Two Prediction Models for Some Economic Indicators of the Russian Arctic Zone

Sergey Chernogorskiy, Konstantin Shvetsov, and Vladimir Parkhomenko(✉)

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia
m39098@mail.ru, shvetsov@inbox.ru, parhomenko.v@gmail.com

Abstract. The goal of this paper is to analyze and predict time series which reflect the dynamics of the gross regional product, the employed population and the total working-age population of the Russian Arctic zone. These tasks are very important to plan the development of the Russian Arctic zone. The ARIMA and VAR prediction models are developed. The VAR model shows better forecasting properties than the ARIMA model for the datasets discussed in the paper.

Keywords: ARIMA model · VAR model · Prediction · Russian arctic zone

1 Introduction

A wide variety of social, business, and economic tasks requires an analysis of time series. An application of two prediction models is discussed in the paper. They are the autoregressive integrated moving average (ARIMA) and the vector autoregressive (VAR) models.

A comparison of one-dimensional and multi-dimensional models is elaborated in [1,2]. Some forecasting results of various economic time series with the help of one-dimensional and two-dimensional autoregressive (vector) models, dynamic models with autoregressive-distributed differences, and the adjusted variations of the models, which coefficients are changing with time, are carried out in the papers.

Paper [3] contains the comparison of various TS and DS time series, where abbreviations TS and DS mean "trend stationary" and "difference stationary", respectively. The main difference between them is included in the response from local extremal values. In the TS case the extremal values do not influence on the rest part of the time series significantly. However in the DS case there is an opposite situation. As a result this leads to the different trends forms. The DS time series look more preferable when the forecasted time series was classified

The paper is based on research carried out with the financial support of the grant of the Russian Scientific Foundation (Project No. 14-38-00009, The program-targeted management of the Russian Arctic zone development). Peter the Great St. Petersburg Polytechnic University.

Y. Bi et al. (eds.), *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016*, Lecture Notes in Networks and Systems 15, DOI 10.1007/978-3-319-56994-9_25

as both DS and TS time series. In Subsect. 3.1 we use the Box-Jenkins method [4], which works with TS time series and assumes a transforming procedure of DS time series into TS type.

A dynamic econometric model describing the economic development of the Russian Arctic zone (RAZ) is obtained in [5,6]. A set of regional indicators is introduced in [7]. The authors take into account the influence of both economic and non-economic features on the RAZ sustainable development.

A review of copula models for economic time series is proposed in [8]. Copula-based multivariate models allow to specify the models for the marginal distributions separately from the dependence structure which links these distributions to form a joint distribution. As a result, this allows the researcher to be free from considering only existing multivariate distributions. Andrew J. Patton surveys estimation and inference methods and goodness-of-fit tests for such models, as well as empirical applications of these copulas for economic and financial time series.

A review and comparative study of time series forecasting for nonlinear and non-stationary processes is proposed in [9]. Authors make a summary comparison of the performance of different prediction models in prior empirical studies which include classic autoregressive models, SVM models, neural networks and neuro-fuzzy type models and other machine learning techniques.

A review of thinning-based models in the analysis of integer-valued time series is provided in [10]. In particular the authors pay attention to models obtained as discrete counterparts of conventional autoregressive moving average and bilinear models, and based on the concept of thinning. They also review manuscripts of the most relevant thinning operators proposed in the analysis of univariate and multivariate integer-valued time series with either finite or infinite support. It's proved in the paper that models with thinning operations are useful in the analysis of many real-world applications ranging from economy and finance to medicine.

The main goal of this paper is to analyse and predict time series which reflect the dynamics of the gross regional product (GRP), the employed population (EMP) and the total working-age population (ACTIVE) of the RAZ. The motivation of the investigation is a lack of working-age population, which leads to the lack of the EMP in the Russian Northern regions within the RAZ. In [11] the authors make a preliminary analysis of these time series.

Datasets are described in Sect. 2. Two new prediction models for the economic indications are developed in Sect. 3. Section 4 is devoted to discussion of the results.

2 Datasets

The RAZ consists of 9 regions [12], three of which are partially included. The data originally provided by [13] is not separated for these three regions, thus the results of our computations slightly exceed the RAZ borders. The aggregated datasets of GRP, EMP and ACTIVE are given in Figs. 1, 2, and 3, respectively. These time series reflect quarterly data from 1995 to 2014 year.

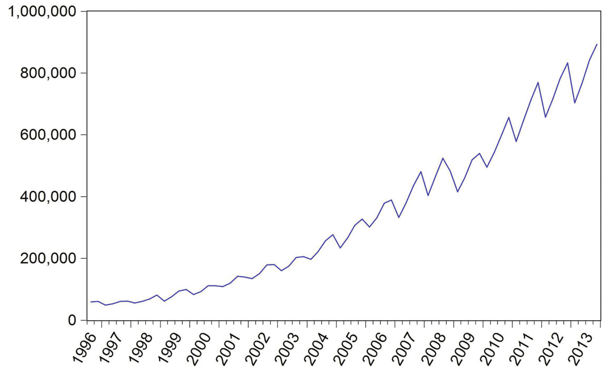

Fig. 1. Time series of GRP 1995–2014, millions of RUR

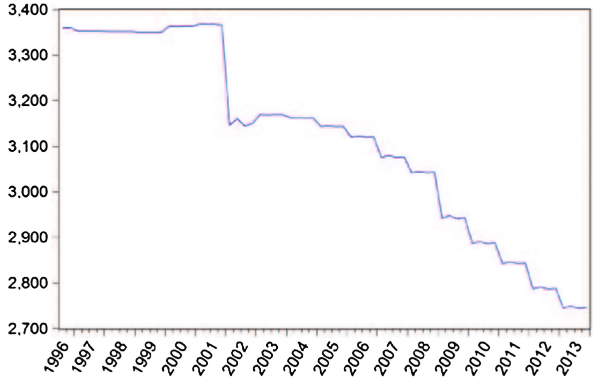

Fig. 2. Time series of ACTIVE 1995-2014, thousands of people

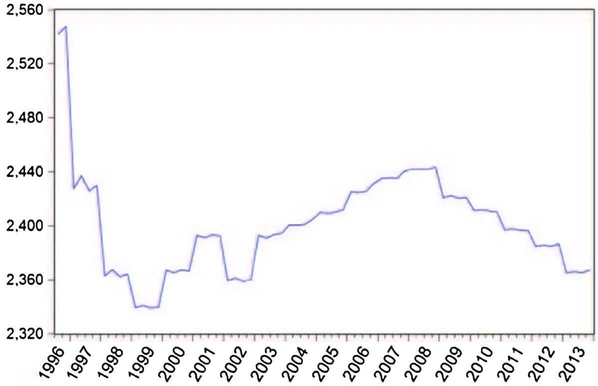

Fig. 3. Time series of EMP 1995-2014, thousands of people

To predict the time series we need to deal with following subtasks:

- Correlogram analysis (ACF and PACF functions) of the time series.
- Testing the time series for unit roots.
- Identifying the ARIMA and VAR models.
- Estimation of the properties of the ARIMA and VAR models.
- Checking the models accuracy.
- Checking the models stability.
- Comparative analysis of the models forecasting quality.
- Applying the models for forecasting.

3 Two New Prediction Models

The visual analysis of the time series in Figs. 1, 2 and 3 leads to the fact that the time series are not stationary. The correlogram analysis lets us conclude, that they are DS time series. To prove this fact we need to accomplish some tests on defining of the unit roots with the following criteria: the augmented Dickey–Fuller test, the Phillips–Perron test, the DF–GLS test, the KPSS test with a constant and a trend and the Ng–Perron test with a constant. The results of the tests for the unit root for testing of the stationarity of the time series are presented in Tables 1 and 4, where abbreviation "R" stands for "Rejected".

Table 1. Test results

Test no.	GRP		EMP		ACTIVE	
	Initial hypothesis					
	DS	TS	DS	TS	DS	TS
1	Not R		Not R		Not R	
2	Not R		Not R		Not R	
3	Not R		Not R		R	
4	Not R		Not R		R	
5	Not R		Not R		Not R	
6		R		R		R
7	Not R		Not R		Not R	

Despite of the fact that the DS-hypothesis of the working-age population is rejected by tests 3 and 4, we classify this time series as DS-type because of the test 6 and the type of correlogram. With the analysis we can classify the time series as DS time series.

3.1 ARIMA Model

Let us create a forecasting seasonal model ARIMA$(p, d, q)(P_s, D_s, Q_s)$ for the analysed time series. The parameters are interpreted as follows: p — the autoregressive parameter, q — the moving average parameter, d — the difference of time series parameter, P_s — the seasonal autoregressive parameter, Q_s — the seasonal moving average parameter, D_s — the seasonal difference of time series parameter. Estimating of the parameters p, q, P_s, Q_s of the model. There were 36 models analyzed where the p and q were varying in the interval from 0 to 2, whereas the P_s and Q_s were taking values 0 or 4.

The parameters of the ARIMA models for the socio-economic indicators of the RAZ are presented in Table 2.

Table 2. Parameters of the ARIMA models

Economic indicator	p, q	P_s, Q_s	BIC criterion
GRP	$p = q = 1$	$P_s = 4$, $Q_s = 4$	−3.201295
EMP	$p = q = 1$	$P_s = 4$, $Q_s = 4$	7.691584
ACTIVE	$p = q = 1$	$P_s = 4$, $Q_s = 4$	9.788278

The ARIMA(1, 1, 1)(4, 1, 4) form is the selected model for each of three time series. The model is presented in (1), (2), and (3) for GRP, EMP, and ACTIVE indicators respectively.

$$\begin{aligned} x_t = {} & 1,807x_{t-1} - 0,807x_{t-2} + 0,633x_{t-4} \\ & - 1,145x_{t-6} + 0,367x_{t-8} - 0,663x_{t-9} \\ & - 0,296x_{t-10} + e_t - 0,948e_{t-1} + 0,929e_{t-4} \\ & - 0,8816e_{t-5} \end{aligned} \tag{1}$$

$$\begin{aligned} x_t = {} & 1,935x_{t-1} - 0,935x_{t-2} + 0,993x_{t-4} \\ & - 1,921x_{t-6} + 0,007x_{t-8} - 0,014x_{t-9} \\ & + 0,007x_{t-10} + e_t - 0,383e_{t-1} + 0,933e_{t-4} \\ & - 0,358e_{t-5} \end{aligned} \tag{2}$$

$$\begin{aligned} x_t = {} & 1,999x_{t-1} - 0,999x_{t-2} + 0,856x_{t-4} \\ & - 1,711x_{t-6} + 0,144x_{t-8} - 0,288x_{t-9} \\ & + 0,144x_{t-10} + e_t - 0,069e_{t-1} - 0,032e_{t-4} \\ & + 0,002e_{t-5} \end{aligned} \tag{3}$$

3.2 VAR Model

The order of the VAR model is determined by the order of the lag. The autoregressive model VAR(p) has the general form (4) for k variables and p lags.

$$\begin{cases} x_{t1} = \alpha_1 + \alpha_{11}^{[1]} x_{t-1,1} + \ldots + \alpha_{1k}^{[1]} x_{t-1,k} \\ + \alpha_{11}^{[2]} x_{t-2,1} + \ldots + \alpha_{1k}^{[2]} x_{t-2,k} + \ldots \\ + \alpha_{11}^{[p]} x_{t-p,1} + \ldots + \alpha_{1k}^{[p]} x_{t-p,k} + \epsilon_{t1}; \\ \\ x_{t2} = \alpha_2 + \alpha_{21}^{[1]} x_{t-1,1} + \ldots + \alpha_{2k}^{[1]} x_{t-1,k} \\ + \alpha_{21}^{[2]} x_{t-2,1} + \ldots + \alpha_{2k}^{[2]} x_{t-2,k} + \ldots \\ + \alpha_{21}^{[p]} x_{t-p,1} + \ldots + \alpha_{2k}^{[p]} x_{t-p,k} + \epsilon_{t2}; \\ \ldots \\ x_{tk} = \alpha_k + \alpha_{k1}^{[1]} x_{t-1,1} + \ldots + \alpha_{kk}^{[1]} x_{t-1,k} \\ + \alpha_{k1}^{[2]} x_{t-2,1} + \ldots + \alpha_{kk}^{[2]} x_{t-2,k} + \ldots \\ + \alpha_{k1}^{[p]} x_{t-p,1} + \ldots + \alpha_{kk}^{[p]} x_{t-p,k} + \epsilon_{tk}. \end{cases} \tag{4}$$

where $\alpha, \ldots, \alpha_k$ — free parameters, $\alpha_{ij}^{[1]}, \ldots, \alpha_{ij}^{[p]}$ — autoregressive parameters, $\epsilon_{t1}, \ldots, \epsilon_{tk}$ — mutually uncorrelated "white noise". The vector-matrix form of (4) see, please, in (5).

$$\vec{X}_t = \vec{\alpha} + A^{[1]} \vec{X}_{t-1} + \ldots + A^{[p]} \vec{X}_{t-p} + \vec{\epsilon}_t \tag{5}$$

For the further modeling we use the same time series, which are applied for creating the ARIMA model. We obtained that these time series are DS time series. So, for creating of the VAR model we take the first differences of the time series. In this case we get the stationary VAR model. In order to determine the number of lags and their sufficiency the Schwartz, BIC and Hennan-Quinn criteria are used. They have shown sufficiency of 4 lags, hence, we use the model VAR(4). Estimation of the VAR(4) parameters was carried out. The stationary VAR(4) model is represented by System (6) of interdependent equations,

Table 3. Comparison of the prediction models

Crit.	ARIMA			VAR		
	GRP	EMP	ACTIVE	GRP	EMP	ACTIVE
R^2	0.9976	0.9311	0.9778	0.9959	0.9466	0.9853
R^2_{adj}	0.9975	0.9269	0.9765	0.9948	0.9320	0.9813
RMSE	4.2150	10.9337	26.574	3.5720	9.0361	22.1450
MAE	3.4070	6.7034	9.3070	2.8873	5.5400	7.7558
MAPE	0.2780	0.2792	0.2943	0.2356	0.2307	0.2453

where GRP, EMP and ACTIVE of the RAZ are denoted by x_{t1}, x_{t2}, and x_{t3} respectively. The standard errors are given in the round brackets.

$$\begin{cases} x_{t1} = \underset{(4{,}13)}{5,5} + \underset{(0{,}13)}{1,51}\, x_{t-1,1} - \underset{(0{,}13)}{0,51}\, x_{t-2,1} \\ + \underset{(0{,}09)}{0,16}\, x_{t-3,1} + \underset{(0{,}10)}{0,54}\, x_{t-4,1} - \underset{(0{,}12)}{1,00}\, x_{t-5,1} \\ + \underset{(0{,}12)}{0,307}\, x_{t-6,1} + \underset{(0{,}97)}{2,08}\, x_{t-4,3} - \underset{(0{,}97)}{2,08}\, x_{t-5,3}; \\ \\ x_{t2} = \underset{(0{,}34)}{0,63} + \underset{(0{,}14)}{1,77}\, x_{t-1,2} - \underset{(0{,}14)}{0,77}\, x_{t-2,2} \\ - \underset{(0{,}08)}{0,125}\, x_{t-3,2} + \underset{(0{,}08)}{0,69}\, x_{t-4,2} - \underset{(0{,}10)}{0,97}\, x_{t-5,2} \\ + \underset{(0{,}10)}{0,41}\, x_{t-6,2} - \underset{(0{,}08)}{0,18}\, x_{t-4,3} + \underset{(0{,}07)}{0,37}\, x_{t-5,3} \\ - \underset{(0{,}07)}{0,185}\, x_{t-6,3}; \\ \\ x_{t3} = \underset{(0{,}71)}{1,175} + \underset{(0{,}01)}{0,04}\, x_{t-4,1} - \underset{(0{,}02)}{0,08}\, x_{t-5,1} \\ + \underset{(0{,}02)}{0,04}\, x_{t-6,1} - \underset{(0{,}15)}{0,307}\, x_{t-4,2} + \underset{(0{,}15)}{0,307}\, x_{t-5,2} \\ + \underset{(0{,}18)}{1,726}\, x_{t-1,3} - \underset{(0{,}18)}{0,726}\, x_{t-2,3}; \end{cases} \tag{6}$$

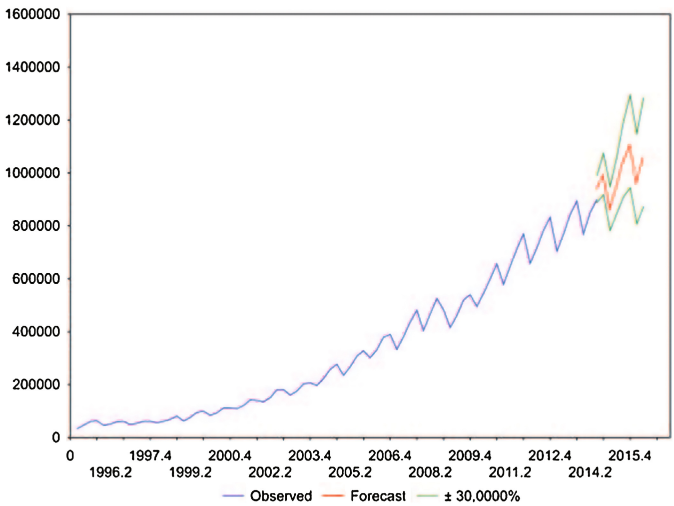

Fig. 4. The GRP prediction till the end of 2015 year, millions of RUR

The test shows that the residuals of the equations are mutually uncorrelated. Moreover, the residuals are normally distributed. Values of the selected determination coefficients for both of the equations indicate a high quality of approximation.

For comparing of the forecasting properties of the ARIMA and VAR models the determination coefficient R^2, the adjusted determination coefficient R^2_{adj}, the square root of mean standard error RMSE, the mean absolute error MAE and the mean absolute percentage error MAPE were calculated. The results are given in Table 3.

According to Table 3 the forecasting properties are better for the VAR model. It can be explained with regard of the mutual influence of the time series in the model. As an example we make a forecast of the GRP, EMP, and ACTIVE with the VAR model till the end of 2015 year, see Figs. 4, 5, and 6, respectively.

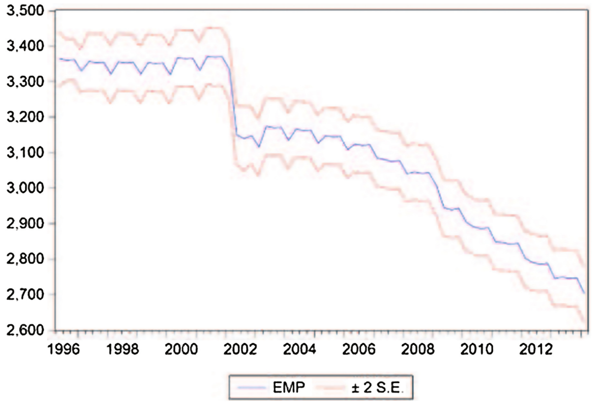

Fig. 5. The EMP prediction till the end of 2015 year, thousands of people

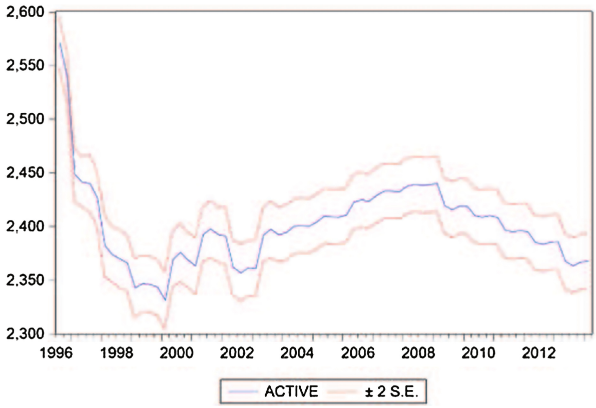

Fig. 6. The ACTIVE prediction till the end of 2015 year, thousands of people

Table 4. Comparison of tests and critical values

No	Test	GRP	EMP	ACTIVE	t_{crit}
		t_ϕ	t_ϕ	t_ϕ	
1	Augmented Dickey-Fuller test with a constant	2.784837	−2.048908	−0.077353	−2.905519
2	Augmented Dickey-Fuller test with a constant and a trend	−1.223917	−1.987994	−2.529907	−3.478305
3	Augmented Dickey-Fuller test	2.326245	−0.253793	−2.443356	−1.94526
4	Phillips-Perron test	5.317244	−1.922118	−3.167592	−1.945024
5	DF-GLS test	0.905988	−0.158285	1.254898	−1.94526
6	KPSS with a constant and a trend	0.198398	0.123711	0.196249	0.216
7	Ng-Perron test with a constant	44.3941	80.6184	60.6747	3.17

4 Conclusion

The investigation is focused on the analysis of the gross regional product, employed population and working-age population of the RAZ for the period of 1995–2014 (quarterly data). The analysis in the previous sections shows that all the time series are of DS type. The forecasting model ARIMA$(1,1,1)(4,1,4)$ is obtained. The economic indicators model is created. The models enables to find the response to the changes of all the variables. The comparative analysis shows better forecasting properties of the VAR(4) model. The prediction for the time series till the end of 2015 is presented.

Acknowledgment. The paper is based on research carried out with the financial support of the grant of the Russian Scientific Foundation (Project No. 14-38-00009, The program-targeted management of the Russian Arctic zone development). Peter the Great St. Petersburg Polytechnic University.

References

1. Clements, M.P., Hendry, D.F.: Forecasting with difference-stationary and trend-stationary models. Econom. J. **4**(1), 1–19 (2001)
2. Stock, J.H., Watson, M.W.: A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. National Bureau of Economic Research, Working Paper 1998, June 6607
3. Entov, R.M., Nosko, V.P., Yudin, A.D., Kodochnikov, P.A., Ponomarenko, S.S.: Prediction problems of some macroeconomic indicators. Institute for the Economy in Transition, Technical report (2002) (in Russian). http://www.iep.ru/files/text/working_papers/46.ZIP

4. Box, G.E., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: Forecasting and Control, vol. 734. Wiley, New York (2011)
5. Didenko, N., Skripnuk, D.: Modeling of sustainable economic development of the Russian Arctic regions with the system of econometric equations. In: Ivanter, V. (ed.) Strategic Priorities of the Development of the Russian Arctic, pp. 63–78. Nauka, Moscow (2014). (in Russian)
6. Didenko, N., Skripnuk, D.: The impact of energy resources on social development in Russia. In: Energy Production and Management in the 21st Century: The Quest for Sustainable Energy, vol. 1, pp. 151–159 (2014). http://www.witpress.com/elibrary/wit-transactions-on-ecology-and-the-environment/190/25672
7. Kozlov, A., Gutman, S., Zaychenko, I.: Theoretical and methodological basics of forming of the regional development indicators for the Russian Arctic zone. In: Ivanter, V. (ed.) Strategic Priorities of the Development of the Russian Arctic, pp. 103–112. Nauka, Moscow (2014). (in Russian)
8. Patton, A.J.: A review of copula models for economic time series. J. Multivariate Anal. **110**, 4–18 (2012). http://www.sciencedirect.com/science/article/pii/S0047259X12000826
9. Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z.J., Bukkapatnam, S.T.: Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. IIE Trans. **47**(10), 1053–1071 (2015). http://www.tandfonline.com/doi/full/10.1080/0740817X.2014.999180
10. Scotto, M.G., Weiss, C.H., Gouveia, S.: Thinning-based models in the analysis of integer-valued time series: a review. Stat. Model. **15**(6), 590–618 (2015). http://smj.sagepub.com/content/15/6/590
11. Chernogorskiy, S., Shvetsov, K.: Problems of modeling and forecasting of the economic development of the Russian Arctic zone. In: Ivanter, V. (ed.) Strategic Priorities of the Development of the Russian Arctic, pp. 121–134. Nauka, Moscow (2014). (in Russian)
12. The Rossiiskaya Gazeta Website. (in Russian). http://www.rg.ru/2014/04/24/arktika-site-dok.html
13. The Russian Federation Federal State Statistics Service Website. http://gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/accounts