Transport Policy 32 (2014) 203–217

Contents lists available at ScienceDirect

Transport Policy

journal homepage: www.elsevier.com/locate/tranpol

Explanatory and prediction power of two macro models. An application to van-involved accidents in Spain

Bahar Dadashova*, Blanca Ramírez Arenas, José McWilliams Mira, Francisco Izquierdo Aparicio

University Automobile Research Institute (INSIA), Technical University of Madrid (UPM), José Gutiérrez Abascal 2, 28006 Madrid, Spain

ARTICLE INFO

Available online 22 February 2014

Keywords:
Traffic accidents
DRAG
Unobserved components model
Box–Cox transformation
Autoregressive errors
Crash prediction

ABSTRACT

The figures representing road safety in Spain have substantially improved during the last decade. However, the severity indicators concerning vans have not improved as favorably as those of other types of vehicles, such as passenger cars and heavy freight transport vehicles. This study is intended to analyze the main factors explaining van accident behavior and to get a further insight into dynamic macro models for road accidents. For this purpose we are using four time series related to the frequency and severity of van accidents on Spanish roads and two types of methodologies applied in the study of traffic accidents: linear regression with Box–Cox transformed variables and autoregressive errors (DRAG), and an unobserved components model (UCM). The four response time series modeled are the number of fatal accidents, the number of accidents with seriously injured victims, the number of fatalities and the number of seriously injured victims. Since the choice of the appropriate macro model for the analysis of road traffic accidents is not a trivial matter, we are considering multiple factors such as goodness of fit and interpretation, as well as the prediction accuracy in order to choose the best model. Overall, the final results make sense and agree with the literature as far as the elasticities and coefficient signs are concerned. It was found that the DRAG-type model yields slightly better predictions for all four models compared to UCM. With these macroeconomic models, the effect of some influential factors (fleet, drivers, exposure variables, economic factors, as well as legislative actions) can be addressed. Estimating the effect of the vigilance and surveillance actions can help safety authorities in their policy evaluation and in the allocation of resources.

1. Introduction

Spain reached the European Union target of achieving a 50% reduction in the number of road fatalities during 2001–2010, following Latvia, Estonia and Lithuania (European Transport Safety Council, 2011). Between 2001 and 2010, the total number of fatal accidents decreased from 4170 to 1953 (53% decrease) and the number of road fatalities from 5516 to 2479 (55%). If the analyses were to be carried out by vehicle type, the number of fatal accidents decreased by 57% and by 56% in passenger cars and large trucks respectively, while in van-involved fatal accidents only a 33% decrease was observed. As for the number of fatalities, there was a 38% decrease in van-involved accidents, while for passenger cars and trucks the figures were 58% and 59% respectively. Comparing in terms of index year rate, it can be observed that the road fatalities involving vans (Light Goods Vehicles – LGV) have not evolved as favorably as the other vehicle types: passenger cars and trucks (Heavy Goods Vehicles – HGV) (Fig. 1). As can be observed, the number of fatalities involving any type of vehicle clearly decreased between 2003 and 2010. Moreover, a slight increase in serious accidents involving vans is noted in 2006 and 2009.

The figures showing the different behavior of van-involved accidents and their consequences were the main motivation for a detailed and thorough study conducted by Spanish researchers led by the Technical University of Madrid, during 2009–2011, aimed at understanding the factors influencing van-involved crashes (FURGOSEG, 2011). The project was focused on the integrated analysis of the different factors that can explain the accident-prone behavior of vans. Data from van-driver surveys, dynamic trials, safety devices installment degrees in vans, the technical maintenance of vehicles and the development of statistical models, etc. were used. This paper is a result of part of the project, the development of macro models for van-involved accidents and their consequences.

* Corresponding author. Tel.: +34 913363014; fax: +34 913365305.
E-mail address: bahar@etsii.upm.es (B. Dadashova).

http://dx.doi.org/10.1016/j.tranpol.2014.01.014
0967-070X

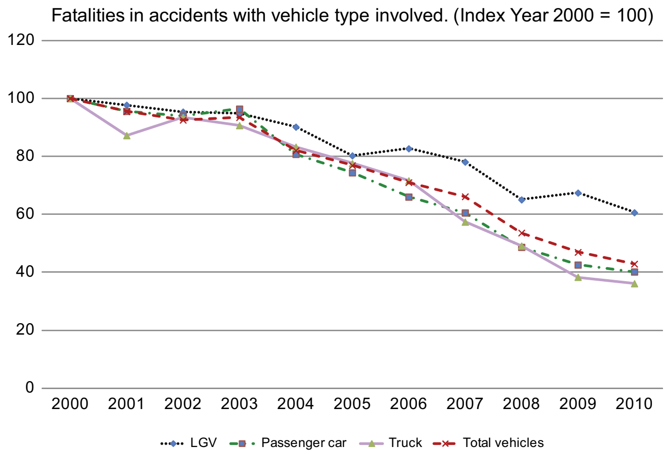

Fig. 1. Fatalities in accidents with vehicle type involved (Index Year 2000 = 100).

Two macro models are considered for the analysis of van-involved accidents: the unobserved components model (UCM – descriptive model) and a special case of the structural explanatory model, i.e. DRAG (Demand for Road use, Accidents and their Gravity) model. The UCM was first developed and applied to traffic data by Harvey and Durbin (1986). The DRAG model was developed by Gaudry (1984). As pointed out by Hakim et al. (1991) an acceptable model is expected to possess characteristics such as description, explanation and prediction of the phenomenon. Thus the two models are compared based on goodness-of-fit measures and prediction accuracy.

The main difference between UCM and DRAG is the specification and separate modeling of unobserved components, i.e. trend and seasonal components, in the case of UCM. These components are captured in the DRAG models by the regressors or the auto-correlated errors. In the road safety literature the UCM are preferred over classical, non-linear regression and DRAG models because as mentioned above, they can be used to explicitly decompose a time series into interesting components such as trend and seasonal effect. They are also flexible and well able to handle the dependencies in time series. Furthermore, they work transparently with missing data and are easily generalized to the multivariate analysis of time series, while the DRAG model is simpler and has a more straightforward interpretation. There is a very close relationship between UCM and ARIMA models, in such a way that one can almost always find, given a specific model of any type, an equivalent in the other one. There is important literature on this issue (Maravall, 1985; Hillmer and Tiao, 1982; Harvey and Scott, 1994). However, in practice, DRAG, which is estimated using the TRIO package (Gaudry et al., 2005), does not cover all the autoregressive integrated moving average (ARIMA) model range since its stationary autocorrelation modeling is restricted to AR structures. Nevertheless, DRAG is quite sophisticated in the treatment of transformations, including the corresponding parameters in the maximum likelihood (MLE) estimation through the TRIO software.

In road safety literature different studies have carried the analysis of macro model comparison based on the forecast results. García-Ferrer et al. (2006) apply different macro models, such as dynamic harmonic regression and casual econometric models to the traffic accident data in Spain. Their results show that none of the applied models is dominant under the forecasting criteria chosen by the authors. Rather, each model produces better prediction accuracy depending on the forecast year. A similar analysis was carried out by Quddus (2008), where Poisson models were compared to Box–Jenkins ARIMA (autoregressive integrated moving average) models using the aggregated and disaggregated traffic accident data. The results of this study show that, depending on the time series, prediction accuracy varies: in the case of the aggregated series the ARIMA model did better while for disaggregated data the best prediction was obtained when the INAR (integer-valued autoregressive) Poisson model was applied. It should be noted that the predictive ability of a given road accident model depends on the set of explanatory variables we include. However, model misspecification does not solely depend on the explanatory variables but also on the error term, which in the case of accident data has a high variance (Scott, 1986) and, for monthly data, it also often depends on the present dependence structure (Hakim et al., 1991), resulting mainly from seasonality and trend effects. The persistence of autocorrelation in the model could result in biased estimates of the coefficients thus affecting prediction. Therefore Box–Jenkins ARIMA models are one of the standard solutions, possibly including transfer functions.

The objective of this study is twofold: firstly, to select a model for the accidental behavior of vans with significant parameter estimates that complies with the existing literature on road safety and presents better prediction accuracy. Secondly, to get an insight into comparative modeling with two frequently applied dynamic models, DRAG and UCM. As far as the authors know, this is the first van-specific macro model and thus the compliance with previous work is studied with respect to general road accident literature.

In the following section the model estimation and selection process are given in detail. In Section 2 we review the two types of models being compared and in Section 3 the data used for the study are discussed. Section 4 presents the results of estimation and prediction and finally, Section 5 concludes this study and provides guidelines for further research.

2. Methodology

2.1. Demand for road use accidents and their gravity model (DRAG)

The DRAG methodology is based on linear regression with autoregressive errors where the dependent and the independent variables are Box–Cox transformed (Box and Cox, 1964). The BCT is applied for achieving normality and homoskedasticity of the output variable. Gaudry (1984) and Gaudry and Lassarre (2000) also incorporate high-order autoregressive errors. This methodology was originally developed for estimating the frequency and severity of road traffic accidents in Quebec, Canada (Gaudry, 1984). Later on, DRAG-inspired models were developed for other parts of the world such as Norway (Fridstrøm, 2000), Stockholm (Tegnèr and Loncar-Lucassi, 1997), France (Jaeger and Lassarre, 2000), Germany (Blum and Gaudry, 2000), California (McCarthy, 2000), Spain (Aparicio Izquierdo et al., 2009), Algeria (Gaudry and Himouri, 2013), which altogether make up the DRAG-family (Table 1). The functional form of DRAG modeling is specified as the following regression model (Liem et al., 2008):

$$Y_t^{(\lambda_Y)} = \sum_{k=1}^{K} \beta_k X_{kt}^{(\lambda_X)} + u_t \tag{2.1}$$

Table 1
DRAG family models.

Country	Authors	Monthly period	Model
Germany	Blum and Gaudry	1968–1989	SNUS
France	Jaeger and Lassarre	1957–1993	TAG
Norway	Fridstrøm	1973–1994	TRULS
Sweden	Tegnèr	1970–1995	DRAG-Stockholm
California	McCarthy	1981–1989	TRACS-CA
Spain	Aparicio et al.	1990–2004	DRAG I-DE Spain
Algeria	Gaudry and Himouri	1970–2007	DRAG-ALZ-1