ISSN: 2688-9323

Forecast and Analysis of Seasonal Fluctuation Series based on SARIMA-GRNN Model

Mingyue Wu*, Chao Zhang

School of Information Science and Engineering, Shandong Agricultural University, Taian, 271001, China

*sdau1211@163.com

Abstract

The paper proposes the SARIMA-GRNN model to study the forecast of the civil aviation passenger traffic sequence. Aiming at the seasonal volatility of the sequence, the paper uses the Product Seasonal Model to predict the linear trend part and the seasonal fluctuation part of the sequence, and applies the Generalized Regression Network Model to the nonlinear residual Part of the prediction analysis, and finally the establishment of the SARIMA-GRNN combination model, fully extracted the sequence seasonal fluctuations, long-term trends, random fluctuations characteristics. In the empirical analysis, the SARIMA-GRNN model has a good prediction effect, and successfully predicted the two growths of China's civil aviation passenger traffic in October 2020 and March 2021, which has certain reference significance for the development of the civil aviation industry.

Keywords

Product Season Model; Generalized Regression Network Model; SARIMA-GRNN Model; Seasonal Fluctuations; Civil Aviation Passenger Traffic.

1. Introduction

With the development of economy, aviation, as an efficient and convenient way of travel, is welcomed by more and more people, and the world's civil aviation industry is developing rapidly, With the alleviation of COVID-19 epidemic, in order to meet the needs of people's lives, the normal operation of the civil aviation industry should be gradually restored. Therefore, this paper takes the passenger traffic of China's civil aviation as an example, according to the seasonal and trend characteristics of civil aviation passenger traffic, combined with historical data Forecast of civil aviation passenger traffic， In the context of entering the post-epidemic era, the passenger traffic forecast model has certain reference significance for the gradual restoration of the civil aviation industry and the restoration of other passenger transportation industries.

Many scholars have tried a variety of methods to predict time series data. X.-Y. Cao et al. [1] used H-P filtering and CensusX12 to decompose the original sequence to eliminate the influence of seasonal fluctuations, and then set up an ARMA model to predict the sequence data; Y.-J. Yu [2] chose to establish an ARIMA-GARCH model to analyze and forecast the interactive influence of the linear trend and random fluctuations of the stock closing price data; J.-Z. Hao [3] established a seasonal SARIMA model based on China's railway passenger traffic data, which effectively fitted the seasonal fluctuation characteristics of railway passenger traffic, and the prediction results were excellent; At the same time, some scholars use artificial neural networks to predict nonlinear sequences. Svitlana [4] uses artificial neural networks to predict exchange rates. The results show that the short-term prediction effect of artificial neural networks is excellent; Mark [5] used a generalized neural network to predict exchange rate

ISSN: 2688-9323

fluctuations, and the results showed that the generalized neural network is better than the multi-layer feedforward neural network and random walk model in fitting nonlinear models; Pan Lin[6] studied the theoretical methods of wavelet analysis and neural network, and proposed a special wavelet neural network for the first time, which has a good prediction effect; L.-M. Huang [7] used the combined model of GARCH model and GRNN to predict the RMB exchange rate, and the results showed that the prediction accuracy of the combined model was better than that of the single model; W.-R. Yan [8] and others established ARIMA- The GRNN model successfully improved the prediction accuracy. Many time series have linear and non-linear composite characteristics. After reading a large number of documents, it is found that from the perspective of seasonal changes in time series, there is less research on combining linear and non-linear models to establish forecasting models. Aiming at the trend and seasonal characteristics of civil aviation passenger traffic, this paper establishes a SARIMA model with predictive advantages for linear models, extracts the linear characteristics of civil aviation passenger traffic, and establishes a GRNN model with predictive advantages for nonlinear models to fit residuals. Difference fluctuation characteristics, the two parts are added to establish a SARIMA-GRNN model combination model to make an empirical analysis of civil aviation passenger traffic, and finally to test the effectiveness of the model.

2. Data Sources and Applications

This paper selects China's civil aviation passenger traffic data from January 2005 to May 2021, including domestic and international routes. The data comes from the official website of the Civil Aviation Administration of China, with a total of 197 data. Use the data from January 2010 to June 2019 as the training set to establish the SARIMA-GRNN prediction model; use the data from July to December 2019 as the test set to test the effectiveness of the model; use the 2020 year after the outbreak The passenger traffic data from January to May 2021 is compared with the forecast data to study the impact of the epidemic on the civil aviation industry.

3. Introduction to SARIMA-GRNN Combined Model

The product seasonal model (SARIMA) is a further analysis of seasonal effects, trend effects, and stochastic fluctuations based on the ARIMA model. The simple season model is suitable for situations where seasonal effects, trend effects, and stochastic fluctuations are easy to separate, and the ARIMA model can be used to fully extract information; If there is a complex interaction between seasonal effects, trend effects, and stochastic fluctuations, the SARIMA model needs to be adopted; Generalized Regression Neural Network (GRNN) is very suitable for solving the problem of nonlinear sequence prediction. By setting the smoothing factor, the transfer function between the neurons of each layer of the neural network is adjusted, to avoid subjectively affecting the prediction results. In the linear space with seasonal fluctuation characteristics, the SARIMA model has good predictive performance, but the SARIMA model cannot extract the nonlinear features in the residuals, thus losing part of the information; The GRNN model has good predictive performance in the nonlinear space. Using the nonlinear residual part of the SARIMA model to train the GRNN model can effectively predict the nonlinear residual.

Therefore, this paper complements the advantages of the SARIMA model and the GRNN model, and divides the civil aviation passenger traffic data S_t into a linear part L_t and a non-linear part e_t. The relationship is as follows:

$$S_t = L_t + e_t \tag{1}$$

Due to the trend of S_t and the characteristics of seasonal interaction, a SARIMA model was established to describe the linear part of S_t, L_t, Then calculate the nonlinear residual e_t of the SARIMA model according to the formula

$$e_t = S_t - L'_t \tag{2}$$

Where L'_t is the predicted value of the linear part. Perform white noise test. If e_t is white noise, it means that the residual e_t has no correlation. The SARIMA model fully extracts the linear information of the sequence, and the residual e_t only contains the nonlinear information of the original sequence. The GRNN model is used to predict the nonlinear residual error, and the predicted value e'_t is obtained. Finally, the linear part and the nonlinear part are added to obtain the predicted value S'_t of the SARIMA-GRNN combined model.

$$S'_t = L'_t + e'_t \tag{3}$$

4. The Linear Part of the SARIMA-GRNN Model

4.1 The Overall Trend and Seasonal Characteristics of the Data

Draw a time sequence diagram of China's civil aviation passenger traffic (e.g., Figure 1). Observing the timing chart, we can see that the passenger traffic of civil aviation has a clear growth trend. Due to the huge impact of the new crown pneumonia epidemic on the civil aviation industry, the passenger traffic volume is much lower than the normal period. To ensure the generality of the establishment of the model, we used the data from January 2010 to December 2019 during the normal period before the outbreak of the new crown pneumonia to establish the model. Calculate the seasonal index of each month (e.g., Figure 2).

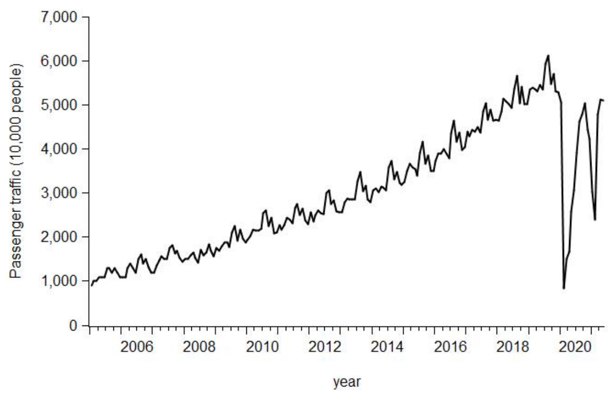


Figure 1. Time series of passenger traffic of China's civil aviation

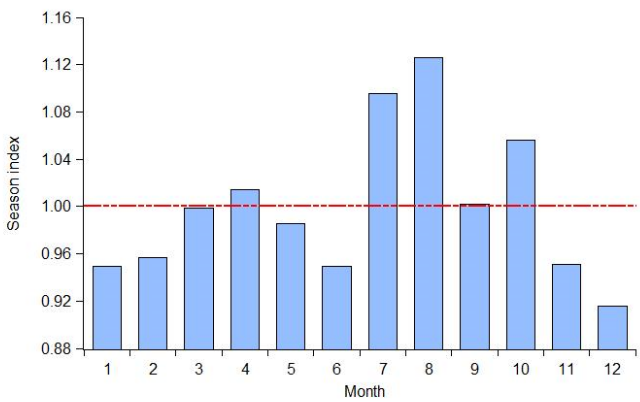


Figure 2. The monthly seasonal index of China's civil aviation passenger traffic

ISSN: 2688-9323

The seasonal index reflects the relationship between the quarter and the overall average. A seasonal index greater than 1 indicates that the value of the quarter is greater than the overall average, and a seasonal index less than 1 indicates that the value of the quarter is less than the overall average, The seasonal index reflects the magnitude of the impact of simple seasonal changes on the event, It can be seen from Figure 2 that the seasonal index in July, August, and October are all greater than 1, indicating that these three months are the peak of travel, and the seasonal index in January, June and December are all less than 1, indicating that people travel less in these three months. In general, civil aviation passenger traffic data is greatly affected by seasonal changes.

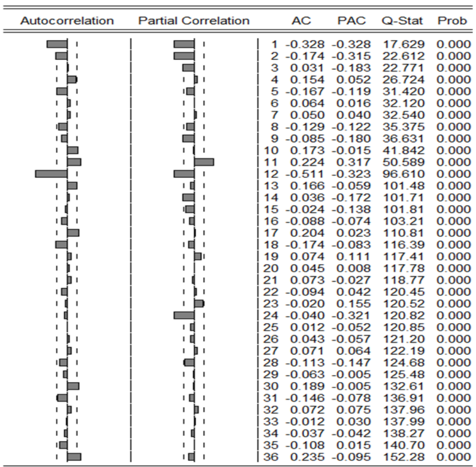
4.2 Stationarity and White Noise Test

It can be seen from Figure 1 that the sequence is a non-stationary sequence and has seasonal fluctuation characteristics. Therefore, this paper takes the logarithm of the civil aviation passenger traffic data S_t, and then performs the first-order 12-step difference, and then performs the ADF unit root test of the difference sequence (e.g., Table 1). For all types of unit root tests, the P value of the t statistic is less than the significance level of 1%, so the log difference series is considered to be stable.

Table 1. Unit root test of logarithmic difference series

Model type	Delay period	T test statistics	P value
No constant mean, no trend	1	-14.1805	<0.001
	2	-11.1809	<0.001
	3	-7.7095	<0.001
Have a constant mean, no trend	1	-14.1474	<0.001
	2	-11.1652	<0.001
	3	-7.6991	<0.001
Have a constant mean and trend	1	-14.1037	<0.001
	2	-11.135	<0.001
	3	-7.678	<0.001

The non-white noise sequence of the time series data is the premise of establishing the product seasonal model. Because the white noise sequence does not contain useful information, the white noise test is performed on the differentiated sequence. The test statistic in this paper is the LB statistic, which makes up for the lack of Q statistic in small sample testing, usually also called LB statistic also Q statistic (e.g., Figure 3).



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.328	-0.328	17.629	0.000
		2	-0.174	-0.315	22.612	0.000
		3	0.031	-0.183	22.771	0.000
		4	0.154	0.052	26.724	0.000
		5	-0.167	-0.119	31.420	0.000
		6	0.064	0.016	32.120	0.000
		7	0.050	0.040	32.540	0.000
		8	-0.129	-0.122	35.375	0.000
		9	-0.085	-0.180	36.631	0.000
		10	0.173	-0.015	41.842	0.000
		11	0.224	0.317	50.589	0.000
		12	-0.511	-0.323	96.610	0.000
		13	0.166	-0.059	101.48	0.000
		14	0.036	-0.172	101.71	0.000
		15	-0.024	-0.138	101.81	0.000
		16	-0.088	-0.074	103.21	0.000
		17	0.204	0.023	110.81	0.000
		18	-0.174	-0.083	116.39	0.000
		19	0.074	0.111	117.41	0.000
		20	0.045	0.008	117.78	0.000
		21	0.073	-0.027	118.77	0.000
		22	-0.094	0.042	120.45	0.000
		23	-0.020	0.155	120.52	0.000
		24	-0.040	-0.321	120.82	0.000
		25	0.012	-0.052	120.85	0.000
		26	0.043	-0.057	121.20	0.000
		27	0.071	0.064	122.19	0.000
		28	-0.113	-0.147	124.68	0.000
		29	-0.063	-0.005	125.48	0.000
		30	0.189	-0.005	132.61	0.000
		31	-0.146	-0.078	136.91	0.000
		32	0.072	0.075	137.96	0.000
		33	-0.012	0.030	137.99	0.000
		34	-0.037	-0.042	138.27	0.000
		35	-0.108	0.015	140.70	0.000
		36	0.235	-0.095	152.28	0.000

Figure 3. Autocorrelation and partial autocorrelation diagrams of logarithmic difference series

It can be seen from Figure 3 that the probability value P of the Q statistic is less than 1%, so the sequence is considered to be a non-white noise sequence at a significance level of 1%, so a product seasonal model can be established for prediction.

4.3 The Order of the Model

This paper attempts to establish a simple seasonal model. After the experiment, it is found that the simple ARIMA model cannot extract the relevant information. The seasonal fluctuations, long-term trends, and stochastic fluctuations of China's civil aviation passenger volume series have complex interaction relationships. Therefore, the product season is established. The model, in order to eliminate the seasonal fluctuation characteristics in the sequence, take the logarithm of the sequence and perform the 1st-order 12-step difference, and draw the autocorrelation graph and partial autocorrelation graph of the sequence (e.g., Figure 3).

First, determine the short-term correlation model according to the characteristics of autocorrelation graphs and partial autocorrelation coefficients within the 12th order of the sequence after the first-order 12-step difference. Figure 3 shows that the autocorrelation coefficients and partial autocorrelation coefficients within the 12th order are not truncated Tail, so try to use the ARMA (1, 1) model to extract the short-term autocorrelation information of the sequence after the difference. Looking at the seasonal autocorrelation characteristics of the difference series, it can be seen from the autocorrelation graph that the delayed 12th order autocorrelation coefficient exceeds twice the standard deviation, and the delayed 24th order autocorrelation coefficient falls within twice the standard deviation, so the first order autocorrelation coefficient is truncated ; From the partial autocorrelation graph, it can be seen that the delay orders of the partial autocorrelation coefficient exceed twice the standard deviation, so the partial autocorrelation coefficient is tailed, and the seasonal autocorrelation model $ARMA(0,1)_{12}$ is obtained, and the possible product season is finally obtained Model $SARIMA(1,1,1)\times(0,1,1)_{12}$.

4.4 Model Parameter Estimation

According to the minimum information criterion, try to compare multiple models, and get the models with the smallest AIC and BIC values as SARIMA $(1,1,1)\times(0,1,1)_{12}$ and SARIMA $(1,1,1)\times(0,1,2)_{12}$, so the two models are tested for parameters (e.g., Table 2, Table 3).

Table 2. Parameter test of $SARIMA(1,1,1)\times(0,1,1)_{12}$ model

Parameter	T statistics	P value
AR1	3.169	0
MA1	-8.56	0
SMA1	-11.913	0

Table 3. Parameter test of $SARIMA(1,1,1)\times(0,1,2)_{12}$ model

Parameter	T statistics	P value
AR1	3.617	0
MA1	-9.546	0
SMA1	-10.127	0
SMA2	2.186	0.015

From the table, the parameters of the ARIMA $(1,1,1)\times(0,1,1)_{12}$ model are significant at the 1% significance level, and the SARIMA $(1,1,1)\times(0,1,2)_{12}$ model. The coefficient of the middle SMA2 term is not significant at the 1% significance level, so SARIMA $(1,1,1)\times(0,1,1)_{12}$ is selected as the fitting model, and the fitting model is as follows:

$$\nabla\nabla^{12} lnX_t = \frac{1-0.8543B}{1-0.4431B}(1-0.7136B^{12})\varepsilon_t \quad (4)$$

Where $\varepsilon_t \sim N(0, 0.0027)$.

4.5 Significance Test of the Model

In order to test the validity of the SARIMA (1,1,1) ×(0,1,1)12 model, white noise test is performed on the residuals of the model (e.g., Table 4).

Table 4. White noise test of residuals

Delay period	P value of Q statistic
6	0.856
12	0.541
18	0.413
24	0.633

It can be seen from the table that the probability value P of the Q statistic is greater than 5%, indicating that the residual of the SARIMA (1,1,1) ×(0,1,1)12 model is significantly white noise, and the residual has no autocorrelation. There are only residual nonlinear fluctuation characteristics, which proves that the model has fully extracted the linear characteristics of the sequence. Finally, the fitting model is used to predict, and the predicted value is inversely first-order difference and anti-digitized to obtain the linear predicted value L_t'.

5. The Nonlinear Part of the SARIMA-GRNN Model

According to the formula $e_t = S_t - L_t'$, the nonlinear residual residual e_t is obtained, and the GRNN model is used to predict and analyze e_t.

This paper uses the data of the previous N days to predict the data of the next M days, so the sample space L is divided into X small sample spaces of length N+M, and X=L-(N+M) +1, each sample is N Day data is used as the input of the generalized neural network, and the data of the next M days is used as the output of the generalized neural network. The residual sequence sample space is L=174. After many attempts, this paper uses the data of the first 4 days to predict the residual of the next day Data, 170 sets of sample spaces are obtained, 168 sets of samples are randomly selected as training samples, and the remaining 12 sets are used as test samples.

Normalize the training samples, and then determine the optimal smoothing factor. After many experiments, the smoothing factor σ is increased in the interval [0.01,0.6] with a step size of $\Delta\sigma$ =0.01, and the average square error between the predicted value of the test interval and the true value is calculated (e.g., Figure 4).

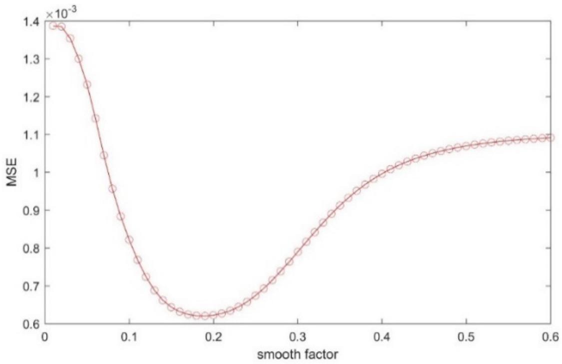


Figure 4. Mean square error of different smoothing factors

It can be seen from the figure that the average square error takes the minimum value when the smoothing factor is 0.18, so the smoothing factor is equal to 0.18, and then use the training set data to train the neural network model, use the trained GRNN model to make predictions, and use the previous normalization Index, restore the predicted value, and finally obtain the predicted value $\boldsymbol{e}_t'$ of the nonlinear residual**.**

6. Model Prediction

From the formula $S_t' = L_t' + e_t'$, obtain the predicted value S_t' of the civil aviation passenger traffic from July 2019 to December 2019 by the SARIMA-GRNN combined model, and compare the predicted value of the combined model, the predicted value of the SARIMA model with the real value (e.g., Figure 5).

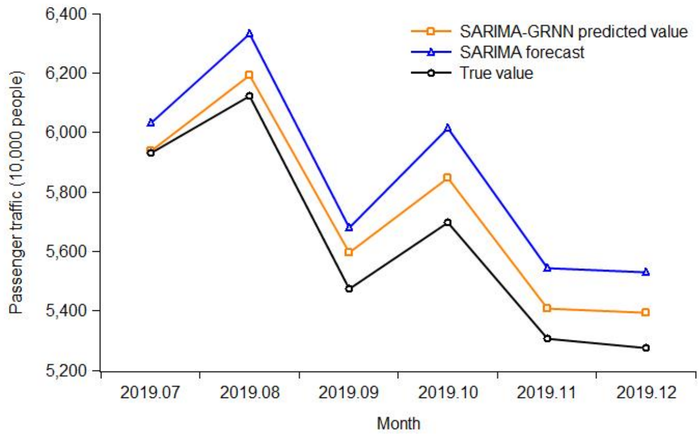


Figure 5. Forecast results of different models

It can be seen from the figure that compared to the predicted value of civil aviation passenger traffic by a single SARIMA model, the predicted value of the SARIMA-GRNN model is closer to the true value, and the prediction effect is better. The SARIMA model can only describe the linear part of the time series, and lacks the description of residual random fluctuations. The linear part of the time series described by the SARIMA model, combined with the nonlinear residual part predicted by the GRNN model, can better describe the real situation and predict the results Closer to the actual situation.

In order to compare the response prediction results more objectively, three prediction performance indicators of Mean Absolute Error (MAE), Mean absolute percentage error (MAPE), and Theil inequality coefficient are selected to measure the pros and cons of the prediction results of the single SARIMA model and the SARIMA-GRNN combined model. Taking into account the large number of civil aviation passenger traffic data, this paper first standardized the civil aviation passenger traffic data, and then calculated various indicators (e.g., Table 5).

Table 5. Comparison of prediction results of different models

Model	MSE	MAPE	Theil inequality coefficient
SARIMA	0.0331	38.2038%	0.0997
SARIMA-GRNN	0.0137	24.1504%	0.0642

Observing the various indicators of the model, the MAE and MAPE of the predicted value of the SARIMA-GRNN combined model are both smaller than the single SARIMA model, and the Theil

ISSN: 2688-9323

inequality coefficient between the predicted value of the SARIMA-GRNN combined model and the true value is less than that of the single SARIMA model, so the SARIMA-GRNN combination The predicted value of the model is closer to the true value, and the predictive ability is better. This not only confirms that the SARIMA-GRNN combined model predicts the trend and the feasibility of seasonal characteristic time series, but also confirms that the SARIMA-GRNN combined model can compensate for the impact of the single SARIMA model ignoring the random fluctuations of the residuals. In terms of running time, the running time of the SARIMA-GRNN combined model is slightly longer than that of the single SARIMA model. In fact, the training time of GRNN is shorter, and the time is mainly used to determine the optimal smoothing factor.

Next, apply the SARIMA-GRNN combined model to predict the civil aviation passenger traffic from January 2020 to May 2021 based on historical data before the outbreak of the new crown pneumonia. (e.g., Figure 6).

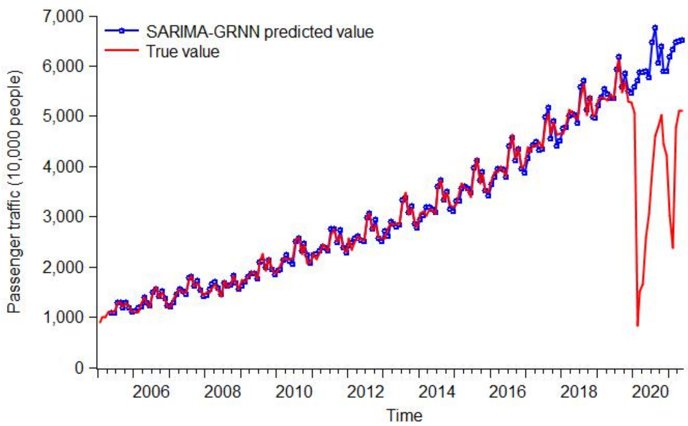


Figure 6. The prediction results of the SARIMA-GRNN model

It can be seen from Figure 6 that the new crown pneumonia epidemic has had a huge impact on the civil aviation industry, and passenger traffic has experienced an overall decline. The SARIMA-GRNN combined model successfully predicted the overall trend of passenger traffic during the epidemic, and successfully predicted October and 2020. The two increases in March 2021 indicate that the SARIMA-GRNN combined model has good applicability.

7. Conclusion

This paper takes the passenger traffic volume of China's civil aviation as an example. First, it analyzes the seasonal variability and increasing trend characteristics of passenger traffic data, selects the product seasonal model to reflect the linear characteristics of the sequence, and trains the residual residuals of the product seasonal model to obtain a generalized neural network model. , Successfully predicted the random fluctuation of the residual, combined the two parts to get the SARIMA-GRNN model, and finally predicted the passenger traffic data from January 2020 to May 2021. The following conclusions were obtained during the research process.

First, by studying the seasonal index of civil aviation passenger traffic, it is found that civil aviation passenger traffic has significant seasonal fluctuations. The peak period is concentrated in July, August, and October, which has certain reference significance for studying other traffic passenger traffic data.

Second, the seasonal effect, trend effect, and random fluctuation of the time series of civil aviation passenger traffic have interactive effects. This paper first established a simple seasonal model, and found that the model fits poorly in the model validity test, and then tried to establish a product seasonal model. The results showed that the parameters of the model were significant, and the residual white noise test passed, proving that the product seasonal model can be sufficient the time-series correlation relationship is extracted, so that there is no correlation in the residuals, and only nonlinear fluctuation characteristics remain.

Third, the time series of civil aviation passenger traffic has both linear and non-linear characteristics. Using a single linear or non-linear model will lose some useful information and lead to deviations in forecasts. The paper decomposes the time series of civil aviation passenger traffic into a linear part and a nonlinear residual part, establishes a product seasonal model from the training set data to extract the linear features of the time series, and trains a generalized neural network to predict and extract the nonlinear residual features. The linear part and the nonlinear the residual part are combined for prediction. The mean absolute error (MAE), the mean absolute percentage error (MAPE), and Theil unequal coefficient are selected as three prediction performance indicators. The prediction results show that the prediction performance of the SARIMA-GRNN combined model is better than that of the single SARIMA model. Predicting residual features can significantly improve the prediction performance of the SARIMA model.

Fourth, with the entry into the post-epidemic era, the overall situation of the epidemic has eased. In order to meet the needs of the people, the normal operation of the transportation industry represented by the civil aviation industry should be gradually restored. The recovery progress of the civil aviation industry needs to consider the normal civil aviation transportation volume. The SARIMA-GRNN combined model proposed in the paper has a good fitting effect, and successfully predicted two growths in October 2020 and March 2021. It has certain reference significance for predicting the overall growth trend of civil aviation passenger traffic and formulating related recovery policies.

Acknowledgments

The author wishes to thank Shandong Agricultural University, this job was supported by the following projects:

1. The Shandong Provincial Department of Education project "Prediction of the incidence of hemorrhagic fever based on a combined model" (project number: 2019GGX101061);

2. The Shandong Provincial Natural Science Foundation of China project "Study on the comprehensive prevention and control of hemorrhagic fever based on spatial metrology and EOF model" (project number: ZR201910220237);

3. "Application of Nonlinear Analysis Method in Garlic Insurance Product Evaluation and Design" project of Shandong Provincial Department of Education, China (project number: 2019GGX101024).

References

[1] X. -Y. Cao. Based on the combination model of X12-ARIMA and LSTM, urban vegetable price fluctuation law and prediction. Journal of Southwest University for Nationalities (Natural Science Edition), 2021, 47(4): 418-425.

[2] Y. -N. Yu. Discussion and research on stock data processing process based on ARIMA-GARCH model. Science Education Journal (Late), 2018(18): 38-41.

[3] J.-Z. Hao, Y.-J. Cui, Forecast of China Railway Passenger Traffic Volume Based on SARIMA Model. Mathematics in Practice and Knowledge, 2015, 45(18): 95-104.

[4] SVITLANA GALESHCHUK. Neural networks performance in exchange rate prediction. Neurocomputing, 2016(172): 446-452.

ISSN: 2688-9323

[5] MARK T LEUNG, A CHEN, HAZEM DAOUK. Forecasting exchange rates using general regression neural networks. Computers and Operations Research,2000(11): 1093-1110.

[6] Lin Pan. Research on the application of stock market prediction based on wavelet analysis and neural network. Wuhan: Wuhan University of Technology, 2006.

[7] L.-M. Huang. Research on GARCH-GRNN Combination Forecast of RMB Exchange Rate. Changsha: Hunan University, 2009.

[8] W.-R. Yan, Yong Xu, et al. Prediction of the incidence of infectious diseases based on the ARIMA-GRNN combined model. Chinese Health Statistics, 2008(1): 82-83.

[9] RODRIGO PEIRANO, WERNER K, MARCEL C MINUTOLO. Forecasting inflation in Latin American countries using a SARIMA–LSTM combination. Soft Computing,2021(25):10851-10862.

[10] CHUKWUDIKE C NWOKIKE, EMMANUEL W OKEREKE. Comparison of the performance of the SANN, SARIMA and ARIMA models for forecasting quarterly GDP of Nigeria. Asian Research Journal of Mathematics,2021,17(3):1-20.

[11] YAN QIN. Seasonal adjustment of CPI time series based on X-12-ARIMA and SARIMA. World Scientific Research Journal,2021,7(5).

[12] Jin-yan Wei, Ru Feng. Research on traffic volume prediction and realization using GRNN model. Journal of Changsha Institute of Communications, 2006(2): 46-50.

[13] Guang-cheng Luo, Jia-jue Gao. Seasonal time series prediction based on the combined model of GRA-LSTM and SARIMA.. Intelligent Computers and Applications, 2021, 11(6): 195-200.