

# Predicting Tourist Number of Guangdong Province Based on Fractal Auto-Regressive Model

Huan Zhao, Weidong Fang

School of Mathematics, South China University of Technology, Guangzhou Guangdong  
Email: 843193045@qq.com, wdfang@scut.edu.cn

Received: Jan. 27<sup>th</sup>, 2019; accepted: Feb. 11<sup>th</sup>, 2019; published: Feb. 19<sup>th</sup>, 2019

---

## Abstract

An effective prediction on tourist number is good for government and related departments taking steps timely to cope with passenger flows. At the same time, it can provide proposal for development, planning and management of local economy. This article adopts monthly data of tourist number coming to Guangdong Province from 2001 to 2017, takes advantage of its long memory, establishes ARFIMA model, and separately predicts twelve months' tourist numbers in 2017. The result shows that the prediction is effective by comparing true value with predicted value.

## Keywords

Long Memory, ARFIMA Model, Fractional-Order Difference, Predict

---

# 基于分形自回归模型的广东省旅游人数预测

赵欢, 方卫东

华南理工大学, 数学学院, 广东 广州  
Email: 843193045@qq.com, wdfang@scut.edu.cn

收稿日期: 2019年1月27日; 录用日期: 2019年2月11日; 发布日期: 2019年2月19日

---

## 摘要

对旅游人数进行有效预测, 有利于政府及相关部门及时采取应对客流量的措施, 同时也为当地旅游业及经济发展、规划、管理提供参考意见。本文采用2001~2016年广东省接待游客人数的月度数据, 利用其具有的长记忆性, 建立ARFIMA (分型差分自回归)模型, 并对2017年12个月的来广旅游人数分别进行预

测, 比较真实值与预测值, 结果表明预测效果较好。

## 关键词

长记忆性, ARFIMA模型, 分数阶差分, 预测

Copyright © 2019 by authors and Hans Publishers Inc.

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

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



Open Access

## 1. 引言

随着我国经济的快速发展, 国民生活水平日益提高, 外出旅游越来越成为人们休闲娱乐的方式之一, 旅游业带来经济快速增长的同时, 景区中大量游客的涌入也给当地交通、环境等造成压力, 因此, 准确预测旅游人数, 以便当地政府以及相关部门及时采取应对措施显得至关重要。关于旅游人数的方法预测已经有很多人做了相关的研究, Granger、Joyeux (1980) [1], Hosking (1981) 分别独立提出 ARFIMA 模型, 在此基础上 Granger [2] 进行了深入研究。冯春山等(2005) [3], 对石油价格数据建立 ARFIMA 模型, 结果表明考虑了长记忆性的 ARFIMA 模型效果好于短记忆性 ARMA (自回归移动平均)模型。姚瑾(2006) [4] 等, 利用 ARFIMA 模型对香港恒生指数进行预测, 结果模型预测结果是失效的, 这是因为模型本身特点以及非线性系统对初始条件敏感性造成的, 翁钢民, 郑竹叶, 刘洋(2009) [5], 利用 ARFIMA 模型对我国入境旅游人数进行了预测, 结果表明 ARFIMA 比 ARIMA (自回归求和滑动平均)和 SARIMA (季节性差分自回归滑动)预测效果好。王振寰, 孙鹏哲(2014) [6]等, 利用 ARFIMA 模型对内蒙古农林牧渔业总产值进行研究, 结果表明拟合较为准确, 误差在可控范围内。刘凯等(2017) [7], 利用分数阶差分对煤炭消费量进行预测, 结果表明 ARFIMA 模型预测偏差更小, 更接近实际值。陈鹏(2018) [8]等, 对黄金价格建立 ARFIMA-GARCH 模型族, 很好地反映了黄金价格波动规律。本文采用 ARFIMA 模型, 对旅游人数序列实行分数阶差分, 能够避免过度差分造成的时间序列数据中信息的损失, 保存原时间序列数据中的有用特征, 减小建模和参数估计中的偏差。

## 2. ARFIMA 模型及预测

### 2.1. ARFIMA 模型定义

如果  $\{x_t\}$  是平稳时间序列, 满足分数阶差分方程  $\phi(B)(1-B)^d X_t = \theta(B)a_t$ , 则称  $\{x_t\}$  服从 ARFIMA( $p, d, q$ )。其中  $B$  为滞后算子,  $\{a_t\}$  是白噪声序列,  $(1-B)^d$  是分数差分算子,  $|d| < 0.5$ ,  $\phi(B)$  与  $\theta(B)$  分别是  $p$ 、 $q$  阶的平稳滞后多项式算子, 且在复平面内根的模大于 1,  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ ,  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ 。

### 2.2. ARFIMA 模型建立

ARFIMA 模型建立具体步骤如下:

- 1) 判断序列平稳性: 根据序列的时序图及自相关图特征初步判断序列的平稳性, 进一步可通过 ADF 检验进行辅助判断。
- 2) 数据预处理及纯随机性检验: 对表现非平稳的序列清理其中的波动趋势, 使之平稳; 对表现平稳

的序列对其进行纯随机性检验。

3) 判断序列的长记忆性及清除短记忆因素: 对得到的平稳非白噪声序列, 画出自相关图, 若函数图象呈多项式速度缓慢递减, 在滞后阶数很大时仍有较强的自相关性, 表明序列具有长记忆性。通过建立自回归 AR( $P$ )模型, 清除短影响因素。

4) 分数阶差分: 通过 R/S 分析法计算长记忆性的平稳非白噪声序列的 Hurst 指数, 利用  $d = H - 0.5$  求出分数阶差分的数值, 对序列进行分数阶差分。

5) ARFIMA 模型定阶: 对分数阶差分后的序列建立 ARMA( $p, q$ )模型, 确定  $p, q$  的值, 即得 ARFIMA( $p, d, q$ )。

6) 估计模型中未知参数  $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$  的值。

7) 检验模型有效性并进行预测。

### 2.3. ARFIMA 模型预测

下面利用 2001~2016 年广东省接待游客过夜人数(万人次)的月度数据[9]建立 ARFIMA 模型(从 2001 年 1 月至 2016 年 12 月共 192 个数据), 并利用模型进行向后 12 步预测, 对 2017 年一共 12 个月旅游人数分别进行预测。建模过程如下:

1) 画出序列  $\{X_t\}$  的时序图以及自相关图, 图 1 是时序图。

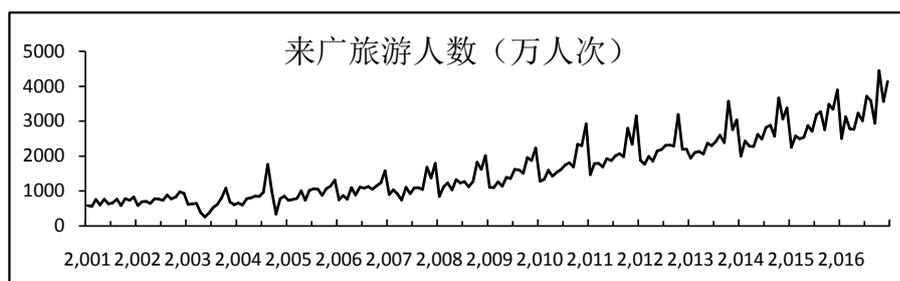


Figure 1. Traveling number sequence diagram

图 1. 旅游人数时序图

由图 2 自相关图可知, 来广旅游人数随时间的变化而变化, 总体表现为上升的趋势。

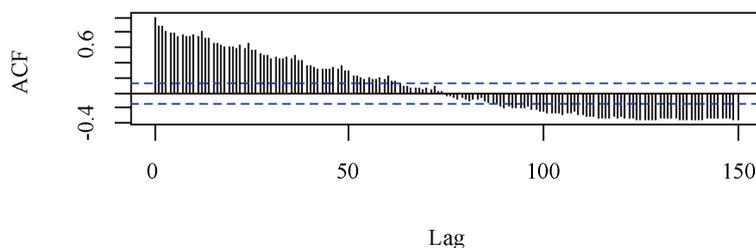
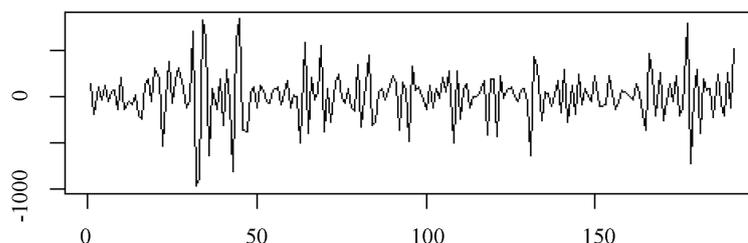


Figure 2. Traveling number autocorrelogram

图 2. 旅游人数自相关图

从自相关图中可以看出序列自相关系数缓慢递减到 0, 在滞后阶数较大的范围内, 自相关系数先正后负, 且表现出三角对称性, 由此可大致判断序列是具有单调趋势且非平稳的。一般地, 考虑对序列进行一次一阶差分, 画出差分后的自相关图发现: 当间隔是周期 12 的倍数时相关性较强。继续对序列进行一次季节差分, 得到差分后  $\{X_t\}$  的时序图  $\{Y_t\}$ , 如图 3 所示。



**Figure 3.** The sequence diagram after the first difference and the first seasonal difference

**图 3.** 一阶差分、一次季节差分后的时序图

此时序列基本围绕一个固定数值波动, 初步判断是稳定的, 为确保准确, 利用 ADF 检验, 结果如图 4 所示。

```

Augmented Dickey-Fuller Test
data: traveldatadiff1s
Dickey-Fuller = -9.7522, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(traveldatadiff1s) : p-value smaller than printed p-value

```

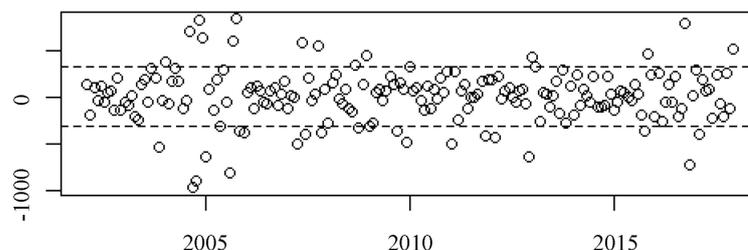
**Figure 4.** ADF test

**图 4.** ADF 检验

$P$  值是小于 0.05, 因此可以判定序列平稳。

2) 对  $\{Y_t\}$  进行纯随机性检验, 计算得出在不同滞后阶数下,  $Q$  检验统计量的  $P$  值均远小于 0.05, 因而是平稳非白噪声序列, 且蕴含的值得提取的信息。

3) 对  $\{Y_t\}$  建立自回归  $AR(P)$  模型, 清除短影响因素, 根据 AIC 最小信息量准则选定  $P = 12$ , 模型参数分别为  $-0.6185, -0.6730, -0.4458, -0.4331, -0.2784, -0.3071, -0.1013, -0.1086, -0.0394, 0.1071, 0.1742, -0.1343$ , 利用  $AR(12)$  估计的值建立置信水平为 95% 的置信区间, 删除掉  $\{Y_t\}$  中 35 个奇异值, 得到新的序列  $\{Z_t\}$ , 过程如图 5 所示。



**Figure 5.** Clear short memory factor sequence diagram

**图 5.** 清除短记忆因素时序图

4) 计算分数阶差分  $d$ : 运用修正的 R/S 分析法计算得到 Hurst [10] 指数为 0.5788384, 根据公式  $d = H - 0.5$ , 即得  $d = 0.0788384$ , 将分数差分后的序列记为  $\{Z_t\}$ , 则有  $Z_t = (1-L)^d Y_t$ , 且满足  $\varphi(L)Z_t = \theta(L)a_t$ , 是均值为 0 的  $ARMA(p,q)$  序列。

对  $\{Z_t\}$  建立 ARMA 模型, 根据自相关和偏自相关图的滞后阶数大致确定  $p, q$  范围, 再利用 AIC 准则, 选择使其值最小对应的模型为 ARFIMA(1,0.0788384,0), 不同模型 AIC 值如表 1 所示。

**Table 1.** Model choose**表 1.** 模型选择

| 模型     | AR(1)   | AR(2)   | MA(1)   | MA(2)   | ARMA(1,1) | ARMA(1,2) | ARMA(2,1) | ARMA(2,2) |
|--------|---------|---------|---------|---------|-----------|-----------|-----------|-----------|
| AIC 数值 | 1999.36 | 2001.28 | 2001.12 | 2001.09 | 2001.29   | 2003.08   | 2001.97   | 2003.85   |

5) 利用 R 软件可得模型的参数方程为  $Z_t = 22.00392 - 0.35Z_{t-1} + \varepsilon_t$ ,

则 ARFIMA 的参数方程为

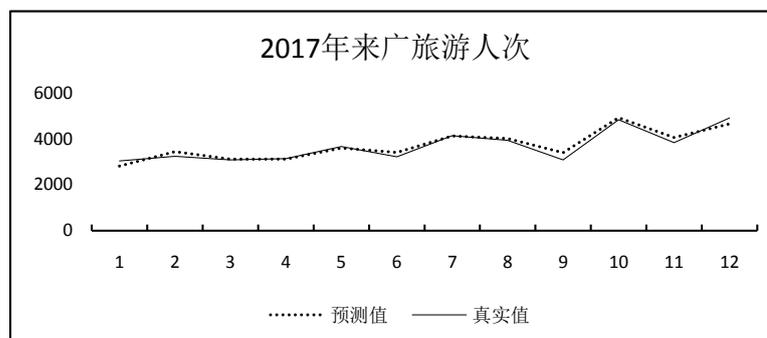
$$(1 + 0.35L)(1 - L)^{0.0788384} (1 - L^2) \nabla^1 X_t = 22.00392 + \varepsilon_t, \text{ 其中 } \varepsilon_t \sim N(0, 20728).$$

6) 对模型的残差进行显著性检验, 得到在各阶延迟下的统计量 LB 统计量  $t$  的  $P$  值都远大于 0.05, 对系数进行显著性检验, 结果表明系数显著非 0, 即模型拟合有效, 利用模型 ARFIMA(1,0.0788384,0) 对 2017 年来广的旅游人数进行预测, 并与真实值进行比较, 如表 2 所示。

**Table 2.** True value and predicted value**表 2.** 实际值与预测值

| 日期     | 实际值     | 预测值     | 相对误差       |
|--------|---------|---------|------------|
| 201701 | 3039.32 | 2814.56 | -7.395075  |
| 201702 | 3247.84 | 3438.42 | 5.8679     |
| 201703 | 3087.63 | 3116.43 | 0.9327542  |
| 201704 | 3135.60 | 3120.64 | -0.4771017 |
| 201705 | 3672.28 | 3604.45 | -1.847081  |
| 201706 | 3218.43 | 3405.35 | 5.807801   |
| 201707 | 4146.63 | 4124.63 | -0.5305513 |
| 201708 | 3942.73 | 4019.62 | 1.950172   |
| 201709 | 3082.35 | 3395.56 | 10.1614    |
| 201710 | 4845.90 | 4925.35 | 1.63953    |
| 201711 | 3844.24 | 4064.24 | 5.722848   |
| 201712 | 4924.02 | 4654.24 | -5.478857  |

由图中可以看出预测的结果总体较为理想, 除 9 月份的相对误差较大以外, 其余月份的相对误差均在较为合理的范围内。图 6 是真实值与预测值的比较图, 以便我们更加方便、直观了解模型拟合效果。

**Figure 6.** Autocorrelation function of agriculture and forestry's index**图 6.** 实际值与预测值拟合

### 3. 结语

随着国民生活水平的日益提高, 以及思想观念的转变, 外出旅游已经越来越成为人们娱乐放松的选择之一, 基于旅游人数具有长记忆性规律, 本文以 2001 年到 2016 年的来广旅游人数月度数据为依据, 建立 ARFIMA(1,0.0788384,0)模型, 并进行了 12 步预测, 得到 2017 年全年 12 个月的来广月度数据, 与真实值比较发现预测结果较好, 为今后旅游业人数预测提供有效依据, 方便政府及部门对旅游业发展做出合理规划及应对措施。

### 参考文献

- [1] Granger, C.W.J. and Joyeux, R. (1980) An Introduction to Long-memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis*, **1**, 15-29. <https://doi.org/10.1111/j.1467-9892.1980.tb00297.x>
- [2] Granger, C.W.J. (1980) Long Memory Relationships and the Aggregation of Dynamic Models. *Journal of Econometrics*, **14**, 227-238. [https://doi.org/10.1016/0304-4076\(80\)90092-5](https://doi.org/10.1016/0304-4076(80)90092-5)
- [3] 冯春山, 吴家春, 蒋馥. 石油价格的 ARFIMA 模型预测研究[J]. 上海理工大学学报, 2005, 27(6): 539-542.
- [4] 姚瑾. 应用 ARFIMA 模型对金融时间序列长期记忆性的研究[D]: [硕士学位论文]. 沈阳: 东北大学, 2006.
- [5] 翁钢民, 郑竹叶, 刘洋. 我国入境旅游预测: 基于 ARFIMA 模型的研究[J]. 商业研究, 2009(6): 1-4.
- [6] 王振寰, 孙鹏哲, 陶伟庆, 邓茗予, 杨婷. 基于 ARFIMA 模型的内蒙古农林牧渔业总产值研究[J]. 内蒙古师范大学学报(自然科学汉文版), 2014(6): 696-698.
- [7] 刘凯, 张晞. 基于分数阶差分 ARIMA 模型的煤炭消费预测[J]. 矿业科学学报, 2017(5): 85-92.
- [8] 陈鹏, 李星野. 基于 ARFIMA-GARCH 模型族的黄金价格预测分析[J]. 电子商务, 2018(5): 42-44.
- [9] 广东省统计信息网[EB/OL]. <http://www.gdstats.gov.cn/tjsj/lyy/>.
- [10] Pai, J.S. and Ravishanker, N. (1996) Bayesian Modelling of ARFIMA Processes by Markov Chain Monte Carlo Methods. *Journal of Forecasting*, **15**, 63-82. [https://doi.org/10.1002/\(SICI\)1099-131X\(199603\)15:2<63::AID-FOR606>3.0.CO;2-5](https://doi.org/10.1002/(SICI)1099-131X(199603)15:2<63::AID-FOR606>3.0.CO;2-5)

#### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [sa@hanspub.org](mailto:sa@hanspub.org)