

# The forecast of the number of tourists received during the National Day period in Beijing under the background of epidemic in 2020

Meng Ying Yang<sup>1</sup>, Xin Liu<sup>2</sup>, Hong Bin Jia<sup>3</sup>

<sup>1</sup> Economics major of Beijing Jiaotong University, Beijing, China

<sup>2</sup> Economics major of Beijing Jiaotong University, Beijing, China

<sup>3</sup> Statistics major of Beijing Jiaotong University, Beijing, China

**Abstract.** COVID-19 suddenly outbreak not only pose a threat to people's life and health, but also interfere with the development of national economy in our country. To investigate the outbreak of the impact of tourism, we choose Beijing as the aim of the research, using autoregressive model, Fourier series model and VAR model. Finally, according to the forecast results, we concluded that outbreak brought direct effects to the Beijing tourism, but with the outbreak under control, under the condition of government's expansionary policies, tourism economy gradually back on track.

## 1 Introduction

In 2020, an unexpected COVID-19 force seriously affected the people's physical and mental health and national economic development in our country. At the peak of the epidemic, almost 2,000 people were diagnosed a day. As of September 25, 2020, COVID-19 has caused 4746 deaths in China. As oversize public health events, COVID-19 quickly caught the attention of the countries, and the government enacted a series of policies to guide people and epidemic prevention against disease.

But the short-term impact of the outbreak on the economy is still inevitable, it has very big effect on importing and exporting trade, catering, culture and entertainment industry, aviation industry. Railways, roads, waterways and civil aviation sent 190 million passengers on a 10-day holiday in 2020, down nearly 73 per cent from the same period last year, according to the ministry of transport; Domestic and international air blockade blocked inbound and outbound tourism. The substantial reduction of travel passengers has directly reduced the source of tourism passenger flow. Travel agencies, hotels, scenic spots, transportation and other industries have experienced a tide of refunds and tickets. But this kind of influence can't reach a long-term effect, not destabilizing our development in the future. Outbreaks in China has stabilized gradually---all walks of life have recovered, and resorts like the scenic spot or cinema also have given back slowly. In order to investigate the impact of the epidemic on the tourism industry, we take the epidemic as the main consideration and predict the number of tourists in Beijing National Day Golden Week in 2020.

Meng Ying Yang: 18241315@bjtu.edu.cn  
Xin Liu: 18241281@bjtu.edu.cn  
Hong Bin Jia: 18271246@bjtu.edu.cn

## 2 Review

Sudden events often have a certain impact on the economy, especially the tourism economy. Lei Fu<sup>[1]</sup> (2002) discussed and constructed the economic and tourism influence mechanism of the Olympic Games according to the principle of festival economic influence, the theory of tourism economy and the characteristics of the Olympic Games itself, and analyzed the tourism effect of the 2000 Sydney Olympic Games. It is considered that the inbound tourism effect of the Olympic Games is more obvious than that of the domestic, and its tourism activities are more concentrated. The regional diffusion effect is not prominent, and the overall strength is general. The prediction of the tourism impact of the Olympic Games should be rational. Xiangming Li<sup>[2]</sup> (2012) believes that maintaining a stable and continuous passenger flow is an important symbol of the success of the World Expo, indicating that as a research indicator, the representative of tourism passenger flow. At the same time, the study proves that the follow-up effect of the World Expo, including investment and consumption, is particularly significant in tourism.

Demand for tourism, and prediction of traffic, academia has been quite a study. Jipeng Li, Mengmeng Chen and Qijian He<sup>[3]</sup> (2018) also use grey theory to predict the national passenger flow from 2018 to 2020 based on the passenger flow data of Xi ' an Rail Transit Line 2. Although there are some discrete data in the prediction process, the prediction model can effectively analyze and process the data under the comprehensive influence of various factors, and has good prediction accuracy and practicability in a certain prediction period.

Based on the theory of BP neural network, Yuwei Zhao, Zhen Zhang, Zibo Guo and Yuanfang Li<sup>[4]</sup> (2020) with artificial neural network technology as the main research methods, predicted Qinhuangdao's numbers of domestic tourists and inbound tourists from 2020 to 2022 using the MATLAB tools to.

Since the outbreak has caused widespread attention at home and abroad. Stefano Maria Iacus, Fabrizio Natale, Carlos Santamaria, Spyridon Spyratos and Michele Vespe<sup>[5]</sup> (2020) implemented a forecasting model based on historical data from January 2010 to October 2019 to set a reference baseline to study the impact of the air travel ban on the aviation industry and its socio-economic impact. In China, Shaohua Wang, Lu Wang, Mengyin Wang and Wei Wang<sup>[6]</sup> (2020) selected Henan Province as the research object. Objective, based on the PSR theoretical model, using questionnaire survey, network text analysis and other methods, the study shows the pressure of tourism 'turnover' caused by the epidemic and analyzes the recovery direction.

### 3 Model Specification

In order to highlight the impact of the epidemic on tourism, this paper selected Beijing, a city with a high level of tourism development, as our research target. We selected the number of tourists received by Beijing National Day in 2008-2019, and predicted the number of tourists received by Beijing National Day in 2020 without the influence of epidemic by constructing the least square linear autoregression model. Compared with the data in previous years, the error was reasonable. Therefore, we selected the predicted number of tourists received by Beijing National Day in 2020 as the final input value. Based on the monthly data of the number of tourists received in Beijing from 2009 to 2009, a trigonometric regression model (the sum of the squares of the prediction errors) is constructed, and the model fits well, so the monthly tourist reception in Beijing from January to August in 2020 without the influence of the epidemic was subtracted from the actual value to get the difference value, which was regarded as the impact value of the epidemic on the number of tourists in Beijing. At the same time, the difference between the predicted value and the actual value in 19 years and from January to August in 2020 was selected to reduce the noise, and the linear regression model was constructed with Baidu index (nucleic acid detection and epidemic situation), and the results showed that the linear fitting degree was good. Then, the monthly data of Baidu Index and the two-dimensional ARMA model were used to predict the data of Baidu Index in October, which were substituted into the linear regression model above to obtain the impact of the epidemic on Beijing tourism in October 2020. Finally, according to the average proportion of 2009-19 National Day holidays and the number of tourists received in October, the monthly influence is converted into the influence of National Day holiday outbreaks, and combined with the predicted number of National Day tourists in Beijing without the influence of the epidemic, the final forecast result is obtained.

### 3.1 Model

First, we split the predicted object  $y_{gq}$  as follows:

$$y_{gq} = x_{gq} - z_{gq} \quad (1)$$

Among them,  $y_{gq}$  is the number of tourists in Beijing in the "National Day" forecast target,  $x_{gq}$  is the predicted value of the number of tourists in National Day without the influence of the epidemic, and  $z_{gq}$  is the predicted value of the impact of the epidemic on the number of tourists.

According to this split, the population prediction can be divided into two steps: firstly, the variation of variables under the assumption that there is no epidemic influence is analyzed, and then the influence items of epidemic are added to correct the predicted target value under the real situation. Following, we will build the model according to the order of  $x_{gq}$ 's prediction and  $z_{gq}$ 's construction.

#### 3.1.1 the Prediction-autoregressive model of $x_{gq}$

First of all, we conducted ADF test on the stationarity of logarithmic data series  $\ln(x_{gq, t})$  to test the stationarity of time series, and the ADF test confidence was 89.0%. We have reasons to believe that this series is stable, and this result will be the basis for the establishment of the autoregressive model below.

On the assumption that there is no epidemic, we use the least-squares linear regression model with time as the variable and the number of tourists in Beijing during the "National Day" in previous years to forecast the number of tourists this year. The model is as follows:

$$\ln(x_{gq, t}) = c_1 \ln(x_{gq, t-1}) + c_2 YB + c_0 \quad (2)$$

Where,  $\ln(x_{gq, t-1})$  represents the autoregressive quantity lagging a time indicator, and Yb is a dummy variable representing the occurrence or not of an event. Considering the abnormal outlier of the number of National Day people in 2009, we assume that the value of Yb in that year is 1, and the value in other years is zero.

The regression coefficients  $c_0$ ,  $c_1$ ,  $c_2$  can be obtained according to the least square estimation method, from which  $x_{gq}$  can be better predicted.

The goodness of fit value of this model is  $R^2 = 0.755$ , and the adjusted goodness of fit is  $0.694 > 0.5$ . The model has relatively good fitting effect. In addition, the t-test p values of  $YB, \ln(x_{gq, t-1})$  are 0.0361 and 0.0266, respectively, which are all less than 0.05, indicating that the regression effect of variables  $YB, \ln(x_{gq, t-1})$  can be considered significant at the significance level of 0.05. As for the significance of the whole equation, we calculated that the p-value of the F statistic of the model was 0.00358, and there was a 99.7% certainty that the equation had a significant effect. The above statistical tests ensure the goodness of fit of the predicted results and the significance of the variables and the model, which shows that the model has certain effectiveness from a statistical point of view.

### 3.1.2 Prediction of epidemic impact size $z_{gq}$

First of all, the number of monthly tourists in Beijing from 2009 to 2019 is decomposed according to the decomposition formula  $y_t = x_t - z_t$  and get  $z_t = x_t - y_t$ , where  $y_t$  is the real number of tourists in Beijing,  $x_t$  is the predicted value of the number of tourists without the impact of the epidemic, and  $z_t$  is the impact value of the epidemic on the number of tourists.

Wherein, the Fourier series regression model is used to fit  $x_t$ :

$$x_t = \sum_{i=1}^k a_i \sin(b_i t + c_i) \quad (3)$$

$t$  is the time index, and  $a$ ,  $b$  and  $c$  are the regression coefficients.

According to this model, by substituting the monthly data of the number of tourists in Beijing from 2009 to 2019, the number of tourists in January-September 2020 without the impact of the epidemic can be obtained under ideal conditions. Thus, the  $z_t$  fitting value of January-September 2020 can be obtained, namely, the impact of the epidemic on the number of tourists:

$$z_t = \sum_{i=1}^k a_i \sin(b_i t + c_i) - y_t \quad (4)$$

So far, in order to make a more accurate prediction of  $Z_t$ , we further analyzed the inducement of  $Z_t$ . Combined with the literature experience, we chose Baidu index -- "epidemic" and "nucleic acid" after the test, established a linear regression model, using the change of Baidu index, that is, the degree of Internet users' attention to these two words to regression  $Z_t$ , and get the predicted value of October.

The linear regression model is as follows:

$$Z_t = \alpha_1 B_{yq} + \alpha_2 B_{hs} + c \quad (5)$$

$B_{yq}$  and  $B_{hs}$  are the monthly data from January to August of the "epidemic" and "nucleic acid", and  $C$  is the constant term.

The goodness of fit value of this model is  $R^2 = 0.882$ , and the adjusted goodness of fit is 0.867, less than 0.5, which is not significantly different from the goodness of fit. The model has a relatively good fitting effect. Additional  $B_{yq}$ ,  $B_{hs}$ 's  $p$  values were 0.0003 and 0.0001, are less than 0.05, shows that under the significance level of 0.05, we can assume that the regression effect of variable  $YB, \ln(x_{gq, t-1})$  is significant. Regarding the significance of the whole equation, we calculated that the  $P$  value of the  $F$  statistic of the model is 0.00358, which is less than 0.05, and there is a certainty of at least 99.7% that the equation has a significant effect. The above statistical tests ensure the goodness of fit of the predicted results and the significance of the variables and the model, which shows that the model has certain effectiveness from a statistical point of view. Thus, Baidu index data of October can be substituted into the model to predict the impact of the epidemic in October. To forecast the Baidu Index in October, we will use the monthly data of the Baidu Index from January to August and the vector autoregressive VAR model.

Model was built as follow:

$$\begin{cases} \ln(B_{yq,t}) = c_{1,1} \ln(B_{hs,t-1}) + c_{1,2} \ln(B_{hs,t-1}) \\ \quad + c_{1,3} \ln(B_{yq,t-1}) + c_{1,4} \ln(B_{yq,t-1}) + c_{1,5} \\ \ln(B_{hs,t}) = c_{2,2} \ln(B_{yq,t-1}) + c_{2,2} \ln(B_{yq,t-1}) \\ \quad + c_{2,3} \ln(B_{hs,t-1}) + c_{2,4} \ln(B_{hs,t-1}) + c_{2,5} \end{cases} \quad (6)$$

Where  $\ln(B_{gq,t-1})$  represents the autoregressive quantity lagging one time indicator, and  $c$  are regression coefficients.

After testing the lagged structure of the VAR model, we found that the reciprocal of the AR root module of the model was 0.5942 (double) and 0.322 (double), both of which were located in the unit circle. This result indicated that the VAR model met the stability test and ensured the feasibility of using the model to predict.

In addition, the goodness of fit value of the model is  $R^2=0.988$ , and the adjusted goodness of fit is 0.939, less than 0.5. The model has a very good fitting effect. Intuitively, this is because the model considers the endogenous influence and lag factors between the two variables at the same time, which is very close to the real situation.

Thus, the above model was substituted to obtain the impact value of the epidemic in October:  $z_{t,10}$

$$z_{t,10} = \alpha_1 B_{yq,10} + \alpha_2 B_{hs,10} + c \quad (7)$$

Then we do the final processing for  $z_{gq}$ . Using the average ratio of the number of tourists with National Day in previous years to the total number of tourists in October, we get the proportion of  $A$  in the number of tourists in October, which is  $\alpha_{gq}$ . So far, we get  $z_{gq} = \alpha_{gq} z_{t,10} = \alpha_{gq} (\alpha_1 B_{yq,10} + \alpha_2 B_{hs,10} + c)$  is the part of the number of tourists affected by the epidemic during National Day.

### 3.1.3 Final model

Based on the conclusions of 3.1.1 and 3.1.2, the prediction model of the number of tourists in Beijing during National Day can be obtained as follows:

$$y_{gq} = e^{c_1 \ln(x_{gq, t-1}) + c_2 YB + c_0} \alpha_{gq} (\alpha_1 B_{yq,10} + \alpha_2 B_{hs,10} + c) \quad (8)$$

## 3.2 Sample selection

### 3.2.1 Monthly number of tourists in Beijing

According to the needs of the model, this paper extracted the number of visitors received in the tourist area of Beijing from Beijing Municipal Culture and Tourism Bureau to represent the number of monthly tourists in Beijing, and the time range is from January 2008 to August 2020. Among them, the data of July and August 2020 are used to test the constructed  $z_{gq}$  model, and the remaining data are used to construct the  $z_{gq}$  model.

### 3.2.2 Baidu index

Based on previous experience and reference to similar articles predicting the impact of the epidemic, the following 10 keywords were selected as initial keywords in this paper: yiqing, xinguanfeiyan, Beijing yiqing,

hesuanjianc yimiao, geli, Covid19, kouzhao, haixian and bianfu. After the initial keyword database is constructed, we searched in the Baidu index database successively, and found that each word was included in the Baidu index, which indicates that the initial keyword database constructed in this paper is reasonable.

By constructing the model of Baidu index of the above-mentioned keywords and  $Z_t$ , yiqing and hesuanjianc were selected as the final keywords in this paper The Baidu index of the two is extracted from Baidu to carry out the next step.

### 3.2.3 Number of tourists in Beijing during the National Day holiday

According to the needs of the model, we collected the total number of tourists in Beijing during the National Day period in the past years through the National Day Holiday Culture and Tourism Market Situation, National Day Holiday Travel Market Situation and "Tourism statistics report of "National Day" released by Beijing Municipal Culture and Tourism Bureau, former National Tourism Administration and National Holiday Office. and the time range of the data is from 2008 to 2019.

## 4 Parameter estimation and prediction results

Based on the above data and the established prediction model, Eviews software is used as the tool to obtain the results:

$$y_{gq} = e^{c_1 \ln(x_{gq, 9}) + c_2 YB + c_0} - \alpha_{gq} (\alpha_1 B_{yq,10} + \alpha_2 B_{hs,10} + c)$$

Where:

$$c_1 = 0.407, c_2 Y \text{ is } 0 \text{ at } 2020, \text{ and } c_0 = 4.18.$$

By the usual data calculates out  $\alpha_{gq} = 0.364$ ,  $\alpha_1 = 0.00151$ ,  $\alpha_2 = 8.23 \times 10^{-5}$ ,  $C = 63.8$ .

In addition, the time series prediction of Baidu index is solved as follows:

$$\begin{cases} \ln(B_{yq,t}) = 0.129 \ln(B_{hs,t-1}) + 0.0038 \ln(B_{hs,t-2}) \\ \quad - 0.699 \ln(B_{yq,t-1}) - 0.451 \ln(B_{yq,t-2}) + 16.6 \\ \ln(B_{hs,t}) = -0.199 \ln(B_{yq,t-1}) + 0.821 \ln(B_{yq,t-2}) \\ \quad + 0.693 \ln(B_{hs,t-1}) - 0.103 \ln(B_{hs,t-2}) + 9.90 \end{cases}$$

Where:

$$B_{yq,10} = 3884093, B_{hs,10} = 492308.1$$

The above results are combined into the  $y_{gq}$  expression. Get:  $y_{gq} = 667.186$

According to the statistical uncertainty of the model itself, we can roughly give the confidence interval of the final predicted value under the condition of certain assumptions.

We assume that the errors caused by the prediction of Fourier series during the model establishment and the errors caused by the VAR model are ignored, and the source of random errors is selected as the errors of the National Day number predicted by the autoregressive model.

The confidence interval was calculated

as  $[e^{6.78}, e^{7.15}]$ . Substituting the decomposition formula above, the final confidence interval was obtained as  $[467.92, 861.95]$ .

After model establishment, data entry and model test, this paper finally concludes that the number of tourists received by Beijing during the National Day period in 2020 will be about 6.67186 million.

## 5 advantages and disadvantages

### 5.1 advantages

This study predicts the number of tourists received by Beijing during the period of National Day in 2020. By using the basic theories of linear regression model, Trigonometric series regression model and ARMA autoregression moving average model, we establish a prediction model for the number of tourists in Beijing during National Day, and successfully predict the number of tourists coming to Beijing during National Day.

It is worth mentioning that, when predicting the difference  $Z_t$  between the real value and the predicted value. Instead of using the traditional method of regression prediction of difference and time  $t$ , we extracted the Baidu index of the keyword database related to this epidemic, conducted regression modeling of difference and Baidu index and made prediction, thus improving the accuracy of prediction.

### 5.2 Disadvantages

First of all, since the COVID-19 outbreak occurred in early 2020, the observation period of the epidemic data is limited, so the data deficiency is inevitable in terms of quantity.

Although we have taken many measures to make up for this defect, for example, when we carry out regression modeling between the difference and the Baidu index, we selected the difference value between the predicted value and the actual value in the whole year of 2019 and from January to August in 2020, so as to mitigate the effects of insufficient sample size. However, the influence of this defect on the prediction accuracy cannot be completely improved.

Second, the forecast is a relatively complex process. Therefore, three steps of prediction were carried out in the process.

As is known to all, the occurrence of errors is inevitable in the forecasting process. Therefore, the errors generated after the three predictions are carried out have a great impact on the accuracy of the forecast.

Finally, there are some factors that cannot be quantified in this forecasting process. For example, a sudden change in policy in response to the epidemic. It is unfortunate that such sudden changes inevitably have some effect on our forecasting process.

## 6 The practical significance

At the beginning of 2020, Covid-19 was rampant all over

the world, and China restricted the flow of personnel for the safety of the people, making the tourism industry undoubtedly one of the industries seriously affected by the epidemic of the new crest and suffering huge economic losses.

Through the establishment of a scientific model, we predict the number of tourists received by Beijing during the National Day period, hoping to provide reliable data for reference.

When the content of this article was completed, the official data of the number of tourists to Beijing during the National Day holiday had been released. Through comparison, it was found that the actual number of tourists to Beijing was greater than the number predicted in this article. It proves that China's epidemic prevention and control work is quite successful, which makes people have strong confidence in the prevention and control of the epidemic. We have every reason to believe that with the right guidance from the government, the tourism industry will gradually recover from the Covid-19 epidemic and return to normal.

## Acknowledgments

This paper is one of the phased achievements of the innovation and entrepreneurship training program for college students in Beijing *The impact of epidemic situation on Beijing Tourism and the Countermeasures* (202110004101)

## References

1. Fu. L. (2002). Research on the impact of the Olympic Games: economy and Tourism <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CDFD9908&filename=2002124103.nh>
2. Li. XM. (2012) The tourism effect of the World Expo in the post Expo era and Its Countermeasures -- Taking Shanghai Expo as an example: Journal of Jiangxi University of Finance and Economics, **01**:28-36.
3. Li. JP, Chen. MM, He. QJ. (2018) Forecast of national day passenger flow of Urban Rail Transit Based on Grey Theory: Qinghai Transportation Science and Technology, **06**:1-4.
4. Zhao. YW, Zhang. Z, Guo. ZB, Li. YF. (2020) A study on the forecast of tourism volume in Qinhuangdao based on BP network: Network Security Technology & Application, **09**:125-126.
5. Iacus, SM, Natale, F, Santamaria, C, Spyratos, S, Vespe, M. (2020) Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact. Safety Science, **129**: 104791 -104791.
6. Wang. SH, Wang. L, Wang. MY, Wang. W. (2020) The impact of new crown pneumonia on Tourism in Henan Province: Areal Research and Development, **39**:1-7.