Malaysian road fatalities prediction for Year 2020

By Rohayu Sarani[1], Sharifah Allyana S.M.R[1] and Wong Shaw Voon[2]

[1] *Researcher, Road Safety Engineering and Environment Research Centre, MIROS,* rohayu@miros.gov.my, allyana@miros.gov.my

[2] *Director General, Malaysian Institute of Road Safety Research (MIROS)MALAYSIA,* wongsv@miros.gov.my

Abstract

Fatality reduction is crucial in road safety management. This paper illustrates fatality models developed in Malaysia in predicting the number of fatalities since the early 1990's to the current figures. All models were based on statistical models, utilising linear models, log linear and time series. The objective of this paper is to highlight the recent road fatalities model, using ARIMA models that have been developed and used by government agencies in planning for road fatality reduction.

Keywords

Road fatalities, Time series, ARIMA

Introduction

The United Nations in the 2030 Agenda for Sustainable Development had highlighted one of the goals which relate to road safety; to halve the number of global deaths and injuries from road traffic accidents by the year 2020. In the year 2013, 1.25 million deaths were recorded globally, and most of these deaths occurred in the low- and middle-income countries (World Health Organization, 2015). Malaysia, as one of the developing countries, has acknowledged road safety as a critical problem that should be addressed. In the 80's, Malaysia recorded an average of 4% annual increment of road traffic fatalities, followed by 5% in the 90's. However, the figure increased at a slower rate of 2% in more recent years (2000–2009) (Rohayu, Allyana , & Wong, 2012). Even though the trend recorded smaller increments over the last nine-year period, the real number of fatalities is increasing. In 2010 alone, 6,872 fatalities were recorded (PDRM, 2011) a 1.8% increase from the previous year (2009: 6,745).

It is often said that if we fail to plan, we are planning to fail. In recent years, the Malaysian government has been actively taking serious steps in reducing road fatalities. Back then in the year 1990, after the Karak crash, which claimed 15 deaths including army personnel, the Cabinet Committee on Road Safety was formed. The Committee set a target of 30% reduction in road deaths by the year 2000. It was based on predicted deaths made by Radin's model (Sohadi, 1998). Through the Ministry of Transport, the Road Safety Department introduced the first Road Safety Plan of Malaysia 2006-2010 (Road Safety Department, 2006). This plan outlined nine strategies to reduce road fatalities through the five-year period. Another target was set up to be achieved by the year 2010. The target was based on a fatality index, similar to indices used in developed countries. It was outlined that by the year 2010, deaths per 10,000 vehicles should reduce to 2.0, and deaths per 100,000 population should reduce to 10. The basis of the target was explained in detail by Law et. al (Law, Radin Umar, & Wong, 2005). Since the last model developed in the year 2005, Malaysia has been focusing on the implementation of road safety initiatives. The establishment of the Malaysian Institute of Road Safety Research (MIROS) in the year 2007 showed government commitment in combating road deaths. As a research institute dedicated for road safety, MIROS established a prediction model to predict road fatalities.

The objective of this paper is to present the current prediction model. As the model developed by Law et al was in the year 2010, there is a need to revise the current state of road safety in Malaysia and establish with the new prediction model. The second Malaysian Road Safety Plan 2014 – 2020 was produced in the year 2014 (Road Safety Department, 2014). The target for year 2020 was based on the predicted number of deaths in the year 2020 (Rohayu, Allyana , & Wong, 2012), which was set to reduce the number of deaths to 5,358. Therefore, it is important to review the model after the year 2015 and develop a new forecast to the year 2020.

Literature Review

Various fatality models have been developed for Malaysian road accident deaths, and used in the national road safety plan. Early works of modelling road deaths in Malaysia started with the development of a simple linear model (Aminuddin, 1990). In the mid-nineties, Karim (Karim, 1995) proposed an improved model, with more predictors. From the model, Karim projected 5,067 deaths in the year 2000, with estimated exposures of a population of 23 million and 10 million vehicles by the year 2000. Radin and Hamid (Radin Umar & Hamid, 1998) proposed that the rate of infrastructure growth in both roads and highways are highly correlated (r= 0.95). In light of improving the model, Radin (Radin Umar, 1998) added two explanatory variables: road length and the effect of standardised accident data.

Radin's model developed in the year 1998 has become part of input in preparing Malaysia's road safety target for year 2000. Radin's model predicted that the number of fatalities for the year 2000 was 9,127. In the year 1998, the number of deaths recorded was 5,740. In the year 1996, Malaysian government established a five-year national road safety target of 30% accident reduction by year 2000. Hence, using Radin's model which predicted 9,127 deaths in the year 2000, Malaysia planned for a 30% reduction target,

Table 1: Fatality models in Malaysia

Author / Year	Predictor Variables	Prediction Model	Predicted fatalities (Year)	Target deaths as set by Road Safety Plan of Malaysia	Actual deaths (Year)
Radin /1998	Population, Number of vehicles, Road length, Effect of standardized accident data	Death = 2289 (exp $^{0.00007veh.pop.road}$) (exp $^{0.2073\ data\ system}$)	9,127 (2000)	6,389	6,035 (2000)
Law et.al / 2005	Vehicle ownership rate	Gompertz Growth ARIMA model	4 deaths per 10,000 vehicles (2010)	4 deaths per 10,000 vehicles	3.4 deaths per 10,000 vehicles (2010)
Rohayu et.al/2012	Annual death series	ARIMA model	10,716 (2020)	5,358	6,674 (2014)

equivalent to 6,389 deaths. In the year 2000, the actual number of deaths was 6,035.

By comparing the actual figures in 1998, and the predicted deaths in the year 2000, the difference is quite significant (difference by 3,387 or 60% increment from the 1998 figure) despite the short time frame. The prediction was based on the growth of population, vehicles, road length and inclusion of accident data from Sabah and Sarawak.

Law et al. (Law, Radin Umar, & Wong, 2005) predicted road accident deaths in the year 2010 by making the projection for the vehicle ownership rate in the year 2010. The study used the Gompertz growth model, proposed by Dargay and Gately (Dargay & Gately, 1997) to project vehicle ownership as a function of per capita Gross Domestic Product (GDP), and Autoregressive Integrated Moving Average (ARIMA) to predict the road accident death rate.

Table 1 summarises fatality models developed in Malaysia, including the latest model.

The time series approach has been used in road safety fields, but mostly to evaluate road safety interventions. For example, Forester, McNown and Singell (Forester, McNown, & Singell, 1984) evaluated speed limit changes, and Garbacz and Kelly (Garbacz & Kelly, 1987) used log linear time series model to evaluate the safety impact of vehicle inspection. In forecasting, Raeside and White (Raeside & White, 2004) used monthly data on fatal and seriously injured victims in traffic crashes in Great Britain. However, the limitation of the data makes their forecasts less reliable as they used eight years of data to forecast for another 10 years ahead. Raeside (Raeside, 2004) then used annual data from 1970 to 2002, employing an autoregressive error model with lagged dependent variables, and forecast fatalities in 2010.

Methodology

The prediction of Malaysian road deaths for the year 2020 was based on a time series model; namely the ARIMA model. The figure for road traffic fatality is based on the official road death fatalities figure reported by the Royal Malaysian Police (PDRM). This annual figure is obtained from the Annual Road Traffic Report. In Malaysia, fatalities are defined as any person who died within 30 days as a result of accident (Royal Malaysian Police 2009). Data is available from 1972 to 2010 (39 annual number of deaths) and contains all road traffic fatalities from all 14 states in Malaysia. In developing the ARIMA model, data from 1972 to 2006 (36 observations) was used. The model developed was used to forecast for year 2007 to 2020 (14 observations).

There are many time series models that cater for different data type and structure, depending on the nature of data. Among the popular time series models is the Box-Jenkins approach, which is synonymous with ARIMA modelling. Auto-Regressive Integrated Moving Average (ARIMA) was developed by Box and Jenkins (Box, Jenkins, & Reisel, 1994), and has been widely used in road safety research. ARIMA modelling is usually applied to time series analysis, forecasting and control. The term ARIMA is a combination of Auto Regressive (AR) Integrated (I) Moving Average (MA) models. There are three stages in developing ARIMA models: model identification, model estimation and model validation. The process of model estimation and validation is iterative, until the best model is found.

The Box-Jenkins model used in the first phase consisted of the autoregressive model (AR), moving average (MA) model and Autoregressive-moving Average (ARMA) models. In the Autoregressive model, the dependent variable, Y_t is defined as a function of its previous value or historical value plus an error term. The Moving Average (MA) model links the current value to random errors that have occurred in the previous periods rather than the actual series themselves (Mohd Alias, 2007). The acronym ARIMA stands for "Auto-Regressive Integrated Moving

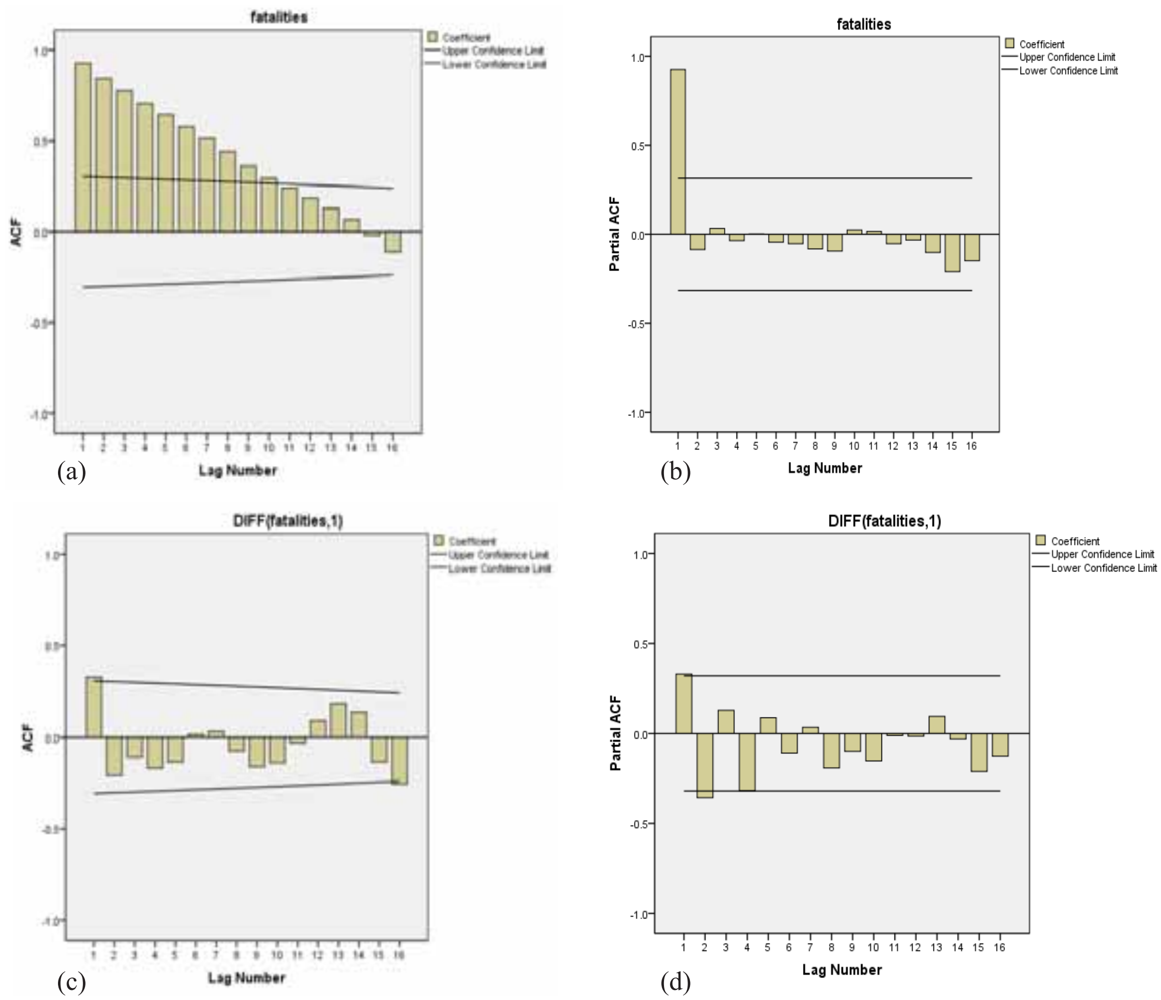


Figure 1: (a) and (b) are correlograms of ACF and PACF of original series, (c) and (d) are correlograms after the first differencing

Average." Lags of the differenced series appearing in the forecasting equation are called "auto-regressive" terms, lags of the forecast errors are called "moving average" terms, and a time series which needs to be differenced to be made stationary is said to be an "integrated" version of a stationary series.

The hardest part in Box Jenkins methodology is to identify which model really describes the data. The identification of the correct model starts with inspecting the autocorrelation coefficients (ACs) and Partial autocorrelation coefficients (PACs). These two measure the degree of interdependence among the observations in the series. In other words, the ACs and PACs measure the degree of interdependence between current and past fatalities figures.

Based on correlograms of ACF and PACF (Figure 1(a) and 2(b)), the most suitable model is identified and further analysis on model fit is conducted. From the correlogram of ACF, the sample ACF values are large and decline rather slowly to zero, therefore it can be concluded that the original series of road fatalities is not stationary. The PACF correlogram, shows that there is a significant large spike at Lag 1, followed by smaller other spikes. These suggest that the original fatalities series can be made stationary after performing the first difference. After inspecting ACF and PACF of original series, it is concluded that no seasonal variation exists in the series. Therefore, a non-seasonal ARIMA should be considered. The ACF and PACF suggest that ARMA (p,d,q) model should be used. The fatalities series is now stationary after the first difference (Figure 1(c) and 2(d)).

ARIMA uses past values and past errors to detect patterns and predict future values. A simple ARIMA (0,1,1) with constant is expressed as below:

$$\widehat{Y}_t = \mu + Y_{(t-1)} - \theta\varepsilon_{(t-1)} \ldots\ldots\ldots\ldots\ldots\ldots \text{ Equation (1)}$$

A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, where:

- p is the number of autoregressive terms,
- d is the number of non-seasonal differences, and
- q is the number of lagged forecast errors in the prediction equation.

The next step is to identify which ARIMA (p,d,q) suits the data best. A list of models to be considered were tested for model fit and error rate employing Expert Modeller function in SPSS version 20.0.

Table 2: Parameter Estimates for ARIMA (0,1,1)

Variable	Parameter estimate	Standard error	t-statistics	Significance value
Constant	0.036	0.016	2.226	0.033
Fatalities (MA Lag 1)	-0.901	0.117	-7.712	0.000

Results

The study uses the ARIMA model in predicting Malaysian fatalities for year 2020, by using historical data from year 1972 until year 2010. Data from year 1972 to 2006 was used in model development and the rest were used as model validation. Results showed that R-squared value is 0.981, really close to 1, indicating that the model is able to explain 98% of variation in the data. Ljung-Box statistics also showed that the error of the model is not correlated. Fatalities for year 2015 and 2020 are 8,760 and 10,716 respectively, for business as usual (BAU) scenario.

The best ARIMA model produced from the data is ARIMA (0,1,1). Table 2 shows parameter estimate for ARIMA (0,1,1) based on Malaysia road deaths annual data. The ARIMA (0,1,1) model with constant has the prediction equation:

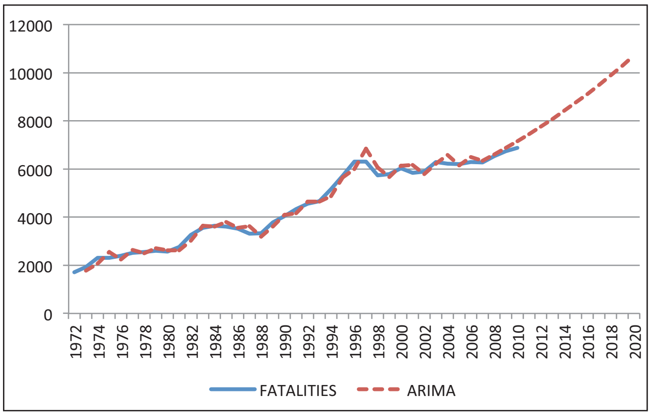
$$\widehat{Y}_t = \mu + Y_{(t-1)} - \theta\varepsilon_{(t-1)}$$

Where Y_t is fatalities at current time, t, $Y_{(t-1)}$, is fatalities a year before, and θ = -0.901

For model fit, R-squared value for ARIMA (0,1,1) is 0.981, in other words, the model suits the data. Ljung Box statistics, which provides an indication of whether the model is correctly specified, Hypothesis is accepted, as the Ljung Box statistics = 0.938 is more than alpha = 0.05. This indicates that residuals of ARIMA models have no auto correlation.

Figure 2 illustrates the original fatalities series as compared to ARIMA model. The solid line indicates the annual number of deaths recorded from the year 1972 to year 2010. On the other hand, the dotted line shows forecasted number of deaths by using ARIMA model. It shows that ARIMA follow closely to the series at the beginning of the forecast. The forecasted figure did not differ much from the original series, up to the year 2010. For each additional year, the predicted figure deviates further, as it forecasted based on historical data.

Figure 2: Fatality forecast for Malaysia



Discussion

Road fatalities prediction is important in this target setting exercise. Utilising a sound predicted figure allows government to intervene by implementing various initiatives to deliver target improvements. In Malaysia alone, there are several models proposed and applied to government since the year 2000. Expected numbers of deaths can be used to establish a national target to offset the rising number of road deaths.

In comparison with other models developed for Malaysia, this univariate ARIMA model, used historical data to predict future road deaths. Adding more variables to the model would allow for sensitivity analysis, but it will require more estimation for each variable, adding risk to accuracy of estimates. For example, if population, road length and GDP were to be added to the equation to predict the number of fatalities, an estimated figure of population, road length and GDP for the year 2020 is needed as an input to the equation, before the number of fatalities could be obtained. Hence, the process would have introduced more errors due to multiple estimations of the independent variables.

One of the important issues in forecasting is to understand the forecast horizon. Forecast horizon means the time frame in which the forecast is valid. In this case, the forecast horizon refers to the year 2011 until 2020. ARIMA model uses historical values. The accuracy of predicted values may be compromised as the forecast horizon becomes longer. This is one of the disadvantages of the model. To overcome this, as new data is available, a new ARIMA model should be developed to generate new forecast figures. Hence, it is important to review the model after the year 2015 and develop a new forecast to the year 2020.

Another important issue is that ARIMA model is sensitive to any structural change. Structural change in road safety may include but is not limited to, the introduction of new effective laws and implementation of new interventions such as an Automated Enforcement System. Such structural changes can have a lagged effect on the series. For instance, if an intervention is introduced in the year 2013, the effect may be seen two years later, which may require the model to be updated depending on the level of impact of the new intervention. In developing the ARIMA model in the study, the author assumes that Business As Usual, which means that if there is no new intervention, the number of deaths will remain at 8,760 for the year 2015 and 10,716 death in the year 2020. Of course, the Malaysian government will push for more interventions to offset the number of deaths, and the model provides a baseline for assessment of disruptive change.

References

Aminuddin, A. (1990). *Road Safety in Malaysia. Proceedings of Road Safety Seminar* (pp. 144-157). Kuala Lumpur: National Road Safety Council.

Box, G., Jenkins, G., & Reisel, G. (1994). *Time series analysis: Forecasting and Control*. Prentice Hall.

Dargay, J., & Gately, D. (1997). Vehicle ownership to 2015: Implications for energy use and emmission. *Energy Policy* (25), 1121-1127.

Forester, T., McNown, R., & Singell, L. (1984). A cost-benefit analysis of the 55 mph speed limit. *Southern Economic Journal*, 50(3), 631-641.

Garbacz, C., & Kelly, J. (1987). Automobile safety inspection: new econometric and benefit/cost estimates. *Applied Economics*, 19, 763-771.

Karim, M. (1995). A macro analysis of road accident trends in Malaysia. *Journal of Eastern Asia Society for Transportation Studies*, 1(3), 941-950.

Law, T., Radin Umar, R., & Wong, S. (2005). The Malaysian Government's road accident death reduction target for year 2010. *IATSS Research* 29(1), 42-49.

Mohd Alias, L. (2007). I*ntroductory Business Forecasting, a practical approach*. Kuala Lumpur: Unversity Publication Centre (UPENA) UiTM.

PDRM. (2011). *Laporan Perangkaan Kemalangan Jalan Raya Malaysia*. Kuala Lumpur: Bukit Aman.

Radin Umar, R. (1998). Model kematian jalan raya di Malaysia: unjuran tahun 2000. *Pertanika Journal of Science & Technology* 6(2), 107-119.

Radin Umar, R., & Hamid, H. (1998). Time Series multivariate traffic accidents and fatality models in Malaysia. *REAAA Journal*, 15-20.

Raeside, R. (2004). Predicting and monitoring casualty numbers in Great Britain. *Journal of Transportation and Statistics*, 7(1), 61-68.

Raeside, R., & White, D. (2004). Predicting casualty numbers in Great Britain. *83rd Annual Meeting of the Transportation Research Board*.

Road Safety Department. (2006). *Pelan Kelesalamatan Jalan Raya Malaysia 2006-2010.* Kuala Lumpur.

Road Safety Department. (2014). *Pelan Keselamatan Jalan Raya Malaysia 2014-2020*. Putrajaya.

Rohayu, S., Allyana , S., & Wong, S. (2012). *Predicting Malaysian Road Fatalities for Year 2020*. Kuala Lumpur: Malaysian Institute of Road Safety Research.

Sohadi, R. U. (1998). Model Kematian Jalan Raya di Malaysia: Unjuran Tahun 2000. *Pertanika Journal of Science & Technology*, Vol 6 No. 2.

World Health Organization. (2015). *Global Status Report on road safety 2015*. Geneva: World Health Organization.