JOINT MODELING OF DAILY CRASH COUNTS AND THE RISK IMPACT OF TRAFFIC EXPOSURE AND WEATHER CONDITIONS IN THE NETHERLANDS

Xanthi Pedeli, *Department of Hygiene, Epidemiology and Medical Statistics, University of Athens, Medical School & Department of Statistics, Athens University of Economics, Greece*
Klea Katsouyanni, *Department of Hygiene, Epidemiology and Medical Statistics, University of Athens, Medical School, Greece*
Dimitris Karlis, *Department of Statistics, Athens University of Economics, Greece*

Background and Aims: Various approaches have been suggested to model time-series crash count data. However, most of them do not explicitly account for the significant autocorrelation present in the data. We consider the joint modeling of daily crash counts in Schiphol, De Bilt and Soesterberg, in the Netherlands for the year 2001. Vicinity of the three areas implies that the data share some similar environment, for example weather conditions and road characteristics. Thus, the accidents series are also expected to be cross–correlated. For the above reasons the joint modeling of the three series can prove to be useful in practice. Moreover, we aim to investigate the risk impact of traffic exposure and weather conditions on the daily crash counts.
Methods: Data were modelled through an extension of the recently popular INteger-valued AutoRegressive (INAR) models to the multivariate setting. The models that arise under alternative distributional assumptions were compared in terms of their fit and plausibility. Estimates of the unknown model parameters were obtained through the method of composite likelihood. A binary variable for weekends vs. weekdays was used as proxy for the day-of-the-week variability in traffic exposure. Weather conditions were expressed through the mean daily temperature in degrees Celsius.
Results: The multivariate INAR(1) regression model with trivariate negative binomial innovations proved to be the best model among those fitted to our data. Our findings are consistent with evidence from earlier research and confirm the greater frequency of crashes on weekdays as compared to weekends ($\bullet$ =-0.375, p<0.001) as well as the negative, highly significant relationship between the absolute temperature and the number of crashes ($\bullet$ =-0.015, p=0.001).
Conclusions: Multivariate INAR processes can provide a useful framework for modeling simultaneously more than one correlated count time series. Additional risk factors, e.g. precipitation, sunshine and visibility, could lead to greater improvement in fit, or even remove the need for a complicated distribution like the multivariate negative binomial.