International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-7 Issue-3, July 2017

Forecast Car Accident in Saudi Arabia with ARIMA Models

Mahmoud Al-Zyood

Abstract: Traffic accidents are the main cause of deaths and injury in Saudi Arabia, this work is a challenge to examine the best ARIMA model for forecast a car accident. Results show that an appropriate model is simply an ARIMA (1, 0, 0, 0) due to the fact that, the ACF has an exponential decay and the PACF has a spike at lag2 which is an indication of the said model. The forecasted car accident cases from 1998 to 2016. The selected model with least AIC value will be selected. We entertained nine tentative ARMA models and Chose that model which has minimum AIC (Akaike Information Criterion).The chosen model is the first one AIC (-0.274306) The selected ARIMA (1, 0) (0, 0), model to forecast for the future values of our time series (car accident). Forecasted for the next 7 years with (95%) prediction intervals The prediction values of traffic accidents show that there will be increasing in deaths and injury coming years

Keywords: Forecasting, ARIMA models, car accident, Akaike Information Criterion (AIC), Bayessian Information Criterion (BIC).

I. INTRODUCTION

Saudi Arabia has among the world's most dangerous roads. The highest rate of road accidents and death toll in the area. Statistics from the General Directorate of Traffic show that the Kingdom has 23 deaths per 100,000 people, with on average 19.1 road fatalities occurring daily. According to a study conducted by Hany Hassan, assistant professor of transportation engineering at King Saud University, there were 600,000 crashes recorded in the Kingdom in 2012, resulting in the death of around 7,638 people.

Having a system by which to predict traffic accidents would have a major impact in the development of appropriate solutions to combat this phenomenon, and can provide important information in accident trends. Based on the Traffic Department statistics Saudi Arabia records 526,000 accidents annually with up to 17 deaths daily. "A total of SR 21 billion is spent annually on road accidents. Saudi Arabia is ranked 23rd on the list of countries witnessing the highest death rates in road accidents in the world. It is second among Arab countries in terms of road deaths,"

There are many different statistical methods to forecast upcoming situations.

This study will use time series analysis; to modeling and forecasting track accident and provide awareness for decision makers to help them to adjust their plans and implement elective appropriate actions plan.

Revised Version Manuscript Received on June 20, 2017.

Mahmoud Al-Zyood, King Abdulaziz University, Finance, College of Business, P.O. Box 344, Rabigh, Jeddah 21911, Saudi Araba, E-mail: amman2121@yahoo.com

II. LITERATURE REVIEW

The Box-Jenkins approach to forecasting was first described by statisticians George Box and Gwilym Jenkins and was developed as a direct result of their experience with forecast problems in the business, economic, and control engineering applications (Box & Jenkins, 1994).

ARIMA processes are a class of stochastic processes used to analyze time series. The application of the ARIMA methodology for the study of time series analysis is due to Box and Jenkins [11]. Meyler et al (1998) drew a framework for ARIMA time series models for forecasting Irish inflation. In their research, they emphasized heavily on optimizing forecast performance while focusing more on minimizing out-of-sample forecast errors rather than maximizing in-sample 'goodness of fit'. Contreras et al (2003) in their study, using ARIMA methodology, provided a method to predict next-day electricity prices both for spot markets and long-term contracts for mainland Spain and Californian markets.

Contreras et al. (2003) used ARIMA models to predict next day electricity prices; they have found two ARIMA models to predict hourly prices in the electricity markets of Spain & California. The Spanish model needs 5 hours to predict future prices as opposed to the 2 hours needed by the Californian model. Datta (2011) used ARIMA model in forecasting inflation in the Bangladesh Economy. He showed that ARIMA (1, 0, 1) model fits the inflation data of Bangladesh satisfactorily.

Al-Zeaud (2011) used ARIMA model in modeling &forecasting volatility. The result shows that best ARIMA models at 95% confidence interval for banks sector is ARIMA (2, 0, and 2) model.

Uko et al. (2012) examined the relative predictive power of ARIMA, VAR & ECM models in forecasting inflation in Nigeria. The result shows that ARIMA is a good predictor of inflation in Nigeria & serves as a benchmark model in inflation forecasting.

III. MATERIALS AND METHODOLOGY

In this research, the main objective was to find the best model to efficiently forecast the car accident in Saudi Arabia applying Box and Jenkins method. The selection strategy for such models was developed and selected by the Box and Jenkins method (Box and Jenkins 1976).

This is a forecasting technique use a historical data and molders it into an Autoregressive technique. Looking at Table in the Appendix, it shows the data of Car accident from 1998 to December 2015,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Totaling 7006698 Car accident, 602623 Injured and 102713 dead. The data were obtained According to traffic Department statistics.

The methodology which used in this paper is Box and Jenkins called the Autoregressive Moving Average (ARIMA).

The ARIMA model is a useful statistical method for analyzing longitudinal data with a correlation among neighboring observations. This method has proven to be very useful in the analysis of multivariate time series. (Sales etal, 1980)

"Experience with real-world data, however, soon convinces one that both stationarity and Gaussianity are fairy tales invented for the amusement of under graduates."(Thomson 1994)

According to Heizer and Render (2009) ARIMA (Autoregressive Integrated Moving Average), is basically using the time series function, which requires a model approach to early identification and assessment of its parameters.

In ARIMA analysis, there are two simple components for representing the behavior of observed time series processes, namely the autoregressive (AR) and moving average (MA) models (Pankratz, A.1988).

ARIMA models are the most general class of models that seek to explain the autocorrelation frequently found in time series data (Hyndman and Athanasopoulos, 2014).

The Suitable forecasting methods were chosen for finding the method that was suitable for more than one year. The proposed forecasting time series process and the steps are shown in Fig.1.

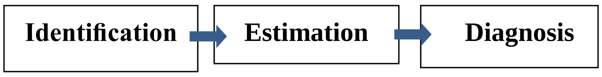


Fig.1. Time series modeling is accomplished through three steps:

These three stages may be repeated to realize the Best model for forecasting.

A. Identification

In the identification stage of model building, we determine the possible models based on the data pattern, But before we can begin to search for the best model for the data, first condition is to check whether the series is stationary or not. (See T. Of ori, 2012). Also if a data series is stationary then the variance of any major subset of the series will be different from the variance of any other major subset only by chance (see Pankratz, 1983).

The stationarity condition ensures that the autoregressive parameters invertible. If this condition is assured then, the estimated model can be forecasted (see Hamilton, 1994).

A time series is said to be stationary when these properties remain constant (Cryer and Chan, 2008, p. 16). To determine the stationarity of the data we used the autocorrelation function (ACF) and partial autocorrelation (PACF).

The final model can be selected using a penalty function statistics such as the Akaike Information Criterion (AIC) or Bayessian Information Criterion (BIC). See Sakamoto et al (1986), Akaike (1974).

B. Estimation

In this stage we find the values of the model factors which provide the fit model to the data.

C. Diagnosis

In this stage we testing the potentials of the model to identify any areas where the model is insufficient. If the model is found to be insufficient, it is essential to repeat Estimation then identify the best model.

IV. RESULTS AND DISCUSSION

A. Model Selection

The methodology of Box-Jenkins's for forecasting needs the series to be stationary.

The stationary condition ensures that the autoregressive parameters invertible. If this condition is assured then, the estimated model can be forecasted (see Hamilton, 1994).

As we see in Fig 2 and Fig 3 and Fig 4 the plot of time series and ACF and PACF of data provide a good indication of a non-stationary series.

Also if a data series is stationary then the variance of any major subset of the series will differ from the variance of any other major subset only by chance (see Pankratz, 1983).

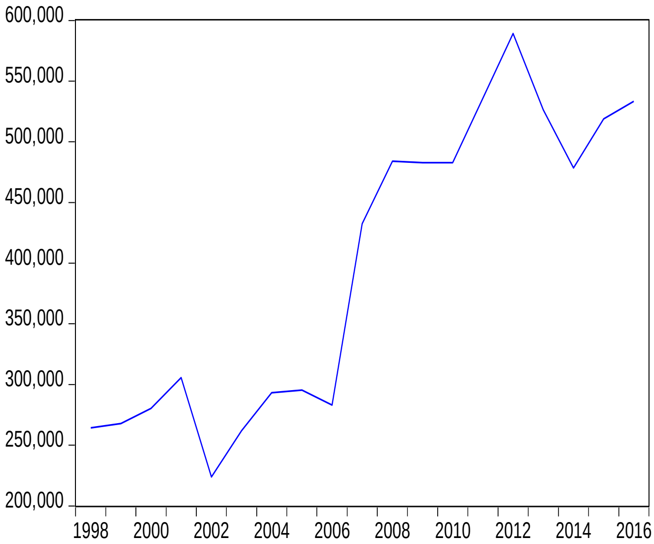


Fig.2: Graph of Number of accidents plot of original series)

The graphs of the sample ACF and PACF were plotted (Fig. 3 and 4).

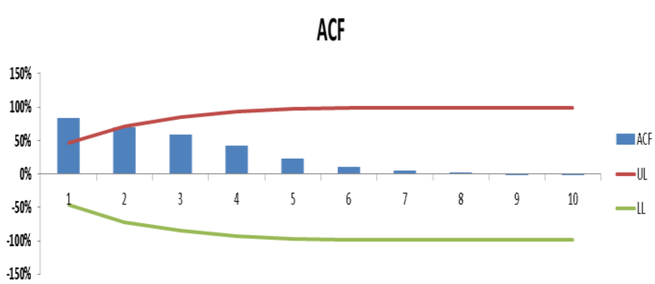


Fig 3: Autocorrelations (ACF) of first differenced series by lag

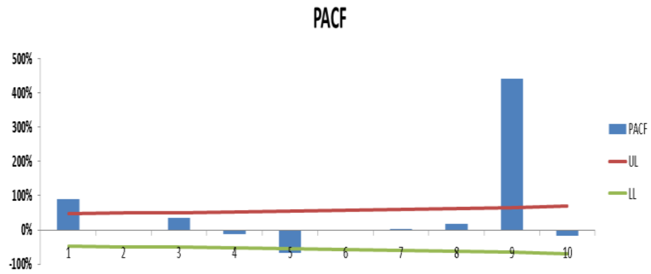


Figure 4: Partial Autocorrelations (PACF) of first differenced series by lag

B. Model Estimation

We will used Akaike information criterion (AIC) [2].According to this the model with least AIC value will be selected. We entertained nine tentative ARMA models and Chose that model which has minimum AIC (Akaike Information Criterion).The chosen model is the first one AIC (-0.274306) see table 1 and Fig 4 and Fig 5

Table 1: AIC and BIC values of fitted ARIMA models

Model	LogL	AIC*	BIC	HQ
(1,0)(0,0)	**4.782988**	**-0.274306**	**-0.143933**	**-0.301103**
(1,1)(0,0)	4.796743	-0.122576	0.051255	-0.158306
(2,0)(0,0)	4.795937	-0.122452	0.051379	-0.158182
(0,2)(0,0)	4.312220	-0.048034	0.125797	-0.083764
(2,1)(0,0)	4.801656	0.030514	0.247803	-0.014148
(1,2)(0,0)	4.799571	0.030835	0.248123	-0.013827
(2,2)(0,0)	5.570777	0.066034	0.326780	0.012439
(0,1)(0,0)	2.467491	0.081924	0.212297	0.055127
(0,0)(0,0)	-0.970710	0.457032	0.543948	0.439167

From table 1 we can clearly observe that the lowest AIC and BIC values are for the ARIMA (1, 0)(0,0), model with (p=1, d=0 and q=0) and henceforth this model can be the best model for making forecasts for future values of our time series data.

Fig 4 Forecast Comparison Graph

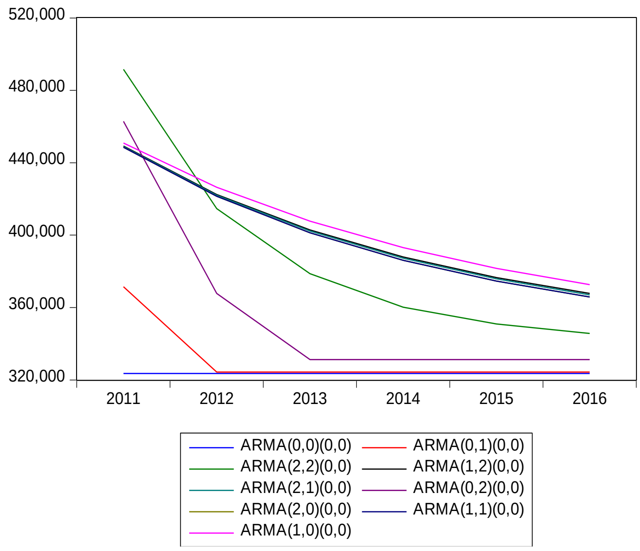
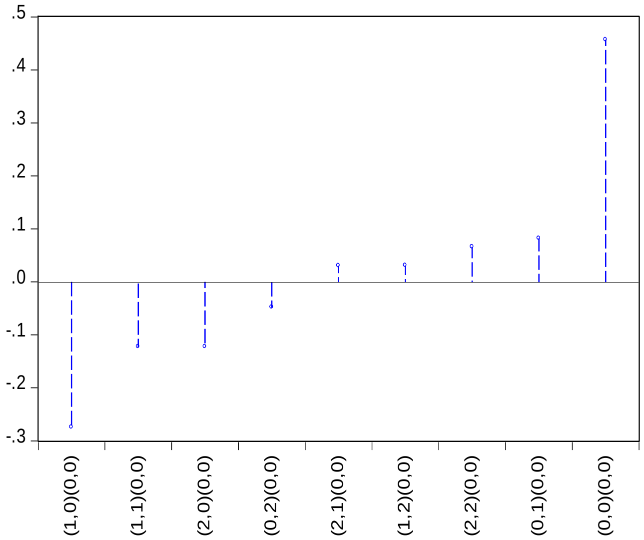


Fig 5 Akaike Information Criteria



C. Forecasting using selected ARIMA model

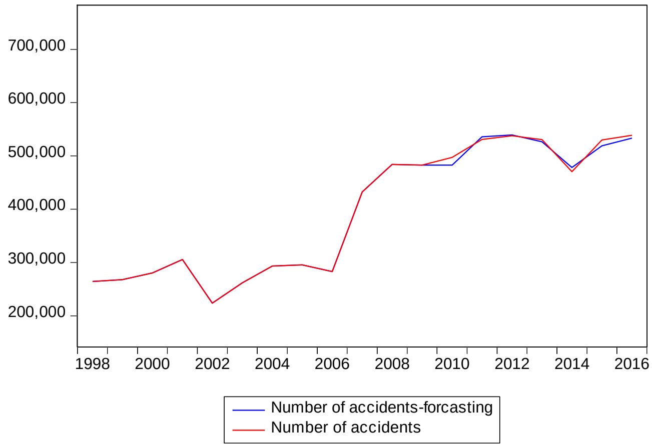
The above selected model ARIMA ((1, 0) (0, 0), which we are the best model to our time Series data.

We now will fit the best ARIMA ((1, 0) (0, 0), model to forecast for the future values of our time series (car accident). See Table 3 shows the forecast for the next 7 years with (95%) prediction intervals:

Table 3: Car accident Forecast data

For 95% confidence intervals		
Year	**Car accident**	**Forecasting Car accident**
1998	264326	NA
1999	267772	NA
2000	280401	NA
2001	305649	NA
2002	223816	NA
2003	261872	NA
2004	293281	NA
2005	295405	NA
2006	283024	NA
2007	432416	NA
2008	484045	NA
2009	482852	NA
2010	482852	497229.1346741808
2011	536055	530894.8687150868
2012	539258	537665.2575975068
2013	526429	530440.7439012367
2014	478450	470442.7310111016
2015	518795	529947.7377273633
2016	533400	538618.5263735236

Fig 4 below show the plot for 7 years' forecast of the Car accident by fitting ARIMA (1, 0, 0, 0) model to our time series data:



(Fig.4: Plot of actual and forecast data)

V. CONCLUSION

In this study, the ARIMA (1, 0, 0, 0) was the best applicant model selected for making forecasts for up to 7 years for the car accident. ARIMA was used for the reasons of its abilities to make Forecasts using a time series data. The prediction values of traffic accidents show that there will be increasing in deaths and injury coming years.

REFERENCES

1. Armstrong, J. S. (2001). Principles of forecasting: A handbook for researchers and practitioners. Boston: Kluwer Academic.
2. Box, G.E. and Jenkins, G.M. (1994) Time Series Analysis: Forecasting and Control. Prentice Hall, Englewood Cliffs.
3. Berube, M. S. (Ed.). (1985). American heritage dictionary (2nd ed.). Boston, MA: Houghton Mifflin.
4. Box, G. E., & Jenkins, G. M. (1994). Time series analysis: Forecasting and control (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
5. Box, G. E., Jenkins, G. M., & Bacon, D. W. (1967). Models for forecasting seasonal and nonseasonal time series. In B. Harris (Ed.), Spectral analysis of time series. New York, NY: John Wiley & Sons.
6. Boylan, J. (2005). Intermittent and lumpy demand: A forecasting challenge. The International Journal of Applied Forecasting, 1, 36-42.
7. Caldwell, J. G. (n.d.) The Box-Jenkins forecasting technique. Retrieved March 3,2012, from http://www.foundationwebsite.org/BoxJenkins.htm
8. Cryer, J.D. and Chan, K.S. (2008) Time Series Analysis with Application in R. Springer, New York. http://dx.doi.org/10.1007/978-0-387-75959-3
9. Hamilton, J. D. (1994). Time series analysis (Vol. 2). Princeton: Princeton university press. commandeur JJ, Bijleveld FD, Bergel-Hayat R, Antoniou C, Yannis G,
10. Papadimitriou E. On statistical inference in time series analysis of the evolution of road safety. Accid Anal Prev. 2013; 60:424–3doi:10.1016/j.aap.2012.11.006. [PubMed: 23260716].
11. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Time Series Analysis Forecasting and Control, Third ed. Englewood Cliffs, NJ: PrenticeHall, 1994.
12. Hannan, E., (1980), The Estimation of the Order of ARMA Process, Annals of Statistics, Vol. 8,pp. 1071-1081.