Prediction of Road Traffic Accident Death Holl Based on Time Series

Xie Huawei

Fujian Police College, Fuzhou 350007, China
momoxhw@163.com

Keywords: ARMA model, time series, road traffic accident, death toll; prediction

Abstract: In order to study the change rule of road traffic accident, and to better evaluate the road traffic safety, road traffic planning and accident prevention and control, the research work is carried out. In this paper, the time series analysis method is adopted to analyze the trend rule in the time series of the road traffic accident death holl. In addition, data preprocessing and model identification and testing were carried out. Finally, ARMA (1,1) prediction model was established. The number of road traffic accident death holl in China from 2003 to 2015 was analyzed and predicted. The results show that the predicted trend is consistent with the actual trend. The maximum error rate is 2.02%, and the comprehensive error rate is only 0.60%. The accuracy is high and the prediction effect is satisfactory. The time series analysis and prediction method highlights the role of time factors in the prediction. It is more suitable for short and medium term forecasting without considering the influence of specific external factors.

1. Introduction

According to the 2015 global road safety status report released by the world health organization, despite improvements in road safety in recent years, about 1.25 million people die each year from road accidents. Road traffic accidents are now the ninth leading cause of death in the world. The ranking is expected to rise to seventh place by 2030.

Road traffic accident prediction can provide safety information for safety managers. According to the predicted results, it can strengthen the weak links, eliminate various hidden dangers and optimize the system security. In addition, the accident rule can be explored through accident prediction. It can provide decision support for road traffic safety evaluation, road traffic planning and accident prevention and control.[1]

The basic idea of time series prediction is to treat accident time series as a sample of random variable series. [2]Probability statistics are used to minimize the effects of random factors, periodic or seasonal changes. The prediction is made by analyzing the general trend of the accident. It has the characteristics of wide application range and small prediction error. In this paper, the ARMA model of road traffic accident is proposed by using stationary time series analysis. Based on the time series data of road traffic accident deaths from 2003 to 2015 in China, this paper explores the sequence characteristics and development rules of road traffic accident deaths.

2. Time series prediction method based on ARMA model

Auto-Regression and Moving Average Model is an important method to study time series. It was proposed by Box and Jenkins in the 1970s. It is based on the "blending" of autoregressive model (AR model) and sliding average model (MA model). It is the most commonly used model to bridge stable sequences. It is divided into AR model, MA model and ARMA model.[3]

2.1 AR model

AR model is also called autoregressive model, and its prediction method is through the linear combination of the observed value in the past and the interference value in the present. The model with the following structure is called p-order autoregressive model, or AR(p).

$$x_t=\phi_0+\phi_1 x_t-1+\phi_2 x_t-2+\ldots+\phi_p x_t-p+\varepsilon_t \tag{1}$$

The value of the random variable Xt is xt-1, xt-2 in the previous p period... Xt-p multiple linear regression. Xt is mainly affected by the sequence values of the past p period. The error term is the current random interference t, which is zero mean white noise sequence. When ϕ 0 = 0, it may be called centralized AR (p) model.

2.2 MA model

The MA model is also called the sliding average model. It is predicted by a linear combination of past and present interference values. Models with the following structures are called q-order autoregressive models, or MA(q) in short.

$$x_t=\mu+\varepsilon_t-\theta_1\varepsilon_t-1-\theta_2\varepsilon_t-2-\ldots-\theta_q\varepsilon_t-q \tag{2}$$

That is, the value of the random variable Xt at time t, Xt is the random disturbance of the previous q period. , which is a multivariate linear function of t minus q. The error term is the random interference t in the current period, which is the zero-mean white noise sequence,and$\mu$is the mean value of the sequence {Xt}. Xt is considered to be mainly affected by the error term of past q period. Whenμ =0, it's called the centralized MA(q) model.

2.3 ARMA model

ARMA model is the most commonly used stationary sequence fitting model. The formula of ARMA model is as follows:

$$x_t=\phi_0+\phi_1 x_{t-1}+\phi_2 x_{t-2}+\ldots+\phi_p x_{t-p}+\varepsilon_t-\theta_1\varepsilon_{t-1}-\theta_2\varepsilon_{t-2}-\ldots-\theta_q\varepsilon_{t-q} \tag{3}$$

The value of the random variable Xt is not only related to the sequence value of the previous p period, but also related to the random disturbance of the previous q period.

When q=0, the RAMA (p,q) model degenerates into AR (p) model.

When p=0, RAMA (p,q) model degenerates into MA (q) model.

RAMA (p,q) model has the property that the autocorrelation coefficient does not truncate and the partial autocorrelation coefficient does not truncate.

3. Establishment of ARMA model

If a sequence of observed values can be determined as a stationary non-white noise sequence through sequence preprocessing, the ARMA model can be used to model the sequence [4]. The ARMA model establishment process is shown in figure 1.

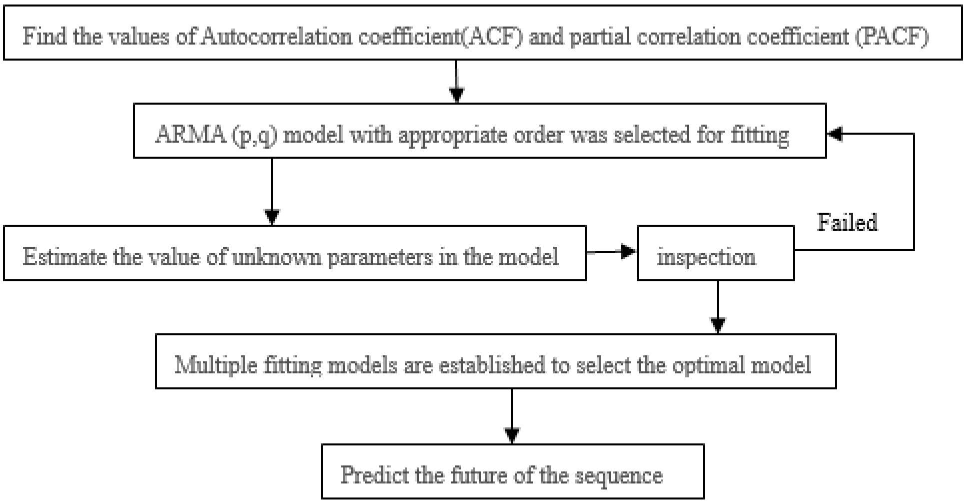


Figure 1. ARMA model establishment process

4. An empirical study on the death toll

4.1 Data source

According to the website of the national bureau of statistics of the People's Republic of China and relevant literature and materials, the death toll sequence of China's road traffic accidents from 2003 to 2015 was established using Eviews7.0, as shown in figure 2.

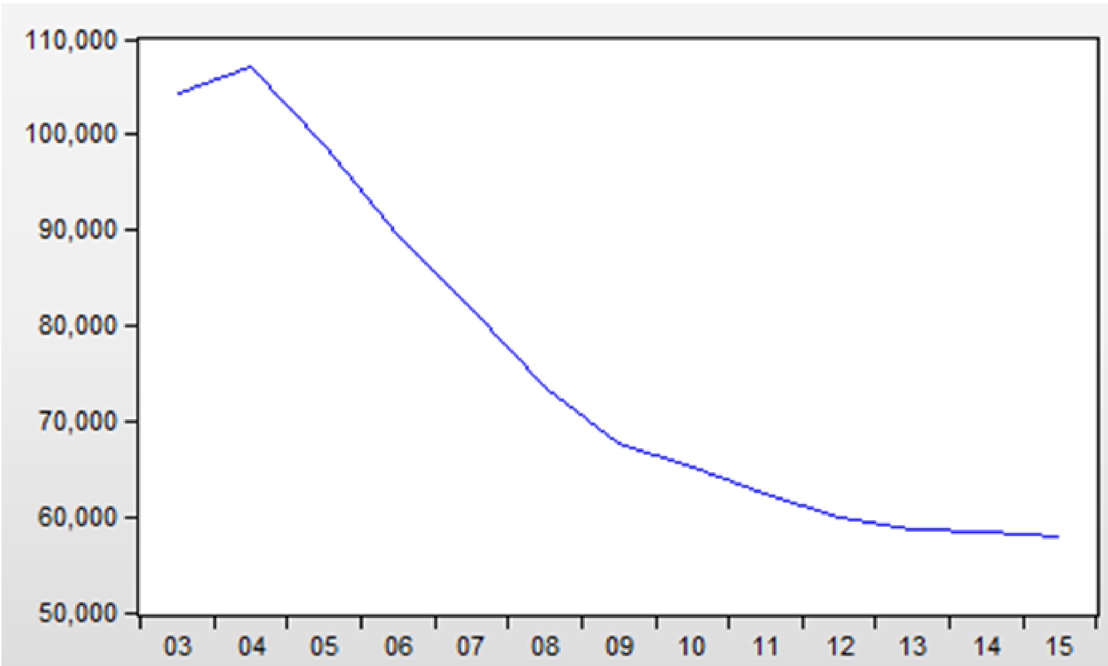
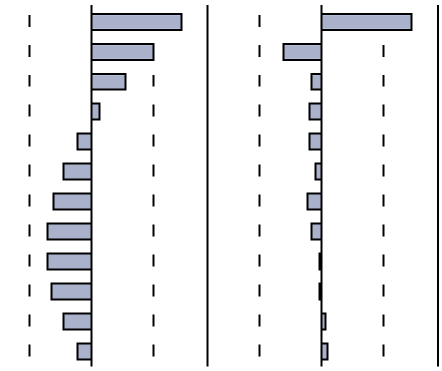


Figure 2. Time series of road traffic accident death roll in China from 2003 to 2015

4.2 Data stability test

To investigate the stability of data, the first step is to examine the time series trend graph. Secondly, the sequence itself is observed. At the same time, the data should be combined with autocorrelation graph, partial autocorrelation graph and unit root. The autocorrelation graph and partial autocorrelation graph of road traffic accident death toll in China from 2003 to 2015 are shown in figure 3. The ADF test results are shown in figure 4.



Correlogram of P

Date: 06/24/18 Time: 10:01
Sample: 2003 2015
Included observations: 13

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.818	0.818	10.877	0.001
		2	0.553	-0.352	16.300	0.000
		3	0.295	-0.098	18.002	0.000
		4	0.070	-0.112	18.108	0.001
		5	-0.115	-0.106	18.432	0.002
		6	-0.246	-0.065	20.125	0.003
		7	-0.339	-0.121	23.868	0.001
		8	-0.397	-0.098	30.008	0.000
		9	-0.401	-0.023	37.845	0.000
		10	-0.354	-0.018	46.011	0.000
		11	-0.258	0.040	52.507	0.000
		12	-0.125	0.055	55.549	0.000

Figure 3. Autocorrelation and partial autocorrelation of road traffic accident death roll from 2003 to 2015

Augmented Dickey-Fuller Unit Root Test on P

Null Hypothesis: P has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic - based on SIC, maxlag=2)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-11.17346	0.0000
Test critical values:	1% level	-4.200056	
	5% level	-3.175352	
	10% level	-2.728985	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 11

Figure 4. ADF test results

According to the ADF test results, the t value of p is -11.17346, prob.=0.0000. The original sequence is stable. Both autocorrelation and partial correlation graphs are trailing, applicable to ARMA model.

4.3 Model identification

ARMA (1, 1) was preliminarily established for model estimation, and the results were shown in figure 5.

Dependent Variable: P
Method: Least Squares
Date: 06/24/18 Time: 11:33
Sample (adjusted): 2004 2015
Included observations: 12 after adjustments
Convergence achieved after 17 iterations
MA Backcast: 2003

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	48350.40	7054.539	6.853800	0.0001
AR(1)	0.835493	0.035845	23.30860	0.0000
MA(1)	0.999640	0.024681	40.50297	0.0000

R-squared	0.996616	Mean dependent var	73404.58
Adjusted R-squared	0.995864	S.D. dependent var	17003.22
S.E. of regression	1093.570	Akaike info criterion	17.04460
Sum squared resid	10763052	Schwarz criterion	17.16583
Log likelihood	-99.26760	Hannan-Quinn criter.	16.99972
F-statistic	1325.135	Durbin-Watson stat	1.752345
Prob(F-statistic)	0.000000		

Figure 5. ARMA (1, 1) model

After many calculations, we determined that the aic value was the minimum when p=1 and q=1. According to information criteria, ARMA (1, 1) was finally established.

4.4 Model test

The residual sequence was tested,Q=7.9950,prob=0.535. It is considered that the residual differential order is classified as white noise in extremely significant cases.

4.5 Model prediction

The ARMA (1,1) model was used to predict the road traffic accident death roll from 2011 to 2015. The predicted results are shown in table 1.

TABLE 1. ARMA (1, 1) model predictive value

Year	**Actual value**	**Predictive value**	**Absolute error**	**Error rate**
2011	62387	62446.04254	59.04254066	0.09%
2012	59997	60018.86578	21.8657832	0.04%
2013	58539	58059.19973	-479.8002731	0.82%
2014	58523	57342.53567	-1180.464334	2.02%
2015	58022	58029.57957	7.579569599	0.01%

5. Results and discussion

This study used ARMA (1,1) model to predict the road traffic accident death roll in China from 2011 to 2015. By comparing with the actual value, it is found that the predicted trend is basically consistent with the actual trend. The maximum error rate is 2.02% and the comprehensive error rate is only 0.60%. The prediction accuracy is high, the prediction effect is ideal and the reliability is high.

However, it should also be noted that the data used in this analysis is the reported value over the years, which is quite different from the real occurrence value. Moreover, the prediction value has certain limitations because it does not take into account the relevant factors influencing the prediction variables, such as road factors and management factors.

The time series analysis and prediction method does not consider the influence of other factors. Only the time factor is emphasized in the prediction. When other factors change greatly, the deviation will be obvious, so it is more suitable for short - and medium-term prediction.

References

[1] *Xie H W. A Study on Prediction of Road Trafic Accident Based on Gray System Theory [J]. Traffic & Transportation, 2015(12): 57-60.*

[2] *Guo W, Zi W, Tai W, et al. Application of Matlab in time series analysis[J]. Applied Science & Technology, 2003.*

[3] *Lütkepohl H. New Introduction to Multiple Time Series Analysis [M]// New introduction to multiple time series analysis. Springer, 2005:88-89.*

[4] *Durbin J, Koopman S J. Time Series Analysis by State Space Methods: Second Edition [J]. Technometrics, 2012, 39(461):255-256.*