

部分线性模型在贵州地区旅游外汇收入分析与预测中的应用

柏庭引

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2022年9月8日; 录用日期: 2022年9月30日; 发布日期: 2022年10月8日

摘要

部分线性模型既具有参数模型的灵活性, 又能避免维数灾难问题, 因此本文考虑用部分线性模型研究贵州地区旅游外汇收入。首先对贵州省2006~2019年旅游外汇收入数据用主成分分析进行降维, 用处理后的数据拟合部分线性模型, 用拟合后的模型和多元线性回归模型分别对贵州2020年旅游外汇收入值进行预测并分别计算预测值与真实值间的绝对预测误差。部分线性模型所得的绝对预测误差为1.5%, 而逐步线性回归方法的绝对预测误差为13.8%。结果表明, 本文所用方法优于多元线性回归模型。

关键词

贵州旅游外汇收入, 主成分分析, 部分线性模型, 预测

Application of Partial Linear Model in the Analysis and Forecast of Tourism Foreign Exchange Income in Guizhou

Tingyin Bo

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Sep. 8th, 2022; accepted: Sep. 30th, 2022; published: Oct. 8th, 2022

Abstract

Partially linear model not only has the flexibility of parameter model, but also can avoid dimensional disaster problem. Therefore, this paper considers using partially linear model to study the foreign exchange income of tourism in Guizhou. Firstly, the dimensionality of tourism foreign ex-

change income data of Guizhou Province from 2006 to 2019 is reduced by principal component analysis. The processed data are used to fit a partial linear model. The fitted model and multiple linear regression model are used to predict the tourism foreign exchange income of Guizhou Province in 2020, and the absolute prediction error between the predicted value and the real value is calculated. The absolute prediction error of partially linear model is 1.5%, while the absolute prediction error of stepwise linear regression method is 13.8%. The results show that the method is superior to the multiple linear regression model.

Keywords

Guizhou Tourism Foreign Exchange Income, Principal Component Analysis, Partially Linear Model, Predict

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

部分线性模型自 Engle [1]研究天气与电力的关系时提出后, 该模型收到大量学者的关注。相关理论文献特别多, 如 Shi 和 Lau [2]研究了部分线性模型的经验似然估计。Holland [3]在部分线性模型的背景下提出了一种惩罚 B 样条估计, 建立了模型中参数和非参数分量估计的渐近分布。Florens 等[4]在协变量是内生变量时, 在工具变量存在下提出识别参数的充要条件。Ma 等[5]在研究超高维函数部分线性模型的分位数回归时提出具有两个非凸惩罚的双惩罚函数分位数目标函数, 并且通过两步法选择所涉及的最优调谐参数。Bazrafkan 等[6]在基于重尾和不对称分布下, 在模拟仿真中证明所提的 EM 型算法在部分线性模型中的有效性。同样, 部分线性模型在应用上也特别多, 孙志华等[7]在研究税收时引入部分线性模型。王琳琳等[8]则是研究该模型在贵州省公路货运量预测的应用。主成分分析是一种多变量分析方法, 通过降维将原来较多变量重新组合成一组新的互相无关的几个综合变量, 同时根据实际需要选出几个较少的综合变量, 并尽可能多地反映原来变量的信息[9]。而主成分分析方法的应用也有很多, 如虞颖和孟彦菊[10]在研究城市空气质量中用主成分对数据进行降维处理。赖志明等[11]利用主成分分析法对不同产地的凉粉草中总黄酮、总酚酸、总糖、紫云英苷、迷迭香酸、咖啡酸含量进行评价, 验证其含量测定方法。近年来, 旅游业作为带动经济发展的重要产业部门, 被确定为国民经济的主要增长点之一。在旅游功能占主导地位的旅游城市, 旅游业已经成为经济发展的主导产业。其中, 入境旅游业作为旅游业的重要组成部分, 其所发挥的作用是不可忽视的。入境旅游业的发展既创造了大量外汇收入, 提供了更多的就业机会, 拉动了产业发展, 又体现了区域的竞争力和对外开放程度, 因而其也是衡量一个国家或地区旅游业综合实力的关键指标[12]。吕欣[13]和丁丹迪[14]在研究影响旅游外汇收入的因素时都采用了多元线性模型, 而王佳慧[15]利用灰色系统模型去分析山西省旅游服务贸易影响因素。

由于部分线性模型既具有参数模型的灵活性, 又具有非参数模型的优点, 即能避免维数灾难问题。而且通过分析可以发现通过主成分降维后所得的主成分, 旅游外汇收入与其中一个主成分是非线性关系, 显然再考虑参数模型对旅游外汇收入就显得不合理。且目前还未有相关文献应用部分线性模型对旅游外汇收入进行研究, 因此, 本文考虑利用部分线性模型分析贵州旅游外汇收入影响因素。

2. 部分线性模型的估计

部分线性模型的表达形式如下:

$$Y = X^T \beta + g(T) + \varepsilon \tag{1}$$

其中 X, T 是协变量, β 是未知参数向量, $g(\cdot)$ 是未知的光滑函数, 模型误差 ε 满足 $E(\varepsilon|X, T) = 0$ 。在给定 T 时, 对式(1)两边同时取期望得到

$$E(Y|T) = E(X|T)^T \beta + g(T) \tag{2}$$

然后由(1)~(2), 整理得

$$(Y - E(Y|T)) = (X - E(X|T))^T \beta + \varepsilon$$

令 $\tilde{Y} = Y - E(Y|T)$, $\tilde{X} = X - E(X|T)$, 则可以得到:

$$\tilde{Y} = \tilde{X}^T \beta + \varepsilon \tag{3}$$

可由最小二乘方法得到参数 β 的估计, 而 $E(Y|T)$ 和 $E(X|T)$ 都是未知的, 为了能估计出参数 β , 分

别用它们的核估计 $\hat{m}_y(T) = \frac{\sum_{i=1}^n Y_i K\left(\frac{T_i - T}{h}\right)}{\sum_{j=1}^n K\left(\frac{T_j - T}{h}\right)}$, $\hat{m}_x(T) = \frac{\sum_{i=1}^n X_i K\left(\frac{T_i - T}{h}\right)}{\sum_{j=1}^n K\left(\frac{T_j - T}{h}\right)}$ 代替。其中 $K(\cdot)$ 是一核函数,

取为高斯核即 $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$, 带宽 $h = 1.06 * sd(T) * n^{-1/5}$ 。则参数 β 的最小二乘估计为:

$$\hat{\beta} = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T \tilde{Y}$$

进一步得到函数 $g(T)$ 的估计如下:

$$\hat{g}(T) = \hat{m}_y(T) - \hat{m}_x^T(T) \hat{\beta}.$$

3. 部分线性模型的应用

3.1. 数据的选取

本文选取的数据来自贵州统计年鉴。由下列指标组成外汇收入来源: 长途交通 X_1 , 游览 X_2 , 住宿 X_3 , 餐饮 X_4 , 商品销售 X_5 , 娱乐 X_6 , 邮电通讯 X_7 , 市内交通 X_8 , 其他服务 X_9 。旅游外汇总收入为 Y , 单位为万美元。

Table 1. Main indicators of foreign exchange income of tourism in Guizhou during 2006~2019

表 1. 2006~2019 年贵州旅游外汇收入主要指标

| 年份 | 旅游外汇收入 | 长途交通 | 游览 | 住宿 | 餐饮 | 商品销售 | 娱乐 | 邮电通讯 | 市内交通 | 其他服务 |
|------|----------|---------|--------|---------|---------|---------|--------|--------|--------|---------|
| 2006 | 11515.66 | 3569.85 | 817.61 | 2360.71 | 1450.97 | 2153.43 | 713.97 | 57.58 | 115.16 | 276.38 |
| 2007 | 12917.56 | 3100.22 | 490.86 | 1459.68 | 1072.16 | 3733.18 | 477.95 | 426.28 | 232.51 | 1924.72 |
| 2008 | 11697.37 | 3661.28 | 432.8 | 1368.59 | 959.18 | 3193.38 | 514.68 | 514.68 | 362.64 | 690.14 |
| 2009 | 11044.4 | 3224.96 | 441.78 | 1366.37 | 949.82 | 3379.59 | 452.82 | 353.42 | 77.31 | 828.33 |
| 2010 | 12957.88 | 5001.74 | 596.06 | 1399.45 | 997.76 | 2410.17 | 583.1 | 362.83 | 220.28 | 1386.49 |
| 2011 | 13507.18 | 5200.26 | 540.29 | 1701.9 | 729.39 | 3241.72 | 634.84 | 324.17 | 108.06 | 1026.55 |
| 2012 | 16893.6 | 5084.97 | 658.85 | 1756.93 | 1114.98 | 4983.61 | 929.15 | 439.23 | 388.55 | 1573.33 |
| 2013 | 20143.41 | 5881.9 | 1087.7 | 2598.5 | 1631.6 | 5398.4 | 725.2 | 543.9 | 745.3 | 1530.9 |

Continued

| | | | | | | | | | | |
|------|----------|----------|---------|---------|---------|---------|---------|--------|---------|---------|
| 2014 | 21671.2 | 5981.26 | 931.86 | 2383.84 | 1625.34 | 5894.57 | 1516.99 | 476.77 | 671.81 | 2188.79 |
| 2015 | 20111.94 | 5144.37 | 1148.28 | 1594.18 | 1091.67 | 6421.88 | 612.98 | 616.73 | 394.65 | 3087 |
| 2016 | 25270.74 | 6463.61 | 1442.81 | 2003.77 | 1372.06 | 8068.14 | 770.68 | 775.73 | 495.26 | 3878.67 |
| 2017 | 28326.58 | 9381.93 | 1690.54 | 2387.49 | 1565.25 | 9446.99 | 848.21 | 814.98 | 525.32 | 1665.87 |
| 2018 | 31762.59 | 12962.31 | 476.44 | 6740.02 | 4872.38 | 2493.36 | 1013.23 | 101.64 | 1191.1 | 1912.11 |
| 2019 | 34503 | 14687.93 | 562.4 | 7255.98 | 3771.18 | 2984.51 | 1255.91 | 113.86 | 1411.17 | 2460.06 |
| | 2110.8 | 425.33 | 66.91 | 568.44 | 368.54 | 415.83 | 66.91 | 26.38 | 134.25 | 38.21 |

3.2. 主成分分析

利用 2006~2019 年贵州省旅游外汇收入指标 $X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9$, 单位万美元, 数据见表 1, 基于协方差矩阵进行主成分分析。由表 2, 可得前三个主成分的累积贡献率为 0.9107 以及图 1 主成分碎石图, 从而选取前 3 个主成分进行分析。

Table 2. Principal component variance contribution rate

表 2. 主成分方差贡献率

| | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 | Comp.8 | Comp.9 |
|------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| Standard deviation | 2.1218 | 1.7765 | 0.7335 | 0.6680 | 0.4456 | 0.2793 | 0.2396 | 0.1461 | 0.0482 |
| Proportion of Variance | 0.5002 | 0.3507 | 0.0598 | 0.0496 | 0.0221 | 0.0087 | 0.0064 | 0.0024 | 0.0003 |
| Cumulative Proportion | 0.5002 | 0.8509 | 0.9107 | 0.9602 | 0.9823 | 0.9910 | 0.9974 | 0.9997 | 1.0000 |

碎石图

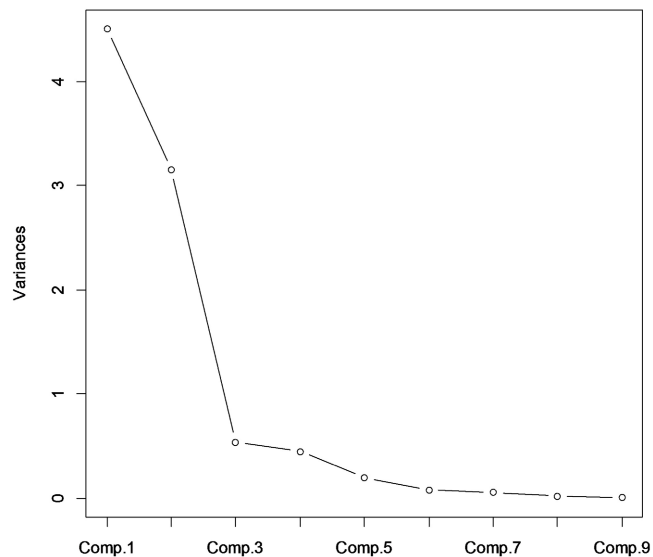


Figure 1. Lithotripsy diagram of principal component analysis

图 1. 主成分分析碎石图

Table 3. Eigenvector matrix
表 3. 特征向量矩阵

| | Prin1 | Prin2 | Prin3 | Prin4 | Prin5 | Prin6 | Prin7 | Prin8 | Prin9 |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| X_1 | -0.4398 | 0.1144 | 0.0817 | -0.2902 | -0.0954 | 0.7453 | -0.0165 | 0.2753 | 0.2490 |
| X_2 | 0.0536 | 0.5087 | -0.1984 | -0.4027 | 0.5783 | -0.1879 | -0.3320 | 0.2348 | -0.0411 |
| X_3 | -0.4636 | -0.0457 | 0.1016 | -0.1714 | 0.0892 | 0.0657 | -0.0745 | -0.4068 | -0.7483 |
| X_4 | -0.4522 | -0.0373 | 0.1521 | -0.2054 | 0.0799 | -0.4743 | 0.6333 | 0.2958 | 0.0942 |
| X_5 | 0.0784 | 0.5433 | -0.1318 | -0.1533 | -0.1013 | 0.0723 | 0.4038 | -0.6486 | 0.2421 |
| X_6 | -0.3341 | 0.1572 | -0.7575 | 0.4952 | -0.0295 | 0.0339 | 0.0884 | 0.1505 | -0.1100 |
| X_7 | 0.2226 | 0.4666 | 0.1119 | -0.0785 | -0.6200 | -0.0597 | 0.0536 | 0.3790 | -0.4238 |
| X_8 | -0.4449 | 0.1159 | 0.0633 | -0.0101 | -0.4051 | -0.4076 | -0.5560 | -0.1678 | 0.3420 |
| X_9 | -0.1378 | 0.4166 | 0.5593 | 0.6380 | 0.2873 | 0.0626 | -0.0070 | 0.0307 | 0.0135 |

根据表 3 特征向量矩阵知所选取的 3 个主成分分别为：

$$\text{prin1} = -0.4398 * X_1 - 0.0536 * X_2 - 0.4636 * X_3 - 0.4522 * X_4 + 0.0784 * X_5 \\ - 0.3341 * X_6 + 0.2226 * X_7 - 0.4449 * X_8 - 0.1378 * X_9$$

$$\text{prin2} = 0.1144 * X_1 + 0.5087 * X_2 - 0.0457 * X_3 - 0.0373 * X_4 + 0.5433 * X_5 \\ + 0.1572 * X_6 + 0.4666 * X_7 + 0.1159 * X_8 + 0.4166 * X_9$$

$$\text{prin3} = 0.0817 * X_1 - 0.1984 * X_2 + 0.1016 * X_3 + 0.1521 * X_4 - 0.1318 * X_5 \\ - 0.7575 * X_6 + 0.1119 * X_7 + 0.0633 * X_8 + 0.5593 * X_9$$

进而，我们可以得到 2006~2019 年贵州旅游外汇收入的三个主成分的值。

3.3. 模型应用

分别做旅游外汇收入值与 3 个主成分的散点图，见图 2~4。

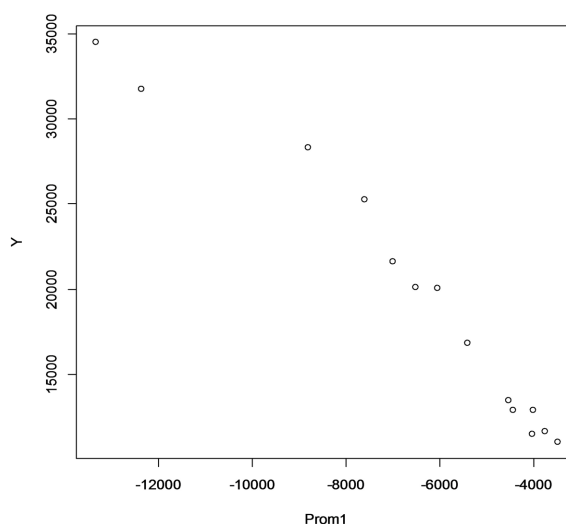


Figure 2. Scatter plot of tourism foreign exchange income and the first principal component

图 2. 旅游外汇收入与第 1 个主成分的散点图

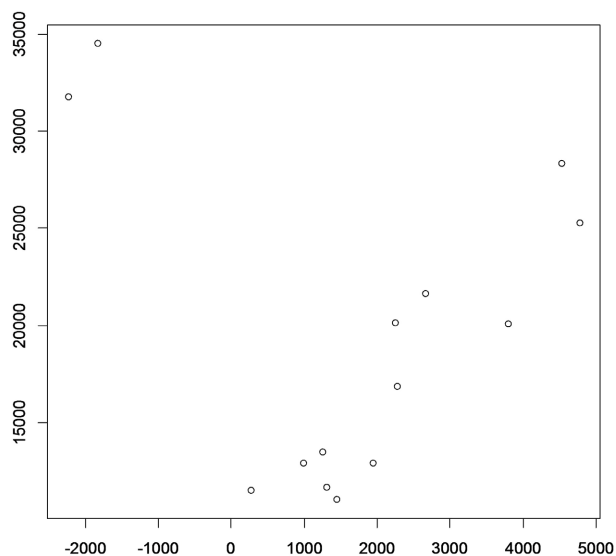


Figure 3. Scatter plot of tourism foreign exchange income and the second principal component

图 3. 旅游外汇收入与第 2 个主成分的散点图

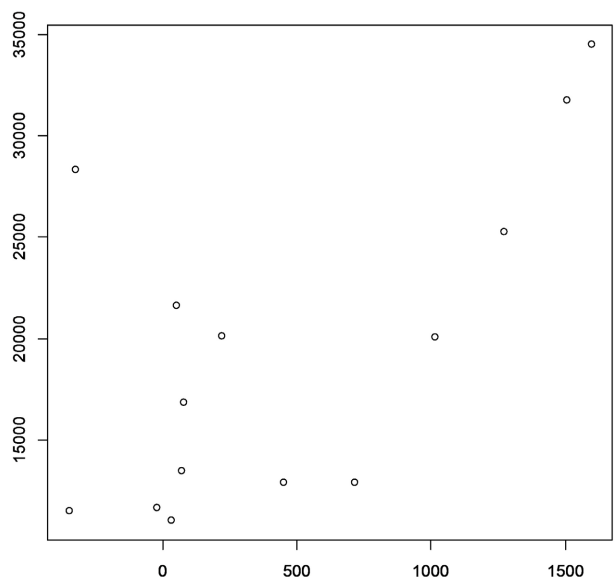


Figure 4. Scatter plot of tourism foreign exchange income and the third principal component

图 4. 旅游外汇收入与第 3 个主成分的散点图

由图 2~4 可以看出：响应变量与第 1 个主成分和第 2 个主成分有着明显的线性关系，而与第 3 个主成分有明显的非线性关系。所以，可以将第 1 个主成分和第 2 个主成分做线性部分的协变量，而把第 3 个主成分作为非参数部分的协变量。具体的模型形式如下：

$$Y = X_1\beta_1 + X_2\beta_2 + g(T) + \varepsilon, E(\varepsilon|X, T) = 0.$$

其中： Y 表示旅游外汇收入总值； X_1 表示第一个主成分； X_2 表示第二个主成分； T 表示第三个主成分； β_1, β_2 表示回归系数； $g(\cdot)$ 表示未知函数； ε 为模型误差。根据本文方法得到参数 β_1, β_2 的估计分别

为 $-2.0074, 0.9495$ ，进而可得到函数 $g(T)$ 的估计为： $\hat{g}(T) = \hat{m}_y(T) - \hat{m}_x^T \beta$ ，其中 $\beta = (-2.0074, 0.9495)$ 。
所以旅游外汇收入值的预测方程为：

$$\hat{Y} = -2.730X_1 + 0.9495X_2 - \hat{g}(T)$$

2020 年旅游外汇收入的有关数据如下：长途交通 X_1 为 425.33，游览 X_2 为 66.91，住宿 X_3 为 568.44，餐饮 X_4 为 368.54，商品销售 X_5 为 415.83，娱乐 X_6 为 66.91，邮电通讯 X_7 为 26.3815，市内交通 X_8 为 134.25，其他服务 X_9 为 38.21，单位为万美元。根据，可以算出第 1 主成分为 -669.70 ，第 2 主成分为 323.20 ，第 3 主成分为 62.61 。所以可求得 2020 年贵州旅游外汇收入的预测值为 2227.05 万美元，而贵州 2020 年旅游外汇收入为 2110.8 万美元，绝对预测误差为 1.5%。而利用逐步线性回归方法进行预测，得到 2020 年贵州旅游外汇收入预测值为 2402.826，绝对预测误差为 13.8%。因此，在此数据集下，部分线性模型的预测效果由于线性模型。

4. 结束语

本文利用主成分分析方法和部分线性模型对贵州省旅游外汇收入进行研究。主成分分析可以很好地进行降维，从而避免维数灾难的出现，而部分线性模型同时具有参数模型和非参数模型的优点。用拟合后的模型和多元线性回归模型分别对贵州 2020 年旅游外汇收入值进行预测并分别计算预测值与真实值间的绝对预测误差。部分线性模型所得的绝对预测误差为 1.5%，而逐步线性回归方法的绝对预测误差为 13.8%。结果表明在此数据集上，部分线性模型效果优于多元线性回归模型的效果。

参考文献

- [1] Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986) Semiparametric Estimates of the Relation between Weather and Electricity Sales. *Journal of the American statistical Association*, **81**, 310-320. <https://doi.org/10.1080/01621459.1986.10478274>
- [2] Shi, J. and Lau, T.S. (2000) Empirical Likelihood for Partially Linear Models. *Journal of Multivariate Analysis*, **72**, 132-148. <https://doi.org/10.1006/jmva.1999.1866>
- [3] Holland, A.D. (2017) Penalized Spline Estimation in the Partially Linear Model. *Journal of Multivariate Analysis*, **153**, 211-235. <https://doi.org/10.1016/j.jmva.2016.10.001>
- [4] Florens, J.P., Johannes, J. and Van Belleghem, S. (2012) Instrumental Regression in Partially Linear Models. *The Econometrics Journal*, **15**, 304-324. <https://doi.org/10.1111/j.1368-423X.2011.00358.x>
- [5] Ma, H., Li, T., Zhu, H. and Zhu, Z. (2019) Quantile Regression for Functional Partially Linear Model in Ultra-High Dimensions. *Computational Statistics & Data Analysis*, **129**, 135-147. <https://doi.org/10.1016/j.csda.2018.06.005>
- [6] Bazrafkan, M., Zare, K., Maleki, M. and Khodadi, Z. (2022) Partially Linear Models Based on Heavy-Tailed and Asymmetrical Distributions. *Stochastic Environmental Research and Risk Assessment*, **36**, 1243-1253. <https://doi.org/10.1007/s00477-021-02101-1>
- [7] 孙志华, 刘金玮, 贾超华. 部分线性模型在北京市税收分析和预测中的应用[J]. 数学的实践与认识, 2011, 41(4): 9-13.
- [8] 王琳琳, 余孝军, 张实. 部分线性模型在贵州省公路货运量预测中的应用[J]. 数学的实践与认识, 2018, 48(21): 234-240.
- [9] 王俊花, 邵林生, 闫建宾, 王瑞钢, 张雪彪, 庞旭, 张沛敏. 基于主成分分析的糯玉米杂交组合农艺性状综合评价[J]. 山西农业科学, 2022, 50(7): 938-944.
- [10] 虞颖, 孟彦菊. 中国 31 个主要城市空气质量的聚类分析和主成分分析[J]. 科技和产业, 2022, 22(5): 246-250.
- [11] 赖志明, 宋晓娟, 魏星任, 莫穗芬, 卢晓莹, 曾唯雅, 严萍, 詹若挺. 基于主成分分析的不同产地凉粉草质量评价[J]. 安徽农业科学, 2022, 50(10): 172-176.
- [12] 李栩萌, 赵艺. 国际旅游外汇收入的影响因素研究——基于省级面板数据的固定效应分析[J]. 商业经济, 2020(9): 95-97. <https://doi.org/10.19905/j.cnki.syj1982.2020.09.033>
- [13] 吕欣. 影响西安旅游外汇收入的多元回归模型[J]. 广西质量监督导报, 2020(4): 204-205.

- [14] 丁丹迪. 探究国际旅游外汇收入的影响因素——基于多元线性回归模型的实证分析[J]. 现代营销(学苑版), 2021(6): 60-62. <https://doi.org/10.19932/j.cnki.22-1256/F.2021.06.060>
- [15] 王佳慧. 山西省旅游服务贸易影响因素研究及发展预测——基于灰色分析的研究[J]. 北方经贸, 2022(5): 150-152.