

# A Study of Prediction Based on ARIMA Model of Road Accidents

Yang Yang

School of Science, Chongqing University of Posts and Telecommunications, Chongqing  
Email: 1638684556@qq.com

Received: Jun. 11<sup>th</sup>, 2017; accepted: Jun. 27<sup>th</sup>, 2017; published: Jun. 30<sup>th</sup>, 2017

---

## Abstract

The number of road accidents is an important aspect on evaluating the performance of traffic management in every country, and a well prediction model of traffic accident could bring a significant improvement to traffic management, traffic control and traffic planning. So, this paper propose an ARIMA model of traffic accident based on the data of traffic accident from 1990 to 2012 in china, and the accuracy is 99.57%. It amply illustrates that this model is efficient and accurate to predict the road accidents in a short run.

## Keywords

ARIMA Model, Road Accidents, Prediction

---

# 基于ARIMA的道路交通事故次数预测研究

杨 阳

重庆邮电大学理学院, 重庆  
Email: 1638684556@qq.com

收稿日期: 2017年6月11日; 录用日期: 2017年6月27日; 发布日期: 2017年6月30日

---

## 摘 要

道路交通事故次数是衡量一个国家交通管理水平的一个体现, 建立一个良好的道路交通事故预测模型对于交通管理、控制和有关规划设计起着极其重要的作用。本文以我国1990~2012年道路交通事故数据为基础, 建立了道路交通事故ARIMA(1,2,2)预测模型, 其预测精度可达99.57%, 表明该模型对我国的交通事故次数预测的短期应用是可行有效的。

## 关键词

ARIMA模型, 道路交通事故次数, 预测

Copyright © 2017 by author 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

随着经济社会的不断发展、人口的增长、道路交通工具数量的增长和道路系统的日益复杂化, 因此对道路交通事故次数进行有效的预测尤为重要, 这对现代道路交通安全管理和有关规划设计等有着重要的意义。

目前, 许多学者对道路交通事故次数预测最常用的方法是时间序列分析, 其中大多运用马尔科夫链模型[1]或者灰色模型[2]等方法对道路事故次数的长期预测比较准确, 但对短期波动预测结果不够准确。本文通过对 1990~2012 年国内道路交通次数序列进行平稳性和白噪声过程的分析, 对道路交通事故次数建立 ARIMA(p,d,q)模型并进行预测, 对国家交通管理水平的发展提供一定的参考价值。

## 2. ARIMA 模型

### 2.1. ARIMA 模型简介

差分自回归移动平均模型(ARIMA)是由 Box 和 Jenkins 提出的一种时间序列建模方法[3]。具有如下结构的模型称为求和自回归移动平均模型, 简记为 ARIMA(p,d,q) 模型:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

式中,  $\nabla^d = (1-B)^d$ , 表示 d 阶差分;  $\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$ , 为平稳可逆 ARMA(p,q) 模型的自回归系数多项式;  $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ , 为平稳可逆模型 ARMA(p,q) 的移动平滑系数多项式。 $\{\varepsilon_t\}$  为零均值白噪声序列。

### 2.2. ARIMA 建立模型步骤

- 1) 获得观察值序列;
- 2) 做出时序图进行初步判断其是否平稳, 再通过观察时间序列的自相关图和偏自相关图来判断时间序列的平稳性。若检验结果为非平稳, 进行从低阶差分运算处理开始, 直至平稳序列;
- 3) 通过观察差分平稳后的自相关图和偏自相关图, 大致确定 ARIMA 中 p, q 的阶数, 并对模型中的 p 和 q 两个参数进行多种组合选择, 两个 AIC 和 SC 函数值达到最小的模型为相对最优的模型;
- 4) 利用最小二乘法对模型进行参数进行估计, 并进行显著性检验, 应剔除不显著参数所对应的自变量并重新拟合模拟模型;
- 5) 进行白噪声检验, 若残差序列是非白噪声序列, 进行拟合 ARIMA 模型, 直至残差序列为白噪声序;
- 6) 根据检验的结果, 确定模型, 并利用该模型进行预测。

### 3. 数据预测处理与 ARIMA 建模

#### 3.1. 数据收集和简单分析

根据中华人民共和国国家统计局《2015 中国统计年鉴》[4]数据查询得到相关数据，统计出中国1990~2015年的国内道路交通事故发生的次数。并根据 Eviews 软件[5] [6]绘制出(图 1)的时间序列图和(图 2)的自相关和偏相关图来进行该序列平稳性的判断。

如图 1 所示我们可以看出 1990~2002 道路交通事故次数呈上升趋势，并且增幅不同。2002~2012 年后道路交通事故次数呈衰减趋势，并且减幅不同。我们可以根据做出该序列的自相关和偏自相关图来判断该序列是否平稳，根据图 2 可看出一期之后还有很多期的系数不在二倍标准差范围之内，并且自相关图没有明显的拖尾性或截尾性，所以我们判断此序列是非平稳的。

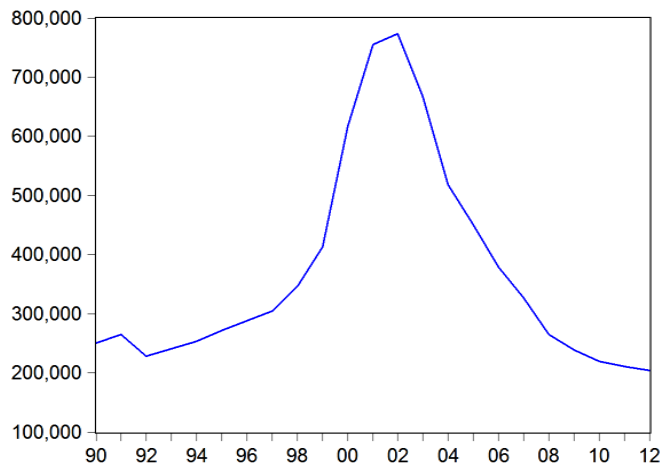


Figure 1. The time series plot of original data  
图 1. 原数据的时间序列图

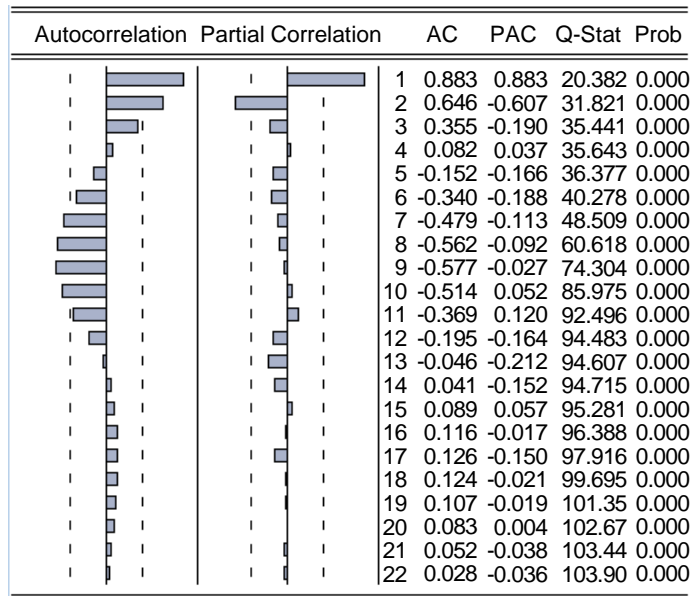


Figure 2. The partial and autocorrelation diagram  
图 2. 原数据的自相关和偏相关图

先由时间序列图可以看出明显的非平稳趋势,先逐渐递增,在2002年又开始呈递减趋势。从时间序列的自相关和偏自相关图中有多处其自相关系数超出标准差范围,所以对序列进行差分处理提取信息。

### 3.2. 序列的差分处理

如图3所示,全国交通事故次数二阶差分后的时间序列图,没有明显的非平稳特征。

二阶差分后,根据其自相关和偏自相关图判断其是否平稳:

如图4所示我们可以看出,1期(包括1期)之后的自相关系数和偏自相关系数均在2倍的标准差范围内。我们初步判断二阶差分后的序列是平稳的。当然高阶差分也会满足序列的平稳性,但是过度差分会丢失大量的相关信息,所以并不是差分阶数越高越好,需要合适。在这里我们最终定为二阶差分即可。

### 3.3. 单位根检验

运用 Eviews 软件对其差分后的序列和原序列进行单位根检验(ADF 检验和 PP 检验)如表1所示,二阶差分的时间序列在90%的置信区间下通过了 ADF 和 PP 单位根检验,可以看做平稳的时间序列。

### 3.4. 运用 AIC 准则和 SC 准则定阶

AIC 准则[7](An information)是由日本统计学家赤池弘次(Akaike)在1973年提出的。该准则既考虑拟合模型的数据接近程度,也考虑模型中所含待定参数的个数,适用于 ARMA 模型检验。

AIC 准则的一般形式可表为:  $AIC = -2\ln(\text{模型最大似然}) + 2(\text{模型独立参数个数})$

对于 ARIMA 模型,若自相关函数在延迟 p 阶拖尾和偏自相关系数在延迟 q 阶拖尾,模型的阶数分别是 p, q。可以初步估计 p 为 1, q 为 2。检验序列为二阶差分时间序列,因此系数 d = 2。进一步得到 4 个 ARIMA(p,2,q)模型的各参数。如表2所示,ARIMA(1,2,2)模型调整后可决系数最大, AIC, SC 参数值最小,可得 ARIMA(1,2,2)模型为最优模型。

### 3.5. 模型的拟合优度检验和残差序列检验

建立模型 ARIMA(1,2,2)模型并估计,见图5。

如图5所示分析: F-statistic 对应的概率值为 0.025764,说明建立的 ARIMA 模型具有较好的线性关系。从残差拟合图(图6),从图中我们可以看出 ARIMA 模型的拟合效果较好。如图7所示我们做出了残差的自相关和偏相关图发现其系数均在两倍的标准差范围之内,我们认为建立的 ARIMA(1,2,2)模型的残差序列是白噪声序列。可得 ARIMA(1,2,2)模型的输出结果得到回归方程为:

$$D_{x_t}^2 = -0.2734D_{x_{t-1}}^2 + \varepsilon_t + 0.8061\varepsilon_{t-1} + 0.9949\varepsilon_{t-2}$$

### 3.6. 模型预测

运用 Eviews 软件做出了其二阶差分后序列的预测图(见图8),以及对原序列全国交通事故次数的预测图(见图9)。

如图8所示我们发现:不等系数为 0.41,方差的比例也比较小,协方差比例为 0.98,由图9所示我们发现:不等系数为 0.05,方差的比例也比较小,协方差比例为 0.81,表明模型的预测能力效果比较好。

利用模型对 2013~2017 年数据进行预测,计算得到 2013~2017 年全国交通事故次数数据以及 2013~2017 年交通事故次数预测的误差率,结果见下表。

从表3可以看出,ARIAM(1,2,2)模型的平均预测误差是 0.85%,平均预测效果高达 99.17%,预测效果理想。

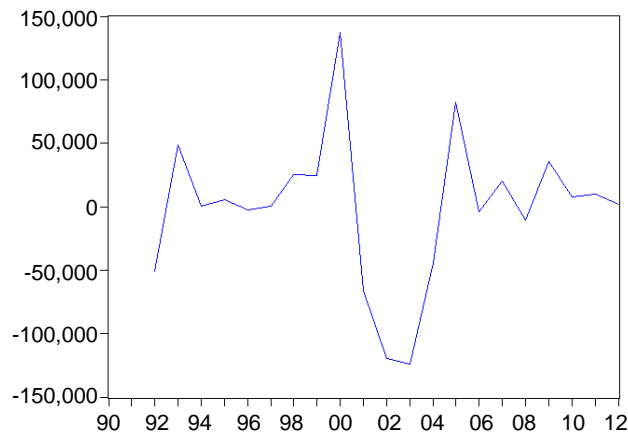


Figure 3. The time series plot of the second order difference about the times of road accidents from 1990 to 2012

图 3. 1990~2012 年全国交通事故次数二阶差分时序图

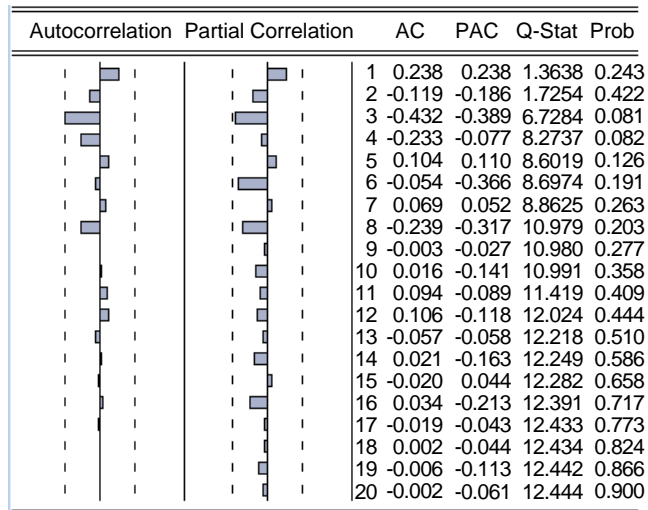


Figure 4. The partial and autocorrelation diagram based on the second order different data

图 4. 二阶差分后的自相关和偏相关图

Table 1. The result of the unit root test about the time series and difference

表 1. 时间序列及其差分的单位根检验结果

检验方法	置信水平	原序列	一阶差分序列	二阶差分序列
ADF 检验	ADF 值	-0.55	-2.54	-3.30
	1% level	-2.68	-3.81	-4.50
	5% level	-1.96	-3.02	-3.66
	10% level	-1.61	-2.65	-3.27
PP 检验	PP 值	-0.64	-2.01	-3.35
	1% level	-2.67	-3.79	-4.49
	5% level	-1.96	-3.01	-3.66
	10% level	-1.61	-2.65	-3.27
平稳性		非平稳	非平稳	平稳

**Table 2.** The parameters table of ARIMA(p,2,q)model  
**表 2.** ARIMA(p,2,q)模型的参数表

模型参数	可决系数	调整可决系数	AIC	SC
(1,2,1)	0.15	0.05	24.91	25.06
(1,2,2)	0.38	0.28	24.50	24.70
(2,2,1)	0.35	0.23	24.76	24.96
(2,2,2)	0.27	0.07	24.99	25.24

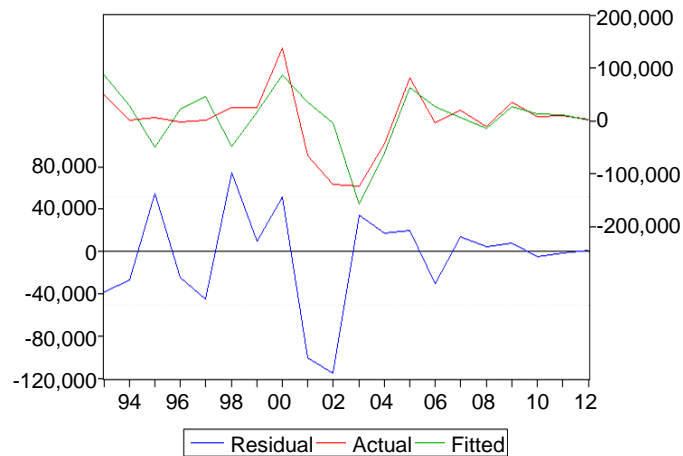
**Table 3.** The predictive result of the road accident times from 2013 to 2017 according to ARIMA(1,2,2) model  
**表 3.** ARIMA(1,2,2)模型预测 2013~2017 年全国交通事故次数结果

年份	二阶差分预测值	人数预测值	人数实际值	预测误差率(%)
2013	3774	201354	198394	1.49
2014	5320	197912	196812	0.56
2015	-8260	186969	187781	0.43
2016	5542	184291		
2017	-2759	179928		

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2571.822	14792.68	-0.173858	0.8638
AR(1)	-0.273387	0.233645	-1.170097	0.2564
MA(1)	0.806114	0.202571	3.979407	0.0008
MA(2)	0.994870	0.059989	16.58408	0.0000
R-squared	0.379207	Mean dependent var	1184.304	
Adjusted R-squared	0.281188	S.D. dependent var	55269.43	
S.E. of regression	46858.97	Akaike info criterion	24.50444	
Sum squared resid	4.17E+10	Schwarz criterion	24.70192	
Log likelihood	-277.8011	Hannan-Quinn criter	24.55411	
F-statistic	3.868680	Durbin-Watson stat	1.987048	
Prob(F-statistic)	0.025764			
Inverted AR Roots	-0.27			
Inverted MA Roots	-0.40-0.91i	-0.40+0.91i		

**Figure 5.** Final result of ARIMA(1,2,2)model

**图 5.** ARIMA(1,2,2)模型的输出结果



**Figure 6.** The residual errors and fitting chart

**图 6.** ARIMA(1,2,2)模型的残差拟合图

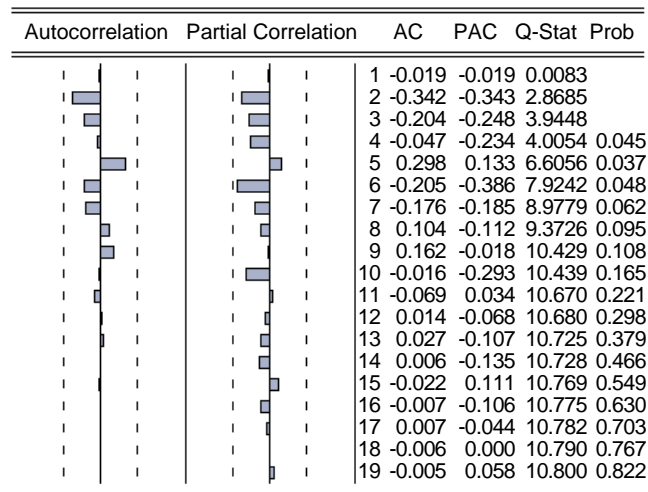


Figure 7. The partial and autocorrelation diagram of residual errors

图 7. 残差的自相关和偏相关图

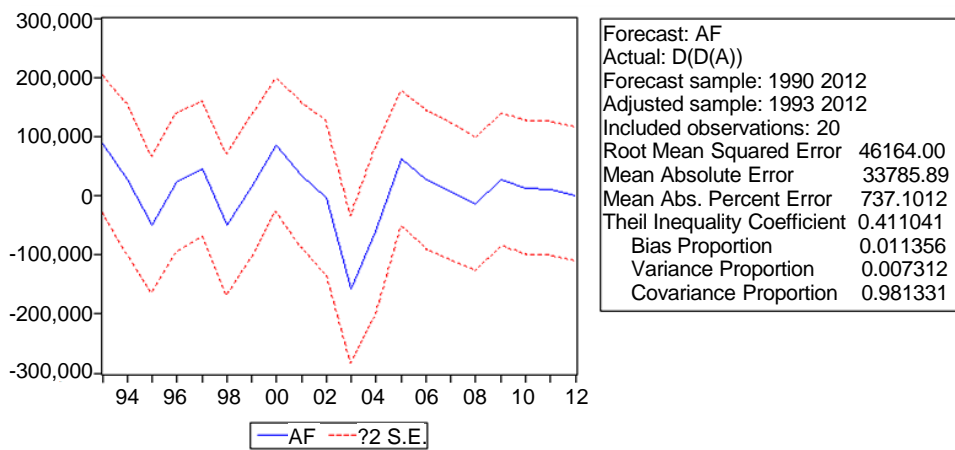


Figure 8. Final chart of the predictive result based on the second order difference series

图 8. 二阶差分序列预测结果输出图

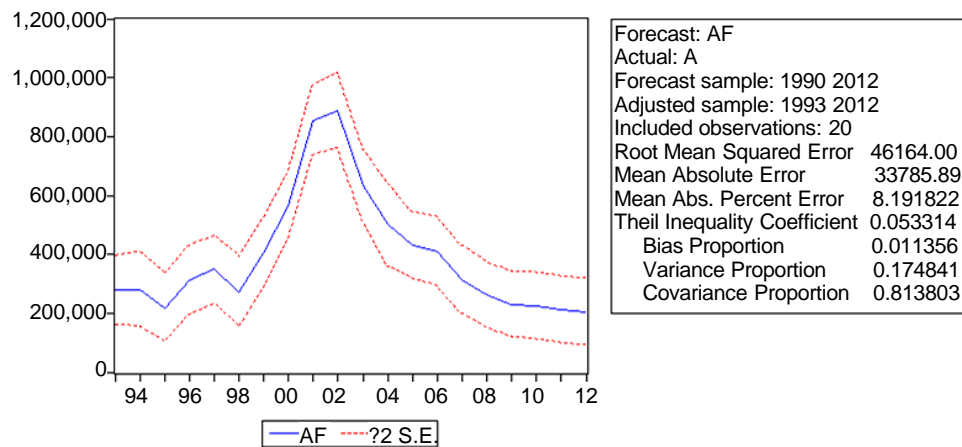


Figure 9. Final chart of the predictive result based on the original data

图 9. 原序列数据预测结果输出图

## 4. 结束语

ARIMA 可以通过差分等数据转换方法将非平稳序列转变成零均值的平稳随机序列, ARIMA 使残差进入模型, 提高了模型的精度, 但是 ARIMA 建模法假定时间序列为未来的发展模式与过去的趋势一致, 因此它往往只适用于作短期预测。基于 1990~2012 年国内道路交通事故次数的时间序列分析, 可以得出 2013~2015 年三年预测值和实际值得误差率平均值为 0.85%, 根据 ARIMA(1,2,2)模型预测出的数据具有一定的科学依据, 可以为增强道路交通管理提供重要的理论依据。

## 致 谢

感谢我的老师鲜思东教授的全程指导。

## 基金项目

重庆市研究生教育教学教委项目(YJG143010), 重庆邮电大学校级教改项目(XJG13028)。

## 参考文献 (References)

- [1] 赵玲. 基于马尔可夫链的道路交通事故预测研究及应用[D]: [博士学位论文]. 西安: 长安大学, 2013.
- [2] 刘淑环. 北京市道路交通事故灰色预测模型的构建与应用[J]. 首都师范大学, 2008, 39(9): 115-122.
- [3] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) Time Series Analysis: Forecasting and Control. *Journal of Marketing Research*, **14**, 199-201.
- [4] 中华人民共和国. 中国统计年鉴 2015 [M]. 北京: 中国统计出版社, 2015.
- [5] 高铁梅. 计量经济分析方法与建模 Eviews 应用及实例[M]. 北京: 清华大学出版社, 2009: 553-556
- [6] 张晓峒. 计量经济学软件 Eviews 使用指南[M]. 第 2 版. 天津: 南开大学出版社, 2004.
- [7] 刘璋温. 赤池信息量准则 AIC 及其意义[J]. 数学的实践与认识, 1980(3): 65-73.

### 期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

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

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