Research on Influencing Factors of Tourism Income in Qinghai Province-Based on error correction model and Granger causality test

Yi Chen\textsuperscript{1a}, Jingqi Liu\textsuperscript{2b}, Songkui Yin\textsuperscript{3c,}\textsuperscript{*}
\textsuperscript{1,2,3} School of Finance and Economics, Qinghai University, 810016, Xining, China

Abstract: As an important growth point of Qinghai’s economic development, tourism has become increasingly prominent in its contribution to the development of the national economy and has become one of the most dynamic industries in the province. Based on the relevant data from 2001 to 2018, this paper explores the influencing factors of tourism revenue in Qinghai Province by constructing an error correction model and combining Granger causality test. The research results show that there is a significant correlation between the total number of tourist visits, per capita GDP and traffic conditions, and the growth of tourism income in Qinghai Province. Among them, every 1% increase in the total number of tourist arrivals drives an average increase of 1.566% in tourism revenue; and the short-term elasticity of tourism revenue to the total number of tourist arrivals is slightly greater than the long-term elasticity.

1 INTRODUCTION

Tourism as an industry associated with high, features a wide range of integrated industry, for many regions, it can become a new economic growth point for the promotion of regional economic development is of great significance. First of all, the tourism industry can optimize the resource allocation system, promote the integration of resources and upgrade the industry; secondly, the development of the tourism industry can promote the development of local economy and tourism-related industries, while also improving the ecological and cultural environment, which not only preserves the "Green water and green mountains", also won the "Golden Mountain and Silver Mountain"\textsuperscript{[1]}.

Yuan Siwei (2014)\textsuperscript{[2]} collected data on the tourism industry from 2004 to 2012 in Jiangxi Province and used grey correlation analysis and BP neural network to point out that GDP and per capita disposable income are the main factors affecting tourism income, as well as the impact on tourism income. Subsequent changes are predicted. Li Guobing (2019)\textsuperscript{[3]} used the grey correlation method and expert scoring method to rank the income influencing factors of the Pearl River Delta cities, and found that their tourism income is affected by three factors (tourists, tourism income, and tourism services). He Zhen (2009)\textsuperscript{[4]} used the grey comprehensive correlation analysis method to analyze 9 factors that have an impact on tourism income. The above studies are all researches on the factors affecting tourism income from the perspective of countries. Ionela MitukoVlad et al. (2016)\textsuperscript{[5]} used Romanian data from 2000 to 2015 to explore the relationship between the average income of tourist accommodation institutions and the number of arrivals in the country through analysis of variance and unary linear regression. Deng Hui, Li Meng (2015)\textsuperscript{[6]} Through the tourism-related data of 22 provinces in my country, a multiple linear regression model was established to obtain three important driving forces for tourism revenue increase: ticket market economy, urban green space, and infrastructure.

For more information related to this original text, if you want to view other translation information, you must enter the corresponding original text. After analyzing the above-mentioned existing research at home and abroad, many scholars adopt traditional regression analysis, gray system, and machine learning methods to conduct research. The problem of false regression that may occur when using time series data and the inaccurate results caused by the instability of economic indicators are solved. At the same time, changes in tourism revenue may be affected by tourism revenue in previous years, but previous studies did not consider the impact of lagging items. To solve the above problems and focus on Qinghai Province as a world-class tourism province with plateau characteristics, and the tourism characteristics of the provinces are very different, and the influencing factors are not the same, this article collects relevant data on the tourism industry of Qinghai Province from 2001 to 2018, Combined with the error correction model and Granger causality test, to explore the influencing factors of tourism income.
2 Overview of Related Theories and Methods

2.1 Cointegration Theory

Cointegration is a statistical description of the long-term equilibrium relationship of non-equilibrium economic variables, and the long-term equilibrium relationship becomes the cointegration relationship \([7]\). If there is a long-term stable relationship between economic variables, the classical regression method can be used to establish the corresponding model, which overcomes the false regression problem caused by unstable economic variables to a certain extent. When there are only two variables, they should be single integers of the same order; if there are more than two variables, the single integer order of the explained variable is less than or equal to the single integer order of the explanatory variable.

2.2 Error correction model

According to Granger's theorem, if several non-stationary variables have co-integration relations, these variables must have error correction model expressions. This model was originally proposed by Sargan and used as a supplementary model of the co-integration model by Hendry-Anderson And Davison et al.'s further improvement \([8]\). The long-term relationship between the series is usually measured by the cointegration model, while the short-term relationship between the series is usually explained by the error correction model.

If both X and Y are single integer variables of order 1 and there is a cointegration relationship, the simplest error correction model can be established:

\[
\Delta Y_t = \beta_1 \Delta X_t - \lambda ecm_{-1} + \epsilon_t
\]  

(1)

2.3 ADF inspection

The ADF test \([8]\) is usually used to test the stationarity of a time series from a statistical point of view using unit-roots. It is usually completed by the following three models:

- **Model I:**
  \[
  \Delta X_t = \delta X_{t-1} + \sum_{i=1}^{n} \beta_i \Delta X_{t-i} + \epsilon_t
  \]  
  (2)

- **Model II:**
  \[
  \Delta X_t = \alpha + \delta X_{t-1} + \sum_{i=1}^{n} \beta_i \Delta X_{t-i} + \epsilon_t
  \]  
  (3)

- **Model III:**
  \[
  \Delta X_t = \alpha + \beta_1 \Delta X_{t-1} + \sum_{i=1}^{n} \beta_i \Delta X_{t-i} + \epsilon_t
  \]  
  (4)

Propose assumption:

\[
H_0: \rho = 1; H_1: \rho < 1.
\]  

(5)

The test is carried out one by one from model III to model I, until the test rejects the null hypothesis.

2.4 Granger Causality Test

It should be statistically explored whether the influence between two variables with a lead-lag relationship in time is one-way or two-way \([9]\). When judging whether X leads to the generation of Y, it is necessary to consider the degree of interpretation of the past value of Y for the current value, and then add the lag value of X to observe whether the degree of interpretation improves. If it does improve and improve, then X is the Granger cause of Y. A prerequisite for Granger causality test is that the time series must have stationarity, otherwise there may be a false regression problem.

Granger causality test mainly uses two regressions:

\[
\begin{align*}
\ln Y_t &= \sum_{i=1}^{m} \alpha_i \ln X_{t-i} + \sum_{i=1}^{m} \beta_i \ln Y_{t-i} + \mu_t \\
\ln X_{t+1} &= \sum_{i=1}^{m} \lambda_i \ln Y_{t-i} + \sum_{i=1}^{m} \delta_i \ln X_{t-i} + \mu_{t+1}
\end{align*}
\]  

(6)-(7)

3 Empirical Analysis Based on Error Correction Model and Granger Causality Test

3.1 Data collection and preprocessing

In this article, through the channels of "China Statistical Yearbook 2019", "Qinghai Statistical Yearbook 2019", Qinghai Provincial Bureau of Statistics, and the official website of Qinghai Provincial Department of Transportation, we have collected GDP per capita in China from 2001 to 2018, the number of inbound tourists and the income of inbound tourism in Qinghai Province. Domestic tourist arrivals, domestic tourism revenue, number of scenic spots, number of star-rated hotels, number of travel agencies, and other data that can reflect the development of tourism in Qinghai Province. Since the number of star-rated hotels, scenic spots, and travel agencies in Qinghai Province did not change significantly during this period, these three indicators were excluded when analyzing the influencing factors. Therefore, the explanatory variables of the model are selected as follows:

Traffic conditions (X1). Generally, places with good traffic conditions can attract tourists, thereby increasing the tourism income of the region. Furthermore, considering that most tourists in Qinghai Province choose chartered self-driving tours, this article uses highway mileage as an indicator to represent traffic conditions.

The Total number of tourists (X2). The total number of tourist arrivals is one of the important indicators to measure the level of tourism development in a country and region. The total number of tourist arrivals is the most direct factor stimulating the growth of tourism income. Therefore, this article takes the total number of tourists as one of the explanatory variables.

Per capita Chinese residents GDP (X3). The reason why we did not select the per capita GDP of Qinghai Province is mainly considering that the main source of tourism income in Qinghai Province is tourists from outside the province. Selecting the national per capita GDP will better explore the impact of economic development on tourism income.

Since most of Qinghai's tourism income comes from domestic tourists (in 2020, the number of domestic tourists in Qinghai Province accounted for 99.71% of the
total number of tourists), so this article will no longer distinguish the two indicators of domestic tourism income and inbound tourism income for analysis. Instead, the total tourism income (Y) is used as the explained variable.

To eliminate the influence of heteroscedasticity and data fluctuation on the analysis results, the index data is processed logarithmically. After taking the natural logarithm, the coefficients in the model represent elasticity and can also reduce the single integral order in the stationarity transformation (Table 1).

### Table 1 Descriptive Statistics Table of Y, X1, X2, X3 Logarithmic Value

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>S.D</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
<th>Minimum</th>
<th>Max</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LnY</td>
<td>4.265</td>
<td>4.180</td>
<td>2.67a</td>
<td>1.178</td>
<td>1.387</td>
<td>0.055</td>
<td>-1.292</td>
<td>2.600</td>
<td>6.100</td>
<td>76.77</td>
</tr>
<tr>
<td>LnX1</td>
<td>10.864</td>
<td>11.055</td>
<td>10.10a</td>
<td>0.196</td>
<td>-0.982</td>
<td>-0.500</td>
<td>10.050</td>
<td>11.320</td>
<td>195.56</td>
<td></td>
</tr>
<tr>
<td>LnX2</td>
<td>7.060</td>
<td>7.060</td>
<td>5.93a</td>
<td>0.748</td>
<td>0.559</td>
<td>0.020</td>
<td>-0.964</td>
<td>5.930</td>
<td>8.340</td>
<td>127.08</td>
</tr>
<tr>
<td>LnX3</td>
<td>8.212</td>
<td>8.335</td>
<td>6.96a</td>
<td>0.779</td>
<td>0.606</td>
<td>-0.347</td>
<td>-1.434</td>
<td>6.960</td>
<td>9.210</td>
<td>147.82</td>
</tr>
</tbody>
</table>

### 3.2 Correlation test

A Correlation test can determine whether there is a correlation between Qinghai Province’s tourism revenue and highway mileage, the total number of tourists, and per capita GDP. It can be seen from Table II that the correlation coefficient between $LNY$ and $LNX_1$ is 0.922, the correlation coefficient between $LNY$ and $LNX_2$ is 0.995, and the correlation coefficient between $LNY$ and $LNX_3$ is 0.981. It can be seen that there is a strong relationship between $LNY$ and $LNX_1$, $LNX_2$, and $LNX_3$. Correlation. However, considering that the time series data may have the phenomenon of spurious regression, that is, the two columns of time series data show a consistent trend of change. Even if they do not have any meaningful relationship, they may show a higher degree of determination when the regression is performed. Coefficient, so it is impossible to establish a regression equation to explore the relationship between variables. Cointegration and causality tests must be carried out to try to avoid the phenomenon of spurious regression.

### Table 2 Correlation Coefficient Table

<table>
<thead>
<tr>
<th>Covariance Correlation</th>
<th>LNY</th>
<th>LNX1</th>
<th>LNX2</th>
<th>LNX3</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNY</td>
<td>1.311</td>
<td>0.453</td>
<td>0.829</td>
<td>0.849</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>0.184</td>
<td>0.530</td>
<td>0.399</td>
</tr>
<tr>
<td>LNX1</td>
<td></td>
<td>0.922</td>
<td>0.291</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.000</td>
<td>0.530</td>
<td>1.000</td>
</tr>
<tr>
<td>LNX2</td>
<td></td>
<td></td>
<td>0.995</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
<td>0.974</td>
</tr>
<tr>
<td>LNX3</td>
<td></td>
<td></td>
<td></td>
<td>0.571</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

### 3.3 Stationarity test

Before performing regression analysis, it is necessary to ensure that the sequence is stationary. Therefore, use Eviews10.0 software to judge the stationarity of $LNY$, $LNX_1$, $LNX_2$, and $LNX_3$ with the graphical method (Figure 1) and the unit root test combined method.
In the graphical method, if the autocorrelation coefficient of the series drops quickly and finally converges to zero, the time series can be considered stationary, otherwise, it is a non-stationary series. It can be seen from Figure 1 that $LNY$, $LNX_2$ and $LNX_3$ have no significant autocorrelation. Therefore, the time series of these three variables are non-stationary. For, lag 1 period, 2 periods...6 periods all have significant autocorrelation, so it can be considered as stable.

Using autocorrelation function graphs to judge the stationarity of a time series is subjective and lacks rigor, and it can only be used to roughly observe a limited number of observations to describe the average process of the series. Therefore, the ADF method is further used to test the stationarity of $LNY$, $LNX_1$, $LNX_2$ and $LNX_3$ (Table III)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$ADF$ value</th>
<th>Critical value (1%)</th>
<th>Critical value (5%)</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LNY$</td>
<td>7.648</td>
<td>-2.708</td>
<td>-1.963</td>
<td>unstable</td>
</tr>
<tr>
<td>$LNX_1$</td>
<td>-10.418</td>
<td>-4.004</td>
<td>-3.099</td>
<td>stable</td>
</tr>
<tr>
<td>$LNX_2$</td>
<td>6.046</td>
<td>-2.708</td>
<td>-1.963</td>
<td>unstable</td>
</tr>
<tr>
<td>$LNX_3$</td>
<td>1.479</td>
<td>-2.718</td>
<td>-1.964</td>
<td>unstable</td>
</tr>
<tr>
<td>$\Delta LNY$</td>
<td>-5.359</td>
<td>-4.886</td>
<td>-3.829</td>
<td>stable</td>
</tr>
<tr>
<td>$\Delta LNX_1$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta LNX_2$</td>
<td>-4.238</td>
<td>-3.920</td>
<td>-3.066</td>
<td>stable</td>
</tr>
<tr>
<td>$\Delta LNX_3$</td>
<td>-0.709</td>
<td>-2.718</td>
<td>-1.964</td>
<td>unstable</td>
</tr>
</tbody>
</table>

It can be seen from the test results that $LNY$ and $LNX_2$ are both first-order single integral non-stationary series, and $LNX_1$, $LNX_3$ are not first-order single integral, so we can only further test whether there is a long-term cointegration relationship between $LNY$ and $LNX_2$.

### 3.4 Cointegration test between variables

Because both $LNY$ and $LNX_2$ are first-order single integer variables, we can then explore whether there is a long-term equilibrium relationship between the two. This paper uses the E-G test to test whether there is a cointegration relationship between Qinghai Province’s tourism revenue and the total number of tourists. Perform OLS regression analysis on $LNY$ and $LNX_2$, and get the result:

$$LnY = -6.794 + 1.566 \times LnX_2 + \epsilon_i$$

$$(-24.987) \quad (40.882)$$

$R^2=0.991$ $\quad$ $DW=1.602$ $\quad$ $F=1671.299$

The regression residual sequence estimation model can be obtained from the above regression equation:

$$\epsilon_i = LNY + 6.794 -1.566 \times LNX_2$$

To verify whether there is a long-term equilibrium relationship between $LNY$ and $LNX_2$, that is, a cointegration relationship, the ADF unit root test needs to be performed on the above regression residual sequence. The test results are shown in the following table

<table>
<thead>
<tr>
<th>Variable</th>
<th>$t$-Statistic</th>
<th>Prob.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Augmented Dickey-Fuller test statistic</td>
<td>-1.711</td>
<td>0.082</td>
</tr>
</tbody>
</table>

Test critical values: 1% level $-2.718$, 5% level $-1.964$, 10% level $-1.606$

It can be seen from the above table that the value of the $ADF$ statistic of the residual sequence is -1.711, but at this time it cannot be compared with the critical value given in the Eviews software. The following critical value should be used according to the cointegration test critical value table. The calculation formula for calculation.
\[ C_p = \varphi_p + \Delta \frac{\Delta LNX}{\Delta t} \]

In the above formula, the critical value obtained by looking up the table according to the number of variables and the significance level of 1% and the sample size is substituted into the critical value of -1.343, which is greater than the ADF value of the residual sequence, so the null hypothesis is rejected, that is, the residual. The difference sequence has no unit root, and the sequence is stationary.

In summary, the tourism income \((Y)\) and total tourist arrivals \((X)\) of Qinghai Province are \((2, 2)\) co-integrated, indicating that there is indeed a long-term stable equilibrium relationship between the two variables, that is, every 1% increase in total tourist arrivals, drove an average increase of 1.566% in tourism revenue \([10]\).

### 3.5 The establishment of error correction model

There is a co-integration relationship between tourism revenue \((Y)\) and total tourist arrivals \((X)\) in Qinghai Province, indicating that there is a long-term equilibrium relationship between the two. But to avoid short-term imbalances, the error term in the cointegration regression model is regarded as an equilibrium error, and an error correction model is established to link the short-term behavior of total tourism revenue with long-term changes. Therefore, the regression model is estimated with \(\Delta LN_Y\) as the explained variable and \(\Delta LNX_2\) as the explanatory variable, and finally, the estimation result of the error correction model is obtained:

\[
\Delta LN_Y = 0.099 + 1.816 \times \Delta LN_{X_2} + 0.655 \times \epsilon_{t-1} + \epsilon_t \tag{11}
\]

\[ (4.346) \quad (6.237) \quad (5.693) \]

\[ R^2=0.971 \quad DW=1.550 \quad F=21.904 \]

The estimation results show that the change of tourism income in Qinghai Province depends not only on the total number of tourist visits but also on the deviation of the previous period’s tourism income from the equilibrium level. The estimated coefficient of error term \(\epsilon_{t-1}\) of 0.655 reflects the deviation Correction, the farther the deviation from the previous period, the greater the amount of correction in this period, that is, the system has an error correction mechanism.

And compared with the long-term equilibrium relationship, it is found that the long-term elasticity of \(LN_Y\) to \(LNX_2\) is 1.566 slightly less than the short-term elasticity of 1.816, indicating that the relationship between the growth of tourism income and the increase of total tourist arrivals in the short-term dynamic is more significant in the short-term than the long-term relationship. The error correction coefficient of 0.655 is relatively large, indicating that the adjustment to short-term fluctuations is relatively strong.

### 3.6 Granger causality analysis

The condition of Granger causality test is that the random variable is a stationary time series. In the above test in this article, \(LN_Y\) and \(LNX_2\) are stable after the first-order difference, and there is a cointegration relationship between the two variables, so Granger is satisfied. The conditions for causality testing. Taking the 5th order lag, the estimation results given by Eviews are shown in the table below.

#### Table 5 Granger Causality Test

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Obs</th>
<th>F-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNX2 does not Granger Cause LN</td>
<td>13</td>
<td>4.101</td>
<td>0.039</td>
</tr>
<tr>
<td>LN Y does not Granger Cause LNX2</td>
<td>2.061</td>
<td>0.358</td>
<td></td>
</tr>
</tbody>
</table>

From the above table, we can get: at the 5% confidence level, \(LNX_2\) is the Granger causality of \(LN_Y\), but \(LN_Y\) is not the Granger causality of \(LNX_2\), that is, the influence between \(LNX_2\) and \(LN_Y\) is considered to be one-way.

Through the Granger causality test, it can be seen that in the long run, the total number of tourist visits has a strong stimulus effect on the tourism income of Qinghai Province, but the increase in tourism income will not increase the total number of tourist visits.

### 4 Conclusion and Inspiration

#### 4.1 Conclusion

This paper establishes an error correction model and combines the Granger causality test, selects three indicators of highway mileage, the total number of tourists, and per capita GDP as variables, and uses the relevant time series data of Qinghai Province from 2001 to 2018 and the time series of national GDP per capita. Based on the data, the influencing factors of tourism income in Qinghai Province were explored; after the time series data passed the stationarity test, the co-integration test was performed to obtain the long-term equilibrium model. Then, the error correction model was established through the \(E-G\) two-step method, and the total number of tourist visits was found to be Tourism income has a significant impact, and the short-term elasticity of tourism income to the total number of tourist visits is greater than the long-term elasticity; then the Granger causality test is used to explore the causal relationship between the total tourist income and the total number of tourist visits.

There is a long-term equilibrium relationship between tourism revenue and total tourist arrivals in Qinghai Province, that is, every 1% increase in total tourist arrivals will drive tourism revenue to increase by 1.566%. As the total number of tourist arrivals in Qinghai Province increases year by year, the tourism income of Qinghai Province will also gradually increase. Similarly, it can be seen from the correlation coefficient matrix that per capita GDP and highway mileage also have a significant correlation with the growth of tourism income.
in Qinghai Province. This shows that with the increase of residents’ disposable income, people begin to attach importance to spiritual life, and travel has become people’s enjoyment of life. The main way. According to the Granger causality test, the increase in total tourist visits can drive the increase in tourism income and promote the rapid growth of tourism income in Qinghai Province; at the same time, tourism income is not the Granger reason for the total tourist visits; the annual tourism income of Qinghai Province The increase will not attract more tourists to Qinghai.

4.2 Inspiration

First of all, according to the results of Granger causality test and co-integration test, it can be seen that the increase in total tourist visits has a certain pulling effect on the increase in tourism income. Therefore, relevant departments should speed up the improvement of the tourism industry system to attract more domestic tourists. Because the total number of tourist visits has a great impact on the total tourism revenue, the relevant departments should adapt to local conditions, rationally develop and utilize local tourism resources, improve the tourism product system, and strengthen the tourism safety and quality assurance system to attract more tourists.

Secondly, the error correction model shows that the relationship between the growth of tourism income and the increase of total tourist arrivals in the short-term dynamic is more significant in the short-term than the long-term relationship; the error correction coefficient is -0.655, which shows that the adjustment of short-term fluctuations is greater. Therefore, relevant departments should avoid the waste of input due to obvious short-term effects.

Finally, the relevant departments should be committed to continuously optimizing tourism routes and tourism service levels to overcome a series of current problems, vigorously build the “Great American Qinghai” cultural brand, and realize the integrated development of “tourism + culture” to increase the total number of visitors to Qinghai. This will promote the growth of tourism income in Qinghai Province.

Acknowledgments

The author’s name is in alphabetical order, and the workload of each author is equivalent.

REFERENCES


