is the fundamentally ambiguous solution of the clustering task. It is known that there is no single or best criterion for clustering quality. Fuzzy clustering means that the researcher uses subjective criteria to set the number of clusters and the exponential weight, which affects the results. If the number of clusters is too small, however, some groups of regions with unique characteristics will not be revealed. An increase in the number of clusters leads to the creation of clusters with close centroids which do not manifest significant differences in terms of the region's demographic potential.

It should be noted that the value of the exponential weight is determined depending on how the clustering results are going to be used. If the level of *m* is too high, the final results are 'excessively averaged'; if it is too low, all regions get the same evaluation as the cluster they belong to with a high membership degree, even if they are actually shared by several clusters.

In further analysis, we can turn from fuzzy clusters to standard hard clusters using a specific list of regions for every cluster, which will enable us to develop a more focused demographic policy. Such a transition can be done in various ways:

– for example, we can use the α-level set. Let us assume that a region belongs to a cluster if its degree of membership in this cluster is higher than the given value α (α∈[0,1]). This method, however, has certain peculiarities which should be taken into account. For example, if the value of α is high enough, some regions may correspond to none of the clusters (if all degrees of membership are smaller than α). If the value of α is low enough, some regions may belong to two or more clusters;

– another way is to use the principle of maximum membership degree. In this case, if a region's membership degree to a certain cluster is maximal, we will consider that it belongs to this cluster.

In order to interpret these findings correctly, we should keep in mind the following.

Firstly, there are different reasons for the imbalanced demographic development of Russian regions:

1. Cultural and religious reasons. Russia is a multinational and multifaith country. For example, in regions with a high share of Muslim population the quantitative indicators of demographic potential tend to be higher. These regions are generally characterized by strict adherence to family traditions, with people preferring home-based care for pre-school children and having strong views about child care and children's physical development and health. In such regions, the potential quality of the population is usually lower than the reproduction potential.

2. Economic reasons. Russia is a country with a high level of differentiation between its regional economies. For example, the maximum GRP per capita in Tumen region exceeds the minimum GRP in Ingushetia by 14 times. Moreover, different Russian regions have different urbanization levels: for example, in Moscow region, the share of the rural population is 7.5% while in the Republic of Altai, it is 70.8% (9.4 times larger). In regions with a comparatively high standard of living, a developed economy, and a high level of urbanization, the potential quality of the population tends to be higher than the reproduction potential. In these regions, the population of reproductive age primarily seeks to realize themselves professionally, which means that childbirth is often delayed, resulting in the inevitable decline of birth rates and reproduction potential.

To improve the demographic situation in Russian regions, it is essential to achieve the right balance between the reproduction potential and the potential quality of the population, while simultaneously enhancing both types of potential. These are the objectives that should underpin the demographic policy of the regions and the country in general. The full-fledged realization of the population's quality potential, together with a high reproduction potential, will provide regions with the amount and quality of human capital necessary for their economic and socio-cultural development.

Secondly, the ranking of the regions according to their reproduction potential and potential quality of their population demonstrates that a special set of demographic measures should be developed and implemented for certain groups of regions. For instance, in regions with high reproduction potential and low potential quality of the population, it is necessary to improve the institutions responsible for the development of human capital, that is, institutions working in the sphere of health care, education, and culture. In some regions, the potential quality is high while the quantitative indicators are low, which requires measures which are traditionally used in Russia to stimulate population growth. It might also be necessary to redistribute resources from the regions which are at the top of the reproduction potential ranking and, therefore, do not need these measures.

CONCLUSIONS

The original methodology applied in this study has several advantages. Firstly, it considerably reduces the experts' workload, as they did not have to provide evaluation for all 79 regions of Russia. What the experts had to do was evaluate five clusters as models of the demographic situation. Secondly, this methodology reduces the subjectivity level of expert evaluation because in the final evaluation of the regions' demographic potential we used the objective membership degrees we received through fuzzy clustering. Thirdly, in comparison to hard clustering, fuzzy clustering enhances the flexibility of evaluation: our assesments of each region do not depend exclusively on the potential of the nearest cluster, as we also take into consideration the region's possible similarities with other neighbouring clusters with different potential. Furthermore, ranking decreases the impact that the clustering parameters have on the final result, which makes the study more objective.

ACKNOWLEDGMENTS

The article is processed as one of the outputs of the research project "Fertility and parenting in Russian regions: models, invigoration strategies, forecasts", supported by the President of Russian Federation, project no. NSh-3429.2018.6.

REFERENCES

Abbasov, A.M. and M.H. Mamedova. 2003. "Application of fuzzy time series to population forecasting". In *Proceedings of CORP 2003* (Vienna, Austria, February 25th-February 28th, 2003).

Bezdek, J.C. 1981. "Pattern Recognition with Fuzzy Objective Function". New York, Plenum Press.

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

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

Kulikova, V.P. and O.A. Nikishina. 2015. "Statistical Modelling and Forecasting of the Region's Demographic Development", *Journal of Omsk Regional Institute,* Vol 1, Issue 1-1, 44-46.

Pastuszka, S. 2017. “Regional Differentiation of The Demographic Potential in Italy and Poland”. *Comparative Economic Research-Central and Eastern Europe,* Vol 20, Issue 3, 137-159.
Russian Regions. Socio-Economic Indicators. 2017. Statistical Book. Rosstat, Moscow.
Sasu, A. 2010. “An application of fuzzy time series to the Romanian population”, *Bulletin of the Transilvania of Brasov*, Vol 3(52), Series III: Mathematics, Informatics, Physics, 125-132.
Seletkov, I.P. and M.A. Martsenyuk. 2016. “Modelling urban growth using matrix fuzzy cellular automata”. In *Proceedings of the Russian Conference of Young Scientists on Mathematics and Interdisciplinary Research (Perm, Russia, May 16th-May 19th, 2016),* 238-241.
Shokin, I.V. and S.V. Fedorov. 2012. “The forecasting method for the mutual influence between dynamic macroeconomical and demographic indicators in regions of Russian Federation (applying some methods of scenario analysis)”, *European Social Science Journal*, Vol 9-2 (25), 382-389.
Shubat, O., Bagirova, A. and I. Shmarova. 2017. “The Use of Cluster Analysis to Assess the Demographic Potential of Russian Regions”. In *Proceedings of the 31th European Conference on Modelling and Simulation (*Budapest, Hungary, May 23rd-May 26th, 2017), 53-59.
Wang, F., Zhao, L. and Zh. Zhao. 2017. “China's family planning policies and their labor market consequences”. *Journal of Population Economics*, Vol 30, Issue 1, 31-68.
Zadeh, L. 1965. “Fuzzy sets”. *Information and Control*, Vol 8, 338-353.

AUTHOR BIOGRAPHIES

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/

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/personal-pages/O.M.Shubat/

ALEXANDER AKISHEV is a PhD student of Institute of Natural Sciences and Mathematics at Ural Federal University (Russia). His research interests include fuzzy clustering and its applications to various applied fields. His email address is: alexander.akishev@urfu.me