

The Analysis of the Selection Criteria of the Optimal Model of the Dynamics in the Case of Extrapolative Forecasting for Short Time Series

E A Polishchuk¹

¹V.I. Vernadsky Crimean Federal University, Simferopol, Russia

E-mail: pea.znu@mail.ru

Abstract. Selection of the optimal dynamics model in the case of extrapolation forecasting by short time series. Article describes the technology of choosing the optimal model of the dynamics of key indicators of the youth labour market of the Russian Federation using extrapolation forecasting by short time series. It is done based on a small amount of retrospective data (2015-2018). Using SPSS data analysis package five trend models were constructed for each indicator, for which basic explanations were prescribed (based on the nature of the traceable dynamics, the considered indicators are divided into separate types: indicators, in the growth dynamics of which a slowdown is observed; indicators, in the growth dynamics of which an acceleration is observed; indicators, in the recession dynamics of which a slowdown is observed; indicators, in the recession dynamics of which an acceleration is observed; linear and polynomial (quadratic) models are competing to approximate the dynamics of the considered indicators of all types; to approximate short time series for extrapolation purposes, polynomial, including quadratic models cannot be used in most cases, because the nature of their behavior is unpredictable even in the short-term forecast), allowing identification of models suitable for each indicator. They allowed subsequently calculating the results of their point-by-point forecasts for 2020. It is concluded that it is important to select the optimal model that will ensure the most probable forecast, while taking into account the fact that it is not always suitable for short-term forecasting.

1. Introduction

The growth of the socio-economic potential of the Russian Federation depends directly on the labour activity of young people, who are a key source of replenishment of resources in the labour market. Therefore, in the context of the strategic development of the state, not only issues related to the study of the characteristics of young people in the labour market for involvement in labour activities in order to develop its labour potential, but also to predict the target indicators of the labour market of young people for the selection of a model that can provide the most likely scenario of development of all possible alternatives are becoming relevant.

One of the key problems of the present study is the small amount of retrospective data (2015-2018) which necessitated the further development of model forecasting tools exclusively taking into account the short-term perspective which differs from the medium-term and long-term higher degree of objective certainty within a relatively short time interval.

2. Relevance, scientific significance

Issues short-term forecasting of selected indicators of development of socio-economic systems and selection of the optimal model were considered in scientific works of many foreign and domestic researchers: Jaumotte F., Lall S., Papageorgiu Ch. [6], Jouini T. [7], Lukoseviciute K., Baubliene R., Howard D., Ragulskis M. [13], Demidova L. Pylkin A., Skvortsov S., Skvortsova T. [2], Zaporozhtsev, I., Sereda, A-V. [20], Ibragimova, Z. [5], Kolmakov I., Koltsov A., Domozhakov M. [8], Kolomeiko M. [9] and others [1, 10-12, 14-19].

A significant contribution to this topic was made by the work prepared under the guidance of professor Shoumetov V.G. in which the authors, on the basis of specific examples, described the process of forecasting by time series, especially short; argued the relevance of the selection of a specific model that can provide the most likely forecast, etc. [18, p. 133-144].

Recognizing the high scientific importance of the analyzed works it should be noted that the forecasts of individual indicators of the labour market have not been properly applied taking into account the promising directions and needs for their effective development.

The purpose of the study is the choice of the optimal model of the dynamics in the case of extrapolative forecasting for short time series, the scientific novelty of which consists in comparing the dynamics of values of the confidence interval for the competing models.

3. Practical significance, research results

Table 1 presents the dynamics of key indicators of the youth labour market in the period from 2015 to 2018 [3-4] on the basis of which their projected values will be calculated in the short term.

Table 1. Parameters of a number the youth labour market indicators in the Russian Federation.

| Name of indicator | Years | | | |
|---|-------|------|------|------|
| | 2015 | 2016 | 2017 | 2018 |
| Youth employment rate (Y_{ey}), % | 57,0 | 58,1 | 59,0 | 59,7 |
| Youth unemployment rate (U_y), % | 15,5 | 14,9 | 14,2 | 13,5 |
| The unemployment registered rate of young people (U_{ty}), % | 3,1 | 2,9 | 2,6 | 2,3 |
| The tension coefficient in the labour market of young people (C_{ty}), units. | 4,3 | 3,8 | 3,4 | 3,2 |
| The coefficient of real youth employment rate (C_{rye}), units. | 0,65 | 0,68 | 0,70 | 0,72 |
| The coefficient of employment rate of young people who have received services that promote competitiveness (C_{ecy}), unts. | 0,22 | 0,26 | 0,33 | 0,39 |

The information presented in table 1 reflects the positive development of the youth labour market: increase in the level of indicators, the desired trend of which is growth; decrease in the level of indicators, the desired trend of which is a decrease.

Despite this, the current values of some of them do not correspond to their thresholds, for example, the coefficient of tension in the youth labour market exceeds the threshold by 3 times, the youth unemployment rate (according to the ILO methodology) – by 1.8 times, etc.

According to the information presented in the table, five trend models (linear, logarithmic, polynomial (quadratic), power, exponential) for each indicator were constructed using the SPSS data analysis package.

On the received models it is necessary to register the following explanations:

1. Based on the nature of the observed dynamics, the indicators should be divided into 4 types: type A must include the indicators in the dynamics of growth which is slowing down (Y_{er} , C_{rye}); the type B should include indicators in the growth dynamics of which there is an acceleration (C_{ecy});

to the type B – indicators in the dynamics of the decline of which there is a slowdown (Cty);
to the type of G – indicators, in the dynamics of the decline of which there is an acceleration (uy, ury).

2. Estimation of the approximating functions allowed to conclude that the following models are competing for the dynamics modeling:

for indicators of type A (Yer, Crye) logarithmic, polynomial (quadratic) and power;

for indicators of type B (Cecy) exponential and polynomial (quadratic);

for indicators of type B (Cty) logarithmic, polynomial (quadratic), exponential and power.

3. Linear and polynomial (quadratic) models are competing to approximate the dynamics of the considered indicators not only of the type G (uy, ury), but also of all other types.

4. For the approximation of short time series for extrapolation polynomial including quadratic models in most cases can not be used because the nature of their behaviour is unpredictable even in the short-term forecast.

5. Along with the coefficient of determination, the criterion for choosing the specification of the model is the level of statistical significance (p-level) of the Fisher criterion, as well as the value of 95 or 90% confidence interval and its dependence on the forecast horizon.

Taking into account the described explanations, it can be concluded that the power and logarithmic models are the most suitable for the indicators ury, Crye and Cty, for the uy – linear model, the ury and the Cecy – exponent.

Table 2 presents the basic characteristics of the observed indicators of the youth labour market of the Russian Federation.

Table 2. The basic components of the youth labour market indicators: the appearance, quality criteria models, the predicted values.

| Indicator | Model | R^2 | F | p | Predicted values for 2020 |
|---------------------------|----------------------------------|--------|----------|-------|---------------------------|
| Y _{er} , % | $x = 56,922246 t^{0,033147}$ | 0,9906 | 211,1291 | 0,004 | 60,41 |
| | $x = 56,915136 + 1,931829 \ln t$ | 0,9893 | 184,8496 | 0,005 | 60,38 |
| U _y , % | $x = 16,200000 - 0,670000 t$ | 0,9986 | 1496,333 | 0,000 | 12,18 |
| | $x = 3,480942 e^{-0,100468 t}$ | 0,9839 | 122,5177 | 0,008 | 1,91 |
| C _{ty} , units. | $x = 4,333681 t^{-0,215581}$ | 0,9913 | 229,8544 | 0,004 | 2,95 |
| | $x = 4,315446 - 0,806086 \ln t$ | 0,9957 | 468,2640 | 0,002 | 2,87 |
| C _{rye} , units. | $X = 0,648610 t^{0,072393}$ | 0,9929 | 281,4866 | 0,003 | 0,74 |
| | $x = 0,648230 + 0,049427 \ln t$ | 0,9901 | 201,4179 | 0,004 | 0,74 |
| C _{ecy} , units. | $x = 0,179629 e^{0,192520 t}$ | 0,9980 | 1019,003 | 0,001 | 0,57 |

The analysis of table 2 data allowed us to conclude that in all the models presented t is a discrete variable counted from the beginning of the studied time range, i.e. in 2015 it takes the value of 1, 2016 – 2, etc. This purpose of the time variable is provided in the Curve Estimation Module of the SPSS

Base software package, while the free coefficient of the power and logarithmic models have a clear interpretation, i.e. the calculated value of the indicator at the initial time.

A clear interpretation is also given to the regression coefficient sign of the presented models: positive – in case of growth of the indicator (for the youth employment level (Y_{er}) and the coefficient of real youth employment (C_{rye})) and negative – in case of its decline (for the coefficient of tension in the youth labour market (C_{ly})).

Less obvious is the interpretation of the regression coefficient of the power and logarithmic models characterizing the rate of growth/decline: as the absolute value of the coefficient increases, the rate of growth/decline increases.

In the course of the study, it was found that the parameters of linear and exponential models are most clearly interpreted, especially when the zero value of a discrete time variable is combined with the beginning or end of the study period. If the variable $t=0$ in the initial year of the period (2015), the free coefficient is interpreted as the calculated value of the indicator at the initial time, if the variable $t=0$ in the final year of the period (2015) – at the final time. Thus, for the youth unemployment rate (U_y), the free coefficient of the linear model in the first case is 15.53, which is very close to the empirical value of 15.5%, in the second case – 13.52, which is also close to the empirical value of 13.5%.

For the level of registered youth unemployment (U_{ry}), the free coefficient of the exponential model in the first case is 3.15, which is very close to the empirical value of 3.1%, in the second case – 2.32, which is also close to the empirical value of 2.3%. Finally, for the employment rate of young people who have received services that enhance competitiveness (C_{ecy}), the free coefficient of the exponential model in the first case is 0.218, which is very close to the empirical value of 2.2 units, and in the second case -0.388, which is also close to the empirical value of 0.39 units.

All this indicates that the transfer of the time variable reference does not entail a change in the quality characteristics of the linear and exponential models, and regression coefficients do not change at the time variable, in turn, their interpretation is very clear (the regression coefficient of the linear model is interpreted as the average annual growth of the indicator, and the regression coefficient of the exponential model – as the average annual growth rate coefficient of the indicator).

The negative value of the regression coefficient -0,100468 of the exponential model of the dynamics of the level of registered youth unemployment is interpreted as the average annual growth rate coefficient of the U_{ry} , i.e. the average annual growth rate of registered unemployment is about 90%. Accordingly, the regression coefficient -0.670000 of the linear model of the dynamics of the level of General youth unemployment is interpreted as a negative average annual increase in total unemployment -0.67%.

Another explanation follows from the analysis of the results of modelling the dynamics of indicators by quadratic models. In the case of modelling the dynamics of the youth employment rate (Y_{er}) by a quadratic polynomial, the width of 95% of the confidence interval was equal to zero in the entire investigated time range, as a consequence of the fact that empirical observations are accurately described by the quadratic function. However, this does not mean that a quadratic model is the best option for forecasting. Evidence of this is implemented in the simulation source data defined with accuracy to one decimal place after the decimal point and ranging within $Y_{er} \pm 0,05$.

4. Conclusion

In a separate comment, the results of the point forecasts of the indicators under consideration for 2020 need to be given, on the basis of which it is established that for competing models the most likely forecast values of the analyzed indicators are close to each other, and their average values should be used in further calculations.

Summing up the results of the study, it should be noted that the results confirm the importance of selecting the optimal model [18, p. 133-134] which allows to provide the most likely forecast, while it is not always suitable for short-term forecasting.

References

- [1] Chekulina T, Lyaskovskaya O, Barbashova E 2017 Methods of dispersion analysis in the research of social and economic processes *OrelIGIET Bulletin* **3(41)** 68–73
- [2] Demidova L, Pylkin A, Skvortsov S, Skvortsova T 2015 Hybrid short time series forecasting models *Hot-line-Telecom* (Moscow)
- [3] GKS. Labor force, employment and unemployment in Russia – 2018 Homepage http://www.gks.ru/bgd/regl/b18_61/Main.htm
- [4] GKS. Russian Statistical Yearbook – 2018 Homepage http://www.gks.ru/bgd/regl/b18_13/Main.htm
- [5] Ibragimova Z 2017 Forecasting the Poverty Levels in the Russian Federation Based on Time Series Methods *The Review of Economy, the Law and Sociology* **4** 44–48
- [6] Jaumotte F, Lall S, Papageorgiu Ch 2013 Rising Income Inequality: Technology, or Trade and Financial Globalization? *IMF Economic Review* **61(2)** 271–309
- [7] Jouini T 2015 Efficient multistep forecast procedures for multivariate time series *Journal of Forecasting* **34(7)** 604–618
- [8] Kolmakov I, Koltsov A, Domozhakov M 2014 Application of the short-term forecasting system to monitor the indicators of the state program of the Russian Federation «Development of Science and Technology» *Intellectualno-analiticheskie metody, modeli i tekhnologii v upravlenii ekonomicheskimi i sotsialnymi sistemami* **4** 105–118
- [9] Kolomeiko M 2014 Analysis of the main trends and forecasting the level of poverty of the population of the astrakhan region on the basis of time series methods *Statistics and Economics* **4** 154–157
- [10] Kopytov V, Tebueva F 2009 Forecasting of exceeding situation of technogenic nature on short temporary rows *Scientific and educational problems of the civil protection* **2** 33–36
- [11] Laikam K 2002 Short-term forecast of differentiation of the income of the population *Society and economy* **8(9)** 68–77
- [12] Lukashin Yu 2003 Adaptive methods for short-term time series forecasting: manual *Finances and statistics* (Moscow)
- [13] Lukoseviciute K, Baubliene R, Howard D, Ragulskis M 2018 Bernstein polynomials for adaptive evolutionary prediction of short-term time series *Applied Soft Computing* **65** 47–57
- [14] Mandel' A 2004 Method of analogs in prediction of short time series: an expert-statistical approach *Automation and remote control* **4** 143–153
- [15] Semakov S, Semakov A 2015 Forecasting and sales management in retail chains *Fismathlit* (Moscow)
- [16] Shirnaeva S 2009 Peculiarities of the system of simultaneous equations for modeling the stabilization processes in the RF economy *Vestnik of Samara State University of Economics* **5(55)** 138–142
- [17] Shirnaeva S 2011 Short-term forecasting of factors of macro economical stabilization of Russian economy *Vestnik of OSU* **1(120)** 133–137
- [18] Shoumetov V, Kryukova O 2013 Methodology and practice of data analysis in management: methods of one-dimensional and two-dimensional analysis *RANEPA Publishing House* (Orel)
- [19] Yarushev S, Yefremova N 2014 Hybrid methods for time series prediction *Gibridnye i sinergeticheskie intellektualnye sistemy* **3** 381–388
- [20] Zaporozhtsev I, Sereda A-V 2015 Numerical method for short-term forecasting of multivariate time series evaluated in nodes of the planar uniform grid *Digital signal processing* **1** 17–24