Analysis of Economic Impact Based on Time Series Model -- Taking the GDP of Two Provinces as an Example

Zhutong Yu*

College of Liberal Arts and Sciences, University of Connecticut, Storrs, Connecticut, United States

*Corresponding author: zhutong.yu@uconn.edu

Abstract. Time series analysis is an important branch of statistics, which is widely used in various fields, as well as in the economic field. In the economic field, a reasonable GDP forecast helps a country to set credible, substantive, and strategic goals. In recent years, Northeast China has seen an obvious decline in the national economic ranking, and the relevant GDP data of Jilin and Heilongjiang provinces are more obvious. This paper is based on time series analysis and support vector regression (SVR) algorithm to build a series model. The model was established by using 60 GDP data of Jilin Province and Heilongjiang Province from 1959 to 2022 with similar economic facts. After calculation and verification, the optimal model of Jilin Province is ARIMA (1, 1,0), the SVR with parameter C of 10 and 2 of 0.007, and the combined model with it as the main model is the optimal model of Heilongjiang Province. The GDP of the two provinces will continue to rise in the next five years, and the GDP of Jilin Province will rise faster, and the GDP gap between the two provinces will gradually narrow.

Keywords: time series model; Economic impact; GDP data; Jilin Province and Heilongjiang Province.

1. Introduction

Time series analysis is an important branch of statistics. It is a subject to study the statistical law of random series by applying stochastic process theory and mathematical statistics methods. It includes the understanding and description of the generation mechanism of the research data, the prediction of future values, and the realization of the optimal control of the system. Reasonable GDP projections help countries set credible, substantive, and strategic goals. In recent years, the northeast of China has seen an obvious decline in the national economic ranking. Heilongjiang Province dropped from 10th in 1959 to 21st. Jilin Province dropped from 15th in 1959 to 22nd. Although the GDP of Heilongjiang Province has been higher than that of Jilin Province, the GDP value of the two provinces has been close. The national GDP rankings of the two provinces have always been close to each other. This paper chooses the GDP data of Heilongjiang Province, whose economic value and geographical location are close to that of Jilin Province, as the comparison and reference object. This paper will build a time series model (ARIMA model, Autoregressive Integrated Moving Averagemodel) for the GDP data of the above two provinces, and compare the known long-term GDP data of the two provinces with the short-term forecast values, respectively, to observe the competitive relationship between the GDP growth rate of the provinces. The comparison of data between provinces makes the two provinces promote each other's development and has practical significance.

The main research ideas of this paper are as follows: the first part of this paper will review and clarify the background and significance of this study, and review the relevant literature. The second part of the paper combs the theory of time series model. The third part of the article takes the GDP data of Jilin Province and Heilongjiang Province as an example to analyze and forecast the economy by using time series data. The fourth part of the paper discusses the results of this study. The fifth part of the article summarizes the research content of this paper.

2. Literature References

In terms of literature, in 2013, based on the time series theory, Li Shouli modeled the GDP data of Zhengzhou from 1981 to 2010, and determined the ARIMA(1,2,1) model as the most suitable model for Zhengzhou's GDP by calculating mean square error with reserved data [1]. Liu Wei established a time series model for the GDP data of Jilin Province from 1952 to 2007 2008, and further predicted the development of the next 10 years after analyzing the GDP data of 56 years with the ARIMA model prediction method in time series analysis, concluding that the GDP of Jilin Province would continue to rise [2]. Song Jing collected the gross regional product (GDP) of Anhui Province from 1978 to 2015 as the basic sample data for modeling, based on different models AIC and SC (Akaike information criterion and Schwarz) The optimal model is determined as ARIMA(1,1,2) by numerical comparison, and the prediction is given [3]. Wang Shasha et al. established a combined model and forecast China's GDP data from 1980 to 2006 [4]. Xiang et al. applied the combined model forecasting method to the forecast of China's GDP, thus achieving the improvement of model forecasting accuracy [5]. Xue Qian et al. applied the combined forecasting method to the GDP data of Chongqing [6]. Chen Lin et al. used Eviews software to model China's GDP data and made a reasonable short-term forecast [7]. Ding Lei established a time series model for the per capita GDP data of Heilongjiang Province from 2008 to 2012. He established the ARIMA model and gave the forecast of per capita GDP in the next five years, concluding that the living standard of people in Heilongjiang Province will continue to grow steadily [8]. Wei Ning selected the GDP of Shaanxi Province and the primary, secondary, and tertiary industries as the basic data, and applied SPSS and Eviews software to establish a simple index model, ARIMA(L,2,1) model, and regression model of the primary, secondary and tertiary industries of Shaanxi Province on GDP [9]. James et al. used time series to analyze the GDP of 150 cities from 1950 to 2015 [10]. Bianic and Tu Rina conducted a time series analysis of Croatia's GDP [11].

3. Model Theory

3.1. Basic Theories Related to Time Series Analysis

The main purpose of time series analysis is to summarize and predict the dynamic changes of the series. The essential features of time series are mainly reflected in two aspects: the first is that the observed values are interdependent or correlated; The other is that the observed values are ordered. The unit root test is used to test the stationarity of a sequence. There are many kinds of unit root tests, among which the Augmented Dickey-Fuller (ADF) test is the most commonly used.

A very important assumption in the process of time series modeling is that the series has stationarity, that is, the properties of the series such as mean and variance do not change with time. If the sequence is not stationary, these properties will of course change, so that the prediction of future properties is prone to error. Therefore, we give some treatment to such non-stationary. The pre-processing of non-stationary time series is to stabilize the non-stationary time series, and only the stationary time series can enter the modeling process.

The pure randomness test is the white noise test, and the white noise test should be carried out on the estimated residual of the model when building the model. After the white noise sequence is obtained, the useful information in the time series has been extracted, and the rest are random disturbance terms. If the residual sequence passes the white noise test, the modeling can be stopped, indicating that no information can continue to be extracted. Suppose the residual sequence is not white noise. In that case, it means that there is still useful information missing in the residual sequence, and the information should be further extracted until the model is optimized.

There are many good criteria to measure the fit of statistical models, among which AIC and BIC (Bayesian Information Criterion) criteria are the most commonly used. Each criterion represents a set of criteria rather than a single, simple formula. These rules describe the loss of information relative to the "real model" after the model is built. Since the trained model is only an approximate model of the real model, these rules cannot explain the accuracy of the model, but they can further select the simulated model.

3.2. Common Models of Time Series

3.2.1 ARMA Model

ARMA is a family of practical and effective models for describing the dynamic properties of a single time series. It was founded by Georage Box, an American scholar, and Gwilym Jenkins, a British statistician. Its basic idea is that a series of time-varying and interrelated timing sequences are approximated by corresponding mathematical models. The analysis and study of the corresponding mathematical model can realize the intrinsic structure and complex characteristics of these dynamic data more essentially, to achieve the best prediction under the meaning of minimum variance.

3.2.2 ARIMA Model

The basic idea behind the ARIMA model is to use historical information about the data itself to predict the future. The value of a label at a point in time is affected both by the value of a label in the past and by chance events in the past, that is, the ARIMA model assumes that: The label value fluctuates around the general trend of time, wherein the trend is formed by the influence of historical labels, and the fluctuation is formed by the influence of accidental events in a period and the general trend itself is not necessarily stable. In short, the ARIMA model is an attempt to extract the time series patterns hidden behind the data through the way of autocorrelation and difference, and then use these patterns to predict future data.

4. Application of the Time Series model in GDP Forecast of Jilin Province and Heilongjiang Province

The data used in this part are the regional GDP data of Jilin Province from 1959 to 2022 and the regional GDP data of Heilongjiang Province from 1959 to 2022. The unit is 100 million yuan. (The data are from the 2022 Statistical Yearbook of Jilin Province and Heilongjiang Province). The selected data contains 64 characters. Since the data value is very large, the corresponding data is logarithmic for analysis. x is the original GDP data of Jilin Province, and y is the logarithmic data of x, as shown in Table 1.

Table 1. GDP data of Jilin Province from 1959 to 2022

Year	x	y	Year	x	y
1959	16.55	2.81	1999	558.06	6.32
1969	31.35	3.45	2009	2348.54	7.76
1979	56.02	4.03	2019	11939.24	9.39
1989	121.67	4.80			

Table 2. GDP data of Heilongjiang Province from 1959 to 2022

Year	h	g	Year	h	g
1959	26.0	3.26	1999	959.70	6.87
1969	54.6	4.00	2009	3637.2	8.20
1979	115.7	4.75	2019	13691.58	9.52
1989	248.4	5.52			

h is the original GDP data of Heilongjiang Province, and g is the logarithmic data of h. The modeling method the article used was based on 60 data from 1959 to 2018 after establishing the model to predict the value of 2019, and adding 61 data of the actual value of 2019 to re-establish the model to predict the value of 2020, as shown in Table 2. And so on, until the 2022 value is predicted, and then the mean square error is calculated using the four years of progressive forecast value and four years of actual data. The models are built with different software, different orders, and different types, but the results are recursively predicted and compared by this process. When modeling recursion step by step, the change of the model means that the newly added data causes the change of the original sequence rule, and the data at this time point changes the rule of the last time series summary to different degrees. Therefore, this recursive prediction method is more accurate in later years when such change data exists.

4.1. Simple Linear Regression Model of GDP of Two Provinces

SPSS has a high reputation for automatic statistical charts, in-depth analysis, complete functions, and easy operation. In this section, SPSS was used to establish a linear regression model.

(1) Establishment of linear regression model in Jilin Province:

As shown in Fig 1, the scatter distribution of data y was obtained after establishing a linear regression model for Jilin Province.

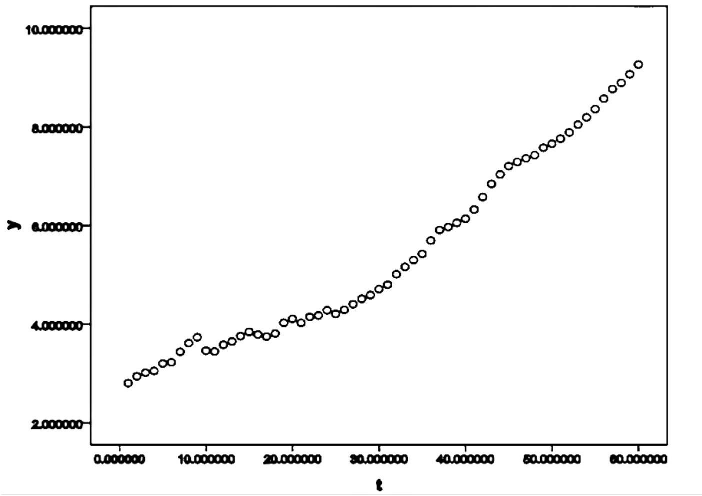

Fig. 1 Scatter distribution of data y (Photo/Picture credit: Original)

Table 3. Results of model test parameters

R	R^2	Durbin-Watson	F	Sig.
0.976	0.953	0.049	1188.348	0.000

The complex correlation coefficient is 0.976, which indicates that y and x do have a high linear relationship over time. The coefficient of determination is 0.953, and the model can explain 95.3% of the variation of the dependent variable, which is a good fit. The Durbin-Watson test value is small and there is a certain distance from 2, and there may be a correlation between the residual and the independent variable. From this parameter, the suitability of the establishment of this model needs to be investigated. $F = 1188.348$, $Sig. = 0.000$, that is, $p < 0.001$, indicating that the linear relationship between the independent variable and the dependent variable is significant, and a linear regression equation can be established. The above data are shown in Table 3.

Table 4 Results of model parameters

	Coefficient	t	$Sig.$
Constant term	2.152	19.737	0.0000
t	0.107	34.472	0.0000

As is shown in Table 4, the constant term of the model is 2.152, and the regression coefficient of the independent variable is 0.107. Since the T-test results all reached the significant level $p < 0.001$, the constant terms and regression coefficients were statistically significant, and the independent variables and dependent variables were linearly significant. In linear regression analysis, the regression coefficients used are non-standardized.

Linear regression equation:

$$y_t = 2.152 + 0.107t \tag{1}$$

According to the 3σ principle, if the absolute value of the normalized and studized residuals is greater than 3, the corresponding observed value is an outlier, so there is no outlier in this data. Most of the studized residuals fall within the interval of absolute value 2. Combined with the statistical table of residuals, there are no outliers in the data. Model 1 is established by applying 60 actual data, and this article uses it to predict the 61st data, which is 8.679. After applying the actual 61 data, the modeling result is $yt = 2.129 + 0.108t$, the 62nd data is predicted to be 8.825, the 62nd actual data is added, and the model does not change; the 63rd prediction result is 8.933, and the model does not change after the actual value of 63 is added. It indicates that the actual data in this stage is stable, and the last forecast result is 9.041. All results are shown in Table 5.

Table 5. Recursive prediction results in yf

Year	yf
2019	8.679
2020	8.825
2021	8.933
2022	9.041

Mean square error:

$$MSE = \frac{1}{4}\sum_{t=61}^{64}(y_t - yf_t)^2 = 0.386561 \tag{2}$$

(2) Establishment of linear regression model in Heilongjiang Province:

The establishment process was the same as that of Jilin Province, and the complex correlation coefficient was 0.983, showing a high linear relationship. The decision coefficient of the model is 0.966, explaining 96.6% of the variation of the dependent variable, and the fit is good. $F = 1650.106$, $Sig. = 0.000$, that is, $p < 0.001$, the linear relationship is significant, and the linear regression equation can be established. The constant term of the model is 2.873, and the regression coefficient of the independent variable is 0.102. Since the results of the T-test were $p < 0.001$, the coefficients were statistically significant and linearly significant. According to the Three Sigma principle, there are no outliers in the data. Known from studying residuals that there are no outliers in the data.

Linear regression model:

$$g_t = 2.873 + 0.102 \tag{3}$$

The forecast results are shown in Table 6.

Table 6. Recurrence prediction results gf

Year	gf
2019	9.095
2020	9.245
2021	9.348
2022	9.451

Mean square error:

$$MSE = \frac{1}{4}\sum_{t=61}^{64}(g_t - gf_t)^2 = 0.099518 \tag{4}$$

In linear fitting, from the two aspects of parameter testing and mean square error, Heilongjiang Province equation 2 is better.

4.2. Simple index model of GDP of two provinces

(1) Simple index model of GDP of Jilin Province:

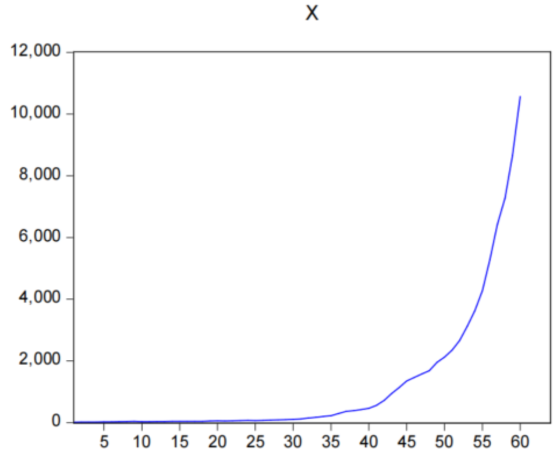

Fig. 2 GDP image of Jilin Province from 1959 to 2018 (Photo/Picture credit: Original).

Observe Fig 2, the data has an exponential trend, and use Eviews to fit a simple exponential model to the logarithmic data y:

Table 7. Fitting results of simple index model

	Coeficient	Std. Error	t-Statistic	Prob.
C(1)	2.721619	0.035332	77.02921	0.0000
C(2)	0.020519	0.000293	70.11026	0.0000
R-squared	0.989712	Akaike info criterion		-0.388037
Durn-Watson stat	0.186712	Sum squares		2.229470

As shown in Table 7, the exponential model can be obtained:

$$y_T = 2.721619 * e^{0.020519T} \tag{5}$$

Table 8. Recursive prediction results

	yf	xf
2019	9.515381	11914.35
2020	9.387586	13963.95
2021	9.873614	15768.18
2022	10.03901	17204.44

As shown in Table 8, the predicted value of the logarithm is yf, and the predicted value of raw data is xf.

Mean square error:

$$MSE = \frac{1}{4}\sum_{T=61}^{64}(y_T - yf_T)^2 = 0.09457529 \tag{6}$$

(2) Simple index model of GDP of Heilongjiang Province:

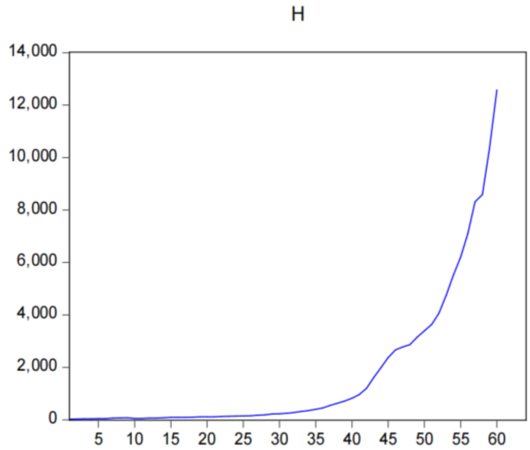

Fig. 3 GDP image of Heilongjiang Province from 1959 to 2011 (Photo/Picture credit: Original).

Observe Fig 3, the data has an exponential trend, and use Eviews to fit a simple exponential model to the logarithmic data:

Table 9. Fitting results of simple index model

	Coeficient	Std. Error	t-Statistic	Prob.
C(1)	3.356513	0.035918	93.44821	0.0000
C(2)	0.017446	0.000248	70.42242	0.0000
R-squared	0.989499	Akaike info criterion		-0.481523
Durbin-Watson stat	0.247700	Sum squared resid		2.030492

As shown in Table 9, the exponential model can be obtained:

$$g_T \ = \ 3.356513 * \ e^{0.017446T} \tag{7}$$

Table 10. Recursive prediction results

Year	gf	hf
2019	9.728958	13474.93
2020	9.878589	15377.64
2021	10.02051	16975.77
2022	10.15386	18269.59

As shown in Table 10, the logarithmic predicted value is gf and the original predicted value is hf. Mean square error:

$$MSE \ = \frac{1}{4}\sum_{T=61}^{64}(g_T - gf_T)^2 \ = \ 0.144221716 \tag{8}$$

About the fitting of the index model of two provinces: Jilin Province model 3 is more suitable for establishing the index model than Heilongjiang Province model 4. By comparing the four test parameters in Table 8 and Table 9, This article believes that Jilin Province model 3 is more consistent with the data, and the mean square error of Jilin Province is smaller. Therefore, the parameter optimal model is the same as the model with the least mean square error, and the model results are reliable.

4.3. Simple Linear Regression Model of GDP of Two Provinces

According to the characteristics of China's economic development and Jilin Province, this chapter will take the GDP data of Jilin Province since China's reform and opening up in 1979 as the basic data, and take four consecutive years as a dimension. In this case, this chapter introduces three dummy variables. We assume that the model expression is:

$$y_t = b_0 + b_1 Q1 + \mathrm{b}_2\mathrm{Q2} + \mathrm{b}_3\mathrm{Q3} + \mathrm{b}_4\mathrm{t} \tag{9}$$

In the multiple regression model, Y is the predicted value, b0 is the intercept, b1 is the influence coefficient of the first year, b2 is the influence coefficient of the second year, and b3 is the influence coefficient of the third year. Q1, Q2, and Q3 only have two values of 0 or 1, and when the three-year dummy variables are all 0, it is the prediction result of the fourth year.

After arranging the relevant data, this paper carries on the time series prediction analysis. First, the following regression results are obtained through regression analysis:

Table 11. GDP regression statistics of Jilin Province from 1979 to 2022

Regression Statistics	
Multiple R	0.987
R Square	0.977
Adjusted R Square	0.963
Standard error	0.280
Observed value	44

As shown in Table 11, the coefficient of determination R2 is 97.7%, the prediction equation can explain 97.7% of the information, and the prediction effect is good. The probability of occurrence (P-value) is less than 0.05, and the prediction model is valid. The values of b0, b1, b2, b3, and b4 were calculated as -2980.71, -155.12, -146.57, -96.01 and 262.35, respectively. Finally, the GDP time series model of Jilin Province from 1979 to 2022 was published as:

$$y_t = 2980.71 - 155.12Q1 - 146.57\mathrm{Q2} - 96.01\mathrm{Q3} + 272.35\mathrm{t} \quad (10)$$

According to the above publicity, the GDP of Jilin Province from 2023 to 2026 is predicted to be 14969.92, 15380.82, 15833.73, and 16332.09, respectively.

Taking the same approach, the GDP time series model of Heilongjiang Province from 1979 to 2022 is publicized as follows:

$$g_t = 3313.03 - 66.05Q1 - 129.84\mathrm{Q2} - 76.66\mathrm{Q3} + 310.65\mathrm{t} \quad (11)$$

According to the above publicity, the GDP of Heilongjiang Province from 2023 to 2026 is predicted to be 15876.23, 16093.09, 16426.92, and 16784.23 respectively.

5. Discussion

The known data of the two provinces from 1959 to 2022 are compared, and the future development situation of the two provinces is predicted and compared horizontally and vertically. The GDP of Heilongjiang Province has been slightly higher than that of Jilin Province in 64 years. In this paper, a new time series is obtained by subtracting the actual GDP value of Heilongjiang Province from the actual GDP value of Jilin Province x. The fluctuation of this time series is observed, and the difference between the two provinces is analyzed.

Longitudinal analysis of the internal GDP of the two provinces: the results of various models of Jilin Province and Heilongjiang Province show that the GDP of the two provinces will continue to rise, in the case of no accidents, this conclusion is beyond doubt.

Horizontal analysis of GDP between the two provinces: From 2019, the GDP of Heilongjiang Province is higher than that of Jilin Province, but the growth rate is not as fast as that of Jilin Province. Looking into the future, all models predict that the gap between them is reduced, that is, Jilin Province leads the GDP growth rate. The ARIMA model is best suited for short-term forecasting, and the information provided by the data beyond 2022 is analytically valuable, but not very reliable, and should only be used for general guidance.

Through the regression analysis of the GDP data of Jilin Province and Heilongjiang Province from 1979 to 2022, the regression equation model is established, and the GDP data of the two provinces from 2023 to 2026 is predicted. It is predicted that the GDP of Jilin Province from 2023 to 2026 will be 14969.92, 15380.82, 15833.73, and 16332.09 respectively. The GDP of Heilongjiang Province from 2023 to 2026 is predicted to be 15876.23, 16093.09, 16426.92, and 16784.23 respectively. The comparative data found that the GDP of the two provinces will continue to rise in the next few years. But in terms of the rising rate, Jilin Province is faster than Heilongjiang Province. The above conclusions have been confirmed. This paper also believes that the GDP of Jilin Province will surpass that of Heilongjiang Province for the first time in the future, which needs more data to confirm.

In addition, it is worth mentioning that compared with the South, the Northeast is mainly dominated by agriculture, energy, industry, etc. What is worrying is its long-term sustainability. The northeast should enrich its industrial structure. In the future research center, scholars can try to analyze the relationship between the GDP of Jilin Province and the three major industries as well as the inter-industry relationship, and further analyze the local situation in Jilin Province.

6. Conclusion

This paper is based on time series analysis and support vector regression (SVR) algorithm to build a series model. The model was established by using 60 GDP data of Jilin Province and Heilongjiang Province from 1959 to 2022 with similar economic facts. This paper also uses SPSS to build a linear regression model, a time series analysis knowledge system, and Eviews software to build an exponential model and ARIMA model, uses the SVR knowledge system and R language software to build support vector regression model and integrates the two parts of the knowledge system to build a combined model. Jilin Province and Heilongjiang Province respectively established these models and gradually predicted the GDP values from 2019 to 2022. The mean square error was calculated by using the predicted values and the known real GDP data for these four years. The model with the smallest mean square error was the most suitable model for the GDP of this province. The optimal GDP model of Jilin Province is ARIMA 1, 1,0), and the optimal GDP model of Heilongjiang Province is a combined model based on support vector machine regression and AR(1).

Using time series to analyze economic data is of great significance, it helps people to understand some information hidden behind economic data and guides people to take some effective measures to promote better economic development in time. In the whole time series research literature, this study will serve as an important theoretical resource to further enrich relevant knowledge.

For China, a huge economy, the study of this paper has a certain reference value, it can provide a reference for other provinces' economic research. For example, in some less developed provinces, analyzing the local economic development trend through time series to find out the subdivisions will help these regions gradually get rid of the backward situation.

After calculation and verification, the optimal model of Jilin Province is ARIMA(1, 1,0), the SVR whose parameter C is 10, 2 is 0.007, and the combined model with it as the main model is the optimal model of Heilongjiang Province. In the regression analysis of the GDP data of Jilin Province and Heilongjiang Province from 1979 to 2022, this paper predicts the GDP data of the two provinces from 2023 to 2026. The comparative data found that the GDP of the two provinces will continue to rise in the next few years. However, in terms of growth rate, Jilin Province is faster than Heilongjiang Province, and the gap between the GDP of the two provinces will gradually narrow. This paper believes that in future further research, the relationship between Jilin Province's GDP and the three industries and the relationship between industries deserves attention. After analyzing GDP through time series, further research on related industries through GDP is helpful to enrich research results and obtain more benefits.

References

[1] Shouli L. Application of time series model in GDP forecast of prefecture-level cities. Zhengzhou University, 2013.

[2] Wei L. Application of time series analysis in Jilin Province GDP forecast. Northeast Normal University, 2008.

[3] Jing S. Application of time series in Anhui Province GDP forecast - based on ARIMA model. Shopping Mall Modernization, 2016, (21): 134-136.

[4] Xiangping Y, Deshan S, Danfeng L. Application of combination model in my country's GDP forecast. Journal of Langfang Normal University (Natural Science Edition), 2010, 10(02): 87-89.

[5] Shasha W, An C, Jing S, et al. Application of combination forecast model in China's GDP forecast. Journal of Shandong University (Science Edition), 2009, 44(02): 56-59.

[6] Qian X, Fengyun M, Zhifeng T. Application of combination forecast method in Chongqing GDP forecast. Journal of Chongqing Technology and Business University (Natural Science Edition), 2017, 34(01): 56-63.

[7] Lin C, Jiachun X, Liangli M. Research on the optimal model of short-term CPI forecasting in China: Selection and optimization based on Several time series models. Journal of Contemporary Economics, 2002, 39(05): 9-16.

[8] Lei D. Time series model and forecast of GDP per capita in Heilongjiang Province. Journal of Natural Sciences of Harbin Normal University, 2010, 26(06): 16-18.

[9] Ning W. Research on time series analysis methods and their application in GDP forecasting of Shaanxi Province. Northwest A&F University, 2010.

[10] James S L, Gubbins P, Murray C J L, et al. Developing a comprehensive time series of GDP per capita for 210 countries from 1950 to 2015. Population health metrics, 2012, 10(1): 1-12.

[11] Bićanić I, Tuđa D. Long-term time series of GDP in Croatia. Privredna kretanja i ekonomska politika, 2014, 23(1 (134)): 37-69.