Proceedings of the 7th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/32/20231614

Assessing How COVID-19 Pandemic Affects the Tourism Industry in Mainland China

Jinhao Li[1,a,†], and Ruihan Xiong[2,a,*,†]

[1]The High School Affiliated to Renmin University of China, Beijing, 100097, China
[2]School of Economics, University of Edinburgh, Edinburgh EH8 9YL, Britain
a. s2050880@edu.ac.uk
**corresponding author*
[†]These authors contributed equally.

Abstract: The study primarily contends that the rise in COVID-19 cases has adversely impacted the volume of tourist visits. This investigation focuses on how the pandemic affected the hotel sector during the outbreak. Throughout the entire economy, the hospitality sector may be among the areas most significantly impacted by the pandemic. Hence, employing linear regression and ARIMA models, this investigation looks at the relationship between the number of COVID-19 instances. and tourism expenditure, overall spending on star-rated hotels, and per capita GDP in mainland China. These factors can represent tourists' budgets and wiliness to spend during the pandemic. Furthermore, the research also considers whole-city quarantine and health codes to show the restriction when tourists travel during the pandemic. According to the test findings, the research reveals that a 1% rise in COVID-19 cases correlates with an estimated 0.108% reduction in tourist numbers.

Keywords: COVID-19, tourism industry, ARIMA model

1. Introduction

The pandemic has profoundly impacted the countries it afflicted, resulting in over 760 million confirmed cases and 6.8 million fatalities, and also seriously impacting the health system of the country [1,2]. In the second part of 2019, Wuhan, China, saw the emergence of it. When the government and medical workers first found its high transmission rate and specific lethality, several policies were published to protect people's health. However, the pandemic had a catastrophic effect on the country's economy and many industries. From January 2020 to January 2022, the sectors most impacted by COVID-19 encompassed airlines, automotive, energy equipment and services, hospitality, dining establishments, recreation, and niche retail [3].

As one of the most significantly affected domains, the tourism sector experienced an unprecedented drop in both international and domestic travel due to travel limitations, border shutdowns, and social distancing protocols. One industry that was also seriously affected is the hotel industry. Safety concerns and policy factors have also led to a decrease in passenger numbers. Due to travel constraints, quarantine measures, and transportation disruptions, arrivals of international tourists have dropped by 22% in the first quarter of 2020, with a potential annual drop ranging between 60-80% [4]. Studies have shown that the number of visitors to Japan has dropped by 99 percent since the outbreak [5]. According to the United Nations World Tourism Organization

DOI: 10.54254/2754-1169/32/20231614

(UNTWO) [6], the hotel and tourism businesses have been significantly more affected by this pandemic issue., resulting in a 60–80% decline in overall activity in 2020. The restriction limited the number of tourists and decreased the demand for living in hotels.

This study aims to demonstrate a connection between the number of infection cases and the overall number of travelers during the outbreak. By emphasizing the pandemic's influence on the hospitality and tourism sectors, the study seeks to offer valuable insights for the government in recovering from the devastating effects on the hotel industry. Some scholars wrote about the significant effect on the whole economy from several sectors of the world economy, which include raw material suppliers like agriculture, manufacturing, and service providers, such as education [7]. Another study mainly argues the tourism and hospitality industry's resilience and how the company copes with the crisis [8]. These rational and feasible solutions allow the study to look at the future possibilities of the hotel industry and pay attention to the cost of the hotel itself. In addition, based on the empirical research. Some illustrates that nearly half of the questioned company's financial problems also provide a sample of the significant impact of the pandemic. Furthermore, it mainly argues the impact of coronavirus on per capita income distribution and mixes the impact of inequalities, which supports the study's view about the impact on the willingness of tourists [5].

To provide a more complete view, this research, on the other hand, principally investigates if there is an inverse link between the escalating COVID-19 cases and monthly tourist arrivals. This study uses the multiple linear regression model and the Auto-Regressive Integrated Moving Average (ARIMA) model, followed by the results showing the number of infected cases and self-quarantine have a significant negative impact on tourism.

2. Methodology

The purpose of this essay is to examine how the epidemic has impacted the Chinese mainland's tourism sector. To test the hypothesis of the monthly number of tourists will be adversely influenced by the increasing number of COVID-19 cases, after controlling for other factors. The methods used in this study include multiple linear regression model and the ARIMA model.

2.1. Data

This study uses number of cases data from the World Health Organization (WHO) provides a record of daily COVID-19 infected cases. Complementary economic and tourism data, including GDP per capita, total tourism expenditure, and the number of international tourists, were collected from the National Bureau of Statistics of China. Additionally, data of Chinese star-rated hotels were obtained from the Ministry of Culture and Tourism, China. These variables were selected for the following reasons. GDP per capita, tourism expenditure, and international tourist volume are crucial indicators of the overall economic landscape within the tourism industry, which enables a thorough evaluation of the pandemic's impact on this sector. Moreover, data regarding Chinese star-rated hotels provide essential insights into the supply aspect of the tourism market, showcasing the accessibility and caliber of lodging options. This data facilitates the analysis of the complex relationship between the pandemic and the tourism infrastructure.

2.2. Linear Regression Model

In this section, the linear regression model is employed to examine the hypothesis that a rise in the number of COVID-19 cases will adversely affect the overall count of tourists.

$$lnTourism=\beta_1 lncases_{it}+\beta_2 lnGDP_{it}+\beta_3 lnSpending_{it}+\beta_4 lnHotel_{it}+\beta_5 Code_i+\beta_6 CQ_i+\beta_7 IQ_i+year_t+u_{it} \quad (1)$$

DOI: 10.54254/2754-1169/32/20231614

The outcome variable lnTourism, is the tourism industry's annual income by using the natural logarithm. As this article aims to find the effect of the number of Covid-19 cases on the number of tourists, the variable lnCases has been selected as the key variable and it indicates the natural logarithm of Covid cases increased daily during the pandemic in Mainland, China. For the control variables, lnGDP, lnSpending and lnHotel has been chosen. lnGDP denotes the natural logarithm of monthly GDP per capita in Mainland, China. lnspending indicates the natural logarithm of the total amount of spending on tourism of domestic tourists, while lnHotel is the natural logarithm of the revenue of Chinese star-rated hotels during the pandemic. Code, CQ, and IQ are dummy variables which indicate whether the tourists need a health code to travel to other cities, whether is a whole city lockdown and whether the individuals need to self-quarantine.

The dataset is describing only the effect of the Chinese tourism industry during the year 2020 – 2022, which were the years that were most influenced by Covid-19. For the sign of coefficient, it has been expected to be negative and at the same time, the change in the number of cases is predicted to have a big impact on how many tourists there are.

2.3. ARIMA Model

In this case, regarding it being a time series dataset, the use of ARIMA model will be considered as it is suitable for both stationary and non-stationary time series data by applying differencing techniques. Predominantly, practitioners employ the ARIMA model for future trend prediction. However, the COVID-19 pandemic, which persisted from 2020 to 2022, significantly altered prevailing circumstances. Consequently, the post-pandemic landscape is expected to deviate from prior patterns, rendering ARIMA-based projections potentially ineffectual for capturing the underlying dynamics in the forthcoming period [9].

In the ARIMA model, there are basically three key factors that need to be determined before applying it, which are the autoregressive component, whether the data need to be differentiated and the moving average component. The autoregressive component operates on the principle that the present value of a time series can be estimated by utilizing its preceding values. The AR(p) model is expressed as a linear combination of past observations with a specified number of lags. Before applying the AR part to the data, the number of lags should be determined. The AR part captures the persistence or autocorrelation structure in the data, which helps in understanding the temporal dependence between observations. Conversely, the moving average component represents the error term as a linear combination of its previous values. It captures the influence of random shocks or noise in the data that cannot be explained by the autoregressive component alone. An MA(q) model signifies the quantity of lagged error terms incorporated into the model. The ARIMA model in this article will mainly use autocorrelation and residual analysis functions.

This study first conducted a unit-root test on the natural logarithm of tourism to detect the stationarity of the time series. The test statistic is -3.618, with a p-value 0.0054, which supports the argument that *lnTourism* is a stationary times series at 5% level and does not require differencing for fitting an ARIMA model.

Before the ARIMA model is operated, this study conducts AIC (Akaike's information criterion) and BIC (Bayesian information criterion) test to select the best number of autoregressive orders, p, and that of moving average order, q. After a series of comparisons on different values of AR orders and MA orders, it has been found that the best-fit specification is ARIMA (1,0,0), which is equivalent to the AR (1) model below.

$$lnT_t=\alpha+\phi_1*lnT_{t-1}+\varepsilon_t \quad (2)$$

DOI: 10.54254/2754-1169/32/20231614

This AR (1) model analyzes the autoregression of the change in the revenue of tourism in one lag and this is the model's final use in the regression section.

3. Results

3.1. Descriptive Statistics

Table 1 below is a summarization of all the variables that take part in the analysis, for both the original data and the data treated by the natural logarithm. This study utilizes the natural logarithm of tourists as the dependent variable in the descriptive statistics analysis. The use of *lnTourism* instead of raw tourism data is justified by the ability of natural logarithms to stabilize variance and provide percentage changes, facilitating a clearer understanding of the relationships between variables. The primary explanatory variables include the natural logarithm on the number of COVID-19 cases, total spending on tourism during the pandemic, total spending on star-level hotels, and GDP per capita in mainland China.

The inclusion of COVID-19 case numbers is driven by the initial hypothesis that an increase in cases will lead to a decline in tourist numbers. The number of cases here is using the natural log of infected cases. Total spending on tourism and star-level hotels was chosen as these factors represent key consumption indicators, reflecting the willingness and activity of tourists during the pandemic. Additionally, GDP per capita is a crucial variable to determine whether tourists possess sufficient funds to travel.

Table 1: Descriptive statistics.

Variables	Mean	Standard Deviation	No. of Observation
Tourists (billion)	0.654	0.237	36
No. of cases (million)	0.377	1.418	36
Total Spending on Tourism(billion)	199.781	31.800	36
Total Spending on Hotel (billion)	10.496	0.732	36
GDP per Capita	6202.111	419.393	36
Log of Tourists	20.228	0.398	36
Log of Cases	12.611	1.997	36
Log of Spending on Tourism	26.009	0.154	36
Log of Spending on Hotel	23.072	0.069	36
Whole City Quarantine	0.39	0.49	36
Self-Quarantine	0.389	0.494	36

Beyond these descriptive statistics, the analysis also considers whole-city quarantine and self-quarantine measures to demonstrate the restrictions faced by tourists when travelling between cities during the pandemic.

3.2. Linear Regression Results

Table 2 below shows the regression result of the linear regression without controls and with controls in two columns. The results in column (1) shows the impact of between *lnCases* on *lnTourism*. The adjusted R-squared (which accounts for the number of predictors in the model) is 0.273, which means after the adjustment there is 27.3% of *InTourism* can be explained by *lnCases*. The coefficient is -0.108 and demonstrates statistical significance at the 1% level. This implies that a 1% rise in COVID-

DOI: 10.54254/2754-1169/32/20231614

19 cases correlates with an estimated 0.108% reduction in tourist numbers, keeping all other variables constant.

The Results in column 2 include additional control variables into the regression model. The adjusted R-squared value of 0.568 indicates that around 56.8% of the variability in *lnTourism* can be explained by the changes in all included independent variables. In column (2) the *lnCases* coefficient is -0.051. However, it lacks statistical significance. This indicates that when accounting for the control variables, the negative impact from the number of COVID-19 cases on tourists is no longer statistically significant. While the dummy variable self-quarantine is statistically significant in this case and it indicates that once the self-quarantine policy is applied, tourism will decrease 71.5 percent, which is a large effect on tourism.

Table 2: Linear regression results (Dependent variable: Log of tourism).

	(1)	(2)
Log of Cases	-0.108***	-0.051
	(0.029)	(0.043)
Log of Total Spending on Tourism		-64.526
		(46.530)
Log of Total Spending on Hotel		142.778
		(103.720)
Whole City Quarantine		-0.095
		(0.099)
Self-Quarantine		-0.715***
		(0.143)
Adjusted R^2	0.273	0.568
F-stats	0.001	0.000
Observations	36	36

Notes: * p<0.05 ** p<0.01 *** p<0.001; standard errors reported in parentheses.

Comparing the two models, it can be observed that Adjusted R^2 is higher in the model in column (2), indicating an improvement in the model's explanatory power.

In column (1), the *lnCases* coefficient is statistically significant and negative, implying a reverse correlation between the COVID-19 case count and the number of tourists. However, this relationship becomes statistically insignificant in the controlled model, possibly indicating that other control variables play a more significant role in explaining the variation in *lnTourism*.

3.3. ARIMA Regression (AR (1) Model)

Table 3 below shows the results of the ARIMA model, which is actually an AR (1) model after the comparison of different numbers of q, d, and q. The log of tourism(t-1) refers to the lagged value of the natural log of tourism. ARMA stands for Autoregressive Moving Average, which represents the non-integrated part of the model. The p-value mentioned below corresponds to the chi-squared test statistic, which seeks to ascertain the presence of a significant association between categorical variables within a sample.

The positive and significant coefficient suggests that a 1% increase in the previous month's *lnTourism* value corresponds to a 0.493% increase in the current month's *lnTourism* value, which is statistically significant at 5% level with a p-value of 0.022.

DOI: 10.54254/2754-1169/32/20231614

Table 3: ARIMA model results (Dependent variable: Log of tourism).

	Coefficient	Standard Error	P > |z|
Log of Tourism(t-1)	0.493*	0.215	0.022
Prob > χ^2	0.022		
No. of observation	36		

Notes: * p<0.05 ** p<0.01 *** p<0.001; standard errors reported in parentheses.

3.4. Discussion

Using linear regression and ARIMA (which is an AR (1)) models on data from 2020 to 2022, the analysis supports the hypothesis that the number of tourists is negatively affected by the number of COVID-19 cases. The multiple linear regression model revealed that the self-quarantine requirement for tourists is the most influential factor, showing statistical significance at the 0.001 level. This finding aligns with the research that demonstrated that quarantine measures are crucial to preventing the spread of COVID-19 [10].

Both the linear regression and ARIMA models' findings that there is a negative correlation between the number of cases and tourism support earlier studies [11]. Such negative impact on tourism and hotel industries could be resulted from factors such as travel restrictions, safety concerns and shifts in consumer behaviour.

This work adds concrete data to the body of literature showing a link between COVID-19 infections and tourism, highlighting the critical importance of self-quarantine measures in this situation. These findings have substantial implications for policymakers and the tourism industry as they navigate the pandemic's challenges. While implementing measures such as self-quarantine regulations, travel limitations, and other safety protocols, concerns about public health must be balanced with the tourism industry's economic resurgence.

4. Conclusion

The hotel and tourism sectors were particularly hard hit by COVID-19 and had substantial crashes. This study looked at how the number of COVID-19 cases affects the tourism industry as first-hand witnesses and within the context of COVID-19. The pandemic has affected hotel spending, indicating its impact on tourists' willingness to travel. Moreover, the government-imposed restrictions have also contributed to the sharp plunge in international travelling in 2020. This study examines the hypothesis using the linear regression and ARIMA models, and the findings indicate that the rising number of COVID-19 cases have an adverse impact on the volume of tourists. According to the results, a one percent increase in the count of COVID-19 cases results in a 0.108% decline in tourist numbers.

This study has some suggestions for improving awareness of unexpected shocks in the future. For example, unexpected economic shocks will undoubtedly significantly impact the tourism and hotel industry. However, there are some limitations in the study. The paper does not consider several factors affecting infection, like the inflation rate during the pandemic and the preference of consumers. Hence, there may be some biased estimation. Some other aspects the research can take into account may include the vast country of infection and higher quality of data.

References

[1] *World Health Organization (WHO): Pneumonia of unknown cause – China. https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229. (2020).*

[2] *Abraham P.: The aftermath of the corona virus pandemic. International Journal of Hospitality Management, (2021).*

DOI: 10.54254/2754-1169/32/20231614

[3] *Vidovic, L.: Industries Most and Least Impacted by COVID-19 from a Probability of Default Perspective. https://www.spglobal.com/marketintelligence/ (2022).*
[4] *Ide, A.: Tourism and ICT Solutions in the COVID-19 Era: A Comparison Between Japan and Sweden. Rev Socionetwork Strat 15, 195–211 (2021).*
[5] *Dreger, C.: Economic impact of the Corona pandemic: Costs and the recovery after the crisis. Asia and the Global Economy 2(1), 2667-1115 (2022).*
[6] *Crespí-Cladera, R., Martín-Oliver, A., Pascual-Fuster, B.: Financial distress in the hospitality industry during the Covid-19 disaster. Tourism Management, 85 (2021).*
[7] *Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., Agha, Riaz.: The socio-economic implications of the coronavirus pandemic (COVID-19): A review. International Journal of Surgery 78, 1743-9191 (2020).*
[8] *Duarte Alonso, A., Kok, S. K., Bressan, A., O'Shea, Mi., Sakellarios, N., Koresis, Al., Buitrago Solis, M. A., Santoni, L. J.: COVID-19, aftermath, impacts, and hospitality firms: An international perspective. International Journal of Hospitality Management 91. 0278-4319 (2020).*
[9] *Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S., Ciccozzi, M.: Application of the ARIMA model on the COVID-2019 epidemic dataset. Data in Brief, 29, 105340 (2020).*
[10] *Altuntas, F., Altuntas, S., & Dereli, T.: Social network analysis of tourism data: A case study of quarantine decisions in COVID-19 pandemic. International Journal of Information Management Data Insights, 2(2), (2022).*
[11] *Škare, M., Soriano, D. R., Porada-Rochoń, M.: Impact of COVID-19 on the Travel and Tourism Industry. Technological Forecasting and Social Change, 163(1), (2016).*