

# 基于时间序列模型的入境者旅游情况的预测研究——以北京市为例

张良玉

北京建筑大学，北京市，102627；

**摘要：**本文依据 2023 年 - 2025 年北京市入境旅游者人数的月度数据，对其序列图、差分序列图、ADF 检验、定阶模型参数、残差项检验等展开研究，并对其趋势进行预测。研究表明，自 2019 年入境旅游者情况受公共安全事件（COVID - 19）大幅影响起，北京市入境旅游市场在 2023 年展现出强劲的复苏态势；经一阶差分处理后，序列趋于平稳，开展自相关和偏相关图分析，最终确定阶数为 ARIMA (1, 1, 0)，随后依据现有数据进行时间模型预测，预计 2025 年之后的月度，北京入境旅游者情况会保持持续增长状态。

**关键词：**入境旅游者；北京市；ARIMA 模型；时间序列

**DOI：**10.64216/3080-1486.25.05.026

## 1 研究现状

2024 年开初，“中国入境游”“China Travel”变成热门的国际流行语，入境旅游展示出中国高度对外开放的程度。不仅能展现文化软实力，更具有推动国际交往、经济复苏、促进双循环发展的功能。与此同时，国家陆续推出的覆盖 53 国“144 小时”过境免签政策与国际航班航线加密，北京、上海等枢纽城市试点“一程多站”旅游产品等，增进入境旅游便利性，打造文化、旅游深度结合新态势。随着旅游业市场活力地逐渐复苏，研究入境者旅游人数现有情况以及未来趋势对政策成效把握和未来方向走势有重要意义。

在研究旅游情况的现有文献中，涵盖的研究方法有时间序列分析、计量经济建模、人工智能算法及专家经验判断等技术路径。康俊锋、符悦、方雷、李咪咪、谢玉静、周朝阳等人采用定量分析的方法，针对疫情前后进入国境人数的时段变化动向与地域变化特性进行探究，并以 ARIMA - LSTM 模型为基础，以最小二乘法加权，预估疫情发生前后入境人次，预估 2024 年 12 月入境人次将恢复至疫情爆发前水准，此前是自疫情完全放开后。黄秀海以机器学习作为理论支撑，引入“分解 - 重构 - 预测 - 集成”理念，构建基于多尺度分析与多源数据驱动的游客量预测模型。对杭州游客量进行分析预测，得到多源的互联网旅游数据所构建的游客量预测模型其预测精度比未加入的基准模型高，多尺度预

测分析有效性显著提高的结论。刘淑慧、张翊基于 ARIMA 模型分别对我国以及山东省入境游客进行分析以及对入境旅游的发展趋势进行预测，发现 ARIMA 模型在进行预测分析中有较好的效果。以上文献研究中，时间序列分析凭借其数据依赖性低、趋势拟合能力强、建模流程标准化等优势，成为旅游人数预测场景中的主流选择。该方法通过挖掘数据自身的周期性、季节性及长期趋势特征，能够有效量化历史数据，挖掘未来趋势，该模型展现出了良好的适用性和可操作性。

## 2 研究方法

### 2.1 ARIMA 模型结构

ARIMA(自回归移动平均)模型是一个经典的模型，被广泛应用于时间序列分析和预测领域。ARIMA 模型将自回归(AR)、积分 I、移动平均 IMA 三个模型的特点融合在一起，用来在时间序列数据中刻画自相关、趋势和随机波动的特点。

ARIMA 模型一般适用于时间序列数据，有一定的自相关性和趋势性。它的基本原理是在数据中引入差分操作，使得时间序列数据在一阶或多阶差分后变得平稳，然后再结合自回归和滑动平均部分来捕捉数据的自相关性和随机波动。

ARIMA(p, d, q)模型的表示为：

$$X_t = c + \sum_{i=1}^p \Phi_i X_{t-i} + \sum_{i=1}^q \theta_i \xi_{t-i} + \xi_t$$

其中,  $X_t$  是当前时刻的观测值,  $c$  是常数项, 表示序列的均值,  $\Phi_1, \Phi_2, \dots, \Phi_p$  是 AR 模型的参数, 表示滞后项的权重,  $\theta_1, \theta_2, \dots, \theta_q$  是 MA 模型的参数, 表示滞后项的权重,  $\xi_t$  是当前时刻的随机噪声, 一般假定它是均值为 0、方差为常数的白噪声。在 ARIMA 模型里,  $d$  代表需进行的差分次数, 用于让时间序列数据趋于平稳。当  $d=0$  时, ARIMA 模型就退化为 ARMA 模型。ARIMA 模型的选择和参数估计通常需要通过 ACF(自相关函数)和观察时间序列的 PACF(偏自相关函数)来确定合适的阶数  $P$ 、 $D$  和  $Q$ 。然后 ARIMA 模型的参数可以用最小二乘估计的方法进行估算。

### 2.2 ARIMA 模型建模过程

ARIMA 模型只适用于分析平滑的序列。当实际应用中的时间序列不是平滑序列时, 不能直接使用 ARIMA 模型, 首先需要验证观察值序列是否为平滑序列, 可以通过绘制可以表示趋势的绘图工具, 如散点图、折线图等进行观察, 同时也需要进行 ADF 单位根检验来验证序列是否平滑。若在观察趋势图及进行 ADF 检验后, 序列表现为平稳, 则可进行模型定阶, 随后展开建模; 若序列未平稳, 则需执行差分运算, 直到差分后的观测值序列达到平稳, 通常在完成一阶差分与二阶差分后, 序列便会显现平稳,  $d$  值即代表差分阶数; 差分会导致数据本身信息部分丢失, 若差分阶数过多, 则表明该序列不适合用 ARIMA 模型进行建模预测, 需考虑重新选择建模模型以进行数据分析。在数据平稳后对其进行白噪声检验, 白噪声检验是指零均值常方差的随机平稳系列; 如果是平稳非白噪声序列就进行 ACF(自相关系数)、PACF(偏自相关系数)的计算, 及 ARMA 等适用模型识别; 根据 ACF 和 PACF 的输出结果以及软件推荐模型定阶, 对  $p$  和  $q$  的取值多次进行不同的组合尝试, 最终根据 AIC 准则进而确定适合的模型参数, 在置信水平为 95%的前提下构建  $Q$  统计量, 如果检出干扰项属于纯随机性质, 则证明提取的模型信息是充足的。可用于预测分析, 在进行误差分析的同时, 对预测值的误差与真实值进行计算。

一般地, 对于给定的时间序列, 模型的建模过程可

以用图 1 ARIMA 建模流程图中所表示:

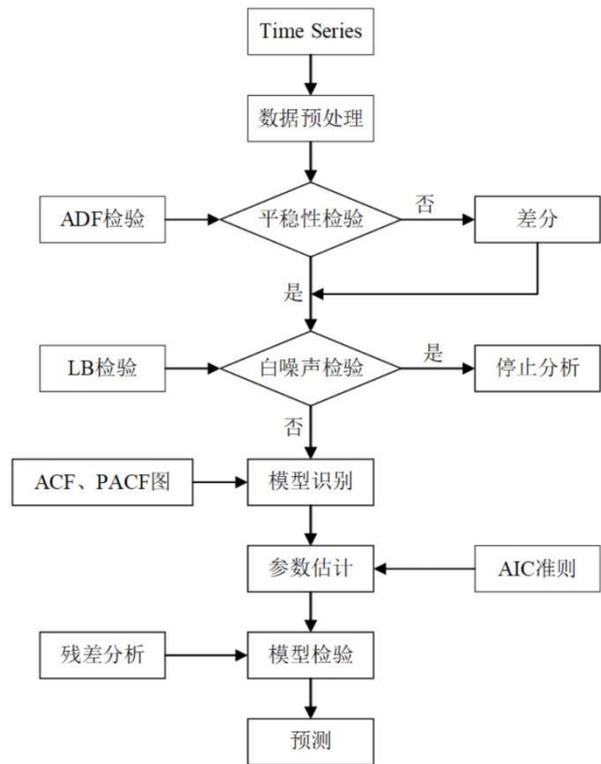


图 1 ARIMA 建模流程图

## 3 以北京市为例入境旅游者人数预测

### 3.1 数据来源

本文研究数据均来自北京市文化和旅游局官方网站政务公开的历年入境旅游者情况, 数据真实准确, 可以公开, 进行相关研究。

### 3.2 ARIMA 模型预测过程

在收集数据的过程中, 由于受疫情影响, 2020 年数据相比较上年大幅度下降, 且包含 2020-2022 均受疫情影响较大, 故取 2023 年至 2025 年共 2 年零 4 个月的北京市入境旅游者(单位: 人)的月度数据, 采用 ARIMA 方法进行分析, 主要分析过程如下。

图形审视, 确认初步模型。首先, 使用图形工具掌握序列特性并初步设定模型参数, 将 2023 年 1 月至 2025 年 4 月北京市入境游客人数的季度数据绘制为时序图, 如图 2 入境旅游者人数序列图所示。自 2020 年疫情爆发, 入境旅游者人数大幅下降, 是近十年旅游业发展的特殊时期, 在此之后, 随着疫情的有效治理, 加之国家推出一系列政策大力复苏旅游业, 游客们出行意愿提高, 入境旅游开始逐渐恢复, 具有明显的长期增长趋势。月度数据经过时序图描述可以看出, 从 2023 年起, 国家

宏观政策不断优化并落地实施，居民生活水平逐渐改善，旅游市场客流稳步回暖，规模有显著回升至疫情前水平的趋势，旅游业也在加快迈入更加多元化的高质量发展阶段，展现出更加强劲的发展活力。

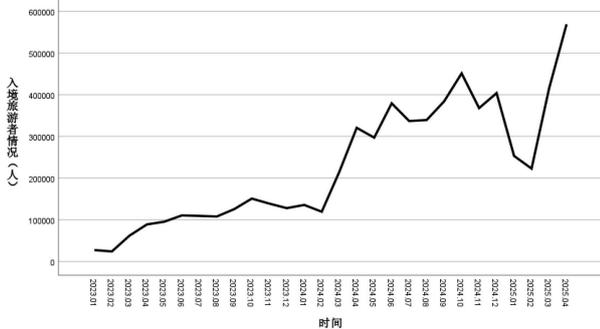


图 2 入境旅游者人数序列图

先对非平滑数列进行一阶逐期差分，结果如图 3 一阶差分时的序图所示，以去除数列中明显的递增趋势，使其其他因素更加明显。从图 3 可以看出，一阶差序经过处理后，在此基础上进行一阶差序的 ADF 测试，序列走势已经完全消失。测试结果如图 4 ADF 测试所示。通过 ADF 检查表，我们知道当差分阶数是 1 点，统计量 ADF 的 P 值是 0.003，低于 0.05，拒绝了原来的假设，确认了一阶差分后的序列是平滑的。因此，ARIMA 模型中的差分阶数确定为 1，之后 ARIMA 模型的定阶将继续进行。

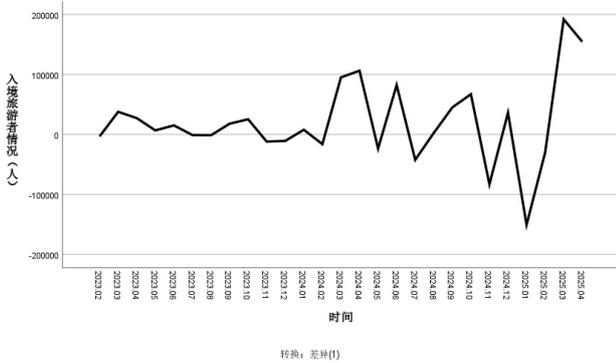


图 3 一阶差分时序图

入境旅游者情况(人)-ADF检验					
差分阶数	t	P	临界值		
			1%	5%	10%
0	1.43	0.997	-3.859	-3.042	-2.661
1	-3.798	0.003	-3.833	-3.031	-2.656

图 4 ADF 检验

接着检验一阶差分后序列的自相关图和偏相关图，根据截尾现象的情况确定自回归阶次p和移动均值阶次q，结果如下图 5 差分后 acf 图、图 6 差分后 pacf 图所示。从输出结果可以看出，自相关图呈一阶截尾，然后根据 AIC 准则可以确定模型是 ARIMA(1,1,0)。

差分后 acf

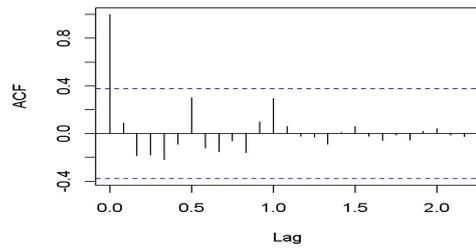


图 5 差分后 acf 图

差分后 pacf

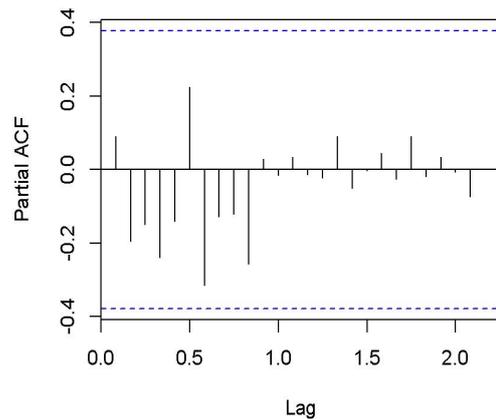


图 6 差分后 pacf 图

根据确定的参数值，通过 SPSS 软件展示本次模型的构建结果，确定本模型公式表示为：

$$y(t) = 19355.155 + 0.101 * y(t-1)$$

此外，ARIMA 模型要求模型误差为白噪声，即误差没有自相关性，能够通过 Q 统计量检验进行白噪声检验，如图 7 白噪声检验所示。我们可以发现其对应的 p 值大于 0.1，这表明模型误差通过白噪声检验，模型拟合效果良好，在 0.1 的显著性水平下，对原假设无法拒绝。

Box-Ljung test

data: fit\$residuals  
X-squared = 0.061231, df = 1, p-value = 0.8046

图 7 白噪声检验

接下来进行模型拟合预测，拟合预测结果如下图 8 模型拟合和预测图所示。图 8 中用不同颜色的线条展示真实值、拟合值以及预测值，由图中可知，拟合效果较好。

残差项的 LM 检验在最后进行，此检验主要用来检验模型的残差序列有无关联性。F 统计量的 P 值为 0.211，大于 0.05，即接近原来的假设，表示序列没有序列相关，因此进一步研究条件异方差等无需使用 ARCH 或 GARCH 模型。综合上述结果，我们可以进行合

理推测，如果没有特殊情况的发生，政策继续利好，北京市入境旅游者人数会保持持续增长状态。

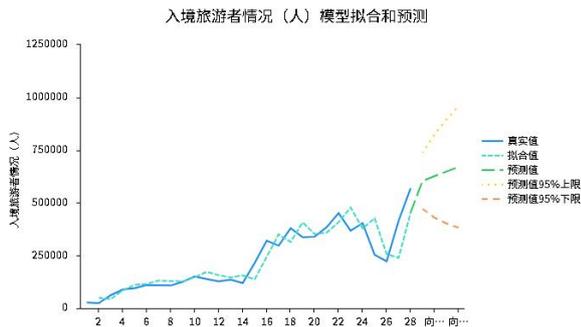


图 8 模型拟合和预测图

### 残差项 LM 检验

F 统计量	1.622	p 值	0.211
T * R <sup>2</sup> 统计量	7.400	p 值	0.193

图 9 残差项 LM 检验图

## 4 结论

本文首先明确研究目的，从北京市文化和旅游局官方网站获取权威公开资料，进一步运用差分方法将非平稳序列进行平稳化处理，对经过一阶差分处理的数据进行 ADF 检验，确保其 p 值小于 0.05，以保证序列是平稳的。然后根据一阶差分后序列的自相关图与偏相关图，来确定模型的 p 值和 q 值参数，从而得出适用于研究北京市入境旅游者人数预测的模型为 ARIMA(1, 1, 0)，并书写该模型公式。随后检验 ARIMA 模型的残差值是否符合白噪声检验，确认本模型残差属于白噪声，说明模型拟合程度较高。最后进行模型拟合预测，拟合预测图展现出本模型具有较高的可行性，拟合程度高，得到北京市入境旅游者人数在没有不可抗因素的特殊情况影响下，会保持持续增长状态的结论。

我国地大物博，且作为全球第二大经济体，拥有者深厚的历史底蕴与现代化智慧城市的双重优势，北京作为我国首都，“双奥之城”，更是当仁不让的多元立体名城。拥有故宫、天坛、长城、“中轴线”等世界物质文化遗产，是中华文明的精髓。2025 年五一假期更是旅客们出游意向较高的城市选择之一，由此可见，预测北京市入境者旅游人数具有多元的重要意义。从政策方面看，有利于动态调整资源配置，落实《北京市推动旅游业高质量发展的实施意见》中“入境游人数年均增长 5%”的目标。从旅游业发展来看，预测可以为景区、酒店、

餐饮等多元从业者提供指导，在大量人流量到达之前提前布局，避免因准备不周而导致供需失调，同时推动智能设备行业打理发展实用产品，如 AI 旅行攻略、智能翻译设备、便捷支付设备等等，让外国友人也体验中国智能。

本文提供的研究方法在在预测中也有不足，在预测过程中，需重点关注以下三大维度：一是外部变量因素，包括国际经济形势、地缘政治变化以及突发事件（如公共卫生事件）的影响，2023 年起的入境游恢复趋势得益于逐渐平稳的社会环境和政策利好，但更全面的预测应该将全球经济形势作为影响因素；二是旅游客群结构因素，每个地区游客可能受不同因素影响关注的旅游重点不一样，应当分析客群结构变化分析原因，分地区给出建议；三是内部变量因素，北京作为我国的政治中心，部分景点在特殊政治活动时有可能进行交通管制，另外，应当考虑景区生态承载能力评估，避免过度开发破坏文化遗产和自然风景，或是人流量超出景区可接待人数上限，造成一系列影响。在接下来的研究中，可以综合上述几点因素进行分析，制定更加符合现实情况的预测模型，进行更加精准的预测，有效提高模型拟合程度，为首都的旅游业发展提供更有力的参考支持。

### 参考文献

[1] 刘淑慧. 对我国入境旅游人数的趋势分析以及预测[J]. 商展经济, 2024, (13): 44-47. DOI: 10. 19995/j. cnki. CN10-1617/F7. 2024. 13. 044.

[2] 康俊锋, 符悦, 方雷, 等. 基于最小二乘法赋权的 ARIMA-LSTM 模型预测入境旅游人数——以上海市为例[J]. 浙江大学学报(理学版), 2023, 50(04): 508-520.

[3] 刘传意. 江西省入境旅游市场分析及人数预测[D]. 江西财经大学, 2022. DOI: 10. 27175/d. cnki. gjxcu. 2022. 001795.

[4] 李玥. 基于 ARIMA 模型的旅游人数预测[D]. 江西财经大学, 2019.

[5] 唐玥. 我国入境旅游需求预测算法研究[D]. 重庆大学, 2019.

[6] 黄晓玉, 王兰会. SPSS 24.0 统计分析[M]. 中国人民大学出版社: 202103. 238.

[7] 涂云东. 时间序列分析[M]. 人民邮电出版社: 202209. 206.