Peer-reviewed papers

Original Road Safety Research

Modelling New Zealand Road Deaths

Colin Morrison[1], Ernest Albuquerque[1]

[1]*Waka Kotahi NZ Transport Agency, Wellington, New Zealand*

Corresponding Author: Colin Morrison, Waka Kotahi NZ Transport Agency, Private Bag 6995, Wellington 6141, New Zealand, colin.morrison@nzta.govt.nz, +6448946681.

This peer-reviewed paper was first presented as an Extended Abstract and Oral Presentation at the 2019 Australasian Road Safety Conference (ARSC2019) held in Adelaide, Australia and first published in the ARSC2019 Proceedings in the form of an Extended Abstract. It was expanded into a 'Full Paper' and underwent further peer-review by three independent experts in the field. It is being reproduced here with the kind permission of the authors and is now only available in this edition of the JRS.

Key findings

- After testing, three models were developed: an ARIMA and two ARDL models
- The variables correlated well: road deaths, GDP, employed, petrol price and young population
- An ARDL model proved the most promising and was used in projecting to 2025
- The ARDL model forecast showed road deaths fall slightly overall from 2018 to 2025.

Abstract

New Zealand is developing an integrated road safety intervention logic model. This paper describes a core component of this wider strategic research carried out in 2018: a baseline model that extrapolates New Zealand road deaths to 2025.

The baseline will provide context to what Waka Kotahi NZ Transport Agency is trying to achieve. It offers a way of understanding what impact interventions have in acting with and against external influences affecting road deaths and serious trauma.

The baseline model considers autonomous change at a macro level given social and economic factors that influence road deaths. Identifying and testing relationships and modelling these explanatory variables clarifies the effect of interventions.

Time-series forecasting begins by carefully collecting and rigorously analysing sequences of discrete-time data, then developing an appropriate model to describe the inherent structure of the series. Successful time-series forecasting depends on fitting an appropriate model to the underlying time-series.

Several time-series models were investigated in understanding road deaths in the New Zealand context. In the final modelling an autoregressive integrated moving average (ARIMA) model and two differing autoregressive distributed lag (ARDL) models were developed. A preferred model was identified. This ARDL model was used to project road deaths to 2025.

Keywords

ARDL, Baseline, Model, Forecast, Road deaths, Time-series.

Glossary

Road deaths: are defined in New Zealand as deaths occurring as the result of injuries sustained in a road crash within 30 days of the crash.

Road crash: is defined as an event involving a vehicle that results in damage to persons or property.

Road: is defined as a place where the public has legal access with a motor vehicle.

Note: In New Zealand prior to 2015 only those road crashes involving a motor vehicle were recorded. Since 2015 cyclist only road crashes resulting in death have also been recorded as road deaths.

Introduction

Waka Kotahi NZ Transport Agency is leading the development of an integrated road safety intervention logic model (2017) with sector partners including NZ Police, Ministry of Transport (MoT) and the Accident Compensation Corporation (ACC). The Integrated Intervention Logic Model (IILM) is a tool to inform strategies aimed at improving safety across the network. New Zealand is working to reduce road trauma by implementing current and proposed interventions. The model allows users to select a suite of actions and activities (the intervention) and prescribe the degree of each (the dose). The model then calculates potential deaths and serious injury (DSI) savings (the response) from that combination of interventions. The dependency, union, dominance or independent nature of the interventions are used in determining the combined effect. The model also accounts for changes in effectiveness of an intervention dependent on the dose and using a projected baseline, the effect of implementing over time.

The main aim of the study presented here is to model a baseline of road deaths in New Zealand from 2018 to 2025. It is considered that time-series models are appropriate for this purpose. However, it needs to be pointed out that time-series modelling is atheoretical: it concentrates on finding a specification that captures the observable dynamic behaviour of a series without necessarily requiring a theoretical explanation for that behaviour. Time-series models can be grounded in theory, and exogenous influences can be incorporated into these models, but these features are not essential. Many time-series models are best regarded as providing a useful approximation to complex, and perhaps not fully understood, real-world phenomena. They offer approximations that help us to predict future realisations and to estimate the impact of disturbances or interventions in the process (Becketti, 2013).

Time-series modelling is a dynamic area that has attracted the attention of researchers over the last few decades. The main aim of time-series modelling is to carefully collect and rigorously study the characteristics of a set of time-series to develop an appropriate model that describes the inherent structure of the data being modelled. This model is then used to generate future values for the series, i.e. to make forecasts. Time-series forecasting can thus be seen as the act of predicting the future by understanding the past. Due to the importance of time-series forecasting in numerous practical fields such as business, economics, finance, science and engineering, proper care should be taken to fit an adequate model to the underlying time-series. Successful time-series forecasting depends on appropriate input data and model fitting. A lot of effort applied by researchers over many years to develop valid models has improved forecasting accuracy.

Methodologies used to model road deaths vary widely and have included simple and multiple regression analyses, Poisson regression analyses, negative binomial regression models, logit and probit models, random parameters models, fuzzy logic models and ARIMA.

There is a large literature on factors that influence the underlying frequency of road trauma. A review of major studies carried out in New Zealand, Australia and selected overseas countries follows.

Literature

Exogenous factors

A change in the economic situation affects individuals' needs and opportunities for travelling, which in turn influence the risk of being killed or injured in a road crash. In order to understand how economic development or activity affects road safety, it is necessary to consider the whole chain of events. From the review of previous modelling above and the literature, several factors are advanced that affect road deaths: these are discussed below.

A downturn in economic activity leads to less travelling and hence less exposure to traffic and a decreased number of crashes. This is partly because people cannot afford to travel as much and partly because the downturn in economic activity leads to higher unemployment rates and therefore fewer journeys to and from work. Most of the reviewed papers concentrate on the travelling of households, but Tay (2001) and Haque (2003) state that freight also follows economic conditions. Likewise, an increase in petrol prices leads to reductions in travelling as households have less disposable income and vice versa.

Several researchers (Hakim et al.1991, Haque. 1993; Tay, 2003; Wagenaar, 1984) have stated that, in addition to a general decrease in travelling, economic factors may influence the distribution of travel across travel types as well as demographic groups. One theory is that travel for recreational purposes and leisure activities is influenced more than journeys to and from work if income is changed. Such trips are also connected with a higher crash risk, since they are more often undertaken in the evenings or at night. Another theory is that travelling in the evening or at night is riskier than travelling during the day due to lack of visibility, higher speeds and a larger proportion of tired, drunken or young drivers on the road. Longer holiday trips are also connected with an increased risk, since they more often than other trip types take place on roads with higher speed limits and in unfamiliar environments (Wiklund et al., 2012).

Tay (2003) states that increased income leads to increased demand for travelling in private vehicles instead of by public transport. This may also have a negative effect on road safety, since public transport travelling is generally safer than other types of travel.

The number of younger people in the population in the age group 15 to 24 has a significant effect on road deaths, as younger drivers are generally more inexperienced, drive less safe cars, have a higher crash risk and may take greater risks around the road network.

Time-series regression models

A number of studies have used a range of time-series regression models to understand road deaths and carry out forecasts. Below is a brief review of some of the studies carried out.

The Australian Bureau of Infrastructure, Transport and Regional Economics (BITRE, 2014) examined the trends in road deaths and injury rates (road deaths and injuries per billion vehicle kilometres travelled (VKT)) in 21 countries around the world. New Zealand was one of the countries investigated. For New Zealand, a model of the fatality rate was constructed using the all-occupant seat belt wearing rate, the percentage of fatally injured drivers over the blood alcohol limit, an index of the speed of the 85th percentile of urban and rural drivers and a dummy variable for the effects of new laws affecting drug users and young drivers. Additional dummies captured the effects of measures taken during the oil crises of 1973 through to 1986.

The BITRE report states that the regression model used produced a reasonable prediction of the change in levels of deaths over time. Regarding New Zealand road deaths, the report concluded that over the period the most significant influence was the increase in seat belt wearing. In addition, unemployment (from the mid-70s), alcohol control (from the mid-80s), and speed control (from the mid-90s) each had an effect.

While the report states that the analysis carried out was robust, there were several concerns with the methodology. It should be noted that the intercept and drug/youth variables were not statistically significant but overall, the model had a high R-squared of 0.98 which may suggest an autocorrelation problem. It is not stated whether autocorrelation tests were carried out and found to be acceptable. Also not stated is whether the variables were checked for stationarity or for being co-integrated.

The analysis carried out showed that road deaths would continue to decline, but with hindsight this was not the case. BITRE suggested that econometric modelling of road deaths or crashes using regression models is extremely difficult as there are many factors that affect road deaths and that identifying them and their inter-relationships (possibly non-linear) is complex.

Another Australian study (Burke & Teame, 2018) showed that low fuel prices and low unemployment were important factors in explaining the rise in Australia's annual road deaths. Using different functional forms, such as ordinary least squares and negative binomial regression models, the study presented macro-level estimates of the factors affecting road deaths in Australia.

It is important to note that the estimated coefficients for petrol prices and unemployment were negative numbers. The authors stated that the reduction of fuel prices might lead to an increase in road deaths via a range of mechanisms. For example, lower fuel prices encourage increased use of motor vehicles. This channel is relevant if the fuel price elasticity of VKT is negative, which is what one would expect. While the proximate causes of any individual crash may be factors such as speeding and distraction, an increase in the total distance driven should be expected to lead to a general increase in the underlying exposure of the population to road crash risks. As Litman (2018) writes "all vehicle travel imposes risks".

Reductions in fuel prices might contribute to increases in road deaths for several reasons relating to road use (Burke & Teame, 2018; Burke & Nishitateno, 2013; Sheehan-Connor, 2015). Conversely, Burke and Teame suggest that there are ways in which lower fuel prices might reduce the number of road deaths. They also point out that electrification and autonomation of the vehicle fleet may cause the link between oil prices and road safety to diminish in the long-term.

An earlier study by Wiklund et al. (2012) also showed that economic factors were important in explaining road deaths in Sweden. The study included a survey of statistical methods used by previous researchers describing the variables they used as indicators of the state of the market (economy). The most common variables were the unemployment rate or the number of unemployed, while level of industrial production measures such as Gross National Product (GNP) were used to a lesser extent. Based on these findings, a model including a time trend, vehicle mileage and the number of unemployed was fitted to a time-series of Swedish data. The results showed that an increase in unemployment was associated with a decrease in the number of road deaths.

A further reason advanced was that the state of the economy affects road users' travel patterns. Results showed that the number of road deaths and number of fatal crashes were higher and collision crashes were more frequent during periods of economic growth. No significant difference was found with respect to time of day, age or gender distribution. Wiklund et al. also investigated the idea that periods of economic growth may induce a higher level of stress in society.

In New Zealand the study by Scuffham and Langley (2002) examined changes in the trend and seasonal patterns in fatal crashes in relation to changes in economic conditions between 1970 and 1994. A structural time-series model was used to estimate quarterly fatal traffic crashes. The dependent variable was modelled as the number of crashes and three variants of the crash rate (crashes per 10,000 km travelled, crashes per 1,000 vehicles and crashes per 1,000 population). Independent variables included in the model were unemployment rate (UER), real gross domestic product (GDP) per capita, the proportion of registered motorcycles, the proportion of young males in the population, alcohol consumption per capita, the open road speed limit, and dummy variables for the 1973 and 1979 oil crises and the seat belt wearing laws. Real GDP per capita, UER and alcohol consumption were all significant and important factors in explaining the short-run dynamics of the model. In the long-run, real GDP per capita was directly related to the number of crashes, but after controlling for distance travelled it was not significant. The road policy factors appeared to have a greater influence on crashes than the role of demographic and economic factors.

Several studies commissioned by the NZ Ministry of Transport have also analysed road deaths in New Zealand (Deloitte, 2017; Infometrics, 2010 & 2013; Keall et al., 2012; Stroombergen, 2013). The Deloitte report for the NZ Ministry of Transport (Ministry of Transport, 2017) modelled the number of casualty crashes and found that the variables VKT, motorcycle registrations, enforcement and speeding were significant. The modelling estimated the impact that a 1% increase in each of the explanatory variables had on the number of crashes (in percentage terms) observed in a given week. For example, a 1% increase in the VKT was associated with a 2.5% increase in the number of crashes. Deloitte stated that the results suggest that crash risk is strongly influenced by VKT as a 1% increase in VKT is associated with a greater than 1% increase in the number of crashes. A 1% increase in the number of motorcycle registrations was also associated with a greater than proportional increase (1.6%) in the number of crashes. This may be the result of motorcyclists' relative vulnerability or simply reflect a buoyant economy. The authors acknowledged that unexplained factors not captured by the model also played a substantial role.

One of the most popular and frequently used stochastic time-series models is the ARIMA model. Commandeur et al. (2013) used an ARIMA model to analyse road deaths in Norway, the UK and France. Their first case study was of road deaths in Norway from 1970 to 2009. An ARIMA (0,1,1) model, without a constant term but including an exogenous variable representing time, was fitted on the log of the annual 1970 to 2009 Norwegian road deaths series. The model parameters were all significant and the residuals of the analysis were considered to be independent, as the diagnostic tests did not reveal any evidence against the assumptions.

In their second case study Commandeur et al. used a multiplicative ARIMA $(0,1,1)(0,1,1)_{12}$ model on the log of the monthly number of car drivers killed or seriously injured (KSI) in the UK. Dummy variables to account for significant events and petrol prices were also included in the model. The model diagnostics were satisfactory in the sense that all parameters were significant and that residuals could be considered to be independent. The values of the regression coefficients for this model indicated that a $100(e^{-0.298} - 1) = -21.8\%$ change in the number of UK drivers KSI from February 1983 onwards was observed and an elasticity of -0.298 in the number of UK drivers KSI compared to the petrol price was suggested. Therefore a 1% rise in the price of petrol was associated with a 0.30% reduction in the number of UK drivers KSI.

The third case study by Commandeur et al. used data from France (January 1975 to December 2000) to demonstrate that an ARIMA model with exogenous (explanatory and intervention) variables can be an efficient tool for analysing the aggregate number of injury crashes and deaths.

From the above cases studies, Commandeur et al. concluded that ARIMA models can be used to analyse the dynamics of a time-series and to extrapolate into the future. Their case studies also showed that explanatory and intervention variables can be included in ARIMA models and the additional corresponding regression coefficients can be estimated and interpreted.

New Zealand context

Developing a New Zealand baseline requires an understanding of local data and trends for example, the recent break in the downward trend in both road deaths and serious injuries. Figure 1. below shows that from 1990 to 2013 road deaths generally declined. However, this trend was reversed, as road deaths increased from 2014 onwards to 2017.

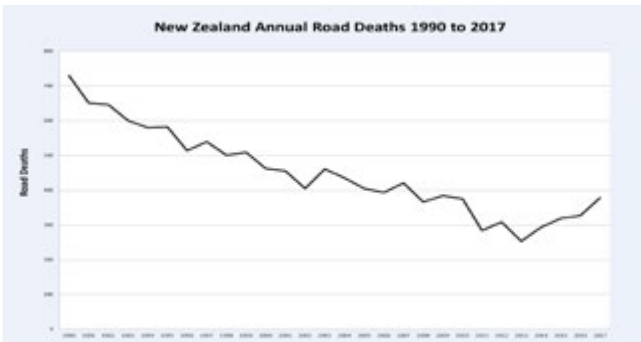

Figure 1: Fatal casualties and road deaths seasonally adjusted 1990 – 2017

From 2014 as DSIs increased, the household savings ratio also fell, suggesting an increase in private consumption and spending. The unemployment rate has decreased and vehicle use (Vehicle Kilometres Travelled, VKT) has risen over the same period.

On call interest rates dropped to a low in 2010. The NZ Treasury has held interest rates at a relatively low rate throughout this period. Building consents rose consistently since 2011 after a period of decline. ANZ Bank economic outlook forecasts of the value of the New Zealand dollar against our major trading currencies (Trade Weighted Index, TWI) were predicted to fall slightly in the next couple of years.

A Reserve Bank of New Zealand bulletin (Delbruck, 2005) suggested that any indirect effects of fuel prices on inflation and economic activity may be relatively large in New Zealand. Diesel has also been the main contributor to increases in New Zealand's oil use over the last decade and a half. It is suggested that the indirect effects of higher fuel prices, through an increase in the costs of providing public transport and other goods and services, may be more significant than in other countries. It was found that indirect effects on the Consumer Price Index (CPI) could potentially be quite sizeable, mostly relating to higher costs to transport services.

In recent years there has been an increase in the number of DSI crashes involving vehicles manufactured between 2004 and 2006. Vehicles older than this are already over-represented in DSI crashes. The change to importation of vehicles when stricter emissions requirements were introduced in 2002 may have had the effect of bringing in newer, but less safe, cheap vehicles that satisfied the emissions standard and were affordable: ongoing analysis at Waka Kotahi NZ Transport Agency suggest that there is evidence to support this.

The involvement of trucks in fatal crashes has changed little over the sample period 1990-2017. There is a wave with a period of around a decade evident in the time-series. This fits with financial cycles that are evident in data: there seems to be a 12-year cycle of financial lows and highs, within which there are regular fluctuations which are roughly either every six years or in three-year cycles.

Motorcyclist deaths remain constant annually with a definite seasonal pattern around summer use. The strong seasonality of motorcycle crashes when removed from fatal crashes makes the remainder of road deaths less seasonal.

The spike in New Zealand road deaths in 2017 was in part due to an increase in the number of cyclist deaths: there were 18 cyclist deaths in 2017 compared to the previous annual average of five. Similarly, pedestrian deaths in 2017 were also the highest annual total for over ten years. These factors combined in contributing to the overall upward trend.

It is not viable to use this level of detail in predictive analysis in the New Zealand context due to the relatively small numbers and random nature of these events. While this level of granularity is not considered in a macro level baseline, it is arguably necessary to be aware of it at least when developing one.

Data

Quality data at set intervals over a reasonable timeframe is necessary for the best predictive models. Much historical New Zealand data is available, collated on a quarterly basis. This was therefore used as the discrete-time interval in the model. Even then, some quarterly data that was available was estimated and often the methodology was not documented, while other datasets did not cover the period sought. Quarterly data was sought from 1990 to 2017, providing 112 data points.

A wide range of data was sourced from official sources:

- Accident Compensation Corporation, motor vehicle accident claims (2018)
- Ministry of Business, Innovation & Employment, energy prices - real and nominal price data relating to New Zealand's energy prices - petrol, diesel, fuel oil, natural gas and electricity (2018)
- Ministry of Health, publicly funded hospital discharges - series (2018)
- Ministry of Transport, RD006 vehicle kilometres travelled on state highways and local roads (2018)
- NZ Police, road policing driver offence data (2018)
- Reserve Bank of New Zealand, exchange rates and trade weighted index (TWI) (2018)
- Stats NZ, estimated resident population (2018)
- The Treasury, fiscal time series historical fiscal indicators 1972-2018 (2018)
- Waka Kotahi NZ Transport Agency, crash analysis system (CAS) (2018)
- Waka Kotahi NZ Transport Agency, transport data - VKT (2018).

An extensive range of data was gathered. Over 50 variables were collected and correlations between pairings of the variables covering the time period were tested.

Note that the criterion of using quarterly data from 1990 excluded the ability to use VKT, as VKT for 1990-1999 is not published. The current method of calculating VKT began in 2001. It is reported as quarterly data but is not seasonal, i.e. it does not reflect quarterly use nor is it seasonally adjusted. Quarterly published VKT is derived from annual vehicle odometer readings. The calculation involves distributing the count across the preceding four quarters while also incorporating an estimate for vehicles newer than three years that do not require annual testing. While national VKT provides a general measure of road use by motor vehicles, in isolation it does not capture the how and why. It was therefore decided to incorporate road use by considering economic and societal factors such as population, employed persons, petrol price and GDP as VKT is strongly correlated to these variables.

Analysis

Correlation

Correlation between pairings of variables was tested in Version 9.4 of SAS, Version 14.3 SAS/STAT (2016). The scatterplots below illustrate the correlation between pairings of those explanatory variables identified for investigative modelling, Figure 2.

The pairings demonstrating the strongest correlation were used in the investigative phase of modelling and are listed below.

The Pearson correlation coefficients, stronger than 0.7 or -0.7, suggest relationships between:

- GDP and road deaths
- Population and road deaths
- Employed (numbers of those in employment) and road deaths
- GDP and employed
- VKT and GDP
- VKT and employed
- Population and petrol price
- Population and GDP
- Population and employed
- Population and percentage of unemployed
- Trade Weighted Index (TWI) and GDP
- TWI and employed.

Variables

The key data used as explanatory variables in the final modelling in this study were:

- Road deaths
- Petrol prices
- Employed
- Young population (number of those persons aged 15 to 24 years)
- GDP expenditure.

Explanations for the relationships to road use and crash outcomes are discussed in the summary of literature above.

Adjustment

It is questionable whether road deaths have a seasonal pattern. Motorcyclist road deaths in New Zealand are clearly seasonal as previously noted, yet when removed from the total the remainder of road deaths have a greatly diminished overall seasonality. After much debate, it was decided to apply Seasonal Adjustment (SA) to road deaths as the other selected variables are seasonally adjusted.

Method

Several time-series models were investigated in understanding road deaths in the New Zealand context.

Dynamic models are constructed using time-series data, with combinations of differing lagged variables.

Figure 2: **Scatterplots of the correlation between pairings of explanatory variables**

The modelling goal is to reflect important interactions among the variables, accurately and concisely. In the final modelling an ARIMA (1,0,1) model and two different ARDL models were developed.

A difference between the ARIMA and ARDL model approaches is that ARIMA is univariate, where the lag of the dependant variable is used as an explanatory variable, whereas ARDL is multivariate, in that additionally it contains lags of the exogenous variables. The ARDL model is based on the belief that an action affects the dependant variable for some time into the future.

The exogenous variables used in the final modelling of road deaths as stated above were petrol price, employed, young population and GDP. All have strong Pearson correlation coefficients.

Testing the data

Before carrying out formal time-series analysis, it is necessary to undertake a series of tests to determine whether the data is stationary or non-stationary. It is important to determine this to avoid spurious regression, as results obtained when using independent non-stationary time-series may indicate a relationship between variables where no relationship exists.

Formally, a time-series Y_t is said to be stationary if (a) its mean and variance are constant over time and (b) the covariance between two values from the series depends on the length of time between them, rather than the actual times at which the values are observed.

Formal testing was carried out to confirm whether the time-series data were stationary or non-stationary. A number of tests can be applied such as the Augmented Dickey-Fuller (ADF) test (Dickey & Fuller, 1981), the DFGLS which calculates a modified Dickey-Fuller test (Elliot, Rothenberg & Stock, 1996) and the Phillips-Peron (1988) test. Of all the tests, studies have shown that the DFGLS test has significantly more power than the others. The DFGLS test was used when fitting our models.

Tests were applied to seasonally adjusted road deaths, petrol prices real and seasonally adjusted, GDP production and GDP expenditure, both real and seasonally adjusted, and employed real and seasonally adjusted. Road deaths seasonally adjusted data were stationary while all the other variables were non-stationary. This meant that the data could be modelled without the need to difference the data.

ARIMA analysis of road deaths

The ARIMA model is often used to model road deaths. Box and Jenkins (1976) developed a practical approach to building an ARIMA model, which best fits to a given time-series and satisfies the parsimony principle. Their concept has fundamental importance in time-series analysis and forecasting.

The Box-Jenkins methodology does not assume any particular pattern in the historical data of the series to be forecasted. Rather, it uses a three-step iterative approach of model identification, parameter estimation and diagnostic checking to determine the best parsimonious model from a general class of ARIMA models. This three-step process is repeated several times until a satisfactory model is finally selected. Then the model can be used for forecasting future values of the time-series.

The ARIMA (1,0,1) model test results showed the p values of the estimated coefficients are all statistically significant. The Q test shows no evidence that the residuals deviate from white noise. The cumulative periodogram generally remains close to the 45-degree line and well within the confidence bands; thus, the residuals do not exhibit any signs of non-random periodicity.

ARDL analysis of road deaths

In modelling road deaths given the data constraints, using stationary and non-stationary data with different levels of integration, a number of different time-series regression models were investigated using Version 14.2 of Stata (2015). The ARDL model was found to be a suitable functional form to explain and forecast road deaths.

In its general form, with *p* lags of *y* and *q* lags of *x*, the ARDL model can be written as:

$$y_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots\dots\dots + \alpha_q x_{t-q} + \beta_0 + \beta_1 y_{t-1} + \dots\dots + \beta p y_{t-p} + \varepsilon_t$$

where ε_t is a random «disturbance» term.

As noted earlier, while road deaths data are stationary, the other variables used previously in the ARIMA model (petrol prices, employed and GDP) are non-stationary. However, it was found in testing the output of the ARDL models that the data were cointegrated. In the ARDL model specification this involved testing whether the residuals of the estimated model are stationary. If the residuals are stationary, the estimated relationships are unlikely to be spurious as they are cointegrated.

The modelling approach adopted was to estimate a model with a set of possible explanatory variables and then successively eliminate statistically insignificant variables while considering the R-squared. Several ARDL model specifications were investigated and two models considered to be the most appropriate were identified.

The first ARDL model used mainly economic variables: petrol prices and unemployment. Both variables were statistically significant. The Engle- Granger test indicated that the variables were cointegrated and various tests carried out showed no autocorrelation (LM, Durbin-Watson and Residual Correlogram). The second ARDL model had in addition a demographic variable, persons in the younger age group between 15 and 24 years. This showed that all the variables were statistically significant, cointegrated and not autocorrelated. The R-squared for ARDL model 1 is 0.7861 and for ARDL model 2 it is 0.8008.

Results

The ARIMA model (1,0,1) forecast indicates that road deaths will continue to rise from 2018 to 2025, influenced by the recent rising trend from 2014 to date and the spike in 2017. These recent factors do not reflect the time-series. The results show that the ARIMA model underestimated road deaths, on average by around five percent.

Both ARDL models forecast a flattening trend in road deaths: a slow rise early in the forecast period followed by a decline. This has now been seen to have happened since completing the modelling.

The ARDL model 2 differs from ARDL model 1 as it includes a demographic variable: population of persons aged 15 to 24. ARDL model 2 shows that, in addition to the economic variable (petrol prices) and socio-economic variable (employed), the number of people aged 15 to 24 in the population is correlated with road deaths. Including this demographic variable produced in ARDL model 2 a higher R-squared value and it is therefore the preferred model.

ARDL model 2 results

Table 1 summarises the ARDL model 2 specification. Tables 2, 3, and 4 show the test results of this model.

Forecasting with the ARDL models

The reliability of the forecast is dependent on the accuracy of forecasting future youth population, petrol prices and employment numbers. The limitations of all these forecasted values of explanatory variables is that they are strongly influenced by exogenous factors that cannot be modelled. Petrol prices are the most volatile. That said, the historical data for 1990 to 2017 shows price trend has moved relatively consistently and variance has been bound by market forces without dramatic disruptive changes.

The approach taken in forecasting these explanatory variables was to use the Holt-Winters algorithm, a seasonal exponential smoothing algorithm (ETS AAA)

Table 1. ARDL model 2 summarised

Variable	Coeff.	Std. Err	P-Value	95% CI
Lag 1 log road deaths seasonally adjusted	0.2952571	0.0921223	0.002	0.1126157 0.4778984
Lag 1 log petrol price real and seasonally adjusted	-0.5380974	0.1360431	0.000	-0.8078161 -0.2683786
Lag 1 log employed seasonally adjusted	-1.171861	0.2040765	0.000	-1.576463 -0.7672598
Lag 1 log population aged 15-24 years	1.308681	0.4113065	0.002	0.4932258 2.124136
Constant	15.62295	2.421833	0.000	10.82143 20.42447
Adjusted R-squared	0.8008			

Table 2. ARDL model 2, Engle-Granger cointegration test

Test Statistic	1% Critical Value	5% Critical Value	10% Critical Value
Z(t) -11.029	-3.507	-2.889	-2.579

Table 3. ARDL model 2, Breusch-Godfrey LM test for autocorrelation

Lags (p)	chi^2	Df	Prob > chi^2
1	1.585	1	0.2080

Table 4. Durbin's alternative test for autocorrelation

Lags (p)	chi^2	Df	Prob > chi^2
1	1.521	1	0.2174

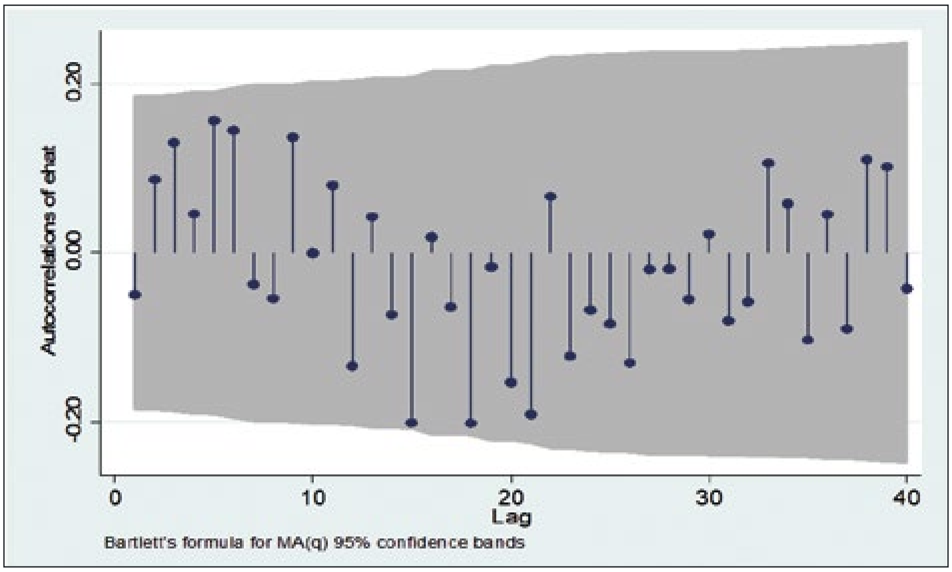

Figure 3. Residual correlogram of ARDL model 2

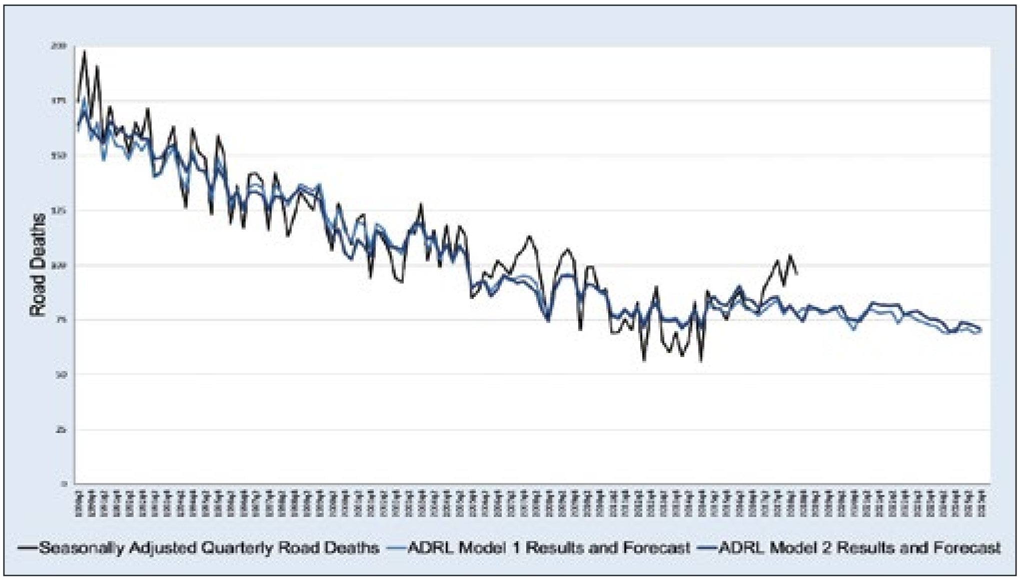

Figure 4. Results of ARDL Models 1 and 2 plus quarterly forecasts 2018 to 2025

and harmonic analysis of unique longer waveform periods of each variable. The forecast variables were used in conjunction with the ARDL model 2 coefficients to forecast quarterly road deaths to 2025.

ARDL model 1 and model 2 quarterly results and forecasts

Table 5. shows the modelled results from 2014 to 2017 and forecast results of both models from 2018 to 2025 for annual road deaths.

For use in Waka Kotahi NZ Transport Agency's IILM, rather than offering various scenarios for a predicted baseline to 2025 based on differing assumptions around future values of the explanatory variables, the approach taken was to forecast the 95%CI scenarios of the variables using ARDL Model 2. This confidence band of quarterly values for the period to 2025 was used in predicting annual road deaths for the baseline.

The values for the IILM baseline from 2018 to 2025 are listed below and illustrated in Figure 6. The values are presented as a range, the span is five percent of their value and they lie within the upper 95%CI model forecast.

2018 343-359 (actual 377)

2019 351-367 (actual 352)

2020 342-358 (actual 318, the Covid-19 NZ-lockdown effect is calculated at -10%)

2021 349-365

Table 5. ARDL Models 1 and 2 Forecast

Year	Annual road deaths	Model 1 Forecast	Model 2 Forecast
2014	293	298	296
2015	319	319	331
2016	327	324	344
2017	378	322	333
2018	*377	316	312
2019	*352	316	319
2020	*318	303	311
2021		313	317
2022		308	322
2023		298	310
2024		281	288
2025		280	290

*Post modelling the counts are now known to 2020, Covid-lockdown effect 2020 is est. -10%.

2022	352-370
2023	341-357
2024	317-330
2025	319-334.

This sequence is to be used as the baseline in Waka Kotahi NZ Transport Agency's integrated road safety intervention logic model.

Conclusion

The view taken in this study was to approach modelling by assuming nothing; to seek quality data that covered a long enough period to be useful; and to test everything.

As one would expect none of the models capture the extreme spikes, peaks and troughs across quarters. However, the fit of all three models across the 28-year time-series was shown to be good. The use of economic and societal factors of youth population, employed persons, petrol price and GDP proved valid in modelling road deaths.

The selected ARDL model was, after much trial and experimentation, identified as the preferred time-series approach for forecasting road deaths in New Zealand.

The predicted results have now been exposed, and with 2018, 2019 and 2020 counts now known, the projection from this ARDL model looks to be a valid baseline for the road safety integrated intervention logic model (IILM) being developed in New Zealand.

The IILM is a tool developed by Waka Kotahi NZ Transport Agency in partnership with key road safety partners. It is used to calculate the potential reductions in deaths and serious injuries that could be achieved through a combination of evidence-based interventions that can be reliably modelled.

A key objective of the model is to show how investment in road safety through the National Land Transport Programme can be optimised, i.e. to give greater assurance that we are investing in the right safety interventions, in the right combination and at the right levels. The

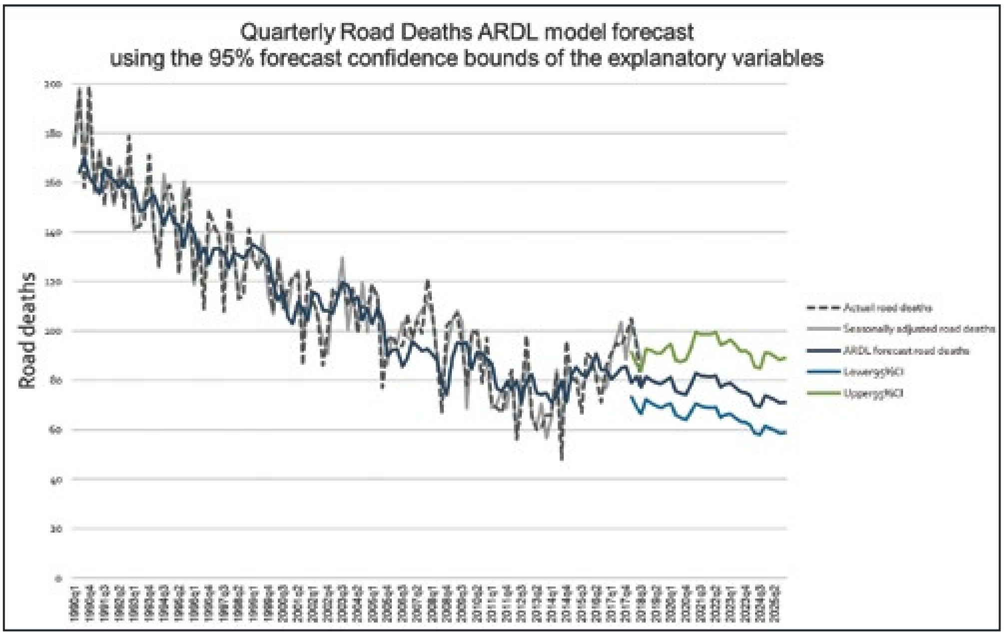

Figure 5. Forecast model using 95%CI scenarios of variables

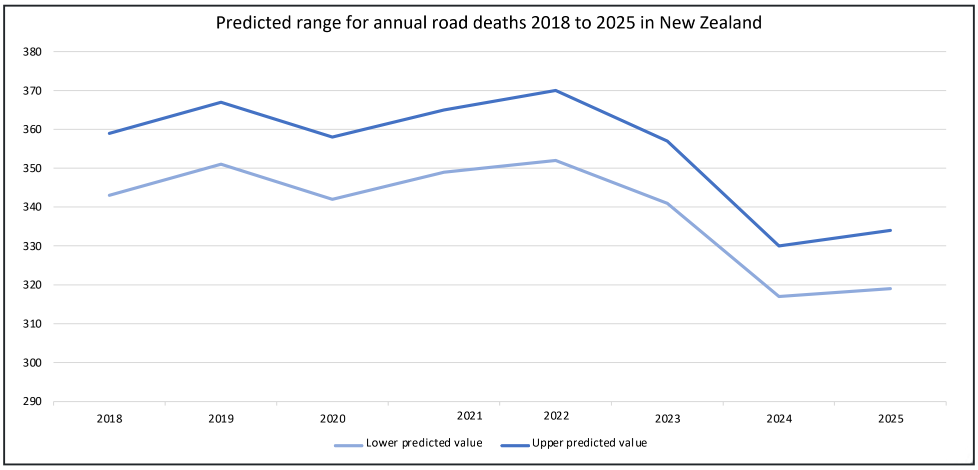

Figure 6. Predicted annual road deaths 2018 to 2025 for a baseline

interventions modelled are in addition to existing levels of investment and effort in road safety.

The IILM is nearing completion and this strategic tool has been used for assessing the DSI reductions and costs from suites of road safety activities, interventions and programmes. It was used in the target setting of the NZ Governments road safety strategy for 2020-2030, Road to Zero.

Acknowledgements

A special mention and thanks to Prof. Keall for undertaking peer review of the research.

Michael Keall, Department of Public Health, BA (Hons), PhD (Injury Epidemiology). Research Associate Professor; Co-director of He Kainga Oranga/Housing and Health Research Programme.

References

Accident Compensation Corporation (ACC). (2018). *Motor Vehicle Accident Claims.* Retrieved from https://catalogue.data.govt.nz/dataset/motor-vehicle-accident-claims

Australian Bureau of Infrastructure, Transport and Regional Economics (BITRE). (2014). *Road Safety: Modelling a Global Phenomenon.* Canberra, ACT, Australia. BITRE.

Becketti, S., (2013). *Introduction to Time-series Using Stata.* Texas, USA: Stata Press.

Box, G. & Jenkins, G., (1976). *Time Series Analysis: Forecasting and Control, Revised Edition*. Holden-Day, San Francisco.

Burke, P.J. & Nishitateno, S., (2013). Gasoline prices, gasoline consumption, and new-vehicle fuel economy: Evidence for a large sample of countries. *Energy Economics. 36.* 363-370

Burke, P.J. & Teame, A., (2018). *Fuel Prices and Road Deaths in Australia.* Centre for Applied Macroeconomic Analysis (CAMA), Crawford School of Public Policy, The Australian National University (ANU).

Commandeur et al. (2013). On statistical inference in time-series analysis of the evolution of road safety. *Accident Analysis and Prevention,. 60,* 424-434.

Delbruck, F., (2005). Oil prices and the New Zealand economy. *Reserve Bank of New Zealand Bulletin, 68 (4),* 5-15

Hakim et al., (1991). A critical review of macro models for road accidents. *Accident Analysis and Prevention, 23 (5),* 379-400.

Haque, R., (2003). *Public Investment in Transportation and Communication and Growth: A Dynamic Panel Approach.* Centre for Growth and Business Cycle Research Discussion Paper Series 31, Economics, The University of Manchester.

Haque. M.O., (1993). Unemployment and road fatalities. *International Journal of Transport Economics, 20,* 175-195.

Litman, T., (2018). *A New Traffic Safety Paradigm.* Victoria Transport Policy Institute, Victoria, BC, Canada.

Ministry of Business, Innovation & Employment (MBIE). (2018). *Energy Prices - Real and nominal price data relating to New Zealand's energy prices — petrol, diesel, fuel oil, natural gas and electricity.* Retrieved from https://www.mbie.govt.nz/building-and-energy/energy-and-natural-resources/energy-statistics-and-modelling/energy-statistics/energy-prices/

Ministry of Health (MoH). (2018). *Publicly funded hospital discharges - series.* Retrieved from https://www.health.govt.nz/nz-health-statistics/health-statistics-and-data-sets/publicly-funded-hospital-discharges-series/publicly-funded-hospital-discharges-series/publicly-funded-hospital-discharges-series

Ministry of Transport, Deloitte (2017). *Qualitative and Quantitative Analysis of the New Zealand Road Toll: Final Report.* Wellington, New Zealand. Ministry of Transport.

Ministry of Transport, Keall, M.D., Stroombergen, A., Sullivan, C., Cameron, M. & Weiss, H., (2012). *An analysis of potential factors behind the 2011 reduction in New Zealand road fatalities.* Wellington, New Zealand. Ministry of Transport.

Ministry of Transport (MoT). (2018). *RD006 Vehicle kilometres travelled on state highways and local roads (billion km)* Retrieved from https://www.transport.govt.nz/mot-resources/transport-dashboard/2-road-transport/rd006-vehicle-kilometres-travelled-on-state-highways-and-local-roads-billion-km/

NZ Police. (2018). *Road Policing Driver Offence Data* Retrieved from https://www.police.govt.nz/about-us/publication/road-policing-driver-offence-data-january-2009-march-2019

Reserve Bank of New Zealand (RBNZ). (2018). *Exchange rates and TWI - B1*. Retrieved from https://www.rbnz.govt.nz/statistics/b1

SAS [Computer software]. (2016). Retrieved from https://www.sas.com/en_us/software/sas9.html

Scuffham, P.A.& Langley, J.D., (2002). A model of traffic crashes in New Zealand. *Accident Analysis and Prevention, 34,* 673–687.

Sheehan-Connor, (2015). Environmental policy and vehicle safety: The impact of gasoline taxes. *Economic Inquiry. 53.*

Stata [Computer software]. (2015). Retrieved from https://www.stata.com/

Stats NZ. (2018) *Estimated Resident Population*. Retrieved from http://nzdotstat.stats.govt.nz/wbos/Index.aspx?_ga=2.73481243.889223470.1561348120-746931749.1558996738

Tay, R., (2001). Fatal crashes involving young male drivers: a continuous time Poisson change-point analysis. *Australian & New Zealand Journal of Public Health, 25(1),* 21-3.

Tay, R., (2003) The efficacy of unemployment rate and leading index as predictors of speed and alcohol related crashes in Australia. *International Journal of Transport Economics, 30,* 363–380.

The Treasury. (2018). *Data - Fiscal Time Series Historical Fiscal Indicators 1972-2018.* Retrieved from https://treasury.govt.nz/publications/information-release/data-fiscal-time-series-historical-fiscal-indicators-1972-2018

Wagenaar, A.C., (1984). Effects of macroeconomic conditions on the incidence of motor vehicle accidents. *Accident Analysis and Prevention, 16,* 191–205.

Waka Kotahi NZ Transport Agency. (2018). *Crash Analysis System (CAS)*. Retrieved from https://www.nzta.govt.nz/safety/safety-resources/crash-analysis-system/

Waka Kotahi NZ Transport Agency. (2017). *NZ Transport Agency Statement of Performance Expectations 2017/18.* Wellington, New Zealand. NZTA.

Waka Kotahi NZ Transport Agency. (2018). *Transport Data – VKT.* Retrieved from http://www.nzta.govt.nz/assets/userfiles/transport-data/VKT.html

Wiklund et al., (2012). *Traffic safety and economic fluctuation. Long-term and short-term analyses and a literature survey.* The Swedish Road Administration, VTI rapport 704A: Sweden.