FORECASTING TOURISM MARKET DEMAND IN HUNAN PROVINCE USING ARIMA MODEL

Quan Qin

Department of mathematics and statistics, Guangxi Normal University
214239164@qq.com

Abstract

Hunan Province is a province with large tourism resources, and it’s tourism occupies a significant position in economic development. In recent years, Hunan province is in a critical period of transition from a major tourist province to a staggering tourist province. Therefore it is of great significance to analyze and predict the development trend of the future tourism market demand in Hunan Province and provide the government and tourism enterprise managers with the scientific market decision. Based on time-series analysis and forecasting theory, this paper attempts to use the time-series data of the number of domestic tourists in Hunan Province which from 2000 to 2019 and construct model autoregressive and integrated moving average (ARIMA) to predict the number of tourists in the Hunan Province in the next four years. The results show that the predicted value is very close to the true value. Therefore, the prediction effect of the model is well, which will provide a theoretical basis for the government to establish relevant policies. However, considering the characteristics of time-series data, the model needs to be further revised and even effectively combined with other models to gain better forecasting results.

Keywords: tourism market, time series, ARIMA model

1. Introduction

With the increase of residents' income and traffic conditions, tourism has gradually become a normal or even rigid demand in people's lives. Hunan Province is a province with large tourism resources, with 2 world natural heritage sites, 3 National historical and cultural cities, 1 cultural heritage site, and 20 national key scenic spots (a People’s Republic of China partnership, 2016; Su, Wall, Wang, & Jin, 2017). The rich tourism resources provide powerful resources for the development of tourism in Hunan Province (Li, Sun, Chen, & Yang, 2012; Wang, Wang, Gan, Chen, & Voda, 2021).

The outline of the 13th Five-Year Plan for the Tourism Industry of Hunan Province points out that Hunan Province will be transformed from a major tourism province into a strong tourism province during the 13th Five-Year Plan period (Zhuo, 2018). by 2020, the total revenue of the tourism industry will reach 900 billion yuan and a total of 900 million tourists will be received. The contribution rate of provincial tourism to GDP exceeds 12%. The rapid development of tourism has a broad market prospect, driving employment growth and generating good economic and social benefits.
However, the rapid development of the tourism industry in Hunan Province has not been smooth sailing. It will also be subject to impact and competition from the tourism market outside the province and overseas. If this province wants to condescend and win the competition, it must accurately analyze and predict the tourism market. It’s momentous for government tourism authorities to formulate scientific and reasonable tourism policy planning, optimize the allocation of tourism market resources and tourism companies to make correct decisions.

Cho (Cho, V.A, 2003) proposed how to judge the accuracy of tourism market forecasting, that is, to judge the accuracy of model forecasting by calculating the error between the predicted value and the actual value. Turner & Witt (2011) (Turner, 2011) established an SEM model of the three associations of vacation tourism, business tourism, and visiting relatives and friends, expanded the types of explanatory variables and analyzed the causal relationship between the three tourism modes. However, the result of model analysis of the impact of tourism market demand needs further discussion. Seetanah (2015) (Seetanah B, Sannassee R, Rojid S, 2015) through the construction of a VAR model to analyze that the country income, average relative cost, development level and tourism infrastructure required by different competitive destinations have an impact on the international tourism market, and relative price also has an impact on the international tourism market. JaeH.Kima (2005) (Kima & Moosab, 2005) used three other methods such as regression model, SARIMA model, and structured time-series model to have carried out a forecast analysis on Australian inbound tourism. He calculated and compared the forecast errors of the three models. The results proved that the forecast accuracy of the seasonal ARIMA model is higher.

The tourism market has the characteristics of complexity, seasonality, and volatility. Traditional forecasting methods such as regression forecasting and gravity model forecasting do not match the volatility data of the modern travel market, making the prediction inaccurate. Therefore, it is very important to choose appropriate methods and models according to the characteristics of the data.

At present, many types of tourism market demand forecasting models can be used (Ariyo A A, Adewumi A O, 2015; Janardhanan & Barrett, 2018; Min & Zhongfeng., 2021; Tiantian, Yanqin, & Le., 2021), including Linear regression model, moving average, exponential smoothing, and so on. However, no model is suitable for all data. In order to obtain a good prediction effect, it is necessary to select an
appropriate and relatively accurate prediction mode according to the data characteristics of the prediction object.

This paper selects a time series model based on the linear characteristics of the number of domestic tourists in Hunan Province, combines quantitative and qualitative analysis methods, comprehensively studies the relationship between different factors and the tourism market. After fully analyzing and studying various factors that affect market changes, analysis of data in the tourism market is expected to provide help for future researchers. In previous studies, scholars have used regression models frequently to make relevant analysis and predictions, thus ignoring the influence of endogenous variables. However, the ARIMA model used in this article takes this problem into account, so as to achieve better analysis and prediction results.

2. Methodological Approach

Now, one of the widely used time series models is ARIMA. The model is relatively simple, which only needs endogenous variables instead of other exogenous variables (Nunohara et al., 2020; Thiruchelvam Loshini, Dass Sarat Chandra, Asirvadam Vijanth Sagayan, Daud Hanita, 2021; Titus, Wanjoya, & Mageto, 2021). An ARIMA model is labeled as an ARIMA model(p,d,q), where “p” is the number of autoregressive terms, “d” is the order of the data stationary, and “q” is the order of the moving average process (Huang, Xiong, Chen, & Zhao, 2020; Rahmanian Vahid, Bokaie Saied, Haghdoost Aliakbar, 2020; Zou, 2020). In the ARIMA model, the future value of a variable is assumed to be a linear function of several past observations and random errors (Huang et al., 2020). That is, the underlying process that generates the time series has the form.

\[ W_t = \nabla^d Y_t \]

\[ W_t = c + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \] (1)

Where \( W_t \) and random perturbation \( e_t \) are the actual value and white noise at time period \( t \), respectively; \( \phi_i, \theta_j \ (i = 1, 2, \cdots, p, \quad j = 1, 2, \cdots, q) \) are model parameters. The above equation entails several important special cases of the ARIMA family of models. If \( q = 0 \), then the equation becomes an AR(Autoregressive) model of order \( p \). When \( p = 0 \), the model reduces to an MA(Moving average) model of order \( q \). The following flow diagram demonstrates the basic methodology of ARIMA development.
This article selects the time series data of the number of domestic tourists in Hunan Province from 2000 to 2019 to establish the ARIMA model. The data comes from the Hunan Provincial Statistical Yearbook and the National Economic and Social Development Bulletin. The corresponding web is http://tjj.hunan.gov.cn/tjfx/tjgb/jjfzgb/.

3.1. **Unit Root Test**

H0: The series has a unit root.

H1: The series does not have a unit root.

<table>
<thead>
<tr>
<th>Tabel 1. Results for unit root test</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-Statistic</td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Augmented Dickey-Fuller test statistic</td>
</tr>
<tr>
<td>Test critical values:</td>
</tr>
<tr>
<td>1% level</td>
</tr>
<tr>
<td>5% level</td>
</tr>
<tr>
<td>10% level</td>
</tr>
</tbody>
</table>

The establishment of the ARIMA model requires the time series to be a stable random process. Therefore, a unit root test is required before the modeling. According to the results of table 1: The t-test statistic is beyond the critical value of the confidence level, so there is a unit root. It means that the sequence is not stationary, so it is necessary to perform difference processing on the sequence. After difference processing $\nabla W_t = W_t - W_{t-1}$ and software calculation, the results are shown in Table 2:

| Tabel 2. Results for the first difference of the data series |
Since the calculated p-value is more significant than the threshold significance level 0.05, the null hypothesis H0 cannot be rejected. The risk of rejecting the null hypothesis H0 while the first difference of the data series is stationary.

### 3.2. Model Identification and order determination

When the first difference of the data series is stationary, the next step is to identify the model.

The Autocorrelation (ACF) and Partial Autocorrelation Function (PACF) were also plotted as to collect more conclusive evidence on its stationary condition. The ACF and PACF graphs can be seen in Figure 2. We can conclude that our model is pure AR.

The model maybe AR(2): \( X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + U_t \) or. At the same time it is possible that the form is AR(3) \( X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + U_t \)

Based on the smallest of the AIC, SC and HQ criterion: we should choose the model for the minimum value of AIC, SC, HQ simultaneously. According to the Table3, the results show that the best model for predicting the number of domestic tourists in Hunan Province is AR(2): \( X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + U_t \).

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>SC</th>
<th>HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR(2)</td>
<td>18.25027</td>
<td>18.39940</td>
<td>18.27551</td>
</tr>
<tr>
<td>AR(3)</td>
<td>18.63949</td>
<td>18.78861</td>
<td>18.66472</td>
</tr>
</tbody>
</table>

### 3.3. Estimation of model’s coefficients
Based on Eviews 10.0 version, the Estimation of the model’s coefficients is as following Table 4. The P-value of each coefficient is less than the significance level, so the model is statistically significant.

Table 4. Estimated parameters of the model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3947.466</td>
<td>2427.659</td>
<td>1.626038</td>
<td>0.0348</td>
</tr>
<tr>
<td>$X_t$</td>
<td>1.285885</td>
<td>0.329192</td>
<td>3.906187</td>
<td>0.0014</td>
</tr>
<tr>
<td>$X_{t-1}$</td>
<td>-0.365405</td>
<td>0.372908</td>
<td>-0.979879</td>
<td>0.0427</td>
</tr>
</tbody>
</table>

R-squared 0.895353  Mean dependent var 4107.221
Adjusted R-squared 0.914423  S.D. dependent var 3213.213
S.E. of regression 1138.660  Akaike info criterion 17.22756
Sum squared resid 19448210  Schwarz criterion 17.42639
Log likelihood -159.6618  Hannan-Quinn criter. 17.26121
F-statistic 42.77953  Durbin-Watson stat 1.753243
Prob (F-statistic) 0.000000

Inverted AR Roots .86 .42

From the above data, we can see that the specific form of the model is as follows:

$$X_t = 313.9024963 + 1.285885X_{t-1} - 0.365405X_{t-2} + U_t$$

$X_t$ stands for the first-order difference sequence of the time series of the number of domestic tourists. At the same time, after the correlation test of the residual series, it is found that the null hypothesis is not rejected at the 5% significance level, indicating that the residual series are not correlated.

3.4. Forecasting

AR(2) was applied to forecast the number of domestic tourists in Hunan Province from 2010 to 2019, and the result is as following Figure 3. The value of the Theil Inequality Coefficient is 0.0238, which is very small, indicating that the model's predictive ability is good.
Moreover, the forecast of the number of domestic tourists in Hunan Province from 2020 to 2023 is shown in Table 5 below:

Table 5. Forecast of the number of domestic tourists in Hunan Province from 2020 to 2023

<table>
<thead>
<tr>
<th>year</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicte value</td>
<td>88436.4</td>
<td>95634.4</td>
<td>13212.8</td>
<td>18625.6</td>
</tr>
</tbody>
</table>

The number of domestic tourists in Hunan Province respectively are 88436.4, 95634.4, 13212.8 and 18625.6. Based on the above analysis, it is known that the domestic tourism market in Hunan Province will further expand and the market prospects are broad. This is consistent with Hunan Province's plans and expectations for the local tourism industry.

3. Conclusion

This paper uses time-series related theories to analyze and predict the number of domestic tourists in Hunan by establishing the AR(Autoregressive) model. From the analysis results, the model can accurately predict the dynamic trend of domestic tourists in Hunan Province, and the prediction effect is better. However, the model has some limitations. For instance, the model requires a long time series and uses its own to make predictions. Therefore, the prediction provided by this technology also still have certain error. At the same time, it should be noted that AR is only suitable for predicting phenomena related to the previous period. Hence, we need a better model for forecasting the number of domestic tourists. However, in general, it also proves the rapid development of tourism in Hunan Province. The predicted value is the same as the
tourism plan of Hunan Province, which provides a theoretical basis for Hunan tourism managers to formulate relevant policies.

In order to continue to maintain the development of tourism in Hunan Province, the relevant management departments could take following advices:
1. Optimizing the operation model and improve the ability to resist risks. 2. Creating innovative tourism products to meet diversified tourism needs. 3. Improving service level and strengthen technological empowerment. 4. Adjusting marketing methods to shape brand image.

4. Acknowledgments

I would like to extend my deep gratitude to all those who have offered practical, cordial and selfless support in writing this thesis.

My deepest gratitude is first to professor Zhang, my mentor, for his constant encouragement and guidance. Without his usual and enlightening guidance, this paper would not have been in its present form.

Secondly, I would like to express my heartfelt thanks to my senior Chen, who has led me to write and modify my paper. Without his help, I don’t know how to grasp the writing logic and ideas of the article.

Finally, I thank my parents for their continued support and encouragement.

References


Quan Qin. Forecasting Tourism Market

centers using LSTM recurrent neural networks and ARIMA models. Internet Technology & Secured Transactions. IEEE.


