Accident Analysis and Prevention 134 (2020) 105324

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

The role of striking and struck vehicles in side crashes between vehicles: Bayesian bivariate probit analysis in China

Quan Yuan[a,b], Xuecai Xu[c,*], Mingchang Xu[a], Junwei Zhao[d], Yibing Li[a]

[a] State Key Laboratory of Automotive Safety and Energy, School of Vehicle and Mobility, Tsinghua University, Beijing, China
[b] Center for Intelligent Connected Vehicles and Transportation, Tsinghua University, Beijing, China
[c] School of Civil Engineering and Mechanics, Huazhong University of Science and Technology, Wuhan, China
[d] School of Automobile, Chang'an University, Xi'an, China

ARTICLE INFO

Keywords:
Side crashes
Striking vehicles
Struck vehicles
Bayesian bivariate probit regression model
Injury severity

ABSTRACT

Objective: Side crashes between vehicles which usually lead to high casualties and property loss, rank first among total crashes in China. This paper aims to identify the factors associated with injury severity of side crashes at intersections and to provide suggestions for developing countermeasures to mitigate the levels of injuries.
Method: In order to investigate the role of striking and struck vehicles in side crashes simultaneously, bivariate probit model was proposed and Bayesian approach was employed to evaluate the model, compared to the corresponding univariate probit model.
Data: Crash data from Beijing, China for the period 2009–2012 were used to carry out the statistical analysis. Based on the investigation with vehicles and data analysis on events, 130 intersection side crash cases were selected to form a specific dataset. Then, the influence of human, vehicles, roadway and environmental variables on crash severity was examined by means of bivariate probit regression within Bayesian framework.
Results: The effects of the factors on striking vehicle drivers and struck vehicle drivers were considered separately and simultaneously to find more targeted conclusions. The statistical analysis revealed vehicle type, lane number, no non-motorized lane and speeding have the corresponding influence on the injury severity of striking vehicles, while time of day and vehicle type of struck vehicles increased the likelihood of being injured.
Conclusions: From the results it can be concluded that there indeed exists correlation between striking and struck vehicles in side crashes, although the correlation is not so strong. Importantly, Bayesian bivariate probit model can address the role of striking and struck vehicles in side crashes simultaneously and can accommodate the correlation clearly, which extends the range of univariate probit analysis. The general and empirical countermeasures are presented to improve the safety at intersections.

1. Introduction

According to World Health Organization (WHO) statistics, about 1.24 million people die in traffic crashes each year all over the world, in which frontal, side and rear-end accounts for about half of all types, while this proportion is as high as two thirds in some countries or regions. Among the top three types, side crashes explain one third. As reported from the U.S. Department of Transportation in 2017, In terms of severe injuries and fatalities occurring on U.S. roads, side crashes represented about 28% of all fatalities for passenger vehicle occupants, coming second only to frontal impacts. Also side crashes accounted for about 25% of all fatal and injury crashes in British Columbia, a Canadian province, 36% of all fatalities in Ontario, and 40% of crashes with serious injury in Australia (Laberge-Nadeau et al., 2009). In China side crashes between vehicles accounted for about 41% of all crashes occurred in 2014, ranking first among all kinds of crash types (Traffic Management Bureau of the Ministry of Public Security of China (TMBMPSC, 2015). Hence, side crashes between vehicles require particular investigation so as to reduce the injury severity and improve the roadway safety.

Currently, there have been a variety of studies focusing on side-related crashes (Viano and Parenteau, 2010; Riley et al., 2012; Golman et al., 2014; Gierczycka and Cronin, 2017; Kelly et al., 2019), and some concentrated on near-side impact (Kapoor et al., 2008), while others focused on the far-side impact (Viano and Parenteau, 2010). From the aspect of side crashes, the rare study by Brumbelow (2012) has

* Corresponding author.
E-mail address: xuecai_xu@hust.edu.cn (X. Xu).

https://doi.org/10.1016/j.aap.2019.105324
Received 7 June 2019; Received in revised form 25 September 2019; Accepted 7 October 2019

specified the potential benefits of underride guards in large truck. So far, most of them have explored side-related crashes from the perspectives of human body or occupants, which obviously makes sense since lives are the most important among the injuries., the studies from the role of striking and struck vehicles are sparse, thus it is challenging to identify the role of striking vehicles and struck vehicles to avoid the injury.

There were many statistical methods employed to examine injury severity (Savolainen et al., 2011; Mannering and Bhat, 2014), and regression analysis has been widely applied to investigate the relationship between crash injury severity and influencing factors. The widely utilized regression approaches, e.g. linear regression, logistic regression and probit regression, assume that all sampling data have the uniform relationship with the influencing factors, which may not be always true, especially in multivariate systems (Zeng et al., 2018; Liu et al., 2018). Therefore, how to release the assumption and address the correlation between multivariate systems have been explored for years.

Bivariate probit analysis, as one special case of multivariate systems, has been applied in various fields, e.g. agriculture (Nkamleu and Adesina, 2000; Torres et al., 2017), social service (Grogan-Kaylor, 2001), economics (Fleming and Kler, 2008; Yue and Zou, 2014), transportation (Lemp et al., 2012; Xu et al., 2014; Guo et al., 2017), tourism (Masiero and Zoltan, 2013), and management science (Yildirim and Dal, 2016). In transportation safety area, bivariate probit analysis provides an alternative to accommodate the correlation between different injury severity levels (Ma and Kockelman, 2006; Xu et al., 2014).

The application of Bayesian approach in transportation safety area has been long-standing from various perspectives. Bayesian approaches, such as Bayesian binary response model (Qin et al., 2006; Yu and Abdel-Aty, 2014; Xu et al., 2016), Bayesian binomial logistic model (Huang et al., 2008; Chen et al., 2016; Yu et al., 2016; Guo et al., 2018a), Bayesian panel data models (Flask and Schneider, 2013; Flask et al., 2014; Yu et al., 2016; Guo et al., 2018b), etc. have demonstrated that this method is very competitive in analyzing crash frequency/severity involving correlation or heterogeneity issue. Bayesian multivariate regression models, such as Bayesian multivariate conditional autoregressive model (Song et al., 2006; Heydari et al;, 2016; Cheng et al., 2017; Zeng et al., 2019), multivariate Poisson-lognormal regression model (Ma and Kockelman, 2006; Ma et al., 2008; Cena et al., 2011; El-Basyouny et al., 2014a, 2014b; Shaheed et al;, 2016; Serhiyenko et al., 2016; Murphy and Xia, 2016), Bayesian spatial model (Xu et al., 2016; Zeng et al., 2017, 2018; Liu and Sharma, 2018; Alarifi et al., 2018; Guo et al., 2018c, 2019), etc., have verified that crashes/crash injury severity can be predicted from different types involving various factors; All of these studies have revealed the potentiality of Bayesian approach for crash frequency and severity analysis.

Side crashes between vehicles at intersection account for a large proportion in total crashes of China, whereas there have been limited amount of studies concentrating on this crash type, and few studies concerned contributing factors involved in these crashes, especially for striking and struck vehicles. To take advantage of the Bayesian approach and multivariate systems, a Bayesian bivariate probit model was proposed by using accident data from 130 side crashes between vehicles at intersection in Beijing to identify the factors that were related to this crash type. The results aim to reveal how different factors affected the severity of accidents and also provide some suggestions for developing countermeasures to mitigate the levels of injuries.

2. Method

When crash happens, the striking and struck vehicles are not completely independent, i.e. they may be correlative to each other. If there are only striking vehicles without struck vehicles, the crash may not occur, while struck vehicles may not exist without striking vehicles. Therefore, from the empirical point of view, two univariate probit models are not appropriate to describe the two types of vehicles, because the classical single equation probit model does not take into account that correlation between dependent variables. To solve this problem, the bivariate probit model emerged, which can consider the correlation between error terms and accommodate the roles of striking and struck vehicles expressed by the injury severity levels correlated. The structure of the bivariate probit model can be described as the followings:

$$y_1^* = x_1^{'}\beta_1 + \varepsilon_1,\ \mathrm{y}_1 = 1\,\mathrm{if}\,\mathrm{y}_1^* > 0,\ 0\,otherwise,$$
$$y_2^* = x_2^{'}\beta_2 + \varepsilon_2,\ \mathrm{y}_2 = 1\,\mathrm{if}\,\mathrm{y}_2^* > 0,\ 0\,otherwise, \tag{1}$$

where y_1 and y_2 represent the injury severity levels of striking and struck vehicles, respectively, $x_1^{'}$ and $x_2^{'}$ denote the influencing factors of two types of vehicles, β_1 and β_2 are the coefficients, and ε_1 and ε_2 are the error terms, the correlation between two equations: $\mathrm{Cov} = [\varepsilon_1,\ \varepsilon_2 \mid \mathrm{x}_1,\ \mathrm{x}_2] = \rho$.

As stated by Greene (2003), maximum likelihood estimation was provided in the bivariate probit model. The cumulative probability of the bivariate normal distribution can be calculated as:

$$Prob(X_1 < x_1, X_2 < x_2) = \int_{-\infty}^{x_2}\int_{-\infty}^{x_1} \varnothing_2(z_1, z_2, \rho)dz_1dz_2 \tag{2}$$

where $\varnothing_2(z_1, z_2, \rho)$, the corresponding probability density function is as follows:

$$\varphi_2(x_1, x_2, \rho) = \frac{e^{-\left(\frac{1}{2}\right)(x_1^2+x_2^2-2\rho x_1x_2)/(1-\rho^2)}}{2\pi(1-\rho^2)^{1/2}} \tag{3}$$

where $\varnothing_2$ and φ_2 represent bivariate normal cumulative distribution and the probability density function. By transferring the probability density function into log function, the estimated values of the coefficients can be obtained from maximizing the likelihood function. More details can be referred to Lemp et al. (2012); Yue and Zou (2014), and Guo et al. (2017).

Next, Bayesian estimation approach is employed due to the following advantages over other methods (Ding et al., 2016): First, the uncertainty is considered in estimating parameters by simulating posterior distribution; second, it is valid in small samples, compared with the asymptotic maximum likelihood method. In maximum likelihood estimation, the true value of the model parameters are considered as fixed but unknown. It maximizes the likelihood of an unknown parameter θ when given the observed data y through the relationship $L(\theta \mid y) \propto p(y \mid \theta)$, whereas Bayesian estimation approximates the posterior density of y, $p(\theta \mid y) \propto p(\theta)L(\theta \mid y)$ where $p(\theta)$ is the prior distribution of θ and $p(\theta \mid y)$ is the posterior density of θ given y. Consequently, the posterior density of y given θ is the product of the prior distribution of θ and the likelihood of the observed data as follows:

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{\int p(y \mid \theta)p(\theta)d\theta} \propto p(\theta)L(\theta \mid y) \tag{4}$$

In this study, non-information priori is adopted because prior information does not exist, while a new class of simulation techniques named Markov Chain Monte Carlo (MCMC) is implemented to compute the joint posterior distribution. Different from 95% confidence interval of maximum likelihood estimation, the results present Bayesian credible interval (BCI) as a probability statement about the parameter itself, i.e. a 95% BCI contains the true parameter value with ~95% certainty. If the 95% BCI of the posterior mean does not include 0, it implies that this effect is statistically significant at the 95% level. More details about the estimation and marginal effects can be referred to Lemp et al. (2012) and Hollenbach et al. (2019).

For model comparison, as provided by many other studies under the Bayesian (Xu et al., 2016; Zeng et al., 2017, 2019), the Deviance Information Criterion (DIC) is used to compare the models, whereas Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are employed to evaluate the goodness-of-fit about maximum likelihood, thus, DIC is used to compare the models abovementioned: