Abstract No: MR-04

Modelling Sri Lankan traffic accident casualties: time series count data model approach

E. S. P. H. Rajadasa* and S. D. Viswakula

Department of Statistics, University of Colombo, Sri Lanka
pramudi5559@gmail.com*

Traffic accident related deaths and fatalities have become the first level global health problem within the past 20 years. Accurate forecasting of the number of traffic accident casualties for a particular geographical area is very important in order to reduce the fatality rate associated with traffic accidents. Health care authorities, hospitals and emergency ambulance services can have a general idea about the number of emergency patients that can be received in given time period. Law enforcement and health care authorities can develop strategic plans to prevent traffic accidents by effectively managing road traffic. Therefore, the ultimate objective of this research was to identify the associated factors, and forecast traffic accident casualties in Sri Lanka with higher accuracy. Usually, count data are modelled with Poisson regression, and time series data are modelled with Gaussian time series modelling techniques. In order to get better forecasting accuracies, both time series and count aspects of traffic accident casualty data should be considered simultaneously. These hybrid models had been rarely used in literature due to the limited awareness and the complexity of the models. Therefore, this research was planned to introduce time series count data modelling approach for Sri Lanka traffic accident data. General time series modelling techniques such as Auto-regressive Integrated Moving Average Models and time series count data modelling techniques such as Time Series Generalized Linear Models have been compared to choose the best model. All island Sri Lankan traffic accident data for years between 2003 and 2016 that was collected by Sri Lanka Police Traffic Head Quarters, has been used to build our traffic accident forecasting model. The data set contained 29 categorical and 11 numerical variables after data cleaning. The time series count data model was able to decrease the mean absolute percentage error by 14.2% The Poisson time series count data model that was fitted using daily accumulated traffic accident casualty time series has become the overall best model. The exploratory analysis shows that there is a strong relationship between number of traffic accident casualties and the variables which indicate the geographical location of the accident such as Province or Police division. Therefore, the forecasting accuracy was further improved by fitting separate Poisson time series count data models for each Police division in Sri Lanka. For example, the root mean squared error was 3.1 for the daily road casualty forecasting model of Nugegoda police division after forecasting for 365 days. Fitting separate models for each police division holds more practical value, since the authorities can get a specific idea about small geographic area. The results of this study have further shown that the variables such as day of the week, time of the day and weather related variables do not have any significant relationship with the number of traffic accident casualties.

Keywords: Traffic accident data, Poisson *tsglm*, INGARCH, Count time series