Time series models in population prediction

A. L. Osipov, email: alosip@mail.ru
E. V. Klepikova, email: e.v.klepikowa@gmail.com

Novosibirsk State University of Economics and Management

Аннотация. *The article examines linear and nonlinear time series models for predicting the population of the Russian Federation in the period from 1946 to 2020. The efficiency of population prediction using linear and nonlinear regression models is compared. Tools and time series models for predicting the population of the Russian Federation have been developed. Computational experiments were carried out, which showed the effectiveness of the proposed linear and nonlinear models.*

Ключевые слова: *Regression analysis; prediction, tools, population size, time series.*

Introduction

In modern conditions, the innovative way of development is considered as one of the main directions of economic modernization. It is almost impossible to compete in the traditional field of activity in the world markets, and only fundamentally new technologies, products, and management innovations will create a new competitive environment for industrial enterprises. In turn, the increase in innovation activity will ensure high rates of economic growth, increase the capitalization of the enterprise on the scale of production. In this regard, it is of particular importance to improve the efficiency of management of innovative activities of industrial enterprises. All of the above can not be done without a thorough analysis and forecasting of the population, which will solve the tasks of reformatting the Russian economy.

In the modern world, for the effective implementation of the socio-economic concept of the country's development, it is necessary to have information about the population. The state is concerned about solving problems related to demography, which is reflected in the Decree of the President of the Russian Federation on approval of the Concept of Demographic Policy of the Russian Federation for the period up to 2025 No. 1351 of 09.10.2007 [1].

The article examines the effectiveness of the developed mathematical models for the implementation of a retrospective population forecast for the period 1946-2020. At present, statistical prediction models are widely used in predicting population dynamics, among which regression modeling and the age translation method are particularly distinguished, which have a high area of use and an effective degree of reliability in the implementation of forecasts. In the practice of predicting population dynamics (PD), nonlinear models are most often encountered [2-4]. In [4], a nonlinear model of the species $Y = e^{a+bt}$ for predicting the population of the Russian Federation for the period from 2001 to 2017 was studied. The model was reduced to a linear form and the coefficients of the model were estimated, which turned out to be equal $a = 4{,}9688$, $b = 0{,}0003$. The article concludes that the model has a good predictive ability. Having carried out the corresponding calculations, we found that in this model the coefficient of determination is very small and is equal to 0,016244. The Fisher criterion was equal to 0,247679. When comparing the Fisher criterion with the table, the above model turned out to be insignificant, which makes it impossible to use it in practice. The obtained statistical characteristics of the model are associated with a small sample size.

The research methods used in the article are related to theoretical and practical aspects: modeling and forecasting; programming systems; theory of effective decision-making; regression and correlation analysis.

1. A mathematical model

The article examines the effectiveness of linear and nonlinear regression models for predicting the population of the Russian Federation based on experimental data presented in table 1. The following time series were studied: linear; exponential; logarithmic; power and polynomial.

Таблица 1

Initial data for analysis

Year	PD, million	Year	PD, million	Year	PD, million
1946	97,548	1971	130,252	1996	148,291638
1947	98,028	1972	130,934	1997	148,028613
1948	98,834	1973	131,687	1998	147,802133
1949	99,706	1974	132,434	1999	147,539426
1950	101,16	1975	133,217	2000	146,890128
1951	102,833	1976	134,092	2001	146,303611
1952	104,439	1977	135,026	2002	145,649334
1953	106,164	1978	135,979	2003	144,96365
1954	107,828	1979	136,922	2004	144,333586
1955	109,643	1980	137,758	2005	143,801046
1956	111,572	1981	138,483	2006	143,236582
1957	113,327	1982	139,221	2007	142,862692
1958	115,035	1983	140,067	2008	142,747535

1959	116,749	1984	141,056	2009	142,737196
1960	118,307	1985	142,061	2010	142,833502
1961	119,906	1986	143,033	2011	142,865433
1962	121,586	1987	144,156	2012	143,056383
1963	123,128	1988	145,386	2013	143,347059
1964	124,514	1989	146,505	2014	143,666931
1965	125,744	1990	147,665081	2015	146,267288
1966	126,749	1991	148,273746	2016	146,54471
1967	127,608	1992	148,514692	2017	146,804372
1968	128,361	1993	148,561694	2018	146,880432
1969	129,037	1994	148,355867	2019	146,78072
1970	129,66	1995	148,459937	2020	146,74859

2. Results of the study

Figure shows a graph of the population by year.

Рисунок. Population of the Russian Federation

A nonlinear model of the form was studied $Y = e^{a+bt}$. The statistical characteristics of the model were calculated using the tools [5-6]. The coefficients of the model are equal: $a = 4,694$; $b = 0,005$. The coefficient of determination was equal to 0,774, the Fisher criterion was equal to 249,443. The model was significant. Indicator $MAE = 7$, indicator $MAPE = 5{,}348$. The accuracy of the forecast calculated by the formula $T = 100 - MAPE$, was equal to 94,65%.

Next, a linear model of the form was investigated $Y = a + bt$. The statistical characteristics of the model were calculated using the tools [5-6]. The coefficients of the model are equal: $a = 109{,}166;\quad b = 0{,}635$. The coefficient of determination was equal to 0,795, the Fisher criterion was equal to 283,222. The model was significant. Indicator $MAE = 6{,}315$, indicator $MAPE = 4{,}865$. The accuracy of the forecast calculated by the formula $T = 100 - MAPE$, was equal to 95,14%.

Next, a power model of the form was studied $Y = at^b$. The statistical characteristics of the model were calculated using the tools [5-6]. The coefficients of the model are equal: $a = 85{,}005;\quad b = 0{,}1317$. The coefficient of determination was equal to 0,9297, the Fisher criterion was equal to 965,407. The model was significant. Indicator $MAE = 3{,}298$, indicator $MAPE = 2{,}533$. The accuracy of the forecast calculated by the formula $T = 100 - MAPE$, was equal to 97,467%.

Then a logarithmic model of the form was investigated $Y = a + b\ln t$. The statistical characteristics of the model were calculated using the tools [5-6]. The coefficients of the model are equal: $a = 78{,}69;\quad b = 16{,}256$. The coefficient of determination was equal to 0,916, the Fisher criterion was equal to 799,163. The model was significant. Indicator $MAE = 3{,}506$, indicator $MAPE = 2{,}766$. The accuracy of the forecast calculated by the formula $T = 100 - MAPE$, was equal to 97,234%.

Then a polynomial model of the form was investigated $Y = a + bt + ct^2$. The statistical characteristics of the model were calculated using the tools [5-6]. The coefficients of the model are equal: $a = 93{,}605;\quad b = 1{,}8473,\ c = -0{,}016$. The coefficient of determination was equal to 0,9833, the Fisher criterion was equal to 4298,257. The model was significant. Indicator $MAE = 1{,}643$, indicator $MAPE = 1{,}201$. The accuracy of the forecast calculated by the formula $T = 100 - MAPE$, was equal to 98,799%.

An additional set of statistical characteristics of time series models is presented in table 2.

Таблица 2

The statistical characteristics of the models

Модель	*MSE*	*RMSE*	R^2	*F*
$Y = e^{4,694+0,005t}$	59,169	7,692	0,774	249,443
$Y = 109,166 + 0,635t$	48,675	6,977	0,795	283,222
$Y = 78,69 + 16,256 \ln t$	19,879	4,459	0,916	799,163
$Y = 85,005 \cdot t^{0,1317}$	16,512	4,063	0,929	965,407
$Y = 93,605 + 1,8473t - 0,016t^2$	3,987	1,997	0,983	4298,257

Of the models shown in Table 2, the best statistical and predictive properties are provided by the polynomial model.

Conclusion

The developed linear and nonlinear regression models based on time series have demonstrated high efficiency, which allows them to be used for practical applications. Linear and nonlinear time series models showing high predictive power are proposed and investigated. The effectiveness of the developed statistical methods based on linear and nonlinear models was tested. Instrumental software tools for predicting the population of the Russian Federation based on the developed models have been developed.

References

1. Decree of the President of the Russian Federation No. 1351 of 09.10.2007 "On approval of the Concept of demographic policy of the Russian Federation for the period up to 2025" // Collection of Legislation of the Russian Federation. - 2007. - No. 42. - St. 5009.
2. Pil, E. A. Forecasting the population for the G7 countries / E. A. Pil // Almanac of Modern Science and Education. - 2011. - No. 7. - pp. 142-144.
3. Tarasova, O. V. On the question of the application of mathematical methods for the study of demographic processes / O. V. Tarasova // Theory and practice of social development. - 2015. - No. 24. pp. 186-188.
4. Kochegarova, O. S. Construction of a statistical model of the total population of the Russian Federation on the basis of a retrospective forecast / O. S. Kochegarova, Yu. V. Lazhauninkas // Modern studies of social problems. - 2017. - Volume 8. - No. 6. - pp. 56-66.

5. Osipov, A. L. Mathematical and instrumental methods of decision-making in economic research / A. L. Osipov, V. P. Trushina // Science of Krasnoyarsk. - Volume 8. - No. 4-2. - 2019. - pp. 58-61.

6. Osipov, A. L. Models of regression analysis for forecasting the population of the Russian Federation / A. L. Osipov, E. A. Rapotsevich // Science of Krasnoyarsk. -Volume 9. - No. 3-4. - 2020. pp. 116-121.