



中经管理文库
管理学精品系列（二）

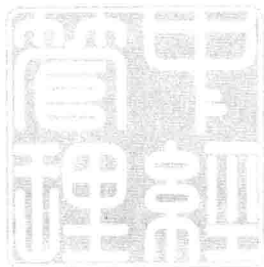
Lüyou Guanli Luncong

旅游管理理论丛

黄先开 / 编



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE



旅游管理论丛

Lüyou Guanli Luncong

黄先开 / 编



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

北 京

图书在版编目 (CIP) 数据

旅游管理论丛 / 黄先开编.

北京: 中国经济出版社, 2014. 7

ISBN 978-7-5136-3322-2

I. ①旅… II. ①黄… III. ①旅游经济—经济管理—文集 IV. ① F590-53
中国版本图书馆 CIP 数据核字 (2014) 第 139938 号

责任编辑 王 建

责任审读 贺 静

责任印制 巢新强

封面设计 华子图文设计公司

出版发行 中国经济出版社

印 刷 者 三河市佳星印装有限公司

经 销 者 各地新华书店

开 本 880mm × 1230mm 1/32

印 张 5.125

字 数 80 千字

版 次 2014 年 7 月第 1 版

印 次 2014 年 7 月第 1 次

书 号 ISBN 978-7-5136-3322-2

定 价 38.00 元

广告经营许可证 京西工商广字第 8179 号

中国经济出版社 网址 www.economyph.com 社址 北京市西城区百万庄北街 3 号 邮编 100037

本版图书如存在印装质量问题, 请与本社发行中心联系调换 (联系电话: 010-68330607)

版权所有 盗版必究 (举报电话: 010-68355416 010-68319282)

国家版权局反盗版举报中心 (举报电话: 12390)

服务热线: 010-88386794

目
录

百度指数与旅游景区游客量的 关系及预测研究·····	1
Research on Urban Tourism Image Positioning : A Case Study of Dengfeng in Henan Province ·····	20
生态旅游形象定位研究·····	44
非物质文化遗产旅游开发中 的三大矛盾探析·····	69
Brand Integration after Cross-border M&A Evidence from China Hotel Industry ·····	93
北京入境旅游市场景气指数研究 ·····	104
An Empirical Study on the Social Network Construction of Enterprises in Beijing High-end Industrial Park ·····	120
中国旅游景区经营管理相关问题探讨·····	135
索引·····	159

百度指数与旅游景区游客量的 关系及预测研究

——以北京故宫为例

黄先开, 张丽峰, 丁于思^①

(北京联合大学旅游学院, 北京 100101)

摘要:

网络搜索数据记录了用户的搜索关注与需求, 为研究旅游经济行为提供了必要的数据库。因此, 本文基于百度指数, 以北京故宫景区为例, 利用计量经济学中的协整理论和格兰杰因果关系分析了百度关键词与北京故宫实际游客量间的关系, 建立了没有百度关键词和加入百度关键词的两种预测模型并进行了预测精度比较, 结果表明: 故宫实际游客量与百度关键词存在长期均衡关系和格兰杰因果关系; 加入百度关键词后的 ARDL 的样本期内的预测精度比

^① 黄先开(1964—), 男, 湖南浏阳人, 教授、博士、博士生导师, 研究方向为旅游经济、数量经济, E-mail: xiankai@bnu.edu.cn; 张丽峰(1969—), 女, 河北秦皇岛人, 副教授、博士, 研究方向为旅游经济、数量经济、低碳经济, E-mail: zhanglifeng999@163.com; 通讯作者: 丁于思, 女, 湖南怀化人, 博士、讲师, 研究方向为酒店管理、游客行为分析, E-mail: lytyusi@bnu.edu.cn。

没有百度关键词的 ARMA 模型提高了 12.4%，样本期外的预测精度提高了 14.5%。运用带有百度关键词的模型，可以达到利用当天及滞后 1 至 2 天的百度指数数据预测故宫当天的游客量的目的，不仅增强了预测的时效性，还可以更加及时、准确地为故宫景区管理部门提供决策依据。

关键词：

百度指数；旅游景区；协整；ARMA 模型；自回归分布滞后模型

中图分类号：

F59 [文献标识码] A [文章编号]

.....►►

一、前言

随着我国旅游事业的迅速发展，国内一些著名的旅游城市、景区景点的游客量日益攀升，特别是“五一”“十一”长假，国内多个地区的旅游景点人满为患，客流量爆棚，远超出景区的承载能力，比如 2012 年“十一”长假期间，陕西华山旅游景区甚至爆发了轰动全国的伤人事件，引起了人们的广泛关注。如何科学合理地对旅游景区的游客量进行预测非常重要，有关部门应结合景区的承载能力，及时对游客量激增的现象进行合理安排并采取应对措施，以避免游客量过多对景区造成破坏，但尤为紧迫的是如何促进景区的可持续发展。

传统预测方法的数据来源于政府及相关统计部门的统计报告，但这些数据的收集和公布往往存在滞后性，并且数据量也难以达到模型预测的要求，这极大限制了预测的有效性。

因此,无论研究者使用的预测方法与预测工具如何先进,也难以从存在着干扰和误差的数据中得到准确的预测结果。因此,寻找质量更高的数据资源对预测结果的改善具有更为重要的意义。搜索引擎的发展为经济学、管理学等学科的预测研究开辟了新的领域。基于网络搜索数据的流感预测、失业率监测、汽车及房地产等行业的销售量预测都取得了较高的准确率,网络数据的即时性能够很好地弥补传统的预测方法的滞后性,具有更强的时效性。因此,利用网络搜索数据对社会经济活动的监测及预测很快引起了众多学者的研究兴趣,它能从大量搜索数据中获取新的关联信息,能从复杂的数据背景中通过数据分析创建合成指数,找出解决问题的有效途径。由于互联网技术在我国日益普及,使得旅游信息的传播不再受时空的限制,因而许多旅游经营者、相关企业以及各地的旅游政府机构都通过互联网这一重要平台发布旅游信息,随之互联网平台也越来越成为广大旅游者出游的重要信息来源。百度和谷歌公司分别推出了百度指数和谷歌搜索解析来分析用户搜索过的条目,通过这项功能,人们可以获取相应时间段内的某关键词分别在百度和谷歌中的关注度趋势和搜索量。这项功能可以直接、客观地反映某特定时间段内的社会热点、用户的兴趣和需求所在。因此,依托网络搜索数据的时效性,挖掘出用户对旅游的网络关注度和实际旅游需求的关系,对于指导旅游管理部门做出科学、合理的决策和安排,促进旅游的可持续发展具有重要的现实意义。

二、国内外研究现状

随着网络信息技术的飞速发展,搜索引擎成为广大网民的重要网络信息获取平台,人们根据自己的需求和兴趣,利用其检索服务功能查询和获取信息,而人们利用搜索引擎查

询和获取信息时留下的庞大的网络搜索数据被网络搜索工具记录下来。有研究表明,这些庞大的搜索数据与现实的社会行为之间存在着一定的相关性。这一研究方法最早应用于流行病监测。Ginsberg 等利用谷歌提供的搜索解析功能,发现部分与流感有关的关键词的网络关注度指数与同期的流发病人数存在较高的相关度。Ginsberg 等由此构建了基于谷歌搜索数据的监测模型,该模型能够比传统监测方法提前 2 周测算出流感的爆发趋势,证明了搜索数据对流感疫情具有一定的预测能力^[1]。之后,这一方法迅速从传染病学向经济领域的各层面扩散,在零售产品销售预测^[2]、房地产市场预测^[3, 4]、股票市场预测^[5, 6]、失业率预测^[7, 8]、消费者信心指数预测^[9, 10, 11]等领域取得了丰富的研究成果。

目前,关于游客量预测的传统的研究方法主要有时间序列法、线性回归模型、指数平滑模型、灰色预测模型、人工神经网络模型等,这些方法主要是利用历史数据进行预测,具有较大的延迟性,很难达到较高的预测精度。基于网络技术对社会行为研究的深入,针对旅游行为的预测也加入到学者们的研究行列中来。目前,基于网络搜索技术的旅游行为预测研究主要集中在网络关注度与游客人数的关系上。路紫等分析了与旅游相关的网络搜索量与实际游客人数之间的相关性及网络信息流对旅游流的导引作用^[12]。路紫等又以澳大利亚旅游网站为研究对象,证实了虚拟网络信息流对现实旅游流的导引作用^[13]。李山等利用百度指数,对我国第一批 5A 级旅游景区的日关注度数据进行了统计和分析^[14]。马丽君等分析了影响客流量及游客网络关注度时空变化的主要因素及其边际效应,构建了国内游客量与游客网络关注度的时空相关模型^[15]。龙茂兴等研究发现区域旅游百度用户关注度与实际旅游客流量具有正相关性,并且前者的变化对后

者的变化具有超前性^[16]。Davidson 等通过实证对比分析以台湾为旅游目的地的旅游网站的信息流与现实旅游流,发现网站信息流对游客旅游行为具有引导潜力^[17]。

综上所述,国外文献主要是利用网络搜索数据对微观、中观和宏观的经济现象进行预测,而利用网络搜索技术对旅游方面进行专门预测的文献较少。国内已有的文献基本上是根据某些旅游网站的信息分析旅游的网络行为,或者只是对网络搜索数据进行分析,而把网络搜索数据和实际数据结合起来,分析两者的相关性并建立模型进行预测的文献还很少。因此,本文以北京故宫为例,利用百度搜索指数,运用计量经济学中的协整与格兰杰因果关系理论,找出网络搜索数据与故宫实际旅游人数间的关系,并运用故宫实际旅游人数建立 ARMA 模型进行预测,然后与加入百度搜索指数解释变量的 ARDL 模型进行预测并比较两者的预测效果,希望能够利用网络搜索数据的时效性来填补传统预测方法的滞后性,以得到更好的预测效果。

三、实证分析

(一) 百度关键词和数据的选取

百度(www.baidu.com)是全球最大的中文搜索引擎。百度公司基于庞大搜索数据推出的百度指数功能,自 2006 年开始可提供不同关键词的网络日搜索量数据。百度指数免费提供以百度网页搜索和百度新闻搜索为基础的海量数据分析服务,用以反映不同关键词在过去一段时间里的“用户关注度”和“媒体关注度”。通过百度指数可以发现、共享和挖掘互联网上最有价值的信息和资讯,直接、客观地反映社会热点、网民的兴趣和需求。百度指数的使用很简单,进入百度指数首页后,在百

度指数搜索栏中输入某个关键词,选择所搜索的时间范围和地区,点击“百度一下”,便可以得到在相应时间段和相应地区内对该关键词的关注度趋势线。鼠标滑过趋势线时,在趋势图中还可以显示鼠标所指出的地区对所键入关键词的日搜索数量。

网络搜索关键词的选取可以说是进行网络搜索与经济行为相关性研究的核心环节,但其选取方法目前尚未取得一致。关键词的选取方法目前主要有三种:技术取词法、直接取词法和范围取词法。技术取词法是利用高性能、大规模的计算机将一切可能的关键词都纳入研究范围内,然后将相关统计模型编成程序并通过运算来选出核心关键词。直接取词法是运用主观经验确定关键词。范围取词法是先确定一个选词的范围,然后在此范围内进行精选以确定关键词。技术取词法的精度虽然高,但是受研究条件的制约极大,需要大量的高速计算机,工作量较大。直接取词法与范围取词法虽然大幅降低了工作量,但是存在遗漏核心关键词的风险^[18]。

旅游活动是大众性的行为活动,随着网络搜索技术的迅猛发展,人们越来越多地通过网络来获取更多的关于旅游的信息。因此,本文以北京故宫博物院为例,运用百度指数,既通过关键词来分析网络搜索数据与实际数据之间的关系,又涉及如何运用网络搜索数据进行预测以提高预测的时效性。根据游客出游前对旅游目的地(景区景点)的旅游信息所关注内容的分析,选出旅游目的地(景区景点)的名称、图片、旅游路线、门票价格等作为与旅游目的地(景区景点)相关的基准关键词。因此,本文选取“故宫”“故宫博物院”“北京故宫”“故宫门票”“故宫门票价格”“故宫图片”等作为基准关键词,然后对这些基准关键词进行百度指数搜索,一方面获得其搜索量,另一方面找出与其相关的关键词。根据这些基准关键词中重复出现次数较多的以及搜索量较大的,最后确定的关键词为“故

宫”“北京故宫”“故宫博物院”和“故宫门票”。同时，本文又运用一些关键词挖掘工具，如站长工具（<http://www.7c.com/keyword/>）、爱站网（<http://www.aizhan.com/>）等进行了检验，进一步验证了以上四个关键词的搜索量排在其他基准关键词的前面。因此，本文将通过这些关键词搜索量的变化来揭示网络搜索数据与故宫实际游客量间的关系。

本文要进行计量经济学模型的建立、分析以及预测，所以模型的样本数据不能太少。本文已得到北京故宫博物院 2007 年 1 月 1 日至 2009 年 12 月 31 日间的实际旅游日接待人数，能够满足样本数据的要求，游客数据来自于 2007 年、2008 年和 2009 年的《故宫博物院年鉴》。与实际数据相呼应，本文分别收集了已确定的四个关键词在相应时间段的百度指数日搜索量。图 1 展示了 2009 年关键词“故宫门票”的日搜索量数据与故宫实际游客量的曲线图，可以看出“故宫门票”日搜索量的变化与实际游客量的走势基本一致。

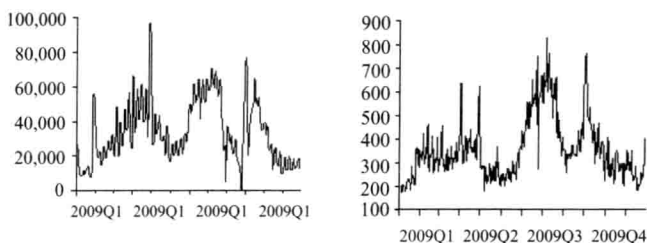


图 1 北京故宫 2009 年实际日游客量和百度关键词“故宫门票”的日搜索量

Fig. 1 Beijing Palace the actual daily travelers and Baidu keyword “the Palace ticket” daily search volume in 2009

(二) 网络搜索数据与实际数据的关系

1. 单位根和协整检验

由于本文的样本数据为时间序列数据,为了保证序列的平稳性和避免伪回归现象,本文在建立计量经济学模型之前要对序列进行单位根检验和协整检验。单位根检验采用 ADF 检验法^[19, 146],检验结果见表 1。

表 1 变量的单位根检验结果

Table 1 The results of variables unit root test

变量	ADF 值	1% 临界值	5% 临界值	10% 临界值	结论
TP	-4.7841	-3.4361 [*]	-2.8639	-2.5681	平稳
BDGG	-4.5239	-3.9665 [*]	-3.4139	-3.1290	平稳
BDBJ	-2.8662	-3.4361	-2.8639*	-2.5681	平稳
BDGP	-5.3413	-3.9665*	-3.4139	-3.1290	平稳
BDGB	-3.6441	-3.9665	-3.4139*	-3.1290	平稳

注: TP 代表故宫实际游客量, BDGG 代表百度关键词“故宫”, BDBJ 代表百度关键词“北京故宫”, BDGP 代表百度关键词“故宫门票”, BDGB 代表百度关键词“故宫博物院”。

由表 1 可知,检验结果表明这几个变量的原序列均是平稳序列,为 0 阶单整,符合协整分析的前提条件。本文下面进行协整分析,目的是检验故宫实际游客量与四个关键词之间是否存在长期均衡关系。

协整检验采用 Engle 和 Granger 提出的两步协整关系检验法,第一步是建立回归方程,建立故宫实际旅游接待人数和四个关键词之间的回归方程。第二步是对方程的残差进行单位根检验,若自变量和因变量之间存在协整关系,回归残差序列应具有稳定性。本文的协整检验属于多变量的协整检验,检验过程中需设置一个变量为因变量,其他变量为自变量,还要进行

OLS 估计并检验残差是否平稳。如果残差不平稳,则需要更换因变量,进行同样的 OLS 估计及相应的残差项检验。当所有的变量都作为因变量被检验之后,仍不能得到平稳的残差项序列,则认为这些变量间不存在协整关系^[19, 155-156]。因此,经过尝试和比较,本文最终确定的协整方程如下(括号中的数值为变量的 t 值):

$$\begin{aligned} TP = & -2609.601 + 0.2058 * BDGG + 11.8062 * BDBJ + 97.6784 * BDGP \quad (1) \\ & (-1.5677) \quad (1.9118) \quad (4.8002) \quad (27.2106) \end{aligned}$$

方程(1)的残差序列的 ADF 值为 -7.5725, 1% 临界值为 -2.5671, 5% 临界值为 -1.9411, 10% 临界值为 -1.6165, 残差的 ADF 值均小于 3 个临界值,表明残差序列是平稳序列。因此,可以判断故宫实际游客量与“故宫”“北京故宫”和“故宫门票”三个百度关键词之间存在协整关系,即长期均衡关系。方程(1)中三个关键词前的回归系数均为正,表明三个关键词与故宫实际游客量间存在正相关的关系,也就是说随着这三个百度关键词搜索量的增加会引起故宫实际游客量的相应增加,只是增加的幅度不同而已,其中关键词“故宫门票”前的回归系数最大,为 97.6784,关键词“北京故宫”的为 11.8062,关键词“故宫”的为 0.2058。长期来讲,百度关键词的搜索指数每增加一个单位,故宫实际游客量会增加相应的单位。

2. 格兰杰因果关系检验

自变量与因变量之间存在协整关系,说明它们之间一定存在着某种形式的格兰杰因果关系。格兰杰因果关系检验能够考察变量之间是否具有统计意义上的因果关系。在检验过程中,由于格兰杰因果关系检验对滞后期的选取有较大的敏感性,所以本文将根据 AIC 和 SC 准则以及残差序列是否存

在相关性来确定最优滞后期。“故宫门票”“故宫”和“北京故宫”三个关键词与故宫实际旅游人数的格兰杰因果关系检验的结果如表 2 所示。

表 2 变量的格兰杰因果关系检验结果

Table 2 The results of variables Granger causality tests

滞后长度	格兰杰因果性	F 值	F 的 p 值	LM(1) 的 p 值	结论
5	TP 不是 BDGG 的 格兰杰原因	0.3822	0.8612	0.0150	不拒绝 拒绝
	BDGG 不是 TP 的 格兰杰原因	1.8559	0.0994	0.7552	
6	TP 不是 BDBJ 的 格兰杰原因	3.9700	0.0006	0.8329	拒绝 拒绝
	BDBJ 不是 TP 的 格兰杰原因	3.4045	0.0025	0.7986	
7	TP 不是 BDGP 的 格兰杰原因	2.8885	0.0054	0.7795	拒绝 拒绝
	BDGP 不是 TP 的 格兰杰原因	11.3376	0.0000	0.9845	

注：LM（1）表示的是模型中残差项的序列相关性检验的一阶滞后的拉格朗日乘子值。

由表 2 可知，关键词“故宫”与故宫实际游客量间存在单向的因果关系，即关键词“故宫”是故宫实际游客量的格兰杰原因的的概率为 90.06%，也就是说关键词“故宫”搜索指数的变化会引起故宫实际游客量的变化。关键词“北京故宫”和故宫实际游客量间存在双向的因果关系，即关键词“北京故宫”是故宫实际游客量的格兰杰原因的的概率为 99.75%，故宫实际游客量是关键词“北京故宫”的格兰杰原因的的概率为 99.94%，故宫实际游客量和关键词“北京故宫”两者间相互

影响。关键词“故宫门票”和故宫实际游客量间也存在双向因果关系，即关键词“故宫门票”是故宫实际游客量的格兰杰原因的概率几乎为 100%，故宫实际游客量是关键词“故宫门票”的格兰杰原因的概率为 99.46%，故宫实际游客量和关键词“故宫门票”两者间也是相互影响的。

（三）预测模型的建立与分析

为了考察网络搜索数据对故宫实际游客量的预测能力，本文首先根据故宫实际游客量的时间序列数据建立 ARMA 模型并进行预测，其次在模型中加入百度关键词作为解释变量再进行预测，最后对这两种模型的预测结果进行比较。

1. ARMA 模型的建立及预测分析

根据故宫实际游客量时间序列的自相关和偏自相关图，初选了几个不同的模型进行比较，再根据模型的拟合优度、AIC 值和 SC 值以及模型预测的均方根误差，最终模型确定为 ARMA(3, 0, 1) 型，此模型具有较小的 AIC 值和 SC 值，模型调整后的拟合优度较高，拟合效果较好，且具有较小的预测均方根误差^[19, 133-143]。为了比较模型的预测能力，此模型的样本期为 2007 年 1 月 1 日至 2009 年 11 月 30 日，2009 年 12 月份的样本作为该模型预测能力的检验数据。具体模型的估计结果如下（括号中的数值为变量的 t 值）：

$$TP=31455.7693+[AR(1)=2.0516, AR(2)=-1.4431, AR(3)=0.3812, MA(1)=-0.8225] \quad (2)$$

$$(9.2858) \quad (39.5447) \quad (-21.4065) \quad (13.5522) \quad (-16.4003)$$

方程（2）调整后的拟合优度为 0.8652，AIC 的值为 20.5564，SC 的值为 20.5798。AR 模型根的倒数 0.97、0.54+.32i 和 0.54-.32i 均在单位圆内，表明模型是稳定的。利用方程（2）对样本期内故宫旅游人数进行预测，预测结果的均方根误差

(Root Mean Squared Error) 为 7006.318。然后再利用方程 (2) 对样本外的 (2009 年 12 月份) 故宫旅游人数进行预测, 实际人数和预测人数的对比图见图 3, 图中 TP 为实际旅游人数, TPF 为预测人数。此时预测结果的均方根误差为 3688.042。

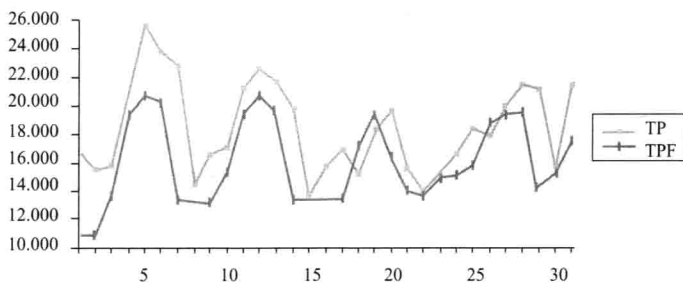


图 2 2009 年 12 月份故宫实际旅游人数和预测人数的对比图

Fig. 2 Comparison of the Palace actual tourist number and the prediction number on December, 2009

2. ARDL 模型的建立及预测分析

由于故宫实际游客量与“故宫”“北京故宫”和“故宫门票”三个关键词之间具有长期均衡关系且存在格兰杰因果关系。因此, 在方程 (2) 的基础上加入这三个关键词作为解释变量进行回归, 考虑到百度搜索指数对实际游客量的滞后影响, 模型中考虑了这三个变量的不同的滞后期数值, 由于模型中既有自变量再的当期及不同滞后期数值, 又有因变量的不同滞后期数值, 因此该模型被称为自回归分布滞后模型 (ARDL 模型) ^[19, 115-117]。方程 (2) 建立的 ARMA 模型可以说是 ARDL 模型的一种特例。本文经过对自变量和因变量不同滞后期数值的反复筛选和测试, 最终确定的回归方程如下 (括号中的数值为变量的 t 值):

$$\begin{aligned}
 TP = & -3390.9961 + 26.5295*BDGP + 4.3578*BDBJ(-2) + 0.0967*BDGG(-1) + \\
 & (-3.1828) \quad (12.3509) \quad (3.1391) \quad (2.6089) \\
 & 1.1452*TP(-1) + 0.3431*TP(-2) \quad (3) \\
 & (39.3844) \quad (-12.0922)
 \end{aligned}$$

方程(3)调整后的拟合优度为0.8759, AIC的值为20.3126, SC的值为20.3407。残差的序列相关性的LM检验值为0.0094, 相伴概率为0.9227, 明显小于5%显著性水平的临界值(为3.84), 表明该模型不存在序列相关性。同方程(2)进行比较可知, 加入百度关键词之后模型的拟合优度提高了, AIC和SC的值变小了, 方程(3)对故宫实际游客量的解释能力得到增强。

同时, 从方程(3)各变量前的回归系数可知, 各关键词搜索指数的当期和各滞后期数值, 以及故宫实际旅游人数的滞后期数值, 对故宫当期实际游客量的影响均为正, 只是影响的幅度不同而已。其中关键词“故宫门票”的影响最大, 其次是关键词“北京故宫”的滞后2期数值, 然后是故宫实际游客量的滞后1期数值和2期数值, 最后是关键词“故宫”的滞后1期数值。根据本模型, 如果知道了北京故宫前1至2天的实际游客量、关键词“故宫门票”的当天搜索指数、关键词“北京故宫”的前2天的搜索指数以及关键词“故宫”的前1天的搜索指数, 就可以预测出故宫当天的实际游客量。因此, 此模型为北京故宫景区管理部门在旅游黄金周期间提前做好应对措施提供了科学的理论依据, 不仅有利于景区游客量的科学安排和对游客量景区资源的合理利用, 而且能够保证游客的安全, 避免景区出现游客量的井喷现象和一些安全事故。

利用方程(3)对样本期内的故宫旅游人数进行预测后的