Accident Analysis and Prevention 113 (2018) 330–340

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Analysis of stationary and dynamic factors affecting highway accident occurrence: A dynamic correlated grouped random parameters binary logit approach

Grigorios Fountas[a], Md Tawfiq Sarwar[b], Panagiotis Ch. Anastasopoulos[c,*], Alan Blatt[d], Kevin Majka[d]

[a] Department of Civil, Structural, and Environmental Engineering, Engineering Statistics and Econometrics Application Research Laboratory, University at Buffalo, The State University of New York, 204 Ketter Hall, Buffalo, NY 14260, United States

[b] National Research Council, Federal Highway Administration, Turner-Fairbank Highway Research Center, 6300 Georgetown Pike, T-210 McLean, VA 22101, United States

[c] Department of Civil, Structural, and Environmental Engineering, Institute for Sustainable Transportation and Logistics, Engineering Statistics and Econometrics Application Research Laboratory, University at Buffalo, The State University of New York, 241 Ketter Hall Buffalo, NY 14260, United States

[d] Public Safety and Transportation Group; CUBRC, 4455 Genesee St., Suite 106, Buffalo, NY 14225, United States

ARTICLE INFO

Keywords:
Accident occurrence
Binary mixed logit
Correlated grouped random parameters
Uncorrelated random parameters
Dynamic and stationary characteristics
Time-varying parameters

ABSTRACT

Traditional accident analysis typically explores non-time-varying (stationary) factors that affect accident occurrence on roadway segments. However, the impact of time-varying (dynamic) factors is not thoroughly investigated. This paper seeks to simultaneously identify pre-crash stationary and dynamic factors of accident occurrence, while accounting for unobserved heterogeneity. Using highly disaggregate information for the potential dynamic factors, and aggregate data for the traditional stationary elements, a dynamic binary random parameters (mixed) logit framework is employed. With this approach, the dynamic nature of weather-related, and driving- and pavement-condition information is jointly investigated with traditional roadway geometric and traffic characteristics. To additionally account for the combined effect of the dynamic and stationary factors on the accident occurrence, the developed random parameters logit framework allows for possible correlations among the random parameters. The analysis is based on crash and non-crash observations between 2011 and 2013, drawn from urban and rural highway segments in the state of Washington. The findings show that the proposed methodological framework can account for both stationary and dynamic factors affecting accident occurrence probabilities, for panel effects, for unobserved heterogeneity through the use of random parameters, and for possible correlation among the latter. The comparative evaluation among the correlated grouped random parameters, the uncorrelated random parameters logit models, and their fixed parameters logit counterpart, demonstrate the potential of the random parameters modeling, in general, and the benefits of the correlated grouped random parameters approach, specifically, in terms of statistical fit and explanatory power.

1. Introduction

Traditional accident analysis typically focuses on exploring and quantifying the impact of influential factors to crash-related implications (accident occurrence, frequency, rate, injury-severity, etc.). A better understanding of the mechanisms involved in accident occurrence can possibly be achieved through the investigation of factors whose effect varies over time or space. In this paper, a fundamental distinction – based on the different nature of the explanatory factors – is made between stationary and dynamic factors affecting accident occurrence. Considering the physical characteristics of a roadway segment, stationary factors remain constant over time and space, and are associated with design or functional elements of the roadway. Dynamic elements on the other hand, can vary over time or space, and include factors that cannot be known *a priori*, such as weather or environmental conditions (Kepaptsoglou et al., 2015). The rigorous investigation of the stationary characteristics in the literature, has provided significant insights pertaining to ongoing advances in roadway design and safety policy. However, the effect of dynamic factors in accident prediction has not been thoroughly investigated. In fact, not controlling for

* Corresponding author.
E-mail addresses: gfountas@buffalo.edu (G. Fountas), mdtawfiq@buffalo.edu (M.T. Sarwar), panastas@buffalo.edu, paxrana79@gmail.com (P.C. Anastasopoulos), blatt@cubrc.org (A. Blatt), majka@cubrc.org (K. Majka).

http://dx.doi.org/10.1016/j.aap.2017.05.018
Received 20 February 2017; Received in revised form 26 April 2017; Accepted 21 May 2017
0001-4575/ Published by Elsevier Ltd.

dynamic factors in accident analysis can result in poor model explanatory power and low-accuracy forecasts. Interestingly, a fair amount of previous work focused on the impact of dynamic factors on accident occurrence, which was, however, characterized by significant dynamic data aggregation (to an extent that cannot reflect the variation and intensity of the related phenomena over their entire range, as disaggregate data would allow for).

A review of past research reveals that the effect of the prominent stationary elements on accident occurrence (Ahmed et al., 2012), frequency (Shankar et al., 1995; Abdel-Aty and Radwan, 2000; Shankar et al., 2002; Anastasopoulos and Mannering, 2009; Abdel-Aty et al., 2009; Wu et al., 2014; Lee et al., 2015), rates (Zhou and Sisiopiku, 1997; Davis, 2000; Hauer, 2001; Hess and Polak, 2003; Anastasopoulos et al., 2008; Anastasopoulos et al., 2012a; Anastasopoulos et al., 2012b; Yu et al., 2013; Anastasopoulos, 2016; Zeng et al., 2017), or injury-severities (Abdel-Aty, 2003; Abdel-Aty and Keller, 2005; Milton et al., 2008 Savolainen and Ghosh, 2008; Rana et al., 2010; Anastasopoulos et al., 2011; Abdel-Aty et al., 2011; Yu and Abdel-Aty, 2014b; Sarwar and Anastasopoulos, 2016; Yasmin et al., 2016; Sarwar and Anastasopoulos, 2017; Fountas and Anastasopoulos, 2017), such as roadway geometrics and aggregate traffic characteristics, has been explored extensively (Hadi et al., 1995; Ivan and O'mara, 1997; Milton and Mannering, 1998; Karlaftis and Golias, 2002; Lee and Mannering, 2002; Anastasopoulos and Mannering, 2009; Anastasopoulos et al., 2011; Park and Abdel-Aty, 2015, 2016; Shi et al., 2016). These studies have shown that an abundance of roadway geometrics (such as number of lanes, horizontal and vertical curve elements, median and shoulder information, cross-section specific characteristics, etc.), traffic characteristics (traffic volume, truck traffic, traffic composition, etc.), and roadway functional characteristics (access control, speed limit, roadway functional class, etc.), constitute influential factors in accident analysis (see Lord and Mannering, 2010, and Mannering and Bhat, 2014, for a comprehensive overview of the methodological approaches and relevant factors).

Furthermore, a significant portion of previous work has been devoted on the simultaneous investigation of stationary and dynamic characteristics, in estimating accident prediction models. In these studies, dynamic elements include weather-related and pavement surface characteristics, whereas stationary elements include roadway geometrics and traffic conditions. For example, a number of studies explored the combined effect of roadway/pavement/traffic characteristics and weather-related (precipitation, and snow depth) factors (see Theofilatos and Yannis, 2014 for an extensive review of such studies), by estimating count data models of accident frequencies (Fridstrøm et al., 1995; Levine et al., 1995; Shankar et al., 1995; Andreescu and Frost, 1998; Shankar et al., 2004; Eisenberg, 2004; El-Basyouny and Sayed, 2006; Abdel-Aty and Pemmanaboina, 2006; Caliendo et al., 2007; Brijs et al., 2008; Yannis and Karlaftis, 2010; Wang et al., 2015; Wang and Abdel-Aty, 2016), or discrete outcome models of accident injury-severity (Yu and Abdel-Aty, 2014a; Aron et al., 2015). However, these studies did not account either for the effect of dynamic (and possibly rapid) variations in weather-related and environmental characteristics, or for the joint effect of the dynamic and stationary characteristics on the accidence occurrence mechanism. Interestingly, in the vast majority of these studies, the effect of weather-related (temperature, precipitation, wind) and environmental (lighting conditions, surface visibility) factors on the accident occurrence mechanism is investigated on the basis of aggregate information corresponding either at the time of the accident, or at a pre-specified extensive time period within the time of the accident. Such considerations cannot account for possible fluctuations of time variant characteristics, leading to significant loss of information, which is critical for the identification of pre-crash dynamic factors. Even though the dynamic nature of such factors is captured through real-time detailed data, the previously employed modeling schemes cannot adequately control for possible unobserved heterogeneity interactions between the stationary and the dynamic characteristics.

As opposed to previous studies that explored stationary and dynamic factors (through aggregated data) on accident frequencies, the intent of this paper is to investigate the simultaneous impact of stationary and pre-crash dynamic factors that may lead to an accident on a specific highway segment, by using highly disaggregate information (which allows capturing the effect of detailed time-varying dynamic weather and environmental characteristics). To account for the dynamic factors and the underlying unobserved heterogeneity, a dynamic random parameters (mixed) logit framework is employed, using data collected from Washington's highway segments with and without accidents, between 2011 and 2013.

A fundamental limitation of the conventional random parameters logit approach arises from the restrictive formulation of the covariance matrix for the random parameters, which does not allow for possible correlations among the explanatory variables to be reflected in the parameter estimates (Greene, 2012; Yu et al., 2015). Not accounting for correlation effects among the dynamic and stationary factors may result in misspecification issues (e.g., biased parameter estimates, inconsistent predictions, and erroneous inferences), because the underlying unobserved heterogeneity will only be partially captured (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009; Mitra and Washington, 2012; Coruh et al., 2015; Sarwar et al., 2016; Anastasopoulos and Mannering, 2016; Anastasopoulos et al., 2016, 2017). To address this limitation and to account for panel effects, a correlated grouped random parameters binary logit model with dynamic and stationary explanatory parameters is estimated – for the first time, to the authors' knowledge – and the results are compared with the fixed and uncorrelated random parameters modeling counterparts, in terms of statistical fit, explanatory power, and forecasting accuracy.

2. Methodological approach

To simultaneously account for stationary factors, for the dynamic nature of the time-varying explanatory parameters (i.e., pre-crash dynamic factors), and for the possibility of systematic variations of the parameter effects across the highway segments (i.e., unobserved heterogeneity), a dynamic binary random parameters (mixed) logit model is employed. Not addressing these modeling specification issues is important, as they can cause significant model misspecification, in terms of inconsistent parameter estimates and outcome probabilities (Washington et al., 2011). To that end, the accident occurrence function, A_{in}, that determines whether an accident, i, occurred or not on a highway segment, n, can be written as:

$$A_{in} = \boldsymbol{\beta}_i \mathbf{X}_n + \boldsymbol{\beta}_i \boldsymbol{D}_{t,n} + \varepsilon_{in} \tag{1}$$

where, $\mathbf{X}_n$ are vectors of the observable stationary (stable over time for the same highway segment, but varying across highway segments) characteristics that determine the accident occurrence, $\boldsymbol{D}_{t,n}$ are vectors of the observable dynamic (variable over time for the same highway segment, and varying across highway segments) characteristics in a specific time interval t that determine the accident occurrence, $\boldsymbol{\beta}$ are vectors of estimable parameters for discrete outcome, i, and ε_{in} is the disturbance term.

The introduction of random parameters in the accident occurrence function allows the estimation of a separate vector of **βs** for each observation, as (Greene, 2012):

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \Gamma \delta_i \tag{2}$$

where, $\boldsymbol{\beta}$ denotes the mean value of the random parameters vector, Γ is a symmetric matrix (also referred to as Cholesky matrix; for further details see: Greene, 2012) whose elements are used for the computation of the standard deviations of the random parameters, and δ denotes a randomly distributed term with mean equal to zero and variance equal to one. In an effort to examine different distributional assumptions with respect to the disturbance term, both probit and logit model