Asian Agricultural Research 2025, 17 (11): 25 – 28 DOI: 10.19601/j.cnki.issn1943 – 9903.2025.11.005

An Empirical Analysis of Factors Influencing the Tourism Economy in Henan Province

Pingping WANG, Liang ZHAO*

School of Business Administration, University of Science and Technology Liaoning, Anshan 114051, China

Abstract [**Objectives**] To study the factors influencing the tourism economy in Henan Province. [**Methods**] Using tourism-related data from Henan Province covering the period from 2000 to 2020, this study constructs a regression model based on multivariate statistical methods to investigate the determinants of the tourism economy. The dependent variable in the model is the domestic tourism revenue of Henan Province, while the independent variables comprise the number of tourist arrivals, total operational railway mileage, the number of travel agencies, and the per capita disposable income of urban residents. [**Results**] Both the total railway mileage and the per capita disposable income of urban residents are the primary factors influencing the development of Henan's tourism economy. [**Conclusions**] It is recommended to reduce uncertainty and liquidity constraints to mitigate residents' precautionary savings behavior, actively expand domestic demand to leverage tourism as an economic driver, and improve infrastructure to support tourism development.

Key words Henan Province, Tourism economy, Influencing factors, Multivariate statistical analysis

0 Introduction

With the continuous advancement of supply-side structural reform in China's tourism sector, the industry has made significant progress, contributing to the national economy at historically high levels. As a result, its role as a pillar industry for economic growth has become increasingly evident[1]. Over the four decades since the launch of reform and opening-up, particularly since the 1990s, China's domestic tourism revenue has grown at an average annual rate of 14.4%, significantly outpacing the GDP growth rate of 9.76% during the same period. In 2019, the comprehensive contribution of tourism to China's GDP reached 11.05%. Meanwhile, tourism has demonstrated a notable economic multiplier effect, effectively promoting the coordinated development of regional economies[2]. Henan Province, situated in the heart of the Central Plains, boasts a rich cultural heritage and a well-developed transportation network. For over 3 000 years, it served as the political, economic, and cultural center of China, a profound historical legacy that has endowed the region with abundant cultural treasures and significant advantages for tourism development. However, despite its wealth of historical and cultural resources, Henan's tourism sector has yet to fully realize its economic potential, especially when compared with provinces such as Shaanxi, which shares a comparable heritage. For Henan, identifying existing challenges in a timely manner and formulating targeted strategies are crucial to addressing current bottlenecks. Thus, in-depth research into the economic factors influencing Henan's tourism industry holds both important theoretical and practical significance.

1 Literature review

Research on China's tourism economy emerged following the country's reform and opening-up policy. Wang Liangju [3] employed principal component analysis to examine 31 provinces and municipalities, revealing that most regions in China experience imbalanced tourism development and weak economic performance in the sector. Shi Peihua [4] proposed that product branding represents a key future direction for tourism development. Yang[5] emphasized that areas abundant in tourism resources should prioritize all-for-one tourism, which facilitates optimal resource allocation across time, space, and sectors. In terms of economic determinants of tourism, Chinese scholars have also achieved significant findings. Wang Guicheng [6], applying grey relational analysis, demonstrated that economic growth substantially boosts tourism revenue in Henan Province. Li Ruyou and Huang Changzhou[7] explored how transportation infrastructure influences regional tourism development. Wu Yuanyuan and Song Yuxiang [8] identified several critical factors, including per capita GDP, tourism service facilities, resource endowment, information technology levels, and transportation capacity. Hao Chen [9] confirmed the significant positive impact of high-speed rail expansion on the tourism economy. Zhong Haofan and Lyu Huaxian [10] established that urban household disposable income and road mileage are key drivers of tourism growth in Guiyang. Similarly, Hu Xinwen *et al.* [11] verified that the number of domestic tourists and railway mileage significantly affect domestic tourism revenue.

2 Variable selection and model construction

2.1 Variable selection According to existing literature and prevailing circumstances, the tourism economy has been markedly affected by external shocks since 2020. Accordingly, this study selects domestic tourism revenue in Henan Province from 2000 to 2020 as the dependent variable. The independent variables include: the number of tourist arrivals in Henan, the number of

Received: August 12, 2025 Accepted: October 15, 2025
Pingping WANG, master's degree candidate, research fields: business administration.
* Corresponding author. Liang ZHAO, doctoral degree, professor, master supervisor, research fields: business administration.

travel agencies, the total operational railway mileage, and the per capita disposable income of urban residents in the province.

2.2 Data sources This study utilizes data from the *Henan Provincial Statistical Bulletin on National Economic and Social Development (2000 – 2020)*. The variables are defined as follows: domestic tourism revenue (Y, in 100 million yuan), tourist arrivals (X_1, in 10 000 persons), number of travel agencies (X_2, units), total operational railway mileage (X_3, in 10 000 km), and urban resident per capita disposable income (X_4, yuan). The corresponding data are presented in Table 1.

Table 1 Tourism revenue and key influencing factors in Henan Province (2000 – 2020)

Year	Y	X_1	X_2	X_3	X_4
2000	809.87	6 464.01	504	1 942	4 766
2001	895.00	7 197.80	610	1 885	5 267
2002	1 011.98	8 063.26	610	1 883	6 245
2003	985.19	7 501.58	701	1 883	6 926
2004	1 219.09	8 967.42	821	1 875	7 705
2005	1 353.54	9 773.64	884	1 924	8 668
2006	1 519.74	10789.33	934	1 862	9 810
2007	1 791.81	12 218.06	964	1 871	11 477
2008	2 029.93	13 585.76	1 045	1 859	13 231
2009	2 383.47	15 455.54	1 109	2 176	14 372
2010	2 964.59	18 141.96	1 292	2 297	15 930
2011	3 931.71	21 067.52	1 349	2 555	18 195
2012	4 807.86	23 911.55	1 624	2 577	20 443
2013	5 708.59	26 753.13	1 810	3 203	22 398
2014	6 801.26	29 405.80	1 984	3 818	24 391
2015	7 976.60	32 779.82	2 150	3 859	25 576
2016	9 200.29	36 200.09	2 345	4 265	27 233
2017	10 667.14	40 739.76	2 639	4 307	29 558
2018	12 253.30	45 252.79	2 962	4 630	31 874
2019	13 740.02	49 409.62	3 276	4 825	34 201
2020	4 528.31	22 590.61	3 425	4 871	34 750

2.3 Model construction This study aims to analyze the variation patterns of tourism economic income with respect to various explanatory variables by establishing the following multiple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \mu_i$$

3 Model validation and adjustment

3.1 Model estimation Based on the variables in Table 1, parameter estimation was performed using the Ordinary Least Squares method with EViews 13 statistical software. The estimated results of the final model are as follows:

$$\hat{Y}_i = -4\,612.386 + (-0.000\,322)X_1 + (-3.547\,051)X_2 + 3.094\,345X_3 + 0.333\,677X_4$$

$$(1\,766.915) \quad (0.001\,282) \quad (3.194\,166) \quad (1.643\,677) \quad (0.239\,966)$$

$$t = (-2.610\,418) \quad (-0.251\,196) \quad (-1.110\,478) \quad (1.882\,574) \quad (1.390\,516)$$

$$R^2 = 0.833\,714 \quad \overline{R}^2 = 0.792\,142 \quad F = 20.054\,89 \quad n = 21$$

3.2 Model validation

3.2.1 Economic significance test. The estimated parameter $\hat{\beta}_1 = -0.000\,322$ indicates that, holding all other variables constant, an increase of 10 000 tourist visits in Henan Province corresponds to an average decrease of 0.000 032 2 billion yuan in tourism revenue. Parameter $\hat{\beta}_2 = -3.547\,051$ suggests that, assuming all other variables remain constant, the addition of one travel agency in Henan Province is associated with an average decline of 354.705 1 million yuan in tourism revenue. Parameter $\hat{\beta}_3 = 3.094\,345$ indicates that, holding other variables constant, each additional kilometer of railway operating mileage in Henan Province increases tourism revenue by an average of 309.434 5 million yuan. Parameter $\hat{\beta}_4 = 0.333\,677$ shows that, with other variables held constant, an increase of 100 million person-km in Henan's tourism transportation volume results in an average increase of 0.033 367 7 billion yuan in tourism revenue. Notably, the results for $\hat{\beta}_1$ and $\hat{\beta}_2$ deviate from theoretical expectations and empirical evidence, as increases in tourist arrivals and the number of travel agencies would generally be expected to enhance tourism consumption rather than reduce tourism revenue.

3.2.2 Statistical tests. (i) Goodness-of-fit test. The regression results show that the model's coefficient of determination, $R^2 = 0.833\,714$, and the adjusted coefficient of determination, $\overline{R}^2 = 0.792\,142$, are both close to 1, indicating that the model provides a good fit to the sample data.

(ii) F-test. The F-statistic for the model is 20.054 89. Considering the multiple explanatory variables and a significance level of $\alpha = 0.05$, for the null hypothesis $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, the critical value of the F-distribution for degrees of freedom $k-1=4$ and $n-k=16$. Since $F = 20.054\,89 > F_\alpha(4, 16) = 3.01$, the null hypothesis is rejected, indicating that the four explanatory variables, "the number of tourists received in Henan Province", "the number of travel agencies", "total operational railway mileage", and "per capita disposable income of urban residents", collectively have a significant effect on the dependent variable.

(iii) t-test. For each null hypothesis $H_0: \beta_j = 0$ ($j = 0, 1, 2, 3, 4$), at a significance level of $\alpha = 0.05$ and degrees of freedom $n-k=16$, the critical value from the t-distribution table is $t_{0.025}(16) = 2.12$. The test results indicate that the absolute values of the t-statistics for all estimated coefficients are less than the critical value, suggesting that none of the coefficients are statistically significant. Holding other explanatory variables constant, X_1, X_2, X_3 and X_4 do not have a significant effect on Y. In addition, the corresponding p-values for the estimates of $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\beta}_4$ are 0.804 9, 0.283 2, 0.078 1, and 0.183 4, respectively. Since all p-values exceed the significance level of 0.05, it is further confirmed that none of the four explanatory variables are statistically significant in the t-test.

At a significance level of $\alpha = 0.10$, the critical value $t\frac{0.10}{2}(16) = 1.746$. The p-value corresponding to $\hat{\beta}_3$ is less than 0.10, indicating that X_3 is statistically significant under this condition. However, the remaining three explanatory variables still fail the t-test.

Thus, while the model exhibits a high coefficient of determination and a clearly significant F-statistic, none of the estimated coefficients of the four explanatory variables are significant at $\alpha = 0.5$, suggesting the potential presence of severe multicollinearity.

3.2.3 Multicollinearity test. Using EViews 13, we obtained the correlation coefficient matrix, as shown in Table 2. The correlation coefficients between X_2 and X_4 and between X_2 and X_3 are above 0.9, indicating strong correlation among the explanatory variables and suggesting multicollinearity in the model.

Table 2 Correlation matrix

Variable	X_1	X_2	X_3	X_4
X_1	1.000 000	−0.080 722	−0.102 811	−0.040 665
X_2	−0.080 722	1.000 000	0.975 734	0.984 465
X_3	−0.102 811	0.975 734	1.000 000	0.962 605
X_4	−0.040 665	0.984 465	0.962 605	1.000 000

Following the preliminary assessment, we employed the Variance Inflation Factor (VIF) method to test for multicollinearity. This was done by regressing each explanatory variable against the others and calculating the corresponding VIFs. The results show that the VIFs for X_2, X_3, and X_4 all exceed 10, confirming severe multicollinearity. To address this issue, a double-logarithmic model is introduced as follows:

$$\mathrm{Ln}Y = \beta_0 + \beta_1 \mathrm{Ln}X_1 + \beta_2 \mathrm{Ln}X_2 + \beta_3 \mathrm{Ln}X_3 + \beta_4 \mathrm{Ln}X_4 + \mu$$

This study applies stepwise regression to mitigate multicollinearity. The procedure commences by identifying the most significant explanatory variable to establish a baseline model. Subsequently, the remaining variables are evaluated for inclusion, with each candidate being subject to a significance test upon introduction; those failing the test are excluded. This iterative process continues until no additional variables contribute significantly. Furthermore, univariate regressions of $\mathrm{Ln}Y$ against $\mathrm{Ln}X_1$, $\mathrm{Ln}X_2$, $\mathrm{Ln}X_3$, and $\mathrm{Ln}X_4$ were performed, the adjusted R^2 values of which are listed in Table 3.

Table 3 Estimation results of simple linear regression

Explanatory variable (LnY vs. LnX_j)	LnX_1	LnX_2	LnX_3	LnX_4
Parameter estimates	0.400 559	1.518 789	2.264 978	1.394 149
t-statistic	2.529 690	13.860 14	10.299 94	15.159 47
R^2	0.251 949	0.909 997	0.848 108	0.923 636
$\bar{R}^2$	0.212 578	0.905 260	0.840 113	0.919 617

The results indicate that $\mathrm{Ln}X_4$ exhibits the highest adjusted coefficient of determination. Consequently, $\mathrm{Ln}X_4$ was selected as the primary variable for conducting a binary regression analysis with the remaining explanatory variables. The findings are listed in Table 4.

Table 4 Binary regression results based on LnX_4

	LnX_1	LnX_2	LnX_3	LnX_4	R^2
LnX_4					
LnX_1	−0.006 896			1.400 778	
Parameter estimate	(0.061 217)			(0.111 281)	0.923 690
(Standard deviation)	−0.112 655			12.587 72	
t-Statistic					
LnX_4					
LnX_2		0.411 291		1.025 360	
Parameter Estimate		(−0.576 110)		(0.524 911)	0.925 739
(Standard Deviation)		0.713 911		1.953 396	
t-Statistic					
LnX_4					
LnX_3			0.719 568	1.011 798	
Parameter Estimate			(0.327 784)	(0.193 334)	0.939 763
(Standard Deviation)			2.195 250	5.233 420	
t-Statistic					

The t-values for $\mathrm{Ln}X_1$, $\mathrm{Ln}X_2$, and $\mathrm{Ln}X_3$ are −0.112 655, 0.713 911, and 2.195 250, respectively. At this stage, only $\mathrm{Ln}X_3$ passes the significance test, while the F-statistic remains significant. In contrast, $\mathrm{Ln}X_1$ and $\mathrm{Ln}X_2$ have t-values below the critical threshold and fail the significance test. Triadic and quadric regression analyses were subsequently conducted based on $\mathrm{Ln}X_4$ and $\mathrm{Ln}X_3$; however, both models exhibited severe multicollinearity due to the inclusion of additional variables. Thus, only $\mathrm{Ln}X_3$ and $\mathrm{Ln}X_4$ were retained, resulting in the following adjusted model:

$$\mathrm{Ln}Y = -7.358\,077 + 0.719\,568\mathrm{Ln}X_3 + 1.011\,798\mathrm{Ln}X_4$$

$$(1.219\,729)\ (0.193\,334)\ (0.327\,784)$$

$$t = (-6.032\,548)\ (5.233\,420)\ (2.195\,250)$$

$$R^2 = 0.939\,763 \quad \bar{R}^2 = 0.933\,070 \quad F = 140.410\,9 \quad n = 21$$

The model's coefficient of determination and adjusted coefficient of determination are exceptionally high, indicating an excellent fit to the observed data. The t-values for all explanatory variables exceed the critical value $t\frac{0.05}{2}(16) = 2.11$, confirming that X_3 and X_4 have a significant effect on the dependent variable. Moreover, the F-statistic substantially exceeds the critical value, confirming that the combined effect of all explanatory variables significantly influences the dependent variable.

3.2.4 Heteroscedasticity test and correction. The White test was employed to examine heteroscedasticity. The results are as follows: F-statistic: 2.251 729; Prob. F (5,15): 0.102 5; Obs* R-squared: 9.003 952; Prob. Chi-Square(5): 0.108 9. At a significance level of $\alpha = 0.05$, $nR^2 = 9.003\,952 < X^2_{0.05}(5) = 11.07$, and the corresponding p-value is 0.102 459 > 0.05. This indicates that no heteroscedasticity is present in the model, eliminating the need for any heteroscedasticity correction.

3.2.5 Autocorrelation test. The Durbin-Watson (DW) test applied to the revised model yielded a DW statistic of 1.278 054, which falls within the inconclusive range (1.13, 1.54), between d_L and d_U. Therefore, the presence of autocorrelation remains uncertain, necessitating further testing using the Lagrange Multiplier (LM) method. An auxiliary regression was performed on the residuals from the least squares estimation, incorporating both the explanatory variables and the residuals from the second lag period. The resulting LM statistic was $LM = TR^2 = 2.260\,841 < X^2_{0.05}(2) = 5.991$, indicating no autocorrelation. The corresponding p-value, $p = 0.322\,9 > 0.05$, further confirms the absence of autocorrelation. Accordingly, these tests indicate that the random error term exhibits no autocorrelation issues.

4 Conclusions and recommendations

4.1 Conclusions After conducting a quantitative analysis of four selected factors influencing Henan Province's tourism revenue and estimating the regression model, two explanatory variables (the number of tourist visits to Henan Province and the number of travel agencies in Henan Province) were removed due to severe multicollinearity after performing multicollinearity correction. After tests for heteroscedasticity and autocorrelation, the final model was established as follows:

$$\mathrm{Ln}Y = -7.358\,077 + 0.719\,568\mathrm{Ln}X_3 + 1.0117\,98\mathrm{Ln}X_4$$

where X_3 represents total operational railway mileage in Henan Province, and X_4 denotes per capita disposable income of urban residents in Henan Province. The model indicates that, holding other explanatory variables constant, a 1% increase in total operational railway mileage in Henan Province leads to an average increase of 0.719 568% in tourism revenue. Similarly, a 1% increase in per capita disposable income of urban residents in Henan corresponds to an average increase of 1.011 798% in tourism revenue. These results suggest that both total operational railway mileage and per capita disposable income of urban residents are positively associated with the level of tourism development in Henan Province.

4.2 Recommendations Based on the foregoing analysis and conclusions, it is recommended to (i) reduce uncertainties and liquidity constraints to curb residents' precautionary savings behavior, thereby lowering the oversensitivity of tourism spending to current income, raising the marginal propensity to consume tourism, and stimulating provincial tourism demand. Efforts should also be made to actively expand domestic demand, leveraging tourism as a key economic driver. (ii) Policies should align with supply-side structural reforms, accelerating the transformation and consumption upgrade of Central Plains and Yellow River cultural products to boost tourism expenditure and raise residents' income, thus fostering a virtuous cycle between income growth and tourism consumption. (iii) The infrastructure should be improved to support tourism development, including enhancing provincial transportation networks, rationally designing tourist routes across cities and counties, and strengthening the construction and operation of Zhengzhou's bus and railway stations to establish it as a core tourist distribution hub. Connectivity between source markets and major scenic spots should be optimized with tailored itineraries, and dedicated shuttle services from airports, railway stations, and high-speed rail terminals to key attractions should be introduced to ensure seamless travel experiences.

References

[1] MA Y, GUO TT. Practicing the "two mountains theory": Core values and implementation pathways for ecotourism development[J]. Tourism Tribune, 2018, 33(8): 16–18. (in Chinese).

[2] ZHANG P, YANG J, ZHOU X. China's tourism development and regional economic growth: Panel data from 254 prefecture-level cities[J]. Economic Management, 2014, 36(6): 116–126. (in Chinese).

[3] WANG LJ. Synthetic evaluation on economic benefit of provincial tourism in China[J]. Commercial Research, 2007(2): 190–193. (in Chinese).

[4] SHI PH. How to advance all-for-one tourism[J]. Thought Lecture Hall. 102 2016/11 (Part 1). (in Chinese).

[5] YANG ZZ. The essence and development stages of all-for-one tourism [J]. Tourism Journal. 2016: 1002–5006. (in Chinese).

[6] WANG GC. Correlation between tourism and socioeconomic development in Henan Province based on grey theory[J]. Journal of Pingdingshan University, 2012, 27(5): 108–112. (in Chinese).

[7] LI RY, HUANG CZ. Research on the impact of china's transportation infrastructure on regional tourism development: evidence from a threshold regression model[J]. Tourism Science, 2015, 29(2): 1–13, 27. (in Chinese).

[8] WU YY, SONG YX. Analysis of spatial pattern evolution characteristics and influencing factors of China's tourism economy[J]. Geographical Science, 2018, 38(9): 1491–1498. (in Chinese).

[9] HAO C. Research on the impact of high-speed railway opening on urban tourism economic growth[D]. Lanzhou: Lanzhou University, 2020. (in Chinese).

[10] ZHONG HF, LYU HX. Analysis of influencing factors in Guiyang tourism industry based on Eviews[J]. Wuhan: Journal of Wuhan Business College, 2020, 34(3): 5–9. (in Chinese).

[11] HU XW, ZHAO ZX, CHEN YL. Quantitative analysis of factors influencing China's tourism economy[J]. Modern Marketing, 2023(24): 4–6. (in Chinese).