

Empirical Study on the Sustainable Development of Domestic Tourism

Shichang Shen

School of Mathematics and Statistics, Qinghai Nationalities University, Xining, China

Email: 13909785766@163.com

How to cite this paper: Shen, S.C. (2020) Empirical Study on the Sustainable Development of Domestic Tourism. *Open Journal of Statistics*, 10, 651-658.
<https://doi.org/10.4236/ojs.2020.104039>

Received: June 29, 2020

Accepted: August 2, 2020

Published: August 5, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Taking the number of domestic tourists as the research object, an appropriate model is established to analyze the sustainable development of China's tourism industry. The number of domestic tourists in China from 1985 to 2015 was taken as a sample, and the ARIMA model was established using Eviews 6.0. Finally, the ARIMA(0, 2, 1) model is established. After testing, the model fits well (MAPE = 7.363) and the prediction accuracy is extremely high (99.56%). The ARIMA model's short-term prediction of the number of tourists is reasonable.

Keywords

Number of Tourists, ARIMA Model, Predictive Analysis

1. Introduction

The increase in people's income, the increase in consumption levels, and national policies have made tourism develop rapidly. During the "Twelfth Five-Year Plan" period, the scale of China's tourism industry continued to expand. The added value of tourism and related industries accounted for 4.33% of GDP. The tourism industry has gradually become a new growth point of the national economy. The development of tourism, which can bring huge economic benefits to the country and individuals and can improve the people's happiness index, cannot be ignored. To promote the healthy and sustainable development of tourism, we must first reasonably predict the number of future tourists. Forecasting the future development trend of the number of tourists will help government tourism departments to formulate tourism development strategies on the one hand, and on the other hand, it will help managers of tourism enterprises to adjust marketing policies to maximize profits. It can also be used as a ref-

erence for those who intend to travel. At present, the ARIMA model has been widely used in other related fields [1] [2] [3], but for tourism, the predecessors mostly used BP neural network model and gray theory [4] [5] for related research. Therefore, this article analyzes the future development trend of the number of tourists in China by using the relevant sample data and the ARIMA model, and puts forward relevant suggestions to promote the sustainable development of China's tourism industry and provide a reference for government tourism departments and tourism enterprise managers.

2. Empirical Analysis

This article takes the number of domestic tourists in China from 1985 to 2015 as a sample and uses Eviews 6.0 to analyze the data [6]. A prediction model is established for the sequence Y_t (Unit: million person-times) of domestic tourists from 1985 to 2011. The number of domestic tourists from 2012 to 2015 is reserved used to test the effect of the model. The data used are from China Statistical Yearbook 2015.

2.1. Data Preprocessing

Observing the trend of domestic tourist numbers from 1985 to 2011 (Figure 1), it can be seen that the overall number of domestic tourists has a certain exponential trend, showing a non-stationary sequence. To confirm the stability of the domestic tourist number sequence Y_t , further ADF unit root test, inspection results are as follows (see Table 1).

From the unit root test results in Table 1, it can be seen that the value of the t statistic is 8.084385, which is larger than the critical values of the confidence levels of 1%, 5%, and 10%, that is, the Y_t sequence has a unit root, so the original sequence Y_t is a non-stationary sequence. In order to smooth the sequence,

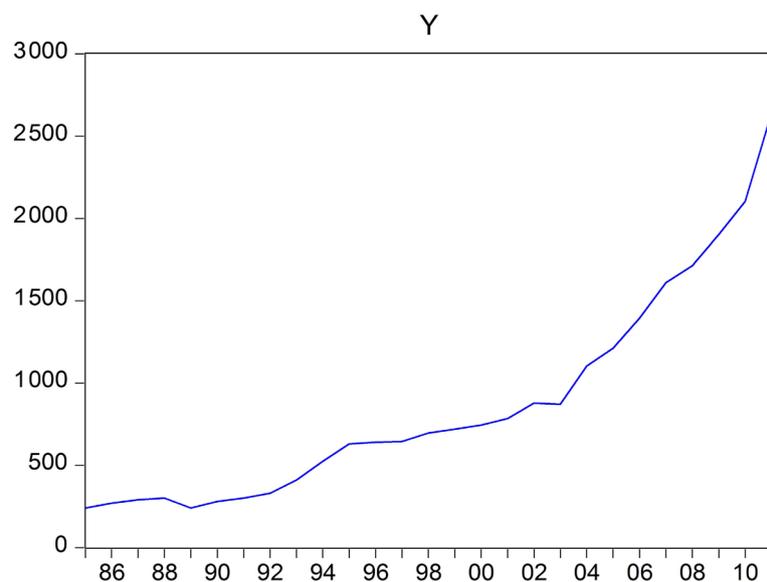


Figure 1. Line chart of Y_t sequence.

logarithmic and second-order difference processing is performed on the original data. The processed sequence is recorded as W_t . ADF unit root test is performed on the W_t sequence to determine that the processed sequence, that is, the W_t sequence is stable and the unit root test results are shown in **Table 2** and **Figure 2**. According to the unit root test results in **Table 2**, the value of the t statistic is -7.857 , which is smaller than the critical values of 1%, 5%, and 10% of the confidence level. In addition, the P value is almost zero. Assume that there is no unit root in the W_t sequence, so the processed sequence W_t is a stationary sequence.

2.2. Model Recognition

Autocorrelation function (ACF) and partial autocorrelation function (PACF) are

Table 1. ADF unit root test of Y_t .

	t -Statistic	Prob.
Augmented Dickey-Fuller test statistic	8.084	1.000
Test critical values	1% level	-2.657
	5% level	-1.954
	10% level	-1.609

Table 2. ADF unit root test for W_t .

	t -Statistic	Prob.
Augmented Dickey-Fuller test statistic	-7.857	0.000
Test critical values	1% level	-3.738
	5% level	-2.992
	10% level	-2.636

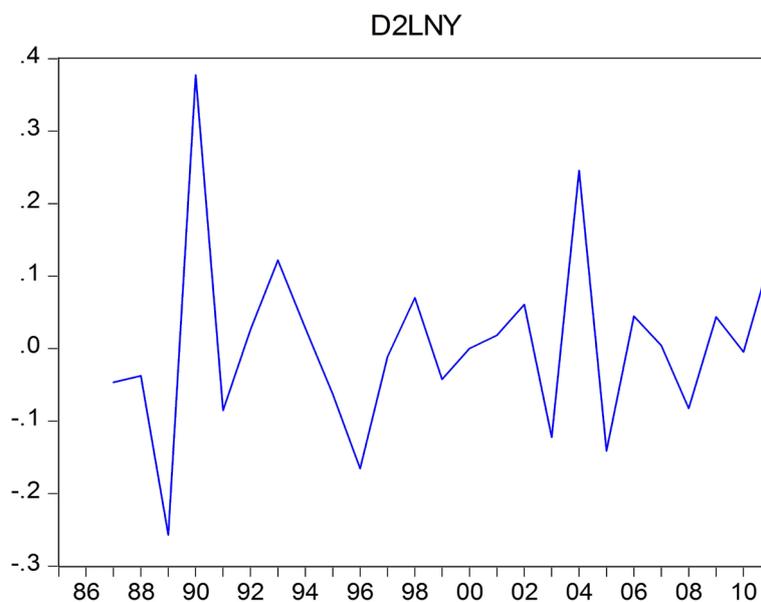


Figure 2. Line chart of W_t sequence.

the most important methods for identifying ARIMA models [7]. In Eviews 6.0, the sample autocorrelation and partial autocorrelation analysis diagrams are usually used to identify and rank models. PAC column φ_{kk} and AC column ρ_k are significantly different from 0 when $k = 1$. Consider $p = 1$ and $q = 1$. At the same time, according to the recognition principle of the ARIMA model, the autocorrelation function ρ_k and the partial autocorrelation function φ_{kk} are calculated as follows:

When $m = 1$ and $k = 1$, The proportion of $\rho_{k+1}, \rho_{k+2}, \dots, \rho_{k+M}$ (where $M = \lceil \sqrt{N} \rceil = \lceil \sqrt{27} \rceil = 5$ and N is the sample size of Y_t sequence) meeting

$$|\rho_{k+i}| \leq \left[\frac{1}{N} \left(1 + 2 \sum_{l=1}^m \hat{\rho}_l^2 \right) \right]^{\frac{1}{2}}, \quad i = 1, 2, \dots, M$$

and the proportion of $\varphi_{k+1}, \varphi_{k+2}, \dots, \varphi_{k+M}$ meeting

$$|\varphi_{k+i}| \leq 2 \left[\frac{1}{N} \left(1 + 2 \sum_{l=1}^m \hat{\rho}_l^2 \right) \right]^{\frac{1}{2}}, \quad i = 1, 2, \dots, M$$

is $100\% > 95\%$, so ρ_k is truncated in one step.

When $m = 1$ and $k = 1$, the percentages of $\varphi_{k+1,k+1}, \varphi_{k+2,k+2}, \dots, \varphi_{k+M,k+M}$ satisfying $|\varphi_{kk}| > \frac{1}{\sqrt{N}}$ and $|\varphi_{kk}| > \frac{2}{\sqrt{N}}$ are respectively 20% and 0%, the former is less than 31.7% and the latter is less than 4.5%, so φ_{kk} is truncated in one step.

In summary, the autocorrelation function ρ_k is truncated in step 1, and the partial autocorrelation function φ_{kk} is also truncated in 1 step, which is consistent with the results obtained by subjectively identifying the correlation diagram of the W_t series. According to the above conclusions, the models that may be suitable are ARIMA(1, 2, 1), ARIMA(1, 2, 0), ARIMA(0, 2, 1).

2.3. Model Establishment

For the ARIMA model, the adjusted R^2 , AIC, and SC criteria are all important factors to consider when choosing a model. When judging the pros and cons of the model according to the AIC and SC criteria, it is generally considered that the model with smaller AIC and SC function values is better. It can be seen from **Table 3** that among the three models, the AIC value and the SC value of the ARIMA(0, 2, 1) model are the smallest. When comparing the adjusted determination coefficient R^2 , the larger its value, the better the model's fitting effect. Among the three models, the adjusted R^2 of the ARIMA(0, 2, 1) model is the

Table 3. Model comparison table.

Model	Inverted Roots	the adjusted R^2	AIC	SC
ARIMA(1, 2, 1)	$ \lambda_i < 1$	0.357	-1.585	-1.486
ARIMA(1, 2, 0)	$ \lambda_i < 1$	0.231	-1.444	-1.395
ARIMA(0, 2, 1)	$ \lambda_i < 1$	0.380	-1.697	-1.648

largest. In summary, the ARIMA(0, 2, 1) model should be selected. In addition, the inverse roots of the lag polynomials of the ARIMA(0, 2, 1) model are all less than 1, which meets the requirements of process stability.

The parameter estimation results of the ARIMA(0, 2, 1) model are shown in **Table 4**.

2.4. Model Residual Sequence Test

After the preliminary judgment of the model as ARIMA(0, 2, 1), it should also be subjected to an adaptive test, that is, the independence test of the model residual a_t series, to determine whether the time series is properly described by this model, and whether the model needs further improvement.

Using Eviews 6.0 software to perform the χ^2 test, the sample size of the residual sequence is 25, and the maximum lag time can be taken [25/10]. Its P value corresponding to the Q test statistic is 0.771, so the residual sequence cannot be rejected. The null hypothesis indicates that the model's residual a_t sequence is purely random and is a white noise sequence.

3. Model Prediction

After the above test model is reasonable, the ARIMA(0, 2, 1) model can be used for short-term prediction. In order to test the prediction accuracy of the model, we first use the mean absolute percentage error (MAPE) and Hill inequality coefficient (TIC) [8] to test the model fitting effect, where

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad \text{TIC} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2 + \frac{1}{n} \sum_{i=1}^n \hat{y}_i^2}},$$

The ARIMA(0, 2, 1) model is used to obtain the predicted value of the number of domestic tourists from 1985 to 2011, which is compared with the real value (see **Figure 3**), and the MAPE and TIC values are calculated. It can be seen from **Figure 3** that the predicted value curve (\hat{X}) and the true value curve (Y)

Table 4. ARIMA(0, 2, 1) model parameter estimation results.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.922	0.053	-17.257	0.000
R-squared	0.380	Mean dependent var		0.004
Adjusted R-squared	0.380	S.D. dependent var		0.129
S.E. of regression	0.102	Akaike info criterion		-1.697
Sum squared resid	0.248	Schwarz criterion		-1.648
Log likelihood	22.213	Hannan-Quinn criter		-1.684
Durbin-Watson stat	1.731			
Inverted MA Roots		0.920		

have a high degree of coincidence, and the trend is basically the same. The residual curve (RESSID) is almost a straight line, and the average absolute percentage. The score error (MAPE) is $7.363 < 10$, which indicates that the model fits well. In addition, $TIC = 0.0398 < 1$, which means that the difference between the predicted value and the true value is very small, and the model is suitable.

At the same time, we will use the ARIMA(0, 2, 1) model to make dynamic predictions of the 4 observations after the Y_t series, and then compare the predicted values with the real values to obtain the dynamic prediction values, absolute errors and relative errors as follows (See Table 5):

It can be seen from Table 5 that the relative errors of the out-of-sample predictions of the model are less than 1%, the average relative error is about 0.44%, the difference between the predicted value and the true value is very small, and the prediction accuracy of the ARIMA(0, 2, 1) model is extremely high.

From the above analysis, it can be seen that the model established for the number of domestic tourists is suitable. Now we make short-term predictions of the number of domestic tourists in China from 2016 to 2020 (see Table 6). In order to more intuitively observe and analyze the changes in the number of tourists, the future forecast of the number of tourists draws a graph together

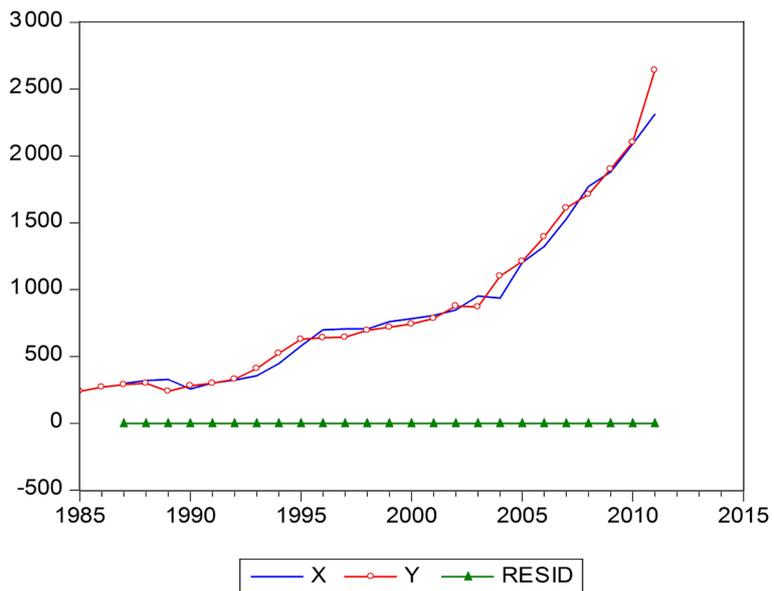


Figure 3. Intra-sample prediction fit of ARIMA(0, 2, 1) model (Unit: million person-times).

Table 5. ARIMA(0, 2, 1) model out-of-sample prediction error analysis table (Unit: million person-times).

Time	Actual value	Dynamic prediction	Absolute error	Relative error (%)
2012	2957	2934.399	22.601	0.764
2013	3262	3260.393	1.607	0.049
2014	3611	3622.603	-11.603	0.321
2015	4000	4025.052	-25.052	0.626

Table 6. Forecast results of the number of domestic tourists from 2016 to 2020 (Unit: million person-times).

Time	2016	2017	2018	2019	2020
Predictive value	4472.211	4969.046	5521.077	6134.435	6815.933

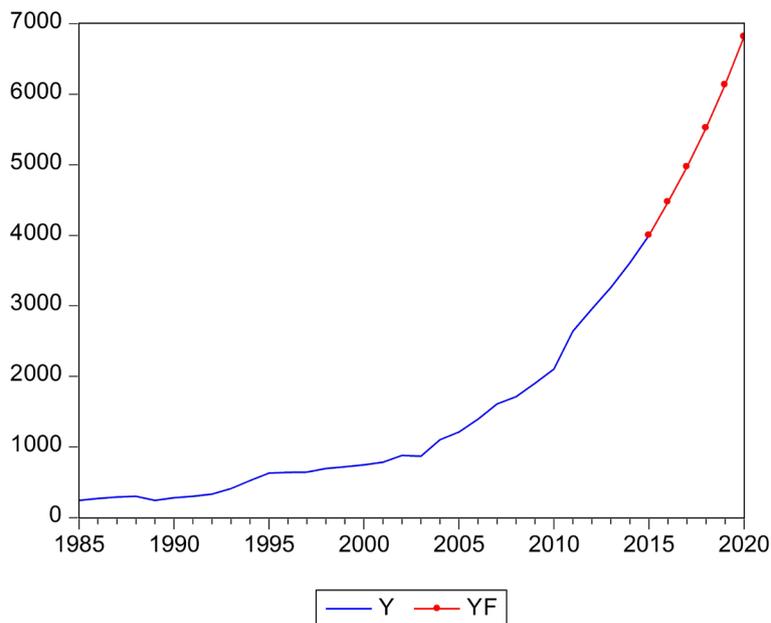


Figure 4. Curve of the previous value (Y) and future forecast value (YF) of the number of tourists.

With previous values (see **Figure 4**). It can be seen from **Figure 4** that in the next five years, the number of domestic tourists will increase historically.

4. Conclusions and Recommendations

1) Based on the sample data of the number of domestic tourists from 1985 to 2015, a $ARIMA(0, 2, 1)$ model was finally established through comparative analysis. The average absolute percentage error MAPE value of the model and the Hill inequality coefficient TIC are 7.363 and 0.0398, respectively. The data from 2012 to 2015 were predicted, and the predicted values were compared with the reserved synchronous values. As a result, it was found that the model prediction accuracy reached 99.56%. These test results mean that the model fits well and the prediction accuracy is extremely high. 2) Forecast the number of domestic tourists from 2016 to 2020 and compare it with historical data (see **Table 6** and **Figure 4**). In the next 5 years, the overall trend of tourist numbers is the same as in previous years. It is a trend growth, and the growth rate is slightly larger than before, which will promote the economic development of our country.

Based on the above conclusions, the author puts forward the following suggestions for reference: 1) Avoid excessive tourists and improve tourist satisfaction in the scenic area. 2) Facing the rapidly increasing number of tourists, the scale of tourist attractions should be expanded, and the activities in the tourist

attractions should be increased to alleviate the pressure of “crowding” in the tourist attractions. 3) Improve tourism service facilities, appropriately reduce the price of products in scenic spots, and promote consumption. 4) Expand the space for tourism development, increase innovation in tourism methods and types, and promote sustainable development of the tourism industry.

Funds

This work is supported by the National Natural Science Foundation of China (No. 11561056) and Natural Science Foundation of Qinghai (No. 2016-ZJ-914).

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Chen, P., Wu, L. and Song, H. (2012) Forecast of Inbound Tourists in Anhui Province Based on ARIMA Model. *Journal of Anhui Agricultural University (Social Science Edition)*, No. 1, 32-35.
- [2] Li, G., Zhao, J.L. and Su, Y. (2012) Prediction of Handheld Order Volume of World Container Ships Based on ARMA Model. *Research on Science and Technology Management*, No. 16, 7-11.
- [3] Liu, L.P. (2011) Time Series Model and Forecast of Total Retail Sales of Social Consumer Goods in China. *Economic Forum*, No. 6, 3-6.
- [4] Guo, Q.C., Kong L.J. and Cui, W.J. (2011) Forecast of Domestic Tourists Based on BP Neural Network Model. *Value Engineering*, No. 27, 7-8.
- [5] Liu, H.M. and Fan, S.H. (2010) Forecast and Analysis of Guangzhou Tourism Reception Number Based on Grey System Theory. *Statistics and Decision*, No. 17, 64-66.
- [6] Yi, D.H. (2008) Data Analysis and E Views Application. Renmin University of China Press, Beijing.
- [7] Wang, Z.L. (2010) Applied Time Series. China Statistics Press, Beijing.
- [8] Amirhan, A. and Liu, W.Z. (2014) Analysis and Forecast of Xinjiang Mutton Yield Based on ARIMA Model. *Heilongjiang Animal Husbandry and Veterinary Medicine (Exploration and Research)*, No. 8, 16-19.