Analytic Methods in Accident Research 18 (2018) 1–14

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

A Comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk

Amir Pooyan Afghari [a,*], Simon Washington [a], Md. Mazharul Haque [b], Zili Li [a]

[a] School of Civil Engineering, Faculty of Engineering, Architecture, and Information Technology, The University of Queensland, St Lucia QLD 4072 Australia
[b] School of Civil Engineering and Built Environment, Science and Engineering Faculty, Queensland University of Technology, Brisbane QLD 4001 Australia

ARTICLE INFO

Article history:
Received 30 May 2017
Received in revised form 23 March 2018
Accepted 23 March 2018

Keywords:
Crash causation mechanism
Data generating process
Random parameters model
Structural equation model
Instrumental variable
Joint model

ABSTRACT

In the safety literature, motor vehicle crashes are modelled predominately using single equation regression models, albeit with a variety of distributional assumptions and econometric enhancements. These models rely on a single linear additive predictive equation—which becomes multiplicative with a log transform—to specify the expected mean crash count conditioned on predictors. The models also specify the distribution of observations around the conditional mean, with common examples including the Poisson, Negative Binomial, and Conway-Maxwell distribution among others. This mainstream probabilistic conceptualization (i.e. model) of motor vehicle crash causation *assumes that* crashes are well-approximated by a single source of risk, wherein several contributing factors exert their collective, non-independent influences on the occurrence of crashes via a linear predictor.

This study first postulates, and then demonstrates empirically, that crash occurrence may be more complex than can be adequately captured by a single equation regression model. The total crash count recorded at a transport network location (e.g. road segment) may arise from multiple simultaneous and inter-dependent sources of risk, rather than one. Each of these sources may uniquely contribute to the total observed crash count. For instance, a site's crash occurrence may be dominated by contributions from driver behaviour issues (e.g. speeding, impaired driving), while another site's crashes might arise predominately from design and operational deficiencies such as deteriorating pavements and worn lane markings. Stated succinctly, this research hypothesises that the unobserved heterogeneity in the accumulation of motor vehicle crashes at transport network locations arises because multiple sources of risk, not one, better captures complexity in the crash occurrence process.

A stochastic multiple risk source methodological approach is developed to correspond with and empirically test this hypothesis. A joint econometric model with random parameters and instrumental variables demonstrates the applicability of the proposed theory and the corresponding methodological approach. The proposed model assumes that complexity of crash occurrence is well approximated using three sources of risk comprised of engineering, unobserved spatial, and driver behavioural factors. It is empirically tested using crash data from state controlled roads in Queensland, Australia. Finally, the multiple risk source model is compared to the traditional single risk source model to assess the viability of the proposed approach based on the sample data.

The multiple risk source model significantly outperformed the single risk source model in terms of prediction ability and goodness of fit measures. In addition, while the single risk source model predicts total crash counts for individual sites, the multiple risk source model

* Corresponding author.
E-mail addresses: a.afghari@uq.edu.au (A.P. Afghari), s.washington@uq.edu.au (S. Washington), m1.haque@qut.edu.au (M.M. Haque), zili.li@uq.edu.au (Z. Li).

https://doi.org/10.1016/j.amar.2018.03.002

predicts crash count proportions contributed by each source of risk, and predicts crashes by risk source. The improvement in fit combined with the theoretical appeal of a multiple risk source model to explain unobserved heterogeneity in crashes suggests—at least for the sample used in the study—that the complexity in crash occurrence is better explained using multiple equation linear predictors. Further research should examine other datasets for repeatability and should further explore and test risk sources.

1. Background

Borrowing from epidemiological methodologies (Rothman et al., 2008), crash risk at a site has traditionally been measured by counting the total number of crashes over a defined time period divided by the exposure to potential crashes (Hauer, 1986). It was soon recognized that the expected total crash count is conditional upon site specific characteristics such as daylight conditions or "physical separation of opposing streams" in addition to exposure (Smeed, 1949). These site-specific characteristics also known as contributing factors to crashes and their effects on safety have been extensively investigated in the literature (Lee and Mannering, 2002; Milton and Mannering, 1998; Mitra and Washington, 2007; Mitra and Washington, 2012; Oh et al., 2004; Poch and Mannering, 1996; Shankar et al., 1995) and can be categorized into multiple distinct sources such as exposure (e.g. traffic flow, entering vehicles, etc.), roadway geometric characteristics, spatial features of the surrounding environment, driver behavioural factors among many others.

Substantial effort is reflected in the literature to search for these sources and their contribution to explaining or predicting total crashes at a site. For example, driver behavioural factors are commonly accepted as the major cause of motor vehicle crashes—estimated to be responsible for between 50% and 90% of total crashes (Washington and Haque, 2013; Rumar, 1985; Sabey and Staughton, 1975). Distraction, fatigue and lack of compliance with driving regulations are behavioural factors frequently identified as contributing factors to crash occurrence. These factors, however, are not easily measured at crash sites and thus have not been directly incorporated into crash models.

Another source of causal factors includes roadway geometric and operational characteristics (also referred to as engineering factors). This source of contributing factors has been a major focus of the literature during the last four decades. Since 1980s, a plethora of studies have been conducted to investigate the effects of roadway geometric and traffic operational characteristics on total crash counts (Abdel-Aty and Radwan, 2000; Anastasopoulos and Mannering, 2009; Fitzpatrick et al., 2010; Geedipally et al., 2012; Joshua and Garber, 1990; Ladron de Guevara et al., 2004; Lee and Mannering, 2002; Malyshkina and Mannering, 2010; Miaou and Lum, 1993; Milton and Mannering, 1998; Mitra and Washington, 2007; Oh et al., 2003; Shankar et al., 1997; Zhu et al., 2010; Persaud et al., 2004; Ma et al., 2008; Poch and Mannering, 1996; Kim et al., 2006; Ye et al., 2009; Savolainen and Tarko, 2005; Schneider IV et al., 2010; El-Basyouny and Sayed, 2009; El-Basyouny and Sayed, 2006; Manuel et al., 2014; Islam et al., 2014; Mehta and Lou, 2013; Caliendo et al., 2007; Montella et al., 2008; Montella and Imbriani, 2015; Oh et al., 2006; Washington et al., 2014; Quddus et al., 2001; Chin and Quddus, 2003a; Chin and Quddus, 2003b; Oh et al., 2004; Mitra and Washington, 2012; Shankar et al., 1995). Factors revealed in these studies include segment length, number of lanes, posted speed limit, lane width, horizontal and vertical alignment of the segment, horizontal and vertical curvature, the presence of shoulder, shoulder width, presence of median, pavement condition, weather conditions, and so forth.

Another source of contributing factors to crash occurrence is spatial features of the surrounding environment including the effects of climate conditions (rainfall, snowfall, wind speed, sun glare, etc.), proximity of schools, bars, pubs and/or hospitals to the roadway segment (Mitra and Washington, 2012; Shankar et al., 1995; Yasmin and Eluru, 2016; Narayanamoorthy et al., 2013; Mitra, 2014; Chiou et al., 2014; Chiou and Fu, 2015; Chen, 2015). As these spatial factors are often unobserved, their effects on crashes cannot readily be estimated. According to the literature, spatial features may be responsible for between 5 and 10 percent of total number of crashes at a site (Mitra and Washington, 2012). The above-mentioned list may be extended to include other sources of risk such as vehicle characteristics and driver attributes.

The historical use of total crash counts as a suitable measure of crash risk at a site has resulted in the embodied assumption that contributing factors to crash occurrence influence the crash risk in a single, linear, additive, predictive function, with unexplained crash risk following a specific residual distribution (such as negative binomial). As shown in Fig. 1, the single risk source linear predictive assumption for crash causation might be depicted as the additive, cumulative effect of several factors in a single chain of events leading to a crash. This figure has been referred to as the Swiss cheese model of crash occurrence (Underwood and Waterson, 2014) depicting a hypothetical sequence of events in a single crash at a single site.

Although the Swiss cheese model has been applied for modelling risk processes in different fields of science including aviation safety (Shappell and Wiegmann, 2012), engineering (Underwood and Waterson, 2014), healthcare (Perneger, 2005), human behaviour (Reason, 1990) and ergonomics (Reason, 1995), it may not be sufficient to facilitate a deep understanding of crash causation particularly from the standpoint of temporally aggregated crash counts occurring on transport network locations. If crashes are aggregated over a period of time (e.g. 1 year), the Swiss Cheese model restricts the subset of causal factors to be the same for all observed crashes (as is done in current crash modelling). Assuming that N crashes occur