Accident Analysis and Prevention 73 (2014) 323–332

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Refined-scale panel data crash rate analysis using random-effects tobit model

Feng Chen [a,1], XiaoXiang Ma [b], Suren Chen [c,*]

[a] Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800Cao'an Road, Shanghai 201804, China
[b] Graduate Research Assistant, Department of Civil & Environmental Engineering Colorado State University, Fort Collins, CO 80523, United States
[c] Department of Civil & Environmental Engineering, Colorado State University, Fort Collins, CO 80523, United States

ARTICLE INFO

Article history:
Received 17 April 2014
Received in revised form 18 August 2014
Accepted 20 September 2014
Available online xxx

Keywords:
Crash rate
Panel data
Refined-scale data
Random-effects tobit
Disaggregate data model
Daytime and nighttime

ABSTRACT

Random effects tobit models are developed in predicting hourly crash rates with refined-scale panel data structure in both temporal and spatial domains. The proposed models address left-censoring effects of crash rates data while accounting for unobserved heterogeneity across groups and serial correlations within group in the meantime. The utilization of panel data in both refined temporal and spatial scales (hourly record and 1-mile roadway segments on average) exhibits strong potential on capturing the nature of time-varying and spatially varying contributing variables that is usually ignored in traditional aggregated traffic accident modeling. 1-year accident data and detailed traffic, environment, road geometry and surface condition data from a segment of I-25 in Colorado are adopted to demonstrate the proposed methodology. To better understand significantly different characteristics of crashes, two separate models, one for daytime and another for nighttime, have been developed. The results show major difference in contributing factors towards crash rate between daytime and nighttime models, implying considerable needs to investigate daytime and nighttime crashes separately using refined-scale data. After the models are developed, a comprehensive review of various contributing factors is made, followed by discussions on some interesting findings.

Published by Elsevier Ltd.

1. Introduction

Traffic accident modeling has long been an important research area for safety researchers and a lot of research efforts have been focused on crash (accident) frequency. In addition to crash frequency, crash rate is another important highway safety indicator with some unique appeals. For example, the adoption of crash rate offers standardized traffic safety measure which can more conveniently assess the relative risks among different road segments (Anastasopoulos et al., 2008). Several major challenges about current crash frequency models have been identified in the comprehensive review study conducted by Lord and Mannering (2010). Among all those, some challenges are also shared by crash rate models, such as time-varying explanatory variables, temporal and spatial correlations.

Crash frequency or crash rate analyses were typically conducted in large scales in both temporal (e.g. yearly) and spatial (e.g. whole road) domains. It is known that the adoption of aggregated explanatory variables in larger scales ignores the within-period variation of explanatory variables, which will result in "the loss of potentially important explanatory information" (Lord and Mannering, 2010) and introduce error in model estimation due to unobserved heterogeneity (Washington et al., 2010). To develop crash risk models in refined scales requires not only the availability of the disaggregated data, but also overcoming some technical challenges, such as correlations by sharing unobserved effects among multiple observations generated from the same road segments and/or time period (Lord and Mannering, 2010; Lord and Persaud, 2000; Shankar et al., 1998; Sittikariya and Shankar, 2009; Ulfarsson and Shankar, 2003). More detailed review of these challenges will be made in Section 2.

As compared to normal driving conditions, refined-scale traffic safety modeling is more critical in adverse driving conditions, under which the explanatory variables often change considerably over different time instants and locations, such as weather, traffic and road surface conditions. As a result, both time-varying and

* Corresponding author. Tel.: +1 970 491 7722; fax: +1 970 491 7727.
E-mail addresses: chenfeng.csu@gmail.com (F. Chen), suren.chen@colostate.edu (S. Chen).
[1] Tel.: +86 021 6958 5721.

http://dx.doi.org/10.1016/j.aap.2014.09.025
0001-4575/Published by Elsevier Ltd.

spatially varying (cross-sectional) data needs to be considered in smaller scales, which make the temporal and spatial correlations even more complex. Panel data with random effects has been used in recent years to deal with the temporal or spatial correlation issues (Qi et al., 2007; Shankar et al., 1998; Ulfarsson and Shankar, 2003). Despite all the progress on advanced statistical models of adopting panel data, most studies were primarily on crash frequency predictions. Crash rate modeling with refined scales, with or without panel data application, has been rarely reported.

In contrast to a large number of literatures about different count-data models for crash frequency prediction (Lord and Mannering, 2010), the studies on crash rate are limited. From the statistical perspective, crash rate modeling involves continuous data, which is usually left-censored at zero. This is because roadway segments without any accident reported over a specified time period will simply yield zero on accident rate data record. In order to handle the censoring problem of crash rate prediction, tobit model was successfully employed by Anastasopoulos et al. (2008) on the highway accident rate data in the state of Indiana. To handle unobserved heterogeneity across observations, Anastasopoulos et al. (2012a) further developed a random-parameters tobit model, which demonstrated superiority over its fixed parameter counterpart in terms of goodness of fit. Xu et al. (2013b) used tobit regression to investigate the endogeneity problem between crash rate and travel speed. Other studies also include the multivariate tobit analysis on crash rate for each injury type (Anastasopoulos et al., 2012b).

Most of existing models on crash frequency or crash rate were developed by combining both daytime and nighttime data as a whole, based on the assumption that traffic safety during daytime and nighttime share same contributing factors and characteristics. It is known that driving environments at daytime and nighttime are actually very different due to varying light, environmental and traffic conditions affecting driver behavior and, eventually traffic safety risks. In recent years, there have emerged a few studies modeling crash frequency for daytime and nighttime separately (Bullough et al., 2013; Dinu and Veeraragavan, 2011; Donnell et al., 2010), which have shown considerable difference in contributing factors towards crash frequency between daytime and nighttime. So far, however, there is no such study reported on crash rate and it remains unclear whether separate modeling of crash rate for daytime or nighttime is needed or not.

The present study aims at conducting crash rate study with refined-scale data in both temporal and spatial domains to identify the relation between crash rate and its contributing factors. The study integrates the strength from both the random effect tobit model and the panel data formulation of detailed driving conditions such as geometry characteristics, traffic flow data, weather conditions, and road surface conditions. As a result, the outlined challenges associated with censoring problem of crash rate, aggregated explanatory variables, temporal and spatial correlation can be appropriately addressed (Lord and Mannering, 2010). To capture different traffic characteristics, two separate models for daytime crashes and nighttime crashes are developed using refined-scale data. Some brief reviews of the topics related to the present study are made in the following section.

2. Literature review

2.1. Studies with refined-scale accident and contributing factors

In most existing studies, aggregated accident data and associated contributing factors were often considered in large time intervals and spatial domains, partially due to the unavailability of detailed data. The contributing factors are normally aggregated into monthly or even annual intervals and over region-wide or long roadway segments. In adverse driving conditions, detailed weather and other environment-related factors (e.g. precipitation, visibility, wind, humidity, pavement characteristics, road surface condition etc.) are often found critical in causing a crash. As a result of adopting aggregating data, some potential explanatory information may be lost and some factors' effects towards crash occurrence can be masked during data aggregation (Lord and Mannering, 2010; Usman et al., 2011). Recently, researchers have started to incorporate more influencing factors in refined scales. Keay and Simmonds (2005) examined the relation between rainfall and daily collision data in Australia. Usman et al. (2010) are among the first to build disaggregated hourly model to study crash frequency incorporating weather and surface data. It was found hourly traffic volume, hourly road surface condition and other hourly weather factors are important for collisions during winter storms. Abdel-Aty et al. (2004) analyzed the effects of traffic characteristics towards freeway crashes using real-time loop detector data. Abdel-Aty and Pemmanaboina (2006) combined the real-time ITS traffic data, the archived weather data and the historical accident data to calibrate the crash prediction model. Other research attempts of building disaggregated data model by incorporating traffic flow, weather data, surface condition and other factors can be found in the studies by Hossain and Muromachi (2013), Usman et al. (2012) and Yu et al. (2013a,b). In contrast to the progress on crash frequency modeling as summarized above, crash rate modeling in refined scales has been rarely reported.

2.2. Panel data models

With refined data in both spatial and temporal domains, serious correlation problems may exist among the records. Panel data (also known as cross-sectional time-series data) models, which have been widely used in econometric, social and behavioral analysis, gained its popularity among safety specialists due to its capacity to address both time-series and cross-sectional variations. In crash modeling arena, panel data models can be mainly generalized into three categories: fixed effects (FE) models, random effects (RE) models, and random parameters (RP) models. A FE model has some appeals in panel data modeling because of not requiring the assumption about unobserved heterogeneity (Greene, 2008). Nonetheless, FE models require estimations of a large number of parameters (e.g. site-specific or time-specific indicator variables), which may dominate the contributing factors. As compared to FE model, RE model is usually more popular in modeling crash data with repeated observations (Aguero-Valverde, 2013; Chin and Quddus, 2003; Qi et al., 2007; Shankar et al., 1998). Recently, random-parameter (RP) model was proposed as a general extension of RE model to allow not only the constant term but also the coefficients to vary across observations (Anastasopoulos and Mannering, 2009; Lord and Mannering, 2010). In addition to these three most frequently used types of panel data models, there were also some other research attempts to account for temporal and spatial correlations in accident occurrence, such as negative multinomial model (Caliendo et al., 2013; Ulfarsson and Shankar, 2003), generalized estimating equation (GEE) model (Lord and Persaud, 2000; Wang and Abdel-Aty, 2006), and time-series model (Quddus, 2008). However, most of these studies focus on crash frequency prediction by applying panel data with repeated yearly observations, rather than data in more refined scales.

3. Methodology

In the following, the base tobit model proposed by Tobin (1958) will be adopted as the starting point to study the left-censored crash rate data. Under panel data formation, repeated observations are given for each group (e.g. roadway segments or intersections).

Download English Version:

https://daneshyari.com/en/article/6965903

Download Persian Version:

https://daneshyari.com/article/6965903

Daneshyari.com