Studying the Influencing Factors of Tourism Economy Among Provinces in Mainland China in the Context of Big Data Based on Spatial Econometric Analysis

Yuting Wang, Wen-Tsao Pan(✉), Yanmei Liang, Han Lv, and Wenjing Dai

School of Business, Guangdong University of Foreign Studies, Guangzhou 510420, China
201630021@oamail.gdufs.edu.cn

Abstract. At the end of 2019, the new crown epidemic spread from a local outbreak to the whole country, causing a huge impact on the tourism industry everywhere. With limited resources, in order to effectively improve the tourism industry input-output ratio, this paper uses a spatial econometric model based on Geoda software to assess the development of tourism efficiency across the country during 2015–2019. The results show that: the number of travel agencies and the number of people working in the tourism industry in the context of the epidemic have a greater impact on tourism efficiency; there is a spatial correlation between tourism efficiency in China, and tourism efficiency is highest in the eastern region. This paper provides a way for provinces to use tourism information data more efficiently, which has practical significance for the development of regionalization and informationization of tourism.

Keywords: tourism efficiency · spatial econometric model · COVID-19 Pandemic

1 Introduction

From 2016 to 2019, the number of tourist trips in China grew gradually from 2016 to the end of 2019, as China's economic growth became more and more pronounced and the people's need to satisfy their well-being became stronger and people would rely on cultural activities such as tourism to enrich their lives. At the end of 2019, the Covid-19 Pandemic swept the world, resulting in a significant decline in Chinese tourist arrivals. In 2020, there is a negative growth with 2.88 billion tourist trips, which is a growth rate of −52.4% compared to 2019. Since the tourism economy is an important indicator of the national economy, the study of the tourism economy is of particular meaning, as there is a small outbreak of the Covid-19 Pandemic that affects not only the tourism industry in the region but also the surrounding areas. Therefore, this paper uses a spatial econometric model to analyze and study the degree of mutual influence of tourism economic efficiency among regions in China, which can help provinces to better use relevant tourism data to improve tourism input and output efficiency, and also make

A. Luqman et al. (Eds.): PMIS 2022, AHCS 6, pp. 427–435, 2023.
https://doi.org/10.2991/978-94-6463-016-9_45

Fig. 1. Trends in the number of Chinese tourists from 2016 to 2020

better use of tourism data as well as communication data in the case of local outbreak of epidemic, so as to avoid the risks caused by the movement of people (Fig. 1).

2 Research Preparation

2.1 Research Methodology

2.1.1 Spatial Econometric Model

Spatial econometrics was first proposed by J. Paelinck at the annual meeting of the Statistical Society of the Netherlands. As scholars continue to pioneer spatial autoregressive models, there are now three classical models in spatial econometric models that are widely used in research, namely, the spatial autoregressive model (SAR), the spatial error model (SEM) and the spatial Durbin model (SDM). Among them, the spatial autocorrelation model aims to investigate whether a variable is correlated over a geospatial region and to what extent, and the spatial autocorrelation coefficient is a commonly used metric. It is divided into global spatial autocorrelation and local spatial autocorrelation. This study uses this method to empirically analyze the global and local spatial autocorrelation of the Chinese territory.

2.1.2 Hausman Test

Hausman test is a general line of testing hypotheses. Hausman test can help to solve the problem of which effect is better to use when there is a significant difference in coefficient estimates between fixed and random effects. The result of the Hausman test provides a choice of options.

2.2 Indicator System Establishment

2.2.1 Indicator Selection

Due to the broad scope of tourism, scholars have chosen different indicators in previous related studies. As traditionally defined in economics, industrial efficiency refers to the

degree of effectiveness in converting inputs into outputs in a given industry. Therefore, tourism efficiency is a measure of the effectiveness of the conversion between tourism-related resource inputs and final outputs. Accordingly, this study divides the indicator system into two major segments: input and output.

Input indicators. The basic factors of production usually include labor, capital, and land factors. Accordingly, this paper selects the reference indicators as follows: ① The number of people employed in the tourism industry. From the tourism industry level, laborers supply labor through employment, so the number of people employed in tourism is chosen as the labor input indicator. ② The number of travel agencies and star-rated hotels. Generally speaking, the abundance of tourism-related industries such as travel agencies also reflects the development of local tourism and the government's attitude towards tourism development. Since it is difficult to obtain specific fixed asset star-rated hotels is chosen to characterize capital investment, this article selects the number of travel agencies and star-rated hotels to represent the capital investment. ③ The number of A-class scenic spots. Land in the factors of production refers to all available natural resources. Whether it is a natural or humanistic attraction, the tourist attraction itself is a presentation of land resources, which is the basic carrier of all tourism development and growth. Combined with the real available statistics, this paper selects the number of A-grade attractions as the land factor input index.

Output indicators. In studies related to tourism efficiency, tourism output mainly refers to the subsequent economic effects brought by the tourism industry. ① Total tourism revenue. Drawing on existing studies, this paper first selects the most intuitive income element as the output indicator. ② Total inbound tourism arrivals. As a branch of tertiary industry, the number of person-trips is an important factor to consider the effectiveness of the tourism industry. Due to the limited data available, this paper selects the total number of inbound tourism arrivals as the reference for the output index. The number of inbound tourists also reflects the prosperity of the local tourism industry to a certain extent.

2.2.2 Data Sources

The data in this paper are obtained from the China Culture, Heritage and Tourism Statistical Yearbook and the China Tourism Statistical Yearbook for the period 2015–2019. Since the data for 2020 were not yet available at the time of writing, they are not taken into account. What's more, due to the lack of data from Hongkong, Macao, Taiwan, these special administrative regions or provinces are not discussed in this article.

3 Empirical Analysis

3.1 Spatial Local Autocorrelation Analysis

3.1.1 Single Moran Index Analysis

Statistics related to tourism efficiency in a total of 31 provinces in China from 2015–2019 were collected from the China Culture, Heritage and Tourism Yearbook and the China Tourism Yearbook. Hong Kong, China, Macao, China, and Taiwan, China, are

not included in this study due to the lack of data and the fact that the epidemic changes, in reality, are significantly different from mainland China.

The 2019 data were selected to calculate the 2019 Moran'I index for each province and city in mainland China with the help of Geoda software.

As the result, the values of the number of star-rated hotels are all less than 0, and it can be obtained that this variable has a negative spatial correlation, while the number of A-rated attractions, the number of travel agencies, and the number of travel employment are all positive values for these three variables, indicating that these three variables have a positive spatial correlation. Among them, the Moran'I of the number of travel agencies and travel employment is significantly positive and the absolute value is greater than most variables, which indicates that the number of travel agencies and travel employment is important factors affecting the efficiency of tourism inputs.

3.1.2 Results of Local Spatial Autocorrelation Analysis

Further, we visualize the local Moran index to better observe the local spatial correlation.

From the Figs. 2 and 3, it can be obtained that the local autocorrelation results of tourism efficiency in mainland China in 2019 under geographical weights mainly show a positive spatial correlation and a negative spatial correlation locally. Therefore, it can be proved that Xinjiang and Inner Mongolia are geographically and spatially clustered

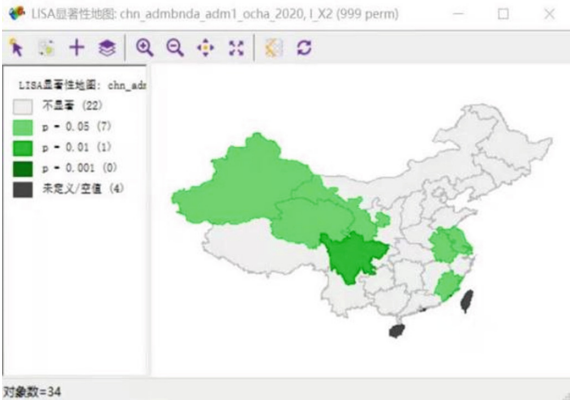

Fig. 2. Lisa salient map of the number of tourists (Photo credit: Original)

Fig. 3. Lisa clustering map of the number of tourists (Photo credit: Original)

with provinces of low tourism efficiency, and there is a positive spatial spillover effect. a positive spatial spillover effect.

The "high-high" clustering of the blue parts of the regions of Guangxi, Guangdong and Fujian provinces, and the p-value of all three provinces is 0.05, reflecting the clustering of the provinces with high tourism efficiency and the positive driving effect of the provinces with high tourism efficiency on the neighboring provinces.

Guangdong Province, Fujian Province and Jiangxi Province are all located in the eastern part of mainland China where economic development is better, but show a "low-high" concentration in the mauve part, probably because Guangdong Province or Fujian Province are more prominent in the tourism economy, absorbing tourism talents from the neighboring provinces that are slightly behind, causing a decrease in tourism efficiency in the neighboring provinces.

This is in line with the reality of mainland China and can also prove the existence of a clear local spatial autocorrelation of tourism efficiency in mainland China.

3.2 Hausman Test

Since the panel model is divided into fixed effects as well as random effects, the Hausman test was performed using R language to find out which effect should be used and the results of the analysis are as follows (Table 1).

Table 1. The results of Hausman test

Chisq	199.67	dF	7	p	2.2E−16

where the Chi statistic is 199.67 and the p-value is close to 0. Thus, the original hypothesis should be rejected at the 1% significance level, so this paper should establish a fixed-effects panel model.

3.3 LM Test

Determine and build a representation of the spatial model based on the LM (error) and lagrange multiplier (lag) as

Table 2. The results of LM test

LM test for spatial lag benefits	9.89 (0.002)	Robust LM test for spatial lag effect	25.44 (0.000)
LM test for spatial error effects	1.47 (0.226)	Robust LM test for spatial error effects	17.01 (0.000)

From the Table 2, the p-values of the LM test are 0.000 and 0.000, both less than 0.05, so the original hypothesis is rejected, indicating that there is a strong autocorrelation between provinces, and this paper studies the influence between provinces, so the

influence of spatial factors is added on this basis, in order to better analyze the influence caused by spatial factors to the local inbound tourism population, so this establishes a spatial error model and a spatial lag model.

The spatial error model refers to the overall correlation between the perturbation term in space and the space, and the perturbation in one space will affect other spaces with the spatial effect, while the spatial lag model refers to the fact that a variable will be affected not only by its own explanatory variables, but also by this variable in other spaces, and both models indicate the influence of spatial factors on the variables, so Matlab was used to combine the 2015–2019 panel data were processed and analyzed, and the results were obtained as follows.

The results of the spatial error model simulation are as follows (Table 3).

Table 3. The results of the spatial error model simulation

Variable	Coefficient	Asymptot t-stat	z-probability
x1	−3.852221	−3.688601	0.000225
x2	0.543440	1.136640	0.255689
x3	−1.908257	−1.500287	0.133540
x4	0.046303	6.059238	0.000000
x5	0.203883	7.261140	0.000000
x6	−0.007287	−1.216489	0.223799
spat.aut.	0.591979	9.429359	0.000000
R-squared	0.7411		

Modeling spatial lags (Table 4).

Table 4. The results of the spatial lag model simulatio

Variable	Coefficient	Asymptot t-stat	z-probability
x1	−3.876296	−3.706854	0.000210
x2	0.551405	1.152295	0.249200
x3	−1.902270	−1.490210	0.136169
x4	0.046305	6.029314	0.000000
x5	0.203499	7.237932	0.000000
x6	−0.007129	−1.187236	0.235135
spat.aut.	0.577980	8.952917	0.000000
R-squared	0.6337		

As seen from the R-squared values, the spatial error model is closer to 1, which indicates a better fit of the spatial error model than the spatial lag model, and also shows

a positive correlation between the tourism population outside each country and each variable. In summary, the tourism population in each region is affected by the number of scenic spots, star-rated restaurants, hotels, and the employed population.

4 Summary

By conducting spatial autocorrelation analysis, Hausman test and building a spatial panel model on the data related to tourism efficiency in mainland China from 2015 to 2019 to explore the key influencing factors affecting its tourism efficiency and which factors change in the context of the new coronavirus sweep will be important in maintaining tourism efficiency, the following conclusions were obtained.

Firstly, the main factors affecting the tourism efficiency of the provinces in mainland China are the number of travel agencies and the number of people working in the tourism industry, and they show an obvious positive promotion effect. Secondly, there is a spatial correlation in China's tourism efficiency, with positive spatial aggregation and spillover effects in provinces with similar tourism factors and similar degrees of the tourism economy, and provinces with geographic proximity and large gaps in the number of travel agencies and the number of people working in the tourism industry. Finally, there is an imbalance in the efficiency of tourism inputs between the western, central and eastern regions of mainland China. Specifically, the eastern region is optimal.

5 Conclusions and Recommendations

Strengthen local management capacity to contain the epidemic in local areas. Since the tourism situation in each region has a condition of mutual influence, most factors are positively correlated, and the Covid-19 Pandemic itself is highly contagious, so when the tourism situation in one region is relatively depressed due to the epidemic, its neighboring regions will also be negatively affected. Through big data communication technology, timely communication of the latest situation of the epidemic to neighboring provinces and municipalities can secure enough time for local coordination and layout, enhance local management capabilities, contain the epidemic in the region, and prevent the spread of the epidemic while also reducing its negative effects on neighboring areas.At the same time, it can respond to the call for the normalization of epidemic prevention and control and long-term preparedness, reducing the cost of future epidemic prevention and the risk of concentrated outbreaks.

Optimize the allocation structure of tourism resources, and make travel agencies a key investment. In the process of univariate Moran index analysis, X2 and X5, represent the number of travel agencies and the number of travel agency employees are the most prominent. And both variables are directly linked to travel agencies, making the spatially significant correlation of the situation in each region. It can be seen that the travel agency-related factors contribute the most to the coordinated development among regions. The travel agency generates the most revenue when the invested capital is the same. Therefore, the proportion of travel agencies can be increased in future tourism investment, thus improving the efficiency of capital investment. In the case of prevention and control

need to occupy part of the cost in the long term, reduce the pressure of other parts of the operation and reduce waste.

Promoting mutual exchanges and strengthening ties between regions with strong linkages, the possibility of regionalized and synergistic development of tourism is greatly increased in the context of big data communication technology, such as the Pearl River Delta, Jiangsu, Zhejiang and Shanghai regions. From the analysis of this paper, there is a mutual influence between the tourism economy of each region, the development of tourism in one region will drive the development of the surrounding economy, and strengthening the linkage between regions can also play the advantages of each region, division of labor, mutual benefit, so that the resources between regions are fully utilized. Therefore, strengthening exchanges between regions with strong linkages is an important way to promote regional development.

By understanding the spatial correlation between provinces in China, information technology such as big data can be used to better deal with outbreak situations due to the mobility of people and achieve regional linkage for epidemic prevention and control. For example, after an epidemic outbreak in a certain province, individual provinces with high correlation with a certain province are understood through relevant factors, and the relevant people who travel closely between the two provinces are retrieved through data to take corresponding epidemic prevention measures to prevent the epidemic from worsening due to factors of high mobility.

Funding. Special Innovation Project of Guangdong University of Foreign Studies (Project number: 299-GK20GS43).

References

1. Bai Y, Liu C (2021) Study on the economic effect of cultural tourism industry – a case study of Zaozhuang City. Hunan Soc Sci 04:55–61
2. Han J, Ming Q, Shi P, Liu A, Xie C (2021) Analysis of the relavant combination of regional tourism economic efficiency and network dominance and the configuration of influencing factors. Hum Geogr 36(04):168–176
3. He H, Wang Y (2021) A study on the impact of bank corporate governance on financial system stability–a regression analysis based on 14 listed banks in China. J Hunan Univ (Soc Sci Edn) 35(04):55–62
4. Li S, Li S, Zhong Y, Ren T, Xiao H, Jia B (2021) Based on the DEA model of regional tourism resources allocation efficiency research. Control Decis 1–10
5. Peng J (2021) Exploring the impact of real estate investment on economic growth–based on panel data of 11 cities in Jiangxi Province. Times Econ Trade 18(05):94–97
6. Shi H (2021) Research on tourism dynamic efficiency in Henan Province based on three-stage DEA-Malmquist model. J Luoyang Norm Univ 40(09):18–22
7. Su H, Bai L (2020) Spatial econometric analysis on efficiency of funds input in higher vocational education and its influencing factors in China. Vocat Tech Educ 41(36):38–45
8. Tao C, Yang X, Wang W (2021) The efficiency evaluation of chongqing tourism industry - based on super-DEA-MI model. J Southwest China Norm Univ (Nat Sci Edn) 46(05):55–63
9. Zhang Y, Tong Y (2021) Analysis of tourism efficiency evolution and variability in key tourism cities in China–based on the super-efficiency EBM model. J Southwest Univ (Nat Sci Edn) 43(04):109–119