Safety Science 72 (2015) 283–292

Contents lists available at ScienceDirect

Safety Science

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

Effects of changes in road network characteristics on road casualties: An application of full Bayes models using panel data

Haojie Li *, Daniel J. Graham, Arnab Majumdar

Centre for Transport Studies, Dept of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 21 May 2014
Received in revised form 20 August 2014
Accepted 1 October 2014
Available online 24 October 2014

Keywords:
Road network
Panel data
Continuous treatment

ABSTRACT

In order to ensure a high level of road safety, road network planning needs to be based on the best knowledge available of the effects of road design on road safety. In this study, we look into how changes in road network characteristics affect road casualties. An approach based on traffic assignment is proposed in order to estimate the traffic exposure at ward level. We apply a widely used approach for before–after evaluation studies, the Bayesian method. We also use a panel semi-parametric model to estimate the dose–response function for continuous treatment variables. The result suggests that there are more casualties in areas with better connectivity and accessibility, where more attention should be paid to safety countermeasures.

1. Introduction

The statistical relationship between road casualties and the characteristics of a road network has been investigated in the literature (e.g. Noland and Quddus, 2004; Wier et al., 2009; Huang et al., 2010; Dumbaugh and Li, 2010; Marshall and Garrick, 2011; Rifaat et al., 2011; Jones et al., 2008; Quddus, 2008). Specifically, road casualties are found to be significantly associated with road network characteristics, such as road length, road density and the number of nodes, i.e. junctions. The objective of this study is to investigate the causal relationship between road traffic crash casualties and road network characteristics, including traditional road network characteristics, the connectivity and accessibility of the road network and the curvature of the road network. A causal framework is established by employing full Bayes models. A panel data semiparametric model is used to estimate dose–response functions for continuous treatment variables.

This research contributes to the literature by addressing two key issues overlooked in previous studies. One constraint in previous research is that, to the best of our knowledge, no panel road network data has been analysed in the UK. This implies that any variance in road network characteristics over time cannot be controlled for and inferences made on the impacts of the road network on road traffic crash casualties could therefore be biased. In contrast, in this study a panel data set for the road network in England and Wales between the years 2001 to 2010 is used to account for time-invariant heterogeneity in the road network.

Another key issue, which is critical in all analyses relevant to road traffic crash casualties, concerns the exposure variable. In analyses at the disaggregate level, the ideal variable used to control for risk exposure is the annual average daily traffic (AADT) for the unit of interest. In terms of aggregate analysis, however, the AADT is unavailable for an aggregate area and therefore typically proxy variables are used. Various such proxy exposure variables, include the usage of cars (Quddus, 2008), aggregated AADT (Jones et al., 2008; Marshall and Garrick, 2011) and a proxy variable derived from the gravity model using data of population and employment (Graham and Glaister, 2003). When applying these variables as exposure variables, however, limitations arise. We will discuss this in detail later. This paper uses the daily traffic trips generated within the study area as an exposure variable, and this is estimated using the origin–destination (OD) data obtained from the Office for National Statistics. A two-stage regression is used for this estimation, whereby the traffic trips are estimated with ordinary least square regression in the first stage and the road casualties are analysed using a panel generalized linear model in the second stage. To account for the composite uncertainty in the two-stage regression, the bootstrap technique is employed.

2. Previous research

Numerous studies have been conducted to develop accident models at aggregate zonal levels. One of the earliest studies by Levine et al. (1995) develops an aggregate model that relates accidents in zones in Hawaii to various factors. However, the model in this study is linear and inappropriate for accident data analysis. Generalized linear models, such as Poisson and NB regression

* Corresponding author. Tel.: +44 (0)20 75942706; fax: +44 (0)20 75946102.
E-mail address: h.li09@imperial.ac.uk (H. Li).

http://dx.doi.org/10.1016/j.ssci.2014.10.001
0925-7535/

models, have become an prevalent approach to accommodating nonnegative discrete data and have been widely used to relate accident frequencies to various characteristics at zonal levels (Hadayeghi et al., 2003; Lovegrove and Sayed, 2007; Abdel-Aty et al., 2011). For example, Hadayeghi et al. (2003) apply a series of macro-level crash prediction models to relate the number of crashes to zonal characteristics in planning zones in the city of Toronto. Recently, several studies have applied such zone-level models to analyse the relationship between road casualties and road network characteristics.

A spatially disaggregated analysis of road casualties in England undertaken by Noland and Quddus (2004) examined the effects of road characteristics and land use on road casualties. Their results suggest that increased length of B roads could increase serious injuries, irrespective of its length, although the coefficients for other types of road were not significant. Marshall and Garrick (2011) investigated how street network characteristics affected road safety in 24 Californian cities from 1997 to 2007. Street network characteristics, such as street network density, street connectivity and street network patterns were controlled for in this study and will be discussed in detail in the following sections. Marshall and Garrick's results suggest that road casualties for all levels of crash severity are correlated with street network characteristics. A higher density of intersection counts (i.e. nodes) is associated with fewer crashes, while street connectivity (link to node ratio) is positively related to crashes. Dumbaugh and Li (2010) also find that miles of arterial roadways and numbers of four-leg intersections were major crash risk factors in Texas, using data from 2003 to 2007. Another study conducted by Rifaat et al. (2011) examined the effect on crash severity of different street patterns, including grid-iron, loops and lollipops, and mixed patterns. Although only pedestrian and bicycle crash data were analysed, the authors found significant effects of street pattern, road features and environmental conditions on the number of crashes. One limitation of all the above research is that they fail to examine the spatial correlation of road traffic crashes. According to LeSage (1998), two problems arise when sample data has a locational dimension: (1) spatial correlation exists between the observations, and (2) spatial heterogeneity occurs in the relationships that are modelled. Traditional zone-level models could violate the traditional Gauss–Markov assumptions, and hence lead to biased inferences.

If spatial correlation is present, appropriate spatial models need to be employed to account for the spatial dependence. Generally, there are two methods for the spatial analysis of road casualty data, traditional econometric models and Bayesian hierarchical models. Although the results from both models are very similar in many cases, it is indicated that inferences derived from traditional spatial models could be misleading when the spatial unit of analysis becomes smaller (Bhati, 2005). Bayesian hierarchical models have been suggested to be more appropriate and have been employed in many studies (Persaud et al., 2010; Aguero-Valverde and Jovanis, 2006; Haque et al., 2010). For example, Aguero-Valverde and Jovanis (2006) compared full Bayes (FB) hierarchical models with traditional negative binomial (NB) models using county-level crash data for Pennsylvania. The existence of spatial correlation in county-level crash data was revealed, although their results from FB hierarchical models were generally consistent with the NB estimates. This similarity in results from the NB and FB models has also been found in other studies (Quddus, 2008; Mitra and Washington, 2007). This is because uncorrelated heterogeneity accounts for most of the variation and the traditional NB models can sufficiently control for this effect. Another probable reason is that when there is sufficient data, the NB method works well and the results from the two methods are similar.

The idea that the values of parameters could arise from distributions is a fundamental feature of Bayesian methods. Bayesian hierarchical models can accommodate distributions such as the hierarchical Poisson–Gamma distribution and the Poisson–Lognormal distribution (Miaou and Lord, 2003; El-Basyouny and Sayed, 2011; Siddiqui et al., 2012; Yanmaz-Tuzel and Ozbay, 2010). Different prior distributions have been discussed by Yanmaz-Tuzel and Ozbay (2010) with their results suggesting that a Poisson–Lognormal model structure with more informative priors and higher levels of hierarchy may reduce the biases in modelling parameters, hence leading to more robust estimates. In this study, hierarchical Poisson–Lognormal models are adopted to compare with the traditional NB models.

Spatial models have been also applied to analyse the relationship between road casualties and road network characteristics in recent years. A study by Jones et al. (2008) used district-level data to investigate the effects of various factors on traffic casualties in England and Wales. The authors found that traffic casualties were significantly related to road length, curvature, junction density and other geographical variables. However, no positive spatial autocorrelation existed at the district level. The study by Quddus (2008), though did find a significant positive spatial correlation among ward-level road traffic crashes in Greater London from 2002 to 2004. Quddus applied both traditional and spatial models with the results from the traditional NB models and the Bayesian hierarchical models being very similar in suggesting that road traffic crashes are associated with the road infrastructure, socioeconomic and traffic conditions. Substantial positive spatial correlation was also found in the analysis of crash data for Florida's 67 counties from 2003 to 2007 (Huang et al., 2010). One reason for the diverse results of spatial correlation tests could be due to the different levels of spatial aggregation in the papers mentioned above.

In addition to spatial correlation, two issues evident in previous studies examining the relationship between the road network and road casualties remain to be adequately addressed. One critical issue is the selection of appropriate traffic exposure variables. In analysis at the unit-level, where the study object is usually road sections or intersections, the AADT or vehicle miles travelled (VMT) is preferred as the traffic exposure variable (Huang et al., 2010; Marshall and Garrick, 2011; Jones et al., 2008). These variables are not always available when the analysis is conducted at the area-level, however. Although proxy variables for traffic exposure have been developed (Graham and Glaister, 2003), there are certain limitations to their use. In this study, a new method to construct the traffic exposure variable is proposed.

The other issue concerns the usage of data for road network characteristics. Detailed data regarding the road network, including road class, road length and nodes information can be obtained from OS Meridian ™ 2. Although this data set has been used in several studies in the UK (Noland and Quddus, 2004; Haynes et al., 2007; Graham and Stephens, 2008; Jones et al., 2008), the data availability is only for a single year and, consequently, the variance in road network over time cannot be accounted for. To overcome this problem, OS Meridian ™ 2 for 2001 to 2010 (except for 2005) is employed in this study. The data set is discussed in the next section.

3. DATA

3.1. Dependent variable

The data used in this analysis includes road casualties recorded in England and Wales from 2001 to 2010. The accident data (i.e. road traffic crash) was collected from the STATS 19 data base and was further classified by severity type. Whilst it is recognized that inaccuracies exist in the STATS 19 data, for the purposes of the research outlined in this paper, it is adequate. The location of an accident was recorded using the British National Grid coordinate