Analytic Methods in Accident Research 7 (2015) 16–28

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Simultaneous equation modeling of freeway accident duration and lanes blocked

Yi-Shih Chung [a], Yu-Chiun Chiou [b,*], Chia-Hua Lin [b]

[a] Department of Logistics and Shipping Management, Kainan University, No. 1, Kainan Rd., Luzhu Shiang, Taoyuan 33857, Taiwan
[b] Department of Transportation and Logistics Management, National Chiao Tung University, 4F, 118, Sec. 1, Chung-Hsiao W. Rd., Taipei 10012, Taiwan

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form
27 April 2015
Accepted 27 April 2015
Available online 9 June 2015

Keywords:
Simultaneous equation modeling
Accident duration
Number of lanes blocked
Traffic accidents

ABSTRACT

Unbiased estimates are crucial to providing correct information that is required in accident management. Whereas earlier studies have proposed various approaches to reduce the adverse effects of unobserved heterogeneity, this study proposes accelerated failure time (AFT) models of accident duration with flexible distributions of the error term and the frailty parameter. A two-stage residual inclusion (2SRI) method for identifying the effects of endogenous variable in the nonlinear AFT models is adopted. The Taiwanese freeway accident databases were used to evaluate the applicability of the proposed approach. The results reveal that AFT models with a non-monotone Log-logistic error term distribution and a gamma frailty parameter distribution outperformed those with other assumed distributions. Number of lanes blocked was found as a significant endogenous variable to accident duration; the endogenous variable effect, if not accounted for or accounted by traditional two-stage least squares estimators, would underestimate the effect of the number of lanes blocked on accident duration.

1. Introduction

Traffic accidents cause not only fatalities and injuries, but also traffic congestion. According to FHWA (2014), non-recurrent incidents, such as stalled vehicles, highway debris, and crashes, are responsible for approximately 60% of the delays that are caused by traffic congestion. Efficiently clearing traffic incidents is thus critical to reducing traffic congestion.

The duration of accidents and the number of lanes blocked are two factors that primarily determine the degree of congestion due to traffic accidents. Accident duration is normally defined as the period from the occurrence of an accident to its clearance (Nam and Mannering, 2000) or to subsequent traffic flow recovery (Chung, 2010). The number of lanes blocked has been found to correlate significantly with traffic capacity reduction, and this effect may vary with the geometry of the road and the lane(s) that were blocked (Smith et al., 2003). Generally, accidents with longer durations or more blocked lanes cause heavier congestion. Therefore, identifying the factors that drive accident duration and the number of blocked lanes is essential to developing countermeasures to reduce accident-induced traffic congestion.

Studies of the factors that are associated with accident duration have been conducted in various areas (Nam and Mannering, 2000; Chung, 2010; Chung et al., 2010; Al Kaabi et al., 2011, 2012; Tavassoli Hojati et al., 2012; Junhua et al.,

* Corresponding author. Tel.: +886 2 23494940.
E-mail address: ycchiou@mail.nctu.edu.tw (Y.-C. Chiou).

http://dx.doi.org/10.1016/j.amar.2015.04.003
2213-6657/

2013); different methods for so doing have been used, including linear regression analyses, non-parametric regression methods, conditional probability analyses, discrete choice models, and artificial neural networks (Tavassoli Hojati et al., 2012). Recently, survival models, especially accelerated failure time (AFT) models, have become popular in accident duration studies, as they support the explicit study of the relationship between the duration of an accident and its instantaneous probability of ending, and the ways in which various factors affect this relationship (Nam and Mannering, 2000). However, only a few studies have examined the number of lanes blocked by traffic accidents. One of the few examples is that of Qi et al. (2009), who developed ordered probit models to determine the number of lanes blocked by traffic accidents based on the empirical data in New York City.

The AFT and regression models that are increasingly used in studies of accident duration and numbers of lanes blocked reveal the importance of understanding the drivers of these two variables, with reference to the direction, magnitude, and significance of their effects on the efficient clearing of accidents. However, this task is not easy. Previous studies have demonstrated that many factors may affect accident duration and the number of lanes blocked; the omission of important variables, called unobserved heterogeneity, in regression models yields biased estimates and, therefore, erroneous information. Prior accident duration studies have attempted to capture unobserved heterogeneity by incorporating additional parameters (such as the frailty parameter) in the model (Al Kaabi et al., 2012). Yet, the incorporation of a number of explanatory variables has not eliminated unobserved heterogeneity.

A closely related issue is endogeneity, which refers to the existence of a correlation between an explanatory variable and the regression error term, that is, the explanatory variable is partly determined by factors within the regression model itself. Some studies of accident duration have shown that the number of lanes blocked is a significant explanatory variable (Garib et al., 1997; Nam and Mannering, 2000; Qi and Teng, 2008; Qi et al., 2009; Junhua et al., 2013) and is itself partially determined by the variance of accident duration (Qi et al., 2009). These findings suggest the possibility that the number of lanes blocked is an endogenous variable in determining accident duration. Restated, the estimates of accident duration models are biased if the endogeneity of the number of lanes blocked is not appropriately addressed.

This study develops simultaneous equation-based models that effectively address both unobserved heterogeneity and endogeneity. The two-stage residual inclusion (2SRI) method, proposed by Terza et al. (2008), is adopted. Unlike conventional instrumental variable (IV) methods, such as two-stage least squares (2SLS), 2SRI produces unbiased estimates by combining nonlinear models, and so is suitable for use herein. The unobserved heterogeneity of accident duration and the number of lanes blocked are modeled using more flexible functional forms than they have been in earlier studies.

The rest of this paper is organized as follows. Section 2 presents a review of the relevant literature, elucidating the details of the methodologies in earlier studies. Section 3 demonstrates the adopted 2SRI method and the proposed models in each individual stage. Section 4 presents the empirical data and variables. Section 5 provides analytical results, which are discussed in Section 6. Section 7 elucidates the limitations of this study and makes suggestions for future research.

2. Earlier research

2.1. Accident duration

The duration of an accident/incident can be roughly divided into four parts—detection/reporting time (from occurrence of incident to notification about incident), preparation time (from notification about incident to dispatch of the IRT (incident response team)), travel time (from the dispatch of the IRT to its arrival), and clearance time (from arrival of the IRT to the cleaning of the incident) (Nam and Mannering, 2000). Some studies have investigated the determinants of the duration of each phase, while others have examined factors that may affect the duration of all phases. For example, Nam and Mannering (2000) adopted a two-year dataset from Washington State's IRT program, and developed models for the duration of each phase. Al Kaabi et al. (2011) used a one-year dataset from Abu Dhabi's federal traffic statistics and collision reports to develop models specifically for the period of accident clearance; in a follow-up study, Al Kaabi et al. (2012) studied the response time on Abu Dhabi's highways. Chung (2010) and Chung et al. (2010) defined accident duration as the time between notification about an accident to its clearance, and developed models of the factors that significantly affect this duration. Studies of individual phases demonstrate that different factors may determine the durations of the different phases (Nam and Mannering, 2000), and countermeasures specific to each phase may be provided by such studies. However, traffic engineers or traffic control centers are typically concerned with the total duration of an accident and studies of “joint phases” of an accident are useful to them.

Like definitions of accident duration, methods for investigating accident duration vary. Numerous earlier studies have used regression and survival analysis (or hazard-based duration models). Traditional regression methods can be found in early studies. For example, Garib et al. (1997) used independent variables that are associated with accident characteristics, traffic characteristics, weather conditions, and geometric characteristics. Khattak et al. (2009) used an ordinary least squares regression model to investigate accident duration in the Hampton Roads area in Virginia.

Survival analysis has become increasingly popular in recent accident duration studies. Nam and Mannering (2000) thoroughly introduced the use of hazard-based duration models in research into incident duration, with various hazard distributions and a means of dealing with unobserved heterogeneity. The primary difference between the use of hazard-based models of duration and traditional regression methods, according to Nam and Mannering (2000), is that the former include duration-dependence, which is the dependence of the probability that an incident will continue on the duration of