The forecast of demographic situation in Russia up to 2033

V V Golubkov¹, T Y Yakovets², O A Malafeyev³, I V Zaitseva³,⁴ and P M Kuschev⁵

¹Federal research center "Informatics and management" of the Russian Academy of Sciences, 119333, Moscow, Vavilova str., 44, building 2, Russia
²Pitirim Sorokin-Nikolai Kondratieff International Institute, 101000, Moscow, Sretensky Boulevard, 6/1, Russia
³Saint-Petersburg State University, 7/9 Universitetskaya nab., Saint-Petersburg, 199034, Russia
⁴Stavropol State Agrarian University, Zootekhnichesky Lane, 12, 355017, Stavropol, Russia
⁵Stavropol branch of the Krasnodar university of Ministry of Internal Affairs of Russia, 43 Kulakov prospect, 355035, Stavropol, Russia

E-mail: golvic@mail.ru, tzag@mail.ru, irina.zaitseva.stv@yandex.ru, o.malafeev@spbu.ru, kushchev70@mail.ru

Abstract. The brief description of the demographic model and modeling technique elaborated and used by the authors for investigating the trends of demographic development in Russia are given. The demographic model and the modelling technique are based on auto regression and regression analysis of time series for demographic indicators, simulation modelling and confidence intervals for model values of demographic indicators with taking into account a priori information and constraints on the model parameters and modelling process. Based on confidence intervals three scenarios of Russia's demographic development were developed: Low, Medium and High. On the time period 2013-2033 being investigated the demographic model to be constructed allows for the scenario variants to calculate one-year age structures for men and women, one-year age fertility, mortality and net migration rates and all necessary for analysis integral demographic indicators. The results of the multivariate analysis of the demographic processes in Russia in 2013-2033-time period to be carried out are described.

1. Demographic model and methodology of multivariate analysis of demographic processes
The demographic model and methodology of modelling of demographic processes used in the work have a number of special features in comparison with other analogous models and methods (for example, [1]). Special features are provided for the specifics of demographic processes in Russia. In particular, in the work a modification [2] of well-known Gamma-model [3] is used, describing one-year age specific fertility rates. The modified Gamma-model for Russia has been developed by one of the authors of this article. As opposed to classical Gamma-model which contains three input (scenario) variables (total fertility rate, mean age of woman at childbearing and variance in age of woman at childbearing) the modified Gamma-model has only two first scenario variables from listed above. These two scenario variables have clear sense for demographers. Since the variance in age of woman at childbearing does not have explicit physical sense for demographers then in the work [2] instead of
the variance there was introduced scenario variable named the most-probable age of woman at childbearing. This variable has explicit physical sense and means the age of woman, at which the probability to bear a child has maximal value. As a result of analysis of Russian statistical data on fertility in the historical period of time a linear dependence of the most-probable age of woman at childbearing from total fertility rate and mean age of woman at childbearing was obtained. This dependence allows to decrease the number of input variables in the modified Gamma-model, which is important for the imitation modelling. It should be noted that introducing of scenario variable most-probable age of woman at childbearing is possible only in case when dependence of specific fertility rates from age has unique maximal value that takes place for most of countries including Russia.

In the work, the modelling of scenario variables of the demographic model was done by probabilistic method based on imitation approach which includes calculation of random realizations of scenario variables. During generation of random realizations of scenario variables authors used the procedure of censoring, and in further calculations only realizations of the scenario variables with values within the assigned limits were used, and their increases for the year in the absolute value did not exceed the given values. The specified limitations were obtained by the authors as a result of the analysis of corresponding Russian statistical data in the historical period of time. This approach allowed to increase the accuracy of modelling. It should be said that the censoring was used in the foreign works (for example, [3]), but only for the foreign countries, in which the limitations on the scenario variables were characteristic for these countries.

In the work one-year sex-age structures of population at the 100-year age interval and the 100-year structures of one-year age-specific mortality rates for men and women were used, which were obtained from the corresponding survival functions. Since the Russian statistical data on survival functions and sex-age structures are given up to 85 years with age step of 5 years, the authors developed a methodology for restoring missing data. In the case of sex-age structures a monotonic cubic spline was used, and in the case of age specific mortality rates a monotonic cubic spline was used and the extrapolation of age specific mortality rates for the older ages with aid of part of Heligman-Pollard mortality model [4] for older ages.

At the basis of the model considered in this article a system of finite-difference equations is laid [5] with one-year time step and one-year age groups for men and women. In order with the aid of the finite-difference equations to calculate for each year of modeling period of time one-year sex-age structures of population, it is necessary to specify these structures for the initial year, as well as to specify as functions of time the share of the born boys, the one-year age specific fertility rates for women, the one-year age-specific mortality rates and one-year age structures of net migrations for men and women.

Since number of listed one-year age indicators, which must be assigned for each year of the modeling period of time, is quite large (in this work there are 36 age-specific fertility rates, 101 age-specific mortality rates and 101 age net migration rates for males and females), then in constructing a demographic model in order to reduce the number of its input (scenario) variables for the corresponding age indicators parametric age models with a small number of parameters are usually used. The parameters of these models, being functions of time, are scenario variables of the demographic model. In this paper, the following parametric age models were used for fertility, mortality and net migration.

1.1. Fertility model
One of the most widespread parametric age models of fertility approximating one-year age-specific fertility rates as functions of age is so called Gamma-model [3]. This model actually has 4 input parameters (scenario variables, depending only from time), one of which is the minimal fertility age. This parameter in Russia usually equals to 15 years. Therefore, actually in our case Gamma-model contains 3 scenario variables. In this work a modification of Gamma-model elaborated by one of the authors for Russia was used as fertility model [2]. Scenario variables of this model are total fertility
rates and mean ages of woman at childbearing, depending only from time, and also the minimal fertility age, equal 15 to years. The model can be written in the form:

\[ b(x,t) = \begin{cases} \frac{\alpha_2(t)(x - \alpha_4(t))^{\alpha_3(t)-1} e^{-\alpha_2(t)(x - \alpha_4(t))}}{\Gamma(\alpha_3(t))}, & x \geq \alpha_4(t) \\ 0, & x < \alpha_4(t) \end{cases} \]

\[ \alpha_1(t) = \beta_1(t) - 1, \alpha_2(t) = \frac{(\beta_2(t) - \beta_4(t))}{\beta_3(t)}, \alpha_3(t) = \frac{(\beta_2(t) - \beta_4(t))^2}{\beta_3(t)}, \alpha_4(t) = \beta_4(t) \]

where \( b(x,t) \) is age fertility rate on the age interval \([x, x+1)\) at time \( t \), \( \Gamma(\alpha) \) - Gamma function, \( \beta_1(t) \) - total fertility rate, \( \beta_2(t) \) - mean age of woman at childbearing, \( \beta_3(t) \) - variance of Gamma distribution and \( \beta_4(t) = 15 \) - minimal fertility age. The variance \( \beta_3(t) \) for Russia is calculated by the formula

\[ \beta_3(t) = [\beta_2(t) - X_m(t)][\beta_2(t) - \beta_4(t)], X_m(t) = q_1 + q_2\beta_1(t) + q_3\beta_2(t) \]

\[ \beta_4(t) = 15 q_1 = 2.8439 q_2 = 0.2499 q_3 = 0.7554 \]

where \( X_m(t) \) - is age of woman, at which the probability to bear child has maximal value. In the formula above the dependence \( X_m(t) \) from total fertility rate \( \beta_1(t) \) and mean age of woman at childbearing \( \beta_2(t) \) was obtained by the authors as a result of analysis of Russian statistical data on fertility at time period 1962-2012.

For the population modeling it is also necessary to specify shares of born boys \( k_m(t) \).

### 1.2. Mortality model

The widespread up to date model of Lee-Carter [6] was used as age parametric model for the natural logarithms of one-year age-specific mortality rates for men and women. This model has one scalar parameter for each sex depending from time and two vector parameter components depending on age only. The number of components of vector parameters equals to the number of one-year age-specific mortality rates. The statistical values of scalar and vector parameters in this model were estimated by statistical data of natural logarithms of one-year age-specific mortality rates for men and women. During the multivariant analysis the vector parameters remained constant, equal to the values, obtained as a result of processing statistical data. Thus, scenario variables in this model of mortality are scalar time parameters for men and women, which under assigned vector parameters are uniquely connected with the life expectancies at birth respectively for men and women. This model in our case is described by the following formulas.

\[
\begin{align*}
\text{Ln}(d_m(x,t)) &= A_m(x) + B_m(x)K_m(t) + \sum_{k=0}^{N} B_m(x) = 1, \sum_{k=0}^{N} K_m(t_0 + k) = 0 \\
\text{Ln}(d_f(x,t)) &= A_f(x) + B_f(x)K_f(t) + \sum_{k=0}^{N} B_f(x) = 1, \sum_{k=0}^{N} K_f(t_0 + k) = 0 \\
x &= 0,1,2,...,100; t = t_0, t_0 + 1, t_0 + 2, ..., t_0 + N
\end{align*}
\]
where \( x \) is age, \( d_m(x,t) \) and \( d_f(x,t) \) are one year age-specific mortality rates respectively for men and women at time \( t \), \( A_m(x) \), \( B_m(x) \) and \( A_f(x) \), \( B_f(x) \) are components of vector parameters for men and women respectively, \( \ln(z) \) is natural logarithm of value \( z \), \( t_0 \) and \( t_0 + N \) are correspondingly the first and terminal time moments of modeling period of time.

By Russian corresponding statistical data on mortality at time period 1994-2012 (\( t_0 = 1994, N = 18 \)) with least squares method estimates \( \tilde{A}_m(x), \tilde{B}_m(x), \tilde{K}_s(t) \) of values \( A_m(x), B_m(x), K_s(t) \) were obtained, where \( S = \{m,f\}; t = t_0, t_0 + N \). Further, according to the Lee-Carter mortality model the values \( \tilde{K}_m(t), \tilde{K}_f(t) (t = t_0, t_0 + N) \) were considered as realizations of random processes \( K_m(t), K_f(t) \), described by the following models of random walk with parameters (shifts) \( c_m \) и \( c_f \).

\[
K_m(t + 1) = c_m + \tilde{K}_m(t) + e_m(t), \quad K_f(t + 1) = c_f + \tilde{K}_f(t) + e_f(t); \quad t = t_0, t_0 + 1, \ldots
\]

where random values \( e_m(t), e_f(t) \) have unbiased normal distributions with variances \( \sigma_m^2, \sigma_f^2 \) correspondingly. Further by the values \( \tilde{K}_m(t), \tilde{K}_f(t) (t > t_0 + N) \) with least squares method estimates \( \bar{c}_m, \bar{c}_f, \bar{\sigma}_m^2, \bar{\sigma}_f^2 \) of parameters \( c_m, c_f, \sigma_m^2, \sigma_f^2 \) were calculated, and by obtained estimates with usage of equations (2) random realizations \( \tilde{K}_m(t), \tilde{K}_f(t) (t > t_0 + N) \) were generated. Then with help of equations (1) by random realizations \( \tilde{K}_m(t), \tilde{K}_f(t) (t > t_0 + N) \) and estimates \( \tilde{A}_m(x), \tilde{B}_m(x) \) random realizations of age-specific mortality rates \( d_m(x,t), d_f(x,t) (t > t_0 + N) \) were generated, by which mean values, low and upper bounds of confidence intervals at modeling period of time were calculated.

1.3. Migration model
In the work for one-year age net migration structures for men and women quite simple model elaborated by the authors was used. This model contains two scalar parameters depending on time (net migration and the ratio of the number of female migrants to the number of male migrants in the total net migration) and vector parameters depending only on age respectively for men and women. The vector parameters are mean values of normalized one-year age net migration structures for men and women obtained by averaging of corresponding statistical data for the last several years prior to the beginning of the modeling time period. In a multivariate analysis, the vector parameters, as well as the ratio of the number of female migrants to the number of male migrants in the total net migration, remained unchanged, equal to the values obtained as a result of processing statistical data. Thus, this migration model has only one scenario variable, the total net migration.

For net migrations for men and women \( S_m(x,t), S_f(x,t) \) the following rather simple model was taken.
\[ S_m(x,t) = \frac{S(t)\bar{S}_m(x)}{1 + k_x(t)} \], \[ S_f(x,t) = \frac{k_x(t)}{1 + k_x(t)}S(t)\bar{S}_f(x) \]

\[ x = 0,1,2,\ldots,100; \ t = t_0, t_0 + 1, t_0 + 2, \ldots \]

where \( x \) - age, \( t \) - time, \( \bar{S}_m(x) \), \( \bar{S}_f(x) \) - time mean values of normalized (in shares) age structures of net migrations respectively for men and women, \( S(t) \) - total net migration (men and women), \( k_x(t) \) - ratio of the number of female migrants to the number of male migrants in total net migration.

The mean normalized age structures \( \bar{S}_m(x) \), \( \bar{S}_f(x) \) were obtained in the following way. Since for every time moment \( t \) statistical data on age migration structures to be at disposal of the authors were five-year age structures of arrivals and departures (both for men and women) at age interval from 0 to 65 and the number of people older 65 years, then on the base of the methodology elaborated by the authors and based on usage of monotonic cubic spline, for every time moment \( t \) one year age structures of arrivals and departures at age segment of 0 to 100 years were calculated. Further by obtained data age normalized structures of net migrations \( \bar{S}_m(x,t), \bar{S}_f(x,t) \) for men and women at time period 2004-2012 were evaluated (the values \( \bar{S}_m(x,t), \bar{S}_f(x,t) \) for earlier moments of time are out of interest, since from 1990 to 2003 there was occurred qualitative change in the dynamics of demographic processes, caused by the demographic transition). As \( \bar{S}_m(x) \) and \( \bar{S}_f(x) \) were taken the values obtained by averaging over time the normalized age structures \( S_m(x,t), S_f(x,t) \) for the last 5 years (for 2008-2012), which turned out to be very close to each other.

Obtained mean values and values of low and upper bounds for confidence intervals of scenario variables on fertility, mortality and net migrations at time period 2013-2033 formed three scenarios of demographic development: Low, Mean and Upper. These scenarios are unequally determined by input data assignment for finite-difference equations describing dynamics of demographic processes.

Depending on the scenario variant, the following data as functions of time were used as input data for these equations (for all scenario variants the sex-age structure of the population is set on the 1-st January of 2013 year to be the initial modeling year).

1.4. Low variant
The share of born boys, the low bounds of confidence intervals for age-specific fertility rates, the upper bounds of confidence intervals for age-specific mortality rates for men and women, the low bounds of confidence intervals for age-specific net migration rates for men and women.

1.5. Mean variant
The share of born boys, the mean of age-specific fertility rates, the mean of age-specific mortality rates for men and women, the mean of age-specific net migration rates for men and women.

1.6. Upper variant.
The share of born boys, the upper bounds of confidence intervals for age-specific fertility rates, the low bounds of confidence intervals for age-specific mortality rates for men and women, the upper bounds of confidence intervals for age-specific net migration rates for men and women.

For the scenario variants of demographic development described above for each year of the modelling period of time 2013-2033. one-year specific fertility and mortality rates were calculated, one-year sex-age structures for men and women, and also a number of integral demographic
indicators. Then the multivariant analysis of future trends in the demographic development of Russia, based on the obtained results of calculations was carried out.

2. Conclusion

As a result of multivariant analysis, carried out with using of the proposed demographic model, the following main conclusions can be drawn regarding Russia's demographic situation in 2013-2033.

- The fertility in the modeling period of time 2013-2033 will decrease. The reduction in fertility rate is generally due to the influence of demographic wave (with increase of time the number of women of fertile age decreases and the peak of the distribution of women by age shifts towards older ages), and also with the decrease in the growth rate of the total fertility rate in case of High scenario variant and decrease in the total fertility rate in case of Mean and Low scenario variants.

- The numbers of births (for all scenario variants) and deaths (for Mean and High scenario variants) will decrease, and the intensity of decrease in the number of births will be substantially greater than that of deaths. In the case of the Low scenario variant, the number of deaths will practically not change. As a result, the annual natural increase of population for Mean and Low scenario variants will have decreasing negative values at the modeling period of time, which means that in this case the mortality rate will exceed the fertility rate. In the case of the High scenario variant, the annual natural increase of population at first will increase from 0 in 2013 to 1.7 in 2016, and then monotonously decrease, tending to almost zero value in 2033.

- The total population will increase from 143.3 million persons in 2013 to 145.7 million and 154.7 million persons in 2033 respectively in cases of Mean and High scenario variants. In case of the Low scenario variant, the total population will decrease monotonically from 143.3 million persons in 2013 to 136.3 million persons in 2033. The growth in total population in the cases of Mean and High scenario variants and a decrease in the rate of population decline in the case of Low scenario variant is mainly due to the net migration increase.

- The population will continue to age. The number of people of in the retirement age will grow, whereas the total number of people in working age and children (children under the age of 16 inclusively) will decrease. The number of children will decrease monotonically from 110 million of persons in 2013 to values that range from 97 to 109.5 million of persons in 2033, while the number of retired persons will increase monotonically from 33 million of persons in 2013 to values lying in the range from 39 to 45 million of people in 2033.

- The life expectancy at birth for men and women will increase monotonically from 64.6 and 75.9 years for men and women in 2012, respectively, to values that range between 65.2-73 years for men and 76.7-81 years for women in 2033. At the same time, the growth of life expectancy rate for men will be slightly higher than for women, and the difference between expected life expectancies of women and men will practically remain constant on the modelling period of time and on the average equal to 10 years.

- The number of able-bodied people will on the average decrease from 86.1 million of persons in 2013 to the values lying in the range of 76.81.4 million of persons in 2033, and the number of disabled people (retired persons and children under 16, inclusive) will on average increase from 57.2 million in 2013 to values lying in the range of 60.3-73.3 million of persons in 2033. So, the coefficient of demographic load on the population (the number of disabled people per 1,000 able-bodied ones) will almost monotonically increase, beginning with values of 664 in 2013 year to values ranging from 793 to 901. From the latter it follows that in 2033 per one person of working age will fall from 1.1 to 1.26 people of working age, that can create socio-economic problems in the future.
The proposed demographic model can be used for a multivariate analysis of demographic processes in conjunction with the macroeconomic model [7]. In this case, it is necessary to construct models describing the dependencies of the scenario variables of the demographic model on economic indicators, including the financial flows allocated for carrying out the socio-demographic programs. The conduction of simulation experiments in this case will allow to predict the consequences of the decisions made at the stage of elaborating solutions related to the state regulation of the socio-demographic processes.

The conclusions obtained as a result of modelling do not contradict to the results of studies carried out by the scientists of Institute of Demography of SRU Higher School of Economics [8] and Institute of Social and Economic Problems of Population of RAS [9, 10].

Acknowledgment
This article was written as a result of fulfillment of the RGF grant in 2016-2018, № 16-02-00229, "Socio-demographic evolution of Russia and other BRICS countries: regularities, trends and prospects". The work is partially supported by the RFBR grant № 18-01-00796.

References
[1] Schoen R 1988 Modeling Multigroup Populations (Plenum Press, New York and London) p 308
[2] Golubkov V V and Krugliakov C V 2010 Construction of the prognostic two-parametric age fertility model by Russian data Dynamics of heterogeneous systems. Proceedings ISA RAN 53(14) 8799
[3] Keilman and Nico and Pham and Dinh Quang 2000 Predictive Intervals for Age-Specific Fertility European Journal of Population 16 41-66
[4] Heligman L and Pollard J H 1980 The age pattern of mortality Journal of the Institute of Actuaries 107 49-75
[5] Rogers A 1975 Introduction to Multiregional Mathematical Demography (New York: Wiley)
[6] Carter L, Lee and Ronald D 1992 Modeling and Forecasting U.S. Mortality: Differentials in Life Expectancy by Sex International Journal of Forecasting 8(3) 393-412
[7] Yakovets T Y 2013 The Demographic Situation of Russia (Germany, Lambert Academic Publishing)
[8] Vishnevsky A G 2012 Demographic breakthrough or circling? Demoskop Weekly 533-4 1-29
[9] Rimashevskaya N 2003 Man and reforms: secrets of survival (Moscow, Institute of Social and Economic Problems of Population of RAS)
[10] Population of modern Russia: risks and opportunities 2013 (Moscow, Economic Education)
[11] Grigorieva X and Malafeev O 2014 A competitive many-period postman problem with varying parameters Applied Mathematical Sciences 8(145-148) 7249-58
[12] Malafeev O and Nemnyugin S 1996 Generalized dynamic model of a system moving in an external field with stochastic components Theoretical and Mathematical Physics 107(3) 770
[13] Pichugin Y and Malafeyev O 2016 Statistical estimation of corruption indicators in the firm Applied Mathematical Sciences 10 (41-44) 2065-73
[14] Kriulina E, Tarasenko N, Miroshnitchenko N, Zaitseva I and Dedyukhina I 2016 Environmental Management in Agriculture: Problems and Solutions Research Journal of Pharmaceutical, Biological and Chemical Sciences 7(3) 1908-12
[15] Zaitseva I, Kriulina E, Ermakova A, Shevchenko E and Vorokhobina Y 2016 Application of Factor Analysis to Study the Labour Capacity of Stavropol Krai Research Journal of Pharmaceutical, Biological and Chemical Sciences 7(4) 2183-6
[16] Zaitseva I, Popova M, Ermakova A, Bogdanova S and Rezenkov D 2016 Determination Prospects Of Development Labor Potential In Agriculture Stavropol Territory Based On Assessment His Condition Research Journal of Pharmaceutical, Biological and Chemical Sciences 7(3) 2592-5
[17] Zaitseva I, Malafeyev O, Strekopytov S, Ermakova A and Shlaev D 2018 Game-theoretical model of labour force training Journal of Theoretical and Applied Information Technology
[18] Zaitseva I, Malafeyev O, Strekopytov S et al. 2018 Mathematic model of regional economy development by the final result of labor resources AIP Conference Proceedings
[19] Zaitseva I, Ermakova A, Shlaev D, Shevchenko E and Lugovskoy S 2017 Workforce planning redistribution of the region's results Research Journal of Pharmaceutical, Biological and Chemical Sciences 8(1) 1862-6
[20] Zaitseva I, Malafeyev O, Poddubnaya N, Vanina A and Novikova E 2019 Solving a dynamic assignment problem in the socio-economic system Journal of Physics
[21] Zaitseva I, Malafeyev O, Marencchuk Y, Kolesov D and Bogdanova S 2019 Competitive Mechanism for the Distribution of Labor Resources in the Transport Objective Journal of Physics: Conference Series
[22] Malafeyev O, Lakhina J, Zaitseva I, Redinskikh N, Smirnova T and Smirnov N 2019 A mathematical model of production facilities location Journal of Physics
[23] Malafeyev O, Redinskikh N, Nemnyugin S, Kolesin I and Zaitseva I 2018 The optimization problem of preventive equipment repair planning AIP Conference Proceedings 1978 100013
[24] Malafeyev O, Awasthi A, Zaitseva I, Rezenkov D and Bogdanova S 2018 A dynamic model of functioning of a bank AIP Conference Proceedings
[25] Malafeyev O, Rylov D, Zaitseva I, Popova M and Novozhilova L 2017 Game-theoretic model of dispersed material drying process AIP Conference Proceedings
[26] Malafeyev O, Saifullina D, Ivaniukovich G, Marakhov V and Zaytseva I 2017 The model of multi-agent interaction in a transportation problem with a corruption component AIP Conference Proceedings
[27] Malafeyev O, Farvazov K, Zenovich O, Zaitseva I, Kostyukov K and Svechinskaya T 2018 Geopolitical model of investment power station construction project implementation AIP Conference Proceedings
[28] Zaitseva I, Malafeyev O, Kolesin I, Ermakova A and Shlaev D 2018 Modeling of the labour force redistribution in investment projects with account of their delay International Conference on Power, Control, Signals and Instrumentation Engineering pp 68-70
[29] Pichugin Y, Malafeyev O, Rylov D and Zaitseva I 2018 A statistical method for corrupt agents detection AIP Conference Proceedings
[30] Malafeyev O, Rylov D, Zaitseva I, Ermakova A and Shlaev D 2018 Multistage voting model with alternative elimination AIP Conference Proceedings