Analysis of Influencing Factors of Jiangsu Province Tourism Based on Eviews

Yang Tianming (Vettel)

International College, Jiangxi University of Finance and Economy

Abstract

Eviews software is adopted for calculation and analysis in the paper. Based on the relevant statistics of tourism industry in Jiangsu Province from 2010 to 2020, this paper uses Eviews software to establish a multiple linear regression model to study the influencing factors of tourism industry in Jiangsu Province.

Keywords: Tourism; EViews; OLS; Regression Analysis

1. Introductions

Jiangsu is the most economically developed province, and it has the second highest GDP and the highest GDP per capita in China. There are 13 cities in Jiangsu province, and they all are among the top 100 cities in China. At the same time, because of developed economy and convenient transportation, Jiangsu's tourism income ranks first among all provinces. Today's tourism is no longer affected by a single economic factor, rather than by environment, history, society, transportation and other factors. In recent years, more and more domestic and foreign tourists are willing to select Jiangsu as tourism destination. One reason for the phenomenon is that people have more disposable income, and another reason is the transportation become more and more convenient. It creates a very good geographical environment because most places in Jiangsu Province are plain and hilly. Many scholars have used many methods to study the influencing factors of tourism income, such as AHP analysis, coordination theory and ECM analysis.

In this paper, through the comprehensive analysis, it is drawn that total tourism income is mainly affected by the number of domestic tourists, the number of foreign tourists, deposits of urban residents, length of roads and length of railways. I use EViews to analyse the correlation between them.

2. Data Collection

Table 1 2005-2020 Tourism consumption data of Jiangsu province

Year	Total Tourism Revenue/ Billion RMB (Y)	Domestic Tourist Numbers/ Million People (X_1)	Foreign Tourist Numbers/ Million People (X_2)	Deposits of Urban Residents/ RMB (X_3)	Length of roads/ Kilometers (X_4)	Length of railways/ Kilometers (X_5)
2005	185.55	172.34	3.78	10493	82739	1598.9
2006	228.43	199.36	4.45	11760	126900	1602.7
2007	273.36	231.99	5.13	13786	134000	1606.9
2008	318.54	261.22	5.44	15781	141000	1642.9
2009	344.95	297.27	5.57	17175	142000	1642.1
2010	462.50	355.19	6.54	19109	150000	1907.8
2011	558.00	410.00	7.37	21810	152000	2304.0
2012	652.40	460.00	7.92	24565	154000	2309.1
2013	719.50	520.00	2.88	26955	156000	2554.1
2014	814.55	570.00	2.97	28844	158000	2632.4
2015	905.01	619.34	3.05	31195	159000	2679.2
2016	1026.36	677.80	3.30	33616	157000	2721.9
2017	1166.22	742.87	3.70	36396	158000	2770.9
2018	1324.73	814.23	4.01	39251	160000	3014.0
2019	1432.16	880.00	4.00	42359	160000	3539.0
2020	825.06	470.00	0.77	43834	161000	3998.0

Source: Jiangsu Statistical Bulletin on National Economic and Social Development 2005-2020

3. Model Creation

The collected data were sorted out according to Table 1. So let's define total tourism income as the explained variable Y and define the number of domestic tourists, the number of foreign tourists, deposits of urban residents, length of roads and length of railways as the explanatory variable X_1, X_2, X_3, X_4, X_5 respectively. The model is:

$$Y_i = \beta_i + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon$$

Using the least square method in the EViews we can get the following result:

Dependent Variable: Y
Method: Least Squares
Date: 05/29/21 Time: 00:24
Sample: 2005 2020
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-46.70367	797.4639	-0.058565	0.9545
X1	0.139795	0.009607	14.55178	0.0000
X2	0.706949	0.508549	1.390129	0.1947
X3	0.127086	0.050179	2.532644	0.0297
X4	-0.018037	0.005808	-3.105256	0.0112
X5	-0.255551	0.507071	-0.503975	0.6252

R-squared	0.997222	Mean dependent var	7023.320
Adjusted R-squared	0.995833	S.D. dependent var	3950.339
S.E. of regression	255.0128	Akaike info criterion	14.20050
Sum squared resid	650315.3	Schwarz criterion	14.49022
Log likelihood	-107.6040	Hannan-Quinn criter.	14.21534
F-statistic	717.8896	Durbin-Watson stat	1.327309
Prob(F-statistic)	0.000000		

Fig 1 Outcome of OLS

According to the data in Figure 1, the estimated result of the model is

$$\widehat{Y}_i = -46.70367 + 0.139795X_1 + 0.706949X_2 + 0.127086X_3 - 0.018037X_4 - 0.255551X_5$$

(797.4639) (0.009607) (0.508549) (0.050179) (0.005808) (0.507071)

$$R^2 = 0.997222 \quad \overline{R^2} = 0.995833$$

From the model we can know R-squared is 0.997222 and adjusted R-squared is 0.995833, showing that the model fits the samples well.

The p-value of X_1's coefficient is 0.0000, which passes the significance test at the significance level of 1%. The p-value of X_2's coefficient is 0.1947, which doesn't pass the significance test at the significance level of 10%. The p-value of X_3's coefficient is 0.0297, which passes the significance test at the significance level of 5%. The p-value of X_4's coefficient is 0.0112, which passes the significance test at the significance level of 5%. The p-value of X_5's coefficient is 0.6252, which passes the significance test at the significance level of 5%. Therefore, the model may have collinearity.

4. Model Analysis

4.1 Multicollinearity Test

Table 2 correlation coefficient

	X1	X2	X3	X4	X5
X1	1.000000	-0.330297	0.895830	0.709920	0.786030
X2	-0.330297	1.000000	-0.523384	-0.114105	-0.575374
X3	0.895830	-0.523384	1.000000	0.739384	0.966068
X4	0.709920	-0.114105	0.739384	1.000000	0.659579
X5	0.786030	-0.575374	0.966068	0.659579	1.000000

The correlation coefficient among the explanatory variables is high and there is multicollinearity.

4.2 Stepwise Regression

Because of the existence of multicollinearity in the model, it is necessary to make a regression analysis of Y for each X separately. The results are then analyzed, and the values of each R are compared.

Dependent Variable: Y
Method: Least Squares
Date: 05/29/21 Time: 18:23
Sample: 2005 2020
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1420.968	292.7307	-4.854181	0.0003
X1	0.175886	0.005561	31.63066	0.0000

R-squared	0.986200	Mean dependent var	7023.320
Adjusted R-squared	0.985214	S.D. dependent var	3950.339
S.E. of regression	480.3461	Akaike info criterion	15.30336
Sum squared resid	3230253.	Schwarz criterion	15.39993
Log likelihood	-120.4269	Hannan-Quinn criter.	15.30830
F-statistic	1000.498	Durbin-Watson stat	0.574672
Prob(F-statistic)	0.000000		

Fig 2 Outcome of OLS (X_1)

The model's R-squared is 0.997222, suggesting that the model fits the samples well. It's p-value is 0.0000 and it passes the significance test at the significance level of 1%.

Dependent Variable: Y
Method: Least Squares
Date: 05/29/21 Time: 18:25
Sample: 2005 2020
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	10601.50	2545.990	4.164001	0.0010
X2	-8.078302	5.335011	-1.514205	0.1522

R-squared	0.140726	Mean dependent var	7023.320
Adjusted R-squared	0.079349	S.D. dependent var	3950.339
S.E. of regression	3790.372	Akaike info criterion	19.43478
Sum squared resid	2.01E+08	Schwarz criterion	19.53136
Log likelihood	-153.4783	Hannan-Quinn criter.	19.43973
F-statistic	2.292818	Durbin-Watson stat	0.543051
Prob(F-statistic)	0.152220		

Fig 3 Outcome of OLS (X_2)

The model's R-squared is 0.140726, suggesting that the model fits the samples bad. It's p-value is 0.1522 and it passes the significance test at the significance level of 10%. So that I avoid using X_2 to do further analysis.

Dependent Variable: Y
Method: Least Squares
Date: 05/29/21 Time: 18:25
Sample: 2005 2020
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1704.646	994.5551	-1.713978	0.1086
X3	0.334943	0.035344	9.476791	0.0000

R-squared	0.865138	Mean dependent var	7023.320
Adjusted R-squared	0.855505	S.D. dependent var	3950.339
S.E. of regression	1501.624	Akaike info criterion	17.58295
Sum squared resid	31568254	Schwarz criterion	17.67952
Log likelihood	-138.6636	Hannan-Quinn criter.	17.58790
F-statistic	89.80956	Durbin-Watson stat	1.407440
Prob(F-statistic)	0.000000		

Fig 4 Outcome of OLS (X_3)

The model's R-squared is 0.865138, suggesting that the model fits the samples well. It's p-value is 0.0000 and it passes the significance test at the significance level of 1%.

Dependent Variable: Y
Method: Least Squares
Date: 05/29/21 Time: 18:25
Sample: 2005 2020
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-13258.64	5646.793	-2.347995	0.0341
X4	0.137994	0.038093	3.622537	0.0028

R-squared	0.483829	Mean dependent var	7023.320
Adjusted R-squared	0.446959	S.D. dependent var	3950.339
S.E. of regression	2937.737	Akaike info criterion	18.92514
Sum squared resid	1.21E+08	Schwarz criterion	19.02171
Log likelihood	-149.4011	Hannan-Quinn criter.	18.93008
F-statistic	13.12278	Durbin-Watson stat	0.662065
Prob(F-statistic)	0.002772		

Fig 5 Outcome of OLS (X_4)

The model's R-squared is 0.483829, suggesting that the model fits the samples poor. It's p-value is 0.0000 and it passes the significance test at the significance level of 1%.

Dependent Variable: Y
Method: Least Squares
Date: 05/29/21 Time: 18:25
Sample: 2005 2020
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3900.095	2000.984	-1.949089	0.0716
X5	4.536785	0.797633	5.687809	0.0001

R-squared	0.697958	Mean dependent var	7023.320
Adjusted R-squared	0.676384	S.D. dependent var	3950.339
S.E. of regression	2247.241	Akaike info criterion	18.38926
Sum squared resid	70701274	Schwarz criterion	18.48584
Log likelihood	-145.1141	Hannan-Quinn criter.	18.39421
F-statistic	32.35117	Durbin-Watson stat	1.045796
Prob(F-statistic)	0.000056		

Fig 6 Outcome of OLS (X_5)

The model's R-squared is 0.697958, suggesting that the model fits the samples not bad.

It's p-value is 0.0000 and it passes the significance test at the significance level of 1%.

Through the help of EViews, I get some following linear regression models:

$$\widehat{Y_i} =- 1420.968 + 0.175886X_1$$
(292.7307) (0.005561)
$$R^2 = 0.997222 \quad \overline{R^2} = 0.995833$$

$$\widehat{Y_i} = 10601.50 - 8.078302X_2$$
(2545.990) (5.335011)
$$R^2 = 0.140726 \quad \overline{R^2} = 0.079349$$

$$\widehat{Y_i} =- 1704.646 + 0.334943X_3$$
(994.5551) (0.035344)
$$R^2 = 0.865138 \quad \overline{R^2} = 0.855505$$

$$\widehat{Y_i} =- 13258.64 + 0.137994X_4$$
(5646.793) (0.038093)
$$R^2 = 0.483829 \quad \overline{R^2} = 0.446959$$

$$\widehat{Y_i} =- 3900.095 + 4.536785X_5$$
(2000.984) (0.797633)
$$R^2 = 0.697958 \quad \overline{R^2} = 0.676384$$

According to the results of the above regression model, X_1's R-squared is the largest among the five models. So that we select the first regression model as a base.

Based on A, other variables were separately added to the model to carry out regression. By this step, we select the best model and add the variables separately again.

Table 3 Stepwise Regression

Model	***Coefficients***						
	C	X_1	X_3	X_4	X_5	R^2	$\overline{R^2}$
$Y = f(X_1)$	1420.968	0.175886				0.997222	0.995833
$Y = f(X_1, X_3)$	-1783.178	0.143348	0.073849			0.994506	0.993660
$Y = f(X_1, X_5)$	-2338.322	0.155904			0.779435	0.994073	0.993161
$Y = f(X_1, X_4)$	-981.5543	0.178276		-0.003770		0.986379	0.984284
$Y = f(X_1, X_3, X_4)$	-244.2457	0.146305	0.086922	-0.013754		0.996630	0.995787
$Y = f(X_1, X_3, X_5)$	-1915.665	0.145678	0.058221		0.177695	0.994543	0.993179

Avoiding models which have negative coefficients or can't pass t-test, we can get a best model.

$$\widehat{Y_i} = -1783.178 + 0.1433485X_1 + 0.073849X_3$$
$$(208.3701) \quad (0.008193) \quad (0.016659)$$
$$R^2 = 0.994506 \quad \overline{R^2} = 0.993660$$

Dependent Variable: Y
Method: Least Squares
Date: 05/29/21 Time: 18:42
Sample: 2005 2020
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1783.178	208.3701	-8.557744	0.0000
X1	0.143348	0.008193	17.49550	0.0000
X3	0.073849	0.016659	4.432996	0.0007

R-squared	0.994506	Mean dependent var	7023.320
Adjusted R-squared	0.993660	S.D. dependent var	3950.339
S.E. of regression	314.5336	Akaike info criterion	14.50742
Sum squared resid	1286108.	Schwarz criterion	14.65228
Log likelihood	-113.0594	Hannan-Quinn criter.	14.51484
F-statistic	1176.530	Durbin-Watson stat	0.534399
Prob(F-statistic)	0.000000		

Fig 7 Outcome of OLS (X_1, X_3)

5. Model Test

5.1 Fitness Test

R-squared is 0.994506 and adjusted R-squared is 0.993660, this model explains more than 99% of samples and fits the samples well.

5.2 T-test

Given significance level α=0.05, the model has 14 degrees of freedom. We can know from the T-distribution table, $t_{\alpha/2}$ =2.145. While X_1's T-Statistic is 17.49550, it passes T test. At the same time X_2's T-Statistic is 4.432996, it passes T test too.

5.3 F-test

F=1176.530, when the significance level α=0.05, the number of variables is 2 and the degrees of freedom is 14, we can find the critical value F_α=3.806. $F>F_\alpha$, so that the regression formula pass the F-test.

5.4 Heteroscedasticity Test

Heteroskedasticity Test: White

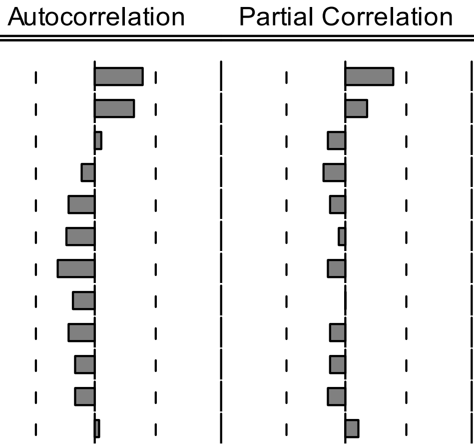
F-statistic	0.710141	Prob. F(5,10)	0.6295
Obs*R-squared	4.192498	Prob. Chi-Square(5)	0.5220
Scaled explained SS	1.166891	Prob. Chi-Square(5)	0.9480

Fig 8 Heteroscedasticity Test

By white test, we can find that all p-values exceed the significance which equals 0.05. It suggests the model doesn't exist heteroscedasticity.

5.5 Autocorrelation Test

Date: 05/29/21 Time: 22:31
Sample: 2005 2020
Included observations: 16



	AC	PAC	Q-Stat	Prob
1	0.398	0.398	3.0390	0.081
2	0.313	0.183	5.0502	0.080
3	0.051	-0.152	5.1072	0.164
4	-0.109	-0.175	5.3932	0.249
5	-0.220	-0.130	6.6591	0.247
6	-0.237	-0.060	8.2781	0.218
7	-0.296	-0.147	11.076	0.135
8	-0.183	-0.006	12.285	0.139
9	-0.213	-0.129	14.145	0.117
10	-0.162	-0.126	15.402	0.118
11	-0.164	-0.143	16.946	0.109
12	0.034	0.110	17.029	0.149

Fig 9 Heteroscedasticity Test

We use partial correlation to test the model, and the results in the figure showed that the partial correlation coefficients did not exceed the dashed line in the figure. So that there was no autocorrelation in the established model. The model passed the autocorrelation test.

6. Summary

The final model is as follows:

$$\widehat{Y}_i = -1783.178 + 0.1433485X_1 + 0.073849X_3$$

$$(208.3701) \quad (0.008193) \quad (0.016659)$$

$$R^2 = 0.994506 \quad \overline{R^2} = 0.993660$$

Among the model, Y is total tourism revenue/ Billion RMB. X_1 is domestic tourist number/ million people. X_3 is the deposits of urban residents/ RMB. The model suggest the number of domestic tourists and deposits of urban residents both have correlations with total tourism revenue. When domestic tourist increase 1 million people, total tourism will increase 0.1433485 billion yuan. And when deposits of urban residents increase 1 yuan, total tourism will increase 0.073849 billion yuan.

7. Conclusion

Multiple regression method was used to analyse the correlations between total tourism income, the number of domestic tourists, the number of foreign tourists, deposits of urban residents, length of roads and length of railways in Jiangsu province, then the collinearity is eliminated by step analysis. The result shows that the number of domestic tourists and deposits of urban residents both have positive effects on total tourism income.

But with the development of economy and society, the government will build more roads and railways. At the same time deposits of urban residents will increase. Not only domestic tourists but also foreign tourists will visit Jiangsu province. Due on above reasons, it is hardly to estimate all collinearity between five independent variables. In addition to this advantage, the COVID-19 caused negative influences on tourism because of travel restrictions and strict VISA. On the whole, by multiple regression stepwise regression we can obtain the most suitable model to fit the samples.

References

[1] Analysis of Influencing Factors of Guiyang Tourism Based on Eviews ZHONG Hao-fan LV Hua-xian 钟皓凡,吕华鲜.基于 Eviews 的贵阳市旅游业影响因素分析[J].武汉商学院学报,2020,34(03):5-9.

[2]Shaohe Zhang,Hui Liu,Pin Wang. Research on the Price Factor Model of Gold and Silver Futures[A]. Wuhan Zhicheng Times Cultural Development Co., Ltd..Proceedings of 2nd International Symposium on Economic Development and Management Innovation (EDMI 2020)[C].Wuhan Zhicheng Times Cultural Development Co., Ltd.:武汉志诚时代文化发展有限公司,2020:7.