Forecasting Trend of Traffic Fatalities in the United Arab Emirates

Ibrahim, M. Abdalla Alfaki

Abstract—This paper utilizes time series methods to fit and forecast rates of traffic fatalities in the United Arab Emirates (UAE). Such data-driven techniques are increasingly used to assess transport safety and the performance of engineering and legislative intervention measures through the use of traffic safety targets. Despite the continuous rise of traffic accidents in the UAE, the country, however, hasn’t yet managed to create a mechanism to setting road safety targets in order to help monitoring the effect of potential safety countermeasures and possible reduction in traffic fatalities. Several competing models are fitted and compared in terms of goodness of fit and forecasting power. The main implications of results obtained from this work are intended to aid the UAE to explore different mechanisms to setting safety targets with the aim of reducing traffic fatality numbers and rates and the creation of sustainable prevention and safety management strategy.

Keywords—Traffic fatality, Safety target, Intervention measures, Fatality trend.

I. INTRODUCTION

ECONOMIC growth, population increase and demographic and cultural changes in the United Arab Emirates (UAE) have contributed in a dramatic increase in the number of vehicles in the last few years. The increase in the number of vehicles on the road, along with emergence of high risk road users have converged in not only a dramatic number of road accidents, but a dramatic number of fatal accidents and fatal injuries, Abdalla [1]-[2]. Compared to western countries and some other countries in the gulf region, the number of road accidents in the UAE is high, in particular, accidents that are caused by wrong driver’s behavior. According to the UAE National Bureau of Statistics report of 2008, 18 percent of the accidents were caused by careless driving, and 43 percent were due to other human mistakes.

Monitoring progress in transport and road safety is essential for effective use of resources and reduction of direct and indirect costs of traffic accidents. Among the many measures that are employed by policy makers and practitioners to measure safety of transport systems is the number of people killed in traffic accidents.

The continuous rise of traffic accidents and death toll in the United Arab Emirates constitute major concerns for public health and economic and social planning (Bener and Crundall [5]; El-Sadig et al. [6]). The country, however, hasn’t yet managed to create a mechanism to setting road safety targets in order to help monitoring the effect of potential safety countermeasures and possible reduction in traffic fatalities.

Three approaches to setting safety targets are discussed in the literature; namely, aspirational targets, model-based targets and extrapolation and evidence-led judgment, Marsden and Bonsall [11]. The aspiration target setting approach is based on setting an arbitrary number that represents the intended achievement in specified period of time, e.g. reaching zero fatalities in specified number of years starting from a specified base year, often without giving any justification on how the number is derived. The model-based approach, on the other hand, calls for setting reduction targets based on available data and a clear understanding of the relationship between traffic safety and other influencing factors. The success of the approach depends on the ability of the model to reflect reality, accuracy of assumptions and the extent of real implementation of assumed policy interventions. In reality it is difficult to model all influencing indicators with sufficient accuracy. Extrapolation and professional judgment can therefore be utilized to resolve this issue, using an evidence-led judgment approach. The success of the evidence-led judgment approach in setting the targets depends on the availability of sufficiently long time series data on accidents and traffic fatalities, the ability to identify the different components of the time series and on availability of substantial evidence on the impact of different policy intervention measures on accident and fatality reductions, Kweon [10].

The main objective of this study is to describe trend and forecast developments in the number and rates of traffic fatalities in the UAE. The main implications of results obtained from this work are intended to aid the UAE as a country to explore different mechanisms to setting safety targets aiming at the reduction of traffic fatality numbers and rates and the creation of sustainable safety management strategy.

II. AN OVERVIEW OF FATALITY PREDICTION MODELS

Statistical models using time series approaches to fitting and forecasting number and rates of traffic casualties are widely discussed in the literature at both aggregate and disaggregate prediction levels (see for example Broughton [16]; Yannis and Antoniou [15]). In several studies allowances
are made to cater for intervention variables and special events, however, the majority are univariate in nature, incorporating limited number of explanatory variables (Raeside and White [17]-[18]).

Earlier in 1949, R. J. Smeed [13] published his famous law, \( F = 0.0003 (V, P)^{1/2} \), for predicting traffic fatalities based on the number of registered vehicles, \( V \), and population size, \( P \). The formula postulates a decrease in traffic fatalities per vehicle as a result of the increase in vehicle ownership, \( F/V = 0.0003 (V/P)^{-2/3} \), it also, however, suggests an increase in the number of fatalities and fatalities per population due to the increase in vehicle ownership, \( F/P = 0.0003 (V/P)^{1/2} \). Several attempts are made to validate Smeed’s law, occasionally with different estimates of the parameters, using data coming from both developed and developing countries (see for example Gharaybeh [7]; Koren and Borsos [9]. The law successfully predicts the increase in traffic fatalities up to certain years, but then the validity of the formula is disputed in several occasions as it failed to match the noticeable decrease in road deaths in some of the developed countries, producing considerable deviations between the expected and the actual number of traffic fatalities (Adams [3]; Koren and Borsos [9]; Bener et al. [4]; Safe Speed [12]).

It is worthy to note that the original relationship used to derive Smeed’s law in the late 60’s (Smeed [14]) is formulated as

\[
F/V = \alpha (V/P)^\beta + \varepsilon, \tag{1}
\]

where \( \alpha \) and \( \beta \) are the model parameters and \( \varepsilon \) are the disturbances which are assumed to be independent and identically normally distributed. Given the nature of traffic fatality data as time series realizations, utilization of Smeed’s model (1) in its current form to fit such data might not account for possible serial correlation of the model’s disturbances, \( \varepsilon \). Attempting to improve model fit and to allow for possible dependence or autocorrelation of disturbances Yannis and Antoniou [15] utilized Smeed’s model (1) as a benchmark to further develop a log-transformed version of model (1) and two further autoregressive non-linear models, defined respectively in (2), (3) and (4) below, based on using vehicle ownership, \( V/P \), as a macroscopic predictor of traffic fatalities. Models (3) and (4) are the autoregressive versions of models (1) and (2) respectively.

\[
\log \left( \frac{F_t}{V_t} \right) = \alpha + \beta \log \left( \frac{V_t}{P_t} \right) + \varepsilon_t, \tag{2}
\]

\[
(F/V)_t = \varphi (F/V)_{t-1} + \alpha (V/P)_{t}^\beta - \varphi \alpha (V/P)_{t-1}^\beta + \varepsilon_t, \tag{3}
\]

\[
\log (F/V)_t = \varphi \log (F/V)_{t-1} + (1-\varphi) \alpha + \beta \log (V/P)_t - \varphi \beta \log (V/P)_{t-1} + \varepsilon_t. \tag{4}
\]

\( \varepsilon_t \) denotes the discrete time \( (t = 1, \ldots, T) \) at which the dependent and the predictor variables are measured. Yannis and Antoniou [15] fitted the four models (1 to 4) using European traffic data spanning the period from 1970 to 2002 using non-linear regression technique. They report an overall superior performance, in terms of goodness of fit and forecasting power, of the autoregressive models compared to the non-linear base model (1) and its similar log-transformed version, respectively. They further indicate that the log-transformed model gives better prediction compared to the non-linear base model (1), but suffered from correlated residuals.

Raeside and White [17]-[18] employed a variation of Broughton [16] model to fit a negative exponential model of the natural logarithm of the number of traffic fatalities, serious injuries and slight casualties in Great Britain covering the period from 1970 to 2010. Using fatal casualties, they present a time series model with an autoregressive and linear trend terms as depicted in equation (5). This model is a deviation from Broughton’s [16] model who used fatality rates (traffic fatalities per vehicle-kilometers) as a dependent variable and an intervention term, seat-belts use as a predictor.

\[
\log (F_t) = \alpha + \beta t + \varphi \log (F_{t-1}) \tag{5}
\]

Several authors adopted the state space methodology or structural time series methods to fit and forecast traffic fatalities. Commandeur and Koopman [19] present a broad overview of the application of this methodology on traffic fatality data. They discuss several scenarios where produced models can adequately describe the data to situations where model explanatory power is increased by incorporating explanatory or intervention terms. They used the technique to adequately fit the annual numbers of traffic fatalities in Finland, 1970 through 2003, using a deterministic level and a stochastic slope model, and to forecast fatalities for the years 2004-2008 with 90% confidence limits.

Hermans et. al. [8] use the unobserved components or structural time series methodology employing the state-space form to discover the long-term time series trend of killed and seriously injured traffic casualties in Belgium 1974 to 1999. The models are designed to allow for the impact of weather conditions, economic factors and several laws and intervention variables. Developed models are used to predict the dependent variable for 2000 and 2001 and to compare results with the actual observations. They later compared the performance of the state-space models with a regression model with ARMA errors, concluding a lot of similarity and the same direction and a comparable magnitude. However, their analysis did not include a measure of exposure to risk of accidents.

Bener et al. [4] employed regression techniques to estimate traffic fatalities using time series data on vehicles and populations of three Mediterranean countries, including the United Arab Emirates (UAE). They reported lower average absolute errors for estimation using regression analysis compared to estimation based on Smeed’s equation. However,
their analysis failed to investigate possible serial correlations in the data and to ascertain whether or not regression residuals are auto-correlated. Presence of positive or negative first order residual autocorrelation, for instance, is a violation of model assumptions and will lead, respectively, to serious under or over-estimation of error variance, resulting in optimistic or pessimistic conclusions and over or under-estimation of the significance of predictors employed in the model.

III. DEVELOPING TRAFFIC FATALITY MODEL USING UAE DATA

Annual traffic fatality data together with population size and number of registered vehicles estimates spanning the period from 1977 to 2008 are considered to describe trend and to predict future numbers and rates of traffic fatalities in the UAE. The data are obtained from official sources in the country; namely, the UAE Bureau of Statistics and the Ministry of Interior Annual Statistical Abstracts together with other relevant sources. Prediction models utilized in this development are macroscopic in nature, depending mainly on annually aggregated data. This kind of models are useful when the overriding objective is the provision of appropriate models that describe trend of traffic fatalities and forecast future values. Indeed, road safety trends and future forecasts can be investigated by studying the effects of various macroscopic variables including motorization level, road expenditure and intervention measures such as implementation of seat-belt use, on traffic fatalities. However, the use of several explanatory variables is usually constrained by the quality and availability of data, particularly in developing countries.

It is worthy to note that accident prediction models are usually constructed with an exposure variable that controls for the total road traffic movements within the road network. In many studies, vehicles-Kilometers travelled is used as an exposure variable. This is particularly true in studies covering mostly developed countries where good traffic and transportation reporting systems exist. In most developing countries the situation might be different and reliance is mainly on other proxy measures of exposure such as the number of registered vehicles and population size. The UAE is no exception, the country lacks a reliable reporting system that accounts for traffic movements in the road network. Therefore, the number of registered vehicles and population size are used to control for exposure in this study.

The time series data employed in this study are partitioned into two datasets. A training dataset, 1977 to 2003, which is used to describe trend and to develop traffic fatality prediction models, and a validation dataset, 2004 to 2008, which is used to validate the developed models. Goodness of fit of developed models and deviation of fitted values from actually observed values (for both within sample: 1977-2003 and out of sample 2004-2008) are measured by estimating the root mean square error (RMSE) as given in equation (6). The lower the value of the RMSE, the better the performance and forecast of the model compared to its counterparts.

\[
RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(F_t - \hat{F}_t)^2},
\]

where \(F_t\) is the observed traffic fatality at time \(t\), and \(\hat{F}_t\) is the fitted annual traffic fatalities obtained from the constructed model.

IV. EXPLORING THE UAE TRAFFIC FATALITY SITUATION

Fig. 1 presents annual UAE road traffic fatalities, 1977 to 2008. It is interesting to note the steady increase in traffic fatalities in the UAE since 1985 with seemingly linear trend. However, the overall pattern between 1977 and 2008 is far from linear. The pre-1985 period witnessed considerable fluctuations in the number of traffic fatalities. The drop in the number, starting 1984 up to 1987, is arguably linked with the collapse in oil prices and the tankers war in the Arabian Gulf erupted between Iran and Iraq, 1986-1987, which had an adverse effect on trade transportation in the whole gulf region.

Evidence shown in Fig. 2 which presents the sample Autocorrelation Function (ACF) of the annual traffic fatalities in the UAE, confirms the presence of serial correlation in the data. Significant autocorrelation coefficients (falling outside the 95% confidence limits) are reported at different lags, \(p-value = 0.0\), based on 16 lags Box-Ljung test with \(\chi^2 = 97.5\). As a result any modeling efforts that does not account for the dependence in the data might render final conclusions useless.
The actual variables that are used in developing the UAE traffic fatality prediction models in this study included fatalities/vehicle, $F/V$, as a dependent variable and car ownership or vehicles/population, $V/P$, as a predictor variable. While the number of traffic fatalities, $F$, in the UAE are showing an increasing trend as displayed in Figure 1, fatality rates per registered vehicles, $F/V$, have sharply decreased in the period from 1977 to 1980, Fig. 3, possibly as a result of several intervention and road safety measures including improvements in infrastructure, enforcement of speed limits and penalties and enactment of a secondary seat belt law. After 1980 a consistently slight decreasing trend is observed, Fig. 3. Vehicles ownership ($V/P$, the dotted line in Fig. 3), on the other hand, is depicting an opposite increasing trend.

The most important observation is that the non-linear base model and the log-transformed model did not pass the

\[ \log((F/V)_t) = \alpha + \beta \cdot t + \varphi \cdot \log((F/V)_{t-1}) \]  

(7)

The analysis in this section is based on Brought [16] model derived according to Raeside [18] as depicted in equation (5). This is an ARMA like model with an autoregressive component (AR) of order 1, and a trend term, $\tau$, as a predictor variable. The model introduced in this study deviates from Raeside’s model in (5) by imposing the number of registered vehicles, $V$, as an exposure variable. The dependent variable, $F/V$, was log-transformed to eliminate non-stationarity of the data. Therefore, the final model developed using the training UAE traffic fatality data, 1977-2003, is defined as

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficient</th>
<th>$\alpha$</th>
<th>$\varphi$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-linear (Smeed)</td>
<td>Estimate</td>
<td>5.8662</td>
<td>-2.2914</td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td>Standard Error</td>
<td>0.1968</td>
<td>0.0867</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-transformed (equation (2))</td>
<td>Estimate</td>
<td>1.6833</td>
<td>-2.1678</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.1437</td>
<td>0.2404</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-linear AR (equation (3))</td>
<td>Estimate</td>
<td>1.8853</td>
<td>0.3398</td>
<td>-0.4818</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.7486</td>
<td>0.0396</td>
<td>0.6366</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.0192</td>
<td>0.0000</td>
<td>0.4569</td>
<td></td>
</tr>
<tr>
<td>Log-transformed AR (equation (4))</td>
<td>Estimate</td>
<td>0.7425</td>
<td>0.6664</td>
<td>-0.7675</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.2706</td>
<td>0.0787</td>
<td>0.4147</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.0155</td>
<td>0.0000</td>
<td>0.0711</td>
<td></td>
</tr>
<tr>
<td>Negative Exponential (equation (7))</td>
<td>Estimate</td>
<td>1.6900</td>
<td>0.8920</td>
<td>-0.0660</td>
</tr>
<tr>
<td>Standard Error</td>
<td>0.5200</td>
<td>0.1030</td>
<td>0.0320</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>0.0030</td>
<td>0.0000</td>
<td>0.0460</td>
<td></td>
</tr>
</tbody>
</table>

The most important observation is that the non-linear base model and the log-transformed model did not pass the
residuals serial correlation test, $p$-value $< 0.01$, based on Box-Ljung criteria as displayed in Table 2, consequently violating the assumption of independent residuals. It remains to acknowledge that the log-transformed model is an improvement over the non-linear base model, providing lower within sample and out of sample RMSE, Table 2. Inclusion of an autoregressive component in both models resulted in eliminating serial correlation and in improving model prediction, Table 2. Results discussed here are similar to conclusions reported by Yannis and Antoniou [15] using European data.

Table 2: Test Model Residuals Autocorrelation (Box-Ljung test, 16 lags) and Model Goodness of fit and Accuracy (RMSE), UAE Fatality Data

<table>
<thead>
<tr>
<th>Model</th>
<th>Box-Ljung</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$X^2$</td>
<td>$p$-value</td>
</tr>
<tr>
<td>Non-linear (Smeed Model, equation (1))</td>
<td>43.4</td>
<td>0.0002</td>
</tr>
<tr>
<td>Log-transformed (equation (2))</td>
<td>42.7</td>
<td>0.0003</td>
</tr>
<tr>
<td>Non-linear AR (equation (3))</td>
<td>16.5</td>
<td>0.4159</td>
</tr>
<tr>
<td>Log-transformed AR (equation (4))</td>
<td>17.3</td>
<td>0.3652</td>
</tr>
<tr>
<td>Negative Exponential (equation (7))</td>
<td>4.7</td>
<td>0.9998</td>
</tr>
</tbody>
</table>

VI. FORECASTING RATES OF UAE TRAFFIC FATALITIES

It was evident from the analysis of the UAE fatality data that all proposed models in this study showed a downward trend in the UAE traffic fatality rates. Forecasts of future fatality rates, 5 years ahead, with lower and upper 95% confidence limits, are generated using the negative exponential model, model (7) above. Forecasts of traffic fatalities for 2013 is 0.4847 traffic fatalities per 1000 registered vehicles, ranging over a 95% prediction interval from 0.2043 to 0.9909 fatalities per 1000 registered vehicles. This is about 30% reduction of the 2008 rate. To convert forecasted fatality rates into fatality numbers, the total number of registered vehicles in that particular year is needed. Several scenarios to forecasting the number of registered vehicles are discussed including the development of a forecasting model for the number of vehicles or alternatively use official vehicles forecasts produced by official agents (Broughton [16]).

Fig. 4: Observed and Predicted Traffic Fatalities with 95% Lower and Upper Confidence Limits

VII. CONCLUDING REMARKS

Many countries have established safety plans and traffic safety targets. Targets provide a clear focus for the work of those involved in transport planning and traffic safety management. There is an overriding consensus that traffic safety targets should be specific, measurable, relevant and time-bound. The UAE is still lagging behind when it comes to the creation of safety plans based on appropriate targets for the reduction of traffic fatalities and injuries. The country is suffering from increased levels of traffic fatalities, however, accounting for the notable increase in the number of registered vehicles in the country together with the positive growth in population size, traffic fatality rates are showing a slightly declining trend. This decline might be attributable to changes in driving behaviors in addition to the introduction of several legislative and engineering intervention measures.

This paper intended to explore and describe the traffic fatality trend in the UAE, and further produce appropriate predictive models capable of forecasting future fatality trend. This is an important step needed for safety target setting. Results presented in the paper identified appropriate models that provided good fit and forecast of the UAE fatality data.

REFERENCES


