Analytic Methods in Accident Research 11 (2016) 48–61

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

Modeling the equivalent property damage only crash rate for road segments using the hurdle regression framework

Lu Ma, Xuedong Yan*, Chong Wei, Jiangfeng Wang

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, PR China

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form
4 July 2016
Accepted 4 July 2016

Keywords:
Equivalent-property-damage-only rate
Hurdle model
Tobit model
Mixed distribution
Accident loss

ABSTRACT

The understanding of the distributional characteristics of the equivalent property damage only (EPDO) crash rate is limited in the existing literature. Models without a proper distribution of EPDO rate could result in biased estimations and misinterpretations of factors. The importance of prediction accuracy and modeling performance for the EPDO rate should be acknowledged since they directly affect the allocation of limited public funds to safety management for road networks. The general objective of this study is to investigate the distributional characteristics of the EPDO rate and accordingly develop proper econometric models for connecting the EPDO rate to explanatory variables. A hurdle framework was proposed in order to accommodate the zero-positive mixed domain of the EPDO rate. For the positive part of the EPDO rate, three representative distributions (lognormal, gamma and normal) were tested and then the three hurdle models were compared against the Tobit model and the random-parameters Tobit model. The empirical results illustrate the lognormal hurdle model's superior modeling performance in comparison to the other four models, and more importantly that conclusion also holds for several different definitions of the EPDO rate under different combinations of property damage only (PDO) equivalency factors.

1. Introduction

The frequency and rate of traffic accidents have been identified as important measures for the safety management of transportation facilities. Many studies have contributed to this field by developing statistical models connecting crash counts to influencing factors including geometric design features and traffic operation characteristics of road segments, and socioeconomic and demographic attributes of traffic analysis zones. The crash frequency and the crash rate have been studied using various types of count models (see studies by Lord and Mannering (2010), Mannering and Bhat (2014), Anastasopoulos et al. (2008, 2012a, 2012b), Gregoriades and Mouskos (2013), Lee et al. (2015), Ma et al. (2015) for more detailed discussions).

Nevertheless, as recently emphasized (Ma et al., 2014; Washington et al., 2014), neither the crash frequency nor the crash rate can accurately reflect the magnitude of traffic risk, because accidents range widely in terms of their severity levels, which should be an important consideration in the conducting of safety evaluations. The single value of the crash frequency

* Corresponding author.
E-mail addresses: lma@bjtu.edu.cn (L. Ma), xdyan@bjtu.edu.cn (X. Yan), chwei@bjtu.edu.cn (C. Wei), wangjiangfeng@bjtu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.amar.2016.07.001
2213-6657/

or crash rate measures the chance of accident occurrence in a specific transportation facility but lacks information on the severity and losses caused by each accident. Therefore, it is necessary to use a criterion that is capable of comprehensively integrating the probability of accident occurrence and the likely severity.

An important approach for combining the information on crash severity with that on crash count is based on equivalent property damage only (EPDO) crashes, in which property damage only crashes – also called "no injury" crashes – are treated as one unit and other types of crashes are weighted with respect to their severity levels. EPDO crashes can also be normalized by certain exposure variables to form the EPDO crash rate, which more clearly reflects the risk of harm due to an accident. EPDO and EPDO rates have been applied in practice and in studies for a long time (e.g. Harkey, 1999; Hunter et al., 2001; PIARC et al., 2004; Campbell and Knapp, 2005; HRPDC, 2006; Felsburg Holt and Ullevig, 2008; Rifaat et al., 2010; Oh et al., 2010; Montella, 2010; UMassSafe, 2011; Boudreau, 2014; Washington et al., 2014). Among these works, a few have focused on statistical characteristics of EPDO-related quantities. For the EPDO crash count, Oh et al. (2010) adopted a negative binomial regression model and Washington et al. (2014) used a nonparametric quartile regression. However, some of the distributional assumptions might still be questionable (Oh et al., 2010). For the EPDO rate, there is still a lack of both an understanding of its distributional characteristics, and analytic models of the statistical analyses. To this end, this study was motivated by the desire to contribute by investigating the distributional characteristics of the EPDO rate and then developing appropriate statistical models.

Generally, the hurdle model is an appropriate approach if the response variable has two different states whose statistical behaviors can be treated hierarchically. For count data modeling, the zero counts can be treated separately from all other positive counts due to the excessive observations of zeros. Similarly, for modeling the EPDO rate, zero observations require special treatment. Thus, the hurdle model is a potential approach for these situations. As censored data, the EPDO rate can also be modeled using the Tobit model. Although, the hurdle model's feature allowing the two states to have different parameters gives it an advantage by enhancing its flexibility, the more essential motivation for this study is that the hurdle model provides a framework allowing almost any choice of distributional assumption for the positive part of the EPDO rate. However, such a property is not available in the Tobit model. Therefore, the hurdle model enables examinations of the performance of different distributions for modeling the positive EPDO rate, whereas the Tobit model can only use the truncated normal distribution.

According to the definition of risks in decision theory (Lehmann and Casella, 1998), the expected EPDO rate is the risk of losses caused by crashes occurring in certain transportation facilities. The PDO equivalency factors are values reflecting the average magnitude of losses due to different accidents, and they might vary according to the types of losses selected. According to the existing literature, the PDO equivalency factors can be societal crash costs (e.g. Oh et al., 2010; Washington et al., 2014), economic losses (e.g. Montella, 2010) or other types (HRPDC, 2006; Felsburg Holt and Ullevig, 2008; UMassSafe, 2011). Therefore, choices of PDO equivalency factors can differ dramatically. Therefore, it would be interesting to understand the influence of different PDO equivalency factors on the performance of different models, something that has generally been missing from previous studies. Hence, another objective of this study is to inspect the impact of different combinations of PDO equivalency factors on the selection of preferred econometric models.

In the following sections, the distributional characteristics of the EPDO rate are discussed first. Then, several competitive econometric structures are introduced and applied to an observed accident dataset using the proposed hurdle-type models (lognormal, gamma and normal distributions for the positive domain) and two Tobit-type models (Tobit and random-parameters Tobit). Next, the models are applied to several different definitions of EPDO rates using PDO equivalency factors from previous studies and practical applications.

2. Method

2.1. An overview of the distribution of the equivalent-property-damage-only rate

A general definition of the EPDO rate is presented in Eq. (1), where N_i is the number of accidents of the ith type of severity, and L_i is the corresponding PDO equivalency factor. Normalizing the weighted sum by certain exposure quantities gives the EPDO rate:

$$EPDO\ rate = \frac{EPDO}{exposure} = \frac{L_1N_1 + L_2N_2 + L_3N_3 \ldots}{exposure} \tag{1}$$

In order to identify an appropriate regression model, it is necessary to understand the distributional characteristics of the EPDO rate. Here, the number of accidents of each severity type could follow some distribution, e.g. Poisson or negative binomial, but the distributional form of the EPDO rate is intractable due to the complicated correlative relationships between the crash counts for different severity levels and the exposure variables. One practical way to deal with such a problem is to try to find approximate parametric distributions.

According to the definition, the EPDO rate is continuous and nonnegative and equals zero if no accident is observed during the study period. Therefore, a certain amount of probability is concentrated on zero EPDO rates, which might be considerable if excessive zero counts of accidents are present. The positive part of the EPDO rate is continuously distributed, with each point of the EPDO rate associated with a probability density. Overall, the EPDO rate actually follows a distribution such that a positively continuous part is mixed with a discrete value of zero.

Thus, the greatest challenge in understanding the distributional pattern of the EPDO rate is to identify an appropriate parametric distribution for its positive domain. Because the distribution of the EPDO rate is intractable, some candidate distributions that are continuous on the positive domain, e.g. lognormal and gamma, have potentials to provide approximations to the positive part of the EPDO rate. The following section will provide an empirical analysis, examining some distributional characteristics of the EPDO rate using observed crash data.

2.2. About the data

This study uses crash data collected in the Pikes Peak Area, Colorado, for the period from July 2006 to December 2010. Based on this dataset, accidents were identified for 983 urban road segments. In choosing an appropriate set of PDO equivalency factors, this study considers the average economic costs caused by accidents of different severity levels. Ozbay et al. (2001) and Miller and Moffet (1993) provided an approximation of such costs, namely 4,113,956, 144,291 and 6783 US dollars for "fatal", "injury" and "no-injury" accidents respectively. The relative magnitudes of 606.5, 21.3 and 1 are therefore adopted as the PDO equivalency factors for the following analyses. The corresponding definition of the EPDO rate is as follows:

$$EPDO\ \ rate = \frac{606.5 \times N_{fatal} + 21.3 \times N_{injury} + 1 \times N_{PDO}}{(\text{vehicle miles traveled}) \times 10^{-6}} \tag{2}$$

Here N_{fatal}, N_{injury} and N_{PDO} are respectively the number of "fatal", "injury" and "no-injury" accidents that occurred in each road segment. Based on Eq. (2), the observed EPDO rates are then calculated for all 983 urban road segments, of which 861 experienced traffic accidents and therefore exhibited positive EPDO rates, while the remaining 122 reported no accidents during the period and therefore have zero EPDO rates. Table 1 summarizes the basic descriptive statistics of the EPDO rate and several explanatory variables.

In order to test the collinearity, we also calculated the variance inflation factor (VIF) for these variables. The VIF measures by how much the variance of estimated regression parameters are inflated as compared to when the predictor variables are not linearly related. The resulting VIF value is 1.045, indicating that the level of multicollinearity is very low. According to the data, roadways are segmented in the way that each road segment has homogeneous characteristics such as number of lanes, pavement conditions, rural/urban indicator etc. Therefore, a road segment could contain zero intersections if it had different road characteristics to both the upstream and downstream sections of the road, and in such a situation it would have a zero intersection density. Fig. 1 shows the histogram as well as a fitted density curve using the observed positive EPDO rate. Its distribution is right-skewed that many road segments have a small EPDO rate whereas few of them have a larger rate. Fig. 2 further presents the quantile-quantile plot of the EPDO rate against the standard normal distribution, which indicates that the positive EPDO rate is far away from being normally distributed. Therefore, a few distributions, e.g. lognormal and gamma, that are flexible enough to reflect the skewness might be suitable candidates.

Table 1
Descriptive characteristics of variables.

Variable	Mean	Std. dev.	Min.	Max.
Response variable				
Positive equivalent-property-damage-only rate (861 observations)	11.220	15.860	0.053	98.680
Zero equivalent-property-damage-only rate (122 observations)	0	0	0	0
Continuous explanatory variables				
Intersection density	2.321	1.878	0	12.255
Segment length	1.747	2.506	0.204	25.890
Annual average daily traffic (AADT) $\times 10^{-4}$	0.752	0.895	0.004	5.370
Pavement condition rating	2.997	0.737	1	5
Categorical explanatory variables		Number of observations		
Rural area[a]	184			
Urban area	779			
Number of through lanes less than or equal to two[a]	661			
Number of through lanes greater than two	322			
Pavement material type at low level[a]	442			
Pavement material type at high level	541			

[a] This category was used as the "reference" in following models.

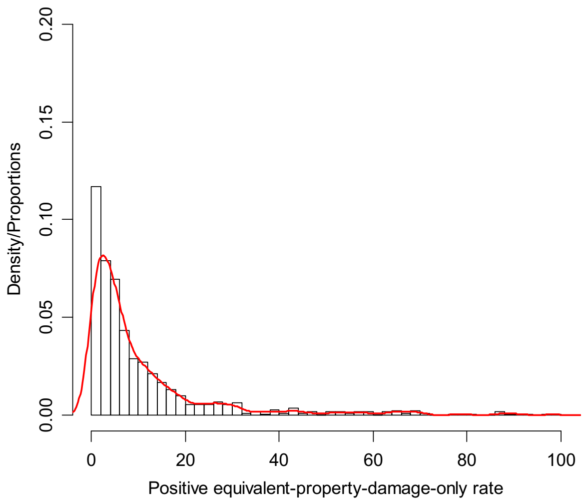

Fig. 1. Distribution of positive equivalent-property-damage-only rate.

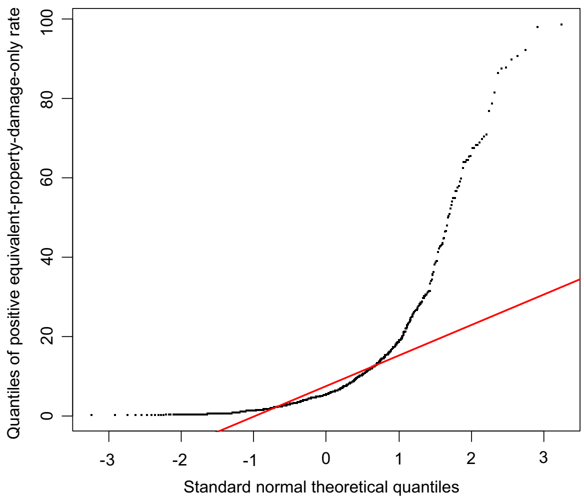

Fig. 2. Quantile-quantile plot of positive equivalent-property-damage-only rate.

2.3. Proposed hurdle models

As was mentioned earlier, the EPDO rate follows a mixed distribution. The zero EPDO rate is the discrete portion, which is associated with a concentrated mass of probability, while the positive EPDO rate is continuously distributed. For such a mixed distribution, an appropriate regressional framework would be the hurdle model (Cragg, 1971), which includes a hurdle between the zero and nonzero outcomes. It also has a flexible framework that relates the hurdle to explanatory variables (Jones, 1989; Martínez-Espiñeira, 2006; Bouchera and Santolino, 2010), and many studies (e.g. Suhaila et al., 2011; Jones, 1989) have verified its appropriateness for capturing a mixed distribution of this type.

The probability mass function for hurdle models can be expressed as follows:

$$f_h(y|\pi, \boldsymbol{\theta}) = \begin{cases} 1 - \pi, & y = 0 \\ \pi f(y|\boldsymbol{\theta}), & y > 0 \end{cases}, \tag{3}$$

where $0 \leq \pi \leq 1$ is the probability that the response Y is positive and $f(y|\theta)$ is the density function conditional on $Y > 0$. The log-likelihood function of the hurdle model is

Table 2
Regression specifications for the three hurdle models.

Distribution	Probability density	Mean	Skewness	Link function
Lognormal	$\frac{1}{\sqrt{2\pi}y\sigma}\exp\left[-\frac{(\ln y-\mu)^2}{2\sigma^2}\right]$	$\exp\left(\mu+\frac{\sigma^2}{2}\right)$	$\left[\exp(\sigma^2)+2\right]\sqrt{\exp(\sigma^2)-1}$	$\boldsymbol{x}_i^T\boldsymbol{\beta}=\mu_i$
Gamma	$\frac{y^{1/\sigma^2-1}\exp\left[-y/(\sigma^2\mu)\right]}{(\sigma^2\mu)^{1/\sigma^2}\Gamma(1/\sigma^2)}$	μ	2σ	$\exp(\boldsymbol{x}_i^T\boldsymbol{\beta})=\mu_i$
Normal	$\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$	μ	0	$\boldsymbol{x}_i^T\boldsymbol{\beta}=\mu_i$

$$\begin{aligned}\text{Log}\, L(\boldsymbol{y};\pi,\boldsymbol{\theta}) &= \sum_{\boldsymbol{y}=0}\log(1-\pi)+\sum_{\boldsymbol{y}>0}\log\left[\pi f(y|\boldsymbol{\theta})\right]=\left[\sum_{\boldsymbol{y}=0}\log(1-\pi)+\sum_{\boldsymbol{y}>0}\log(\pi)\right]+\sum_{\boldsymbol{y}>0}\log f(y|\boldsymbol{\theta})\\ &= \log L_1(\pi|\boldsymbol{y})+\log L_2(\boldsymbol{\theta}|y).\end{aligned} \tag{4}$$

The log-likelihood function is broken down into two parts (first and second layer) that have disjointed parameters. Therefore, it can be maximized by maximizing the two layers individually (Jones, 1989). In a hierarchical manner, the first layer treats all positive observations as one category and the second layer is conditional on all the observations being positive. In the first-layer model, a logistic regression is adopted, as shown in Eq. (5):

$$\boldsymbol{x}_i^T\boldsymbol{\delta}=\text{logit}(\pi_i)=\log\left(\frac{\pi_i}{1-\pi_i}\right). \tag{5}$$

Here $\boldsymbol{x}_i$ is the vector of explanatory variables associated with observation i, and $\boldsymbol{\delta}$ is the vector of parameters for the logistic regression. In the second-layer model, the lognormal, gamma and normal distributions were chosen for inspection. Table 2 provides the detailed specifications for these models, in which the lognormal and gamma distributions are appropriate for positively skewed data. Although the normal distribution seems inappropriate for such data, it was considered in this study for comparison purposes.

2.4. Tobit-type models

The EPDO rate can be treated as left-censored data, for which the Tobit-type model is also a potential approach. In this study, a traditional Tobit model (Tobin, 1958) and a random-parameters Tobit model (Anastasopoulos et al., 2012a) were developed and applied to the observed data. According to Eq. (6), the Tobit model assumes a normally distributed latent variable y_i^* whose probability space is partitioned into two parts to represent the corresponding zero and positive states of the EPDO rate:

$$y_i=\begin{cases}0 & y_i^*\le 0\\ y_i^* & y_i^*>0\end{cases},\ \text{where}\ y_i^*=\boldsymbol{x}_i^T\boldsymbol{\beta}+\varepsilon_i,\ \text{and}\ \varepsilon_i\sim N\left(0,\sigma^2\right). \tag{6}$$

For the Tobit model, the log-likelihood function is presented in Eq. (7), where $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the cumulative distribution function and probability density function of the standard normal distribution. The maximum likelihood estimation (MLE) method can be readily used to obtain the parameter estimations.

$$LL_1=\sum_{y=0}\log\left\{\Phi\left(\frac{-\boldsymbol{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right\}+\sum_{y>0}\log\left\{\frac{1}{\sigma}\phi\left(\frac{y_i-\boldsymbol{x}_i^T\boldsymbol{\beta}}{\sigma}\right)\right\} \tag{7}$$

The Tobit model is appropriate in reflecting the zero-altered distribution of EPDO rates, but it is inadequate if some important factors were unobserved. Accidents are the results of complex processes involving numerous unobserved factors that affect both the likelihood of an accident and its injury severity (Mannering et al., 2016; Ma et al., 2016). It indicates heterogeneous relationships between EPDO rates and observed factors. One of the ways for dealing with the unobserved heterogeneity is to allow coefficients to be randomly distributed (Barua et al., 2016). The consequent random-parameters Tobit model was first developed by Anastasopoulos et al. (2012a). In this model, the parameter vector $\boldsymbol{\beta}$ is assumed to be random and follows a multivariate distribution. As suggested, a multivariate normal distribution $\boldsymbol{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}},\boldsymbol{\Sigma})$ is adopted in this study in which the variance-covariance matrix $\boldsymbol{\Sigma}$ is assumed to be a diagonal matrix, representing that these random parameters are independent. The corresponding log-likelihood function of the random-parameters Tobit model is given in Eq. (8):

Table 3
Estimation results from the hurdle models.

Explanatory variables	First-layer model		Second-layer model					
			Lognormal		Gamma		Normal	
	Coef.	t stat.	Coef.	t stat.	Coef.	t stat.	Coef.	t stat.
Intercept	–	–	1.184	11.037	2.253	33.422	9.990	10.878
Intersection density	0.063	2.368	0.072	6.823	0.055	6.014	0.626	5.227
Segment length	0.860	4.694	–	–	–	–	–	–
AADT	1.556	4.246	−0.252	−4.976	−0.262	−6.745	−3.100	−5.128
Pavement condition rating	0.205	2.631	–	–	–	–	–	–
Urban area	–	–	0.267	2.125	–	–	–	–
Number of through lanes greater than two	1.528	3.250	–	–	–	–	–	–
High-level pavement material type	−1.153	−5.128	–	–	–	–	–	–

The converged log-likelihood values and modeling performance measures are presented in Table 5.
– indicates that the parameter is statistically insignificant.

$$LL_2 = \sum_{y=0} \log\left\{ \int_{R^k} \Phi\left(\frac{-\boldsymbol{x}_i^T\beta}{\sigma}\right) f\left(\beta|\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}\right)d\beta \right\} + \sum_{y>0} \log\left\{ \int_{R^k} \frac{1}{\sigma}\phi\left(\frac{y_i - \boldsymbol{x}_i^T\beta}{\sigma}\right) f\left(\beta|\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}\right)d\beta \right\} \tag{8}$$

Here k is the number of random parameters, and R^k is the k-dimensional real-number space. Obtaining the parameters using a regular MLE procedure is difficult, since the likelihood function cannot be evaluated analytically. The maximum simulated likelihood estimation approach (Bhat, 2001; Anastasopoulos et al., 2012a; Anastasopoulos, 2016) was therefore adopted to solve the problem.

3. Results and analyses

3.1. Results for hurdle models

Based on hurdle models, Table 3 presents the explanatory variables having statistically significant effects on the EPDO rate. The first-layer model focuses on the binary choice of zero/positive EPDO rate for which the dependent variable is 1 if the accident occurred and 0 otherwise. Thus, a positive parameter means that being in the relevant state (for dummy variables) or increasing the magnitude (for continuous variables) will increase the probability of having positive EPDO rates. For the second-layer model, we conducted regression analyses using the lognormal distribution, the gamma distribution and the normal distribution. These models indicate that the explanatory variables affect the EPDO rate through both the first-layer and the second-layer parameters. Since the behavioral effects of the factors are consistent throughout the three models, the following section will provide interpretations with respect to the lognormal hurdle specification.

In this study, a road segment is defined as a continuous stretch of a particular roadway. Longer segments could connect several intersections. Although, intersection-related accidents have been removed from the dataset, we introduce the intersection density to examine the impact of intersections on non-intersection-related accidents that occurred in the road segments. Intersection density is defined as the number of intersections per mile of road segment, and the results show that it has positive parameters in both layers of the model. Therefore, road segments with higher intersection densities are more likely to have higher EPDO rates. This indicates that intersections can affect EPDO rates for road segments. In fact, traffic flow may be less ordered and drivers will tend to be preparing to make maneuvers when approaching or leaving intersections. For example, drivers usually need to take in additional road sign information prior to passing through intersections. These actions increase the mental workload of drivers, especially pertinent for older drivers (Caird et al., 2008), and increase the accident risk.

Segment length and annual average daily traffic (AADT) are components used for measuring the vehicle miles traveled of each road segment. For road segments that experienced accidents during the study period, the vehicle miles traveled is an effective proxy for the exposure quantity in the denominator when computing the EPDO rate. However, for road segments with zero accidents, the vehicle miles traveled is non-informative because the observed EPDO rate is always zero irrespective of its magnitude. Therefore, segment length and AADT appear as explanatory variables representing exposure information in the first-layer model. As expected, the two variables exhibit a positive impact on the odds of having a positive EPDO rate.

It should be noted that the second-layer model also adopts segment length and AADT as explanatory variables in an attempt to extract additional information, even though the two variables were used in calculating the positive EPDO rate. In the second-layer model, AADT exhibits a significant impact. The negative sign indicates that road segments with a high volume of traffic tend to have low EPDO rates. Such a situation might be attributable to two aspects: First, road segments with high volumes of traffic tend to have low crash rates (Anastasopoulos et al., 2008). Second, higher traffic volumes often lead to less severe accidents due to lower vehicle speeds.

Pavement condition rating is a continuous criterion ranging from 0 to 5. Specifically, a larger rating represents smoother pavements with fewer cracks and patches, while a rating near to zero indicates pavements that are in an extremely deteriorated condition. It should be noted that, in this dataset, the worst pavement condition rating observed is 1. In the first-layer model, this variable's impact is significant with a positive parameter, while in the second-layer model the parameter is negative. In other words, road segments with better pavement conditions are more likely to have higher EPDO rates. However, conditional on a positive EPDO rate, better pavement conditions will produce smaller EPDO rates. One of the possible reasons is that people may drive faster (Mannering, 2009) and pay less attention when there are good pavement conditions, compatible with the results from the first-layer model. However, conditional on the occurrence of accidents, worse pavement conditions may produce more severe accidents and consequently lead to a larger EPDO rate.

The location (urban/rural) of the road segments is only significant in the second-layer model. Urban segments are more likely to have higher EPDO rates than rural segments. The number of lanes of a roadway is a commonly used variable connected to accident counts in existing studies. In this study, road segments with more than two lanes are defined as the reference category. According to the first-layer model, road segments with more than two lanes are more likely to have higher EPDO rates than roads with fewer lanes if other conditions remain the same. This is in line with several previous studies (e.g. Abdel-Aty and Radwan, 2000) and it is probably due to the fact that, with an increasing number of lanes, the number of conflicting actions due to lane changing is increased (Kononov et al., 2008), which raises the chance of accidents. However, conditional on a positive EPDO rate, road segments with more through lanes are inclined to have smaller EPDO rates. Therefore, the overall impact of this variable is not fixed, but actually depends on the magnitude of all the other variables. This situation has been referred to as counterintuitive (Shankar et al., 1997), and it in fact reflects the non-monotonic nature of the influence this variable has on the EPDO rate. In addition, prior research has concluded that road segments with more lanes are more likely to be the scenes of less severe accidents (Zhu and Srinivasan, 2011), which could also explain the unfixed influence this variable has on the EPDO rate.

Pavement material type also shows a significant impact on the EPDO rate. Two types of pavements are defined in terms of the rigidity of the road material. The reference category represents road pavements with a thin bituminous surface that is less rigid than the other category, termed high-level pavements. The parameter is negative in both layers of the model. This indicates that rigid pavements are inclined to have smaller EPDO rates.

3.2. Results of Tobit-type models

The Tobit model and a random-parameters Tobit model were also applied to the observed data for comparison purposes. Table 4 illustrates the estimated parameters for the two models. The behavioral interpretations of the explanatory variables are in generally consistent between the two Tobit-type models and also with the lognormal hurdle model. The random-parameters Tobit model is able to capture the heterogeneous influence of factors on the EPDO rate, and is expected to have a higher maximum likelihood value than the Tobit model. For the random-parameters Tobit model, only intersection density was revealed to have significant random parameter. This means that considerable uncertainty exists regarding the impact of intersection density on the EPDO rate.

3.3. Model comparisons

We used information criteria including Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), goodness-of-fit measures including Maddala R-squared (Maddala, 1983), and prediction performances including hit rate for zero/positive states and mean absolute percentage error (MAPE) of positive observations, to assess these models. The hit rate for zero/positive states is defined as the correct classification rate of zero/positive states of EPDO rate, as in Eq. (9):

Table 4
Estimation results from the Tobit model and the random-parameters Tobit model.

Explanatory variables	**Tobit model**		**Random-parameters Tobit model**			
			Mean parameters		Standard deviation	
	Coef.	t stat.	Coef.	t stat.	Coef.	t stat.
Intercept	5.598	4.490	5.403	7.596	–	–
Intersection density	0.648	4.962	0.712	7.316	0.745	7.357
Segment length	0.839	2.312	0.847	3.458	–	–
AADT	−1.750	−2.258	−1.497	−3.432	–	–
No. of through lanes greater than two	3.453	2.365	2.760	4.560	–	–
High-level pavement material type	−3.136	−2.738	−3.197	−5.375	–	–

The converged log-likelihood values and modeling performance measures are presented in Table 5.
– indicates that the parameter is statistically insignificant.

Table 5
Summary of the modeling performance of the five models.

Statistics	Hurdle models			Tobit-type models	
	Lognormal	Gamma	Normal	Regular	Random-parameters
Log-likelihood at convergence	−3065.155	−3107.993	−3783.588	−3652.222	−3648.739
AIC	6152.310	6235.986	7587.176	7318.443	7313.478
BIC	6205.880	6284.686	7635.877	7352.534	7352.438
Maddala R-squared	7.670×10^{-5}	7.177×10^{-5}	5.396×10^{-5}	1.306×10^{-5}	1.504×10^{-5}
Hit rate for zero/nonzero states	0.818	0.818	0.818	0.651	0.652
MAPE of positive observations	1.728	3.854	3.845	4.905	4.672

$$hit\ \ rate=\sum_{i=1}^{n}\frac{P\left(\hat{y}_i=0\right)I\left(y_i=0\right)+P\left(\hat{y}_i>0\right)I\left(y_i>0\right)}{n} \tag{9}$$

Here $I(\cdot)$ is the indicator function, which takes the value 1 if the argument statement is true and 0 otherwise. Eq. (10) defines the MAPE for positive observations, where n_p is the total number of positive observations:

$$MAPE=\sum_{i=1}^{n_p}\frac{\left|\hat{y}_i-y_i\right|}{n_p y_i} \tag{10}$$

These measures comprehensively reflect the models' ability to represent the underlying distributional characteristics of the EPDO rate and to extract information from the data. Table 5 presents the values of these measures for the five models. Out of these hurdle models, the lognormal hurdle model performs the best according to all the criteria and it is clear that the lognormal (binomial-lognormal hurdle) and gamma (binomial-gamma hurdle) are significantly better than the normal hurdle model. Out of the two Tobit-type models, since the regular Tobit model is nested within the random-parameters Tobit model, a larger maximum log-likelihood is expected for the latter. The random-parameters Tobit model exhibits better performance than the regular Tobit model with respect to all measures. However, the improvement is limited in this study. Random parameters allow heterogeneous influential patterns of variables, and for different realizations of a particular random parameter, they actually still serve a single model. However, the two sets of parameters in the presented hurdle model belong to two different models, capturing two different stages of impacts from the explanatory variables.

Based on the current data, the lognormal hurdle model is superior to the other four models and is significantly better than the normal hurdle model and the two Tobit-type models, which involve normal error terms. Thus, the distributional pattern is crucial for modeling the EPDO rate, and an appropriate form for the positive EPDO rate might surpass the effect of the econometric form of the model. As mentioned earlier, the positive EPDO rate is right-skewed, implying that distributions like lognormal and gamma might be preferable for modeling the EPDO rate. The hurdle model provides a flexible framework, allowing a variety of distributional assumptions to be made for the positive EPFO rate, whereas the existing Tobit-type models only accepts the normal distribution for the latent variable. To be noted, these models might have an omitted variables bias (Mannering and Bhat, 2014) due to the absence of some important explanatory variables in the dataset.

4. Demonstration of an application on identifications of high-accident locations

One of the functions of the EPDO rate is to serve as the criterion for identifying high-accident locations and it is able to consider the heterogeneous losses caused by accidents of different injury levels. In identifying high-accident locations, segments in a given network are ranked by their excess EPDO rate, which is defined as the difference between the observed EPDO rate y_i and the expected EPDO rate $\hat{y}_i$ estimated using given models:

$$Excess\ \ EPDO\ \ rate_i=y_i-\hat{y}_i \tag{11}$$

Segments having larger excess EPDO rates are considered hazardous and may require further audits. An example of the identification of high-accident locations is presented here using the recommended lognormal hurdle model. Eq. (11) provides the criterion for ranking segments. Using the observed data, Fig. 3 highlights the top 10% of segments as identified in the ranking and Table 6 lists the basic characteristics of the top 10 segments.

The purpose of using excess EPDO rate is to identify locations having higher EPDO rate than the average-level EPDO rate sustained by similar road segments. Fig. 4 illustrates the relationship between ranks of excess EPDO rate and observed EPDO rate. It indicates that the two different ranks are highly consistent for the cases with large EPDO rates. Therefore, locations with extremely large observed EPDO rate are more likely to be identified as high-accident locations, whereas for other cases, there might be substantial discrepancies between the two ranks.

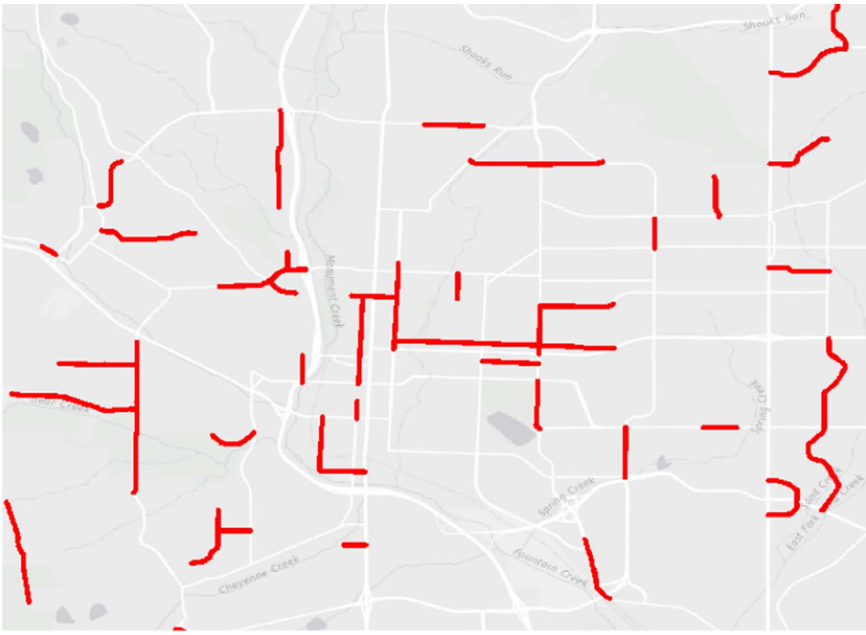

Fig. 3. Identified high-accident locations.

Table 6
Top 10 identified traffic high-accident locations.

No.	Intersection density	Segment length	AADT	Pavement condition rating	Location	No. of lanes	Pavement material type	No injury	Injury	Fatal
1	1.619	0.618	6100	3.5	Rural	≤ 2	Low	4	0	1
2	5.674	0.176	1000	2.5	Urban	> 2	High	7	1	0
3	2.761	0.724	1400	2.5	Urban	> 2	High	23	6	0
4	8.282	0.604	1000	2.5	Urban	> 2	High	6	4	0
5	9.467	0.211	1400	4	Urban	> 2	Low	1	2	0
6	10.684	0.374	12,000	3.5	Urban	≤ 2	Low	9	1	1
7	12.435	0.241	700	3.2	Urban	> 2	High	3	1	0
8	15.686	0.191	5200	1.5	Urban	≤ 2	Low	34	5	0
9	3.628	2.756	2800	4	Urban	> 2	Low	18	4	0
10	9.591	1.147	4900	2.6	Urban	≤ 2	Low	38	3	1

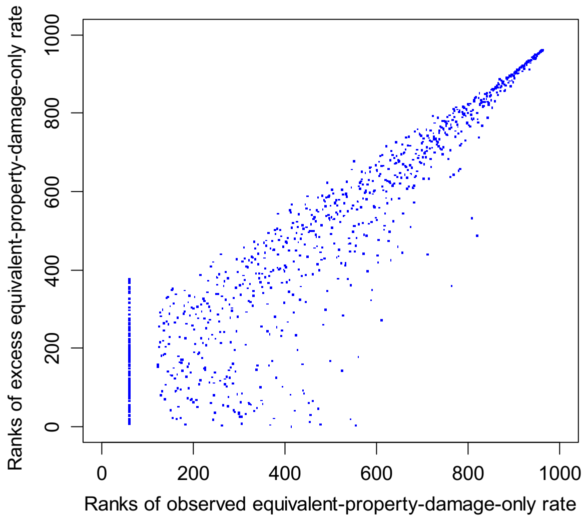

Fig. 4. Relationships between ranks of observed and excess equivalent-property-damage-only rates.

Table 7
Property-damage-only equivalency factors adopted in previous studies and practical applications.

Studies and practical applications	Property-damage-only equivalency factors				
	Fatal injury	Incapacitating injury	Non-incapacitating injury	Possible injury	No injury
Harkey, 1999	76.8	76.8	8.4	8.4	1
Hunter et al., 2001	76.8	76.8	8.4	8.4	1
Ozbay et al., 2001	606.5	21.3	21.3	21.3	1
HRPDC, 2006	12	3	3	3	1
Felsburg Holt and Ullevig, 2008	12	5	5	5	1
Rifaat et al., 2010	9.5	3.5	3.5	3.5	1
Oh et al., 2010	1330	949	11	11	1
Montella, 2010	771	35	35	35	1
UMassSafe, 2011	9.5	4.5	3.5	2.5	1
Boudreau, 2014	10	5	5	5	1
Washington et al., 2014	1330	949	11	11	1

5. Examining other combinations of property-damage-only equivalency factors

The magnitude of EPDO crashes relies heavily on the values of the PDO equivalency factors associated with each accident severity level. Table 7 summarizes the PDO equivalency factors adopted in the existing literature. It is obvious that they vary substantially across different studies and practical applications. Such discrepancies actually result from the ways in which "injury" or "fatal" accidents are scaled, using "no-injury" accidents to represent one unit. For example, Oh et al. (2010) used the average societal crash costs associated with different type of accidents to form the PDO equivalency factors; Montella (2010) adopted the economic losses associated with different types of accidents; Several other studies have used the values designed by state departments of transportation (e.g. Harkey, 1999; Hunter et al., 2001; Boudreau, 2014) or other agencies (e.g. HRPDC, 2006; Felsburg Holt and Ullevig, 2008).

For those studies that consider the economic losses or social costs in determining the PDO equivalency factors, the resulting factors associated with severe accidents, especially fatal ones, are significantly higher than in other studies, because the economic losses include many different aspects, such as medical expenses, emergency services, lost market productivity, ambulance costs, the costs of hospital treatment etc. (Oh et al., 2010; Montella, 2010).

In fact, it is meaningful to use any PDO equivalency factors for calculating the EPDO rate if these factors reflect certain relative losses or harms associated with different types of accidents. However, different choices of PDO equivalency factors could change the distributional patterns of the EPDO rate and further affect the appropriateness of different econometric models. The following analysis inspects the performance of the five models under several different combinations of PDO equivalency factors. In Table 7, more than half of the studies adopted the same categorization as this study (fatal, injury, and no-injury). The rest all used four categories of injury severity, which are slightly out of alignment with this study. For the latter situation, the PDO equivalency factors for "non-incapacitating injury" in those prior studies are simply adopted as the PDO equivalency factor for "injury" in our analysis. Table 8 provides the AIC values of the five models using different EPDO rates.

Importantly, Table 8 indicates the superior performance of the lognormal hurdle model, consistently across the different definitions of EPDO rates. Therefore, the lognormal distribution seems to have great potential to represent a variety of EPDO rates defined based on different PDO equivalency factors. In fact, the lognormal distribution has previously been identified as an appropriate candidate for modeling loss-related quantities. For example, Bolancé et al. (2003) suggested a lognormal distribution as a baseline for the actuarial loss analysis of traffic accident claim values. Eling (2012) also concluded that the lognormal distribution was appropriate for modeling losses. The EPDO rate introduced here actually forms a type of losses, since PDO equivalency factors can be treated as relative losses associated with different levels of accident severity.

Table 8
AIC values of the five models with different sets of property-damage-only equivalency factors.

Property-damage-only equivalency factors			Hurdle models			Tobit models	
Fatal	Injury	No-injury	Lognormal	Gamma	Normal	Regular	Random-parameters
76.8	8.4	1	5317.197	5446.322	7183.085	6898.897	6844.142
606.5	21.3	1	6152.310	6235.986	7587.176	7318.443	7313.478
12	3	1	4544.232	4737.432	6554.652	6265.442	6212.146
12	5	1	4833.451	4996.791	6814.907	6526.821	6466.570
9.5	3.5	1	4610.968	4794.359	6601.421	6313.391	6253.764
1330	11	1	5488.315	5702.552	7245.690	6930.151	6898.746
771	35	1	6463.982	6477.873	7645.690	7383.590	7386.956
10	5	1	4826.221	4988.734	6802.306	6514.507	6451.178

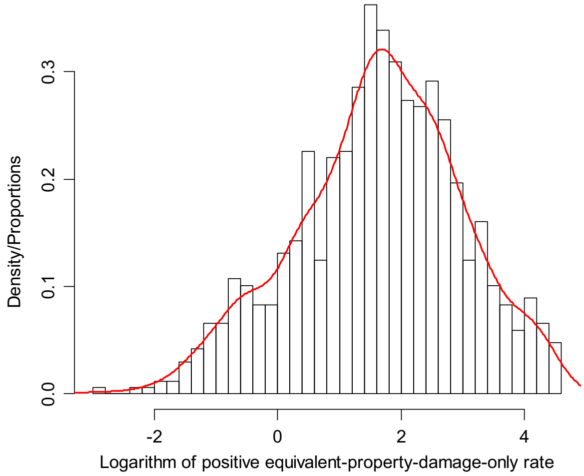

Fig. 5. Distribution of the logarithm of positive equivalent-property-damage-only rate.

6. Discussion and conclusions

6.1. Further inspection of the lognormal distribution for modeling the positive equivalent-property-damage-only rate

The hurdle regression model has illustrated the appropriateness of the lognormal distribution for reflecting the positive part of the EPDO rate in the observed crash data. This section will strengthen that conclusion by inspecting the distributional pattern of the positive EPDO rate on a logarithmic scale, as this is equivalent to examining the normality of the logarithm of the EPDO rate. Figs. 5 and 6 shows the histogram and the quantile-quantile plot against the normal distribution for the log-transformed data. The distribution appears symmetric after the logarithmic transformation. In fact, a logarithmic transformation is especially useful for representing right-skewed samples (Vilar et al., 2009; Eling, 2012).

Overall, the lognormal distribution could be a good approximation for modeling the distribution of positive EPDO rates. Based on the observed data, the lognormal distribution is able to capture the skewness of the data as well as providing relatively high modeling performance for regression analysis.

6.2. Relationships between equivalent-property-damage-only rate and crash rate

If all PDO equivalency factors are equal to one, then the EPDO rate is exactly equal to the crash rate. It is believed that the crash rate and the EPDO rate could produce different results in ranking the risks of accidental losses. Interestingly, it is found that the two criteria are relatively consistent at extremely small or extremely large values, whereas the discrepancies are larger for most other cases. That is, for low-risk or high-risk road segments, the EPDO rate performs similarly to the crash rate, but there are substantial differences for medium-risk road segments. If the losses caused by serious injuries are

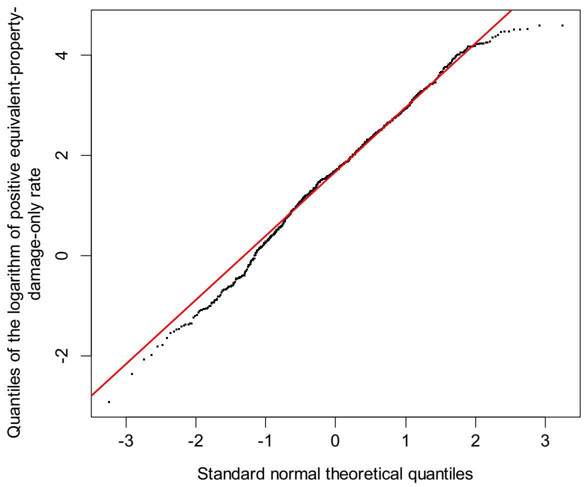

Fig. 6. Quantile-quantile plot of the logarithm of positive equivalent-property-damage-only rate.

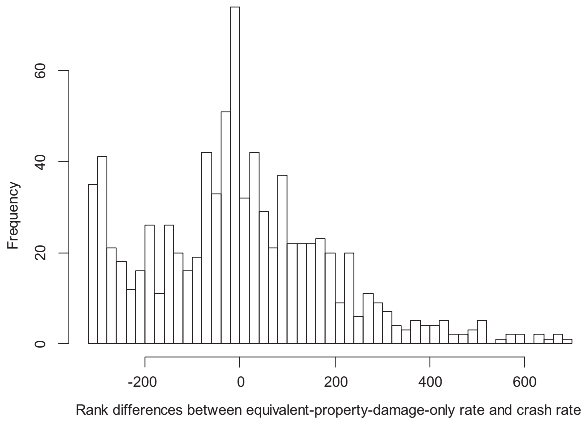

Fig. 7. Rank discrepancy between equivalent-property-damage-only rate and crash rate.

emphasized in road-safety evaluations, it is necessary to use the EPDO rate as the evaluation criterion, especially for the medium-risk cases, which cover the majority of road segments.

Fig. 7 illustrates the distribution of the differences between the ranks of the crash rate and the ranks of the EPDO rate using the observed data, demonstrating an asymmetric phenomenon whereby the crash rate tends to underestimate the EPDO rate (the maximum rank difference is 676) for some cases, while it overestimates the EPDO rate to a relatively smaller extent (the minimum rank difference is -371) for the rest of the cases. Therefore, using the crash rate as a criterion by which to rank the accident risk of road segments might greatly undervalue the harmfulness of severe crashes.

6.3. General conclusions and future directions of research

The motivation for using EPDO crashes for evaluating the safety status of road segments was that doing so allows us to integrate information on accident occurrence chances and the overall harmfulness of accidents. This study defined EPDO crashes by monetizing accident injury-severities, which is subjective in some extent, and hence there are potentials to involve other factors in normalizing different accidents with respect to property-damage-only accidents in the future.

Other than weighting accidents by different PDO equivalency factors prior to the regression, it is possible to model the multivariate distribution of crash rates categorized according to different severity levels. Under such an approach, the EPDO-type quantities can be calculated after the regression. For example, Ma and Kockelman (2006) used a multivariate Poisson regression to model the accident counts classified by injury levels. Anastasopoulos et al. (2012b) estimated models of injury-severity rates and frequencies of accidents, using multivariate Tobit and multivariate negative binomial models, respectively. Chiou and Fu (2013) then applied a multinomial-generalized Poisson model to analyze the crash frequencies for different severity levels. These analyses have undeniably enhanced the understanding of correlative relationships among crash counts/rates at different levels of severity, but two issues might impede the method's practical application for the evaluation of roadway safety. First, in safety evaluation, very severe injuries, especially fatal injuries, are important as they account for huge losses. Yet, these crashes are rare, and hence the insufficient observations in the data could cause difficulties for multivariate models or introduce a large amount of variance into the estimators. Second, the identification process is sometimes required for ranking road segments using certain quantile estimations of the EPDO rate (Washington et al., 2014). The models introduced in this study in fact return the whole distribution of the EPDO rate. Hence, both the mean value and any quantile values are easy to calculate, whereas it is difficult to obtain quantile values for EPDO crashes from the results of multivariate models. Overall, the EPDO rate introduced in this study provides an intuitive and simple index for evaluating the safety status of road segments.

This study is devoted to constructing a regression model for connecting the EPDO rate to explanatory variables. Because the EPDO rate follows a mixed distribution consisting of a discrete probability mass on zero blended with a continuous probability density on the positive domain, a hurdle framework was proposed to deal with this situation.

Another purpose of this study was to inspect the distributional patterns of the EPDO rate, especially its positive part, which is crucial for constructing a specific hurdle model. According to the regression analyses and a distributional inspection of the log-transformed EPDO rate, the lognormal distribution is revealed to be an appropriate parametric choice for reflecting the positive part of the EPDO rate. More importantly the superiority of the lognormal hurdle model was verified using several EPDO rates defined by different combinations of PDO equivalency factors. The dataset adopted in this study has limited number of explanatory variables, and it is also possible that a better model specification could further highlight the presented model.

Such a conclusion is also in line with several prior studies (e.g. Bolancé et al., 2003; Eling, 2012) that recommend the use of the lognormal distribution for modeling losses. A parallel study (Ma et al., 2015) investigated the distributional characteristics of the crash rate by fitting several candidate parametric distributions. The results also suggested that a lognormal distribution and a hurdle model would be appropriate for modeling the crash rate. These empirical findings could imply that a particular mechanism might underlie the data-generating process of the EPDO rate, approximating to a lognormal distribution. Therefore, it would be interesting to theoretically reveal the nature of the distribution of the EPDO rate in the future.

Prior studies and practical applications have adopted various PDO equivalency factors due to different considerations of losses brought about by accidents of different severity levels. In fact, fatal accidents are very rare events and their occurrence could be attributable to many chance factors other than the characteristics of the transportation facilities. For example, the physical situation of a driver or passenger could lead to a fatal crash. Therefore, a very large PDO equivalency factor for fatal accidents might greatly raise the variance of the estimated EPDO rates and hence lead to inaccurate identifications of high-accident locations. Therefore, it is necessary that future studies look at the impact on modeling performance of very large PDO equivalency factors for fatal injuries.

Other than the hurdle models, capturing multi-level impacts from the explanatory variables could be readily done with a finite-mixture Tobit model, which is of great values for future study. In addition, a finite-mixture Tobit model can also be generalized to provided random parameters within mixtures as has been done with ordered probit models (see Xiong and Mannering (2013) for more details).

Findings in this study are data-specific, and therefore they require further examinations using else datasets. We believe that this work takes an important step toward comprehensive analyses of safety management and identifications of high-accident locations. In future studies, there is also potential for more flexible distributions, such as the Tweedie distribution or nonparametric specifications, to be considered.

Acknowledgment

This research was supported by National Natural Science Foundation of China, Nos. 51208032 and 71210001. The authors also acknowledge valuable feedback provided by two reviewers on an earlier draft of this manuscript.

References

Abdel-Aty, M., Radwan, E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32 (5), 633–642.
Anastasopoulos, P., 2016. Random parameters multivariate tobit and zero-inflated count data models: addressing unobserved and zero-state heterogeneity in accident injury-severity rate and frequency analysis. Analytic Methods in Accident Research 11, 17–32.
Anastasopoulos, P., Tarko, A., Mannering, F., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 40 (2), 768–775.
Anastasopoulos, P., Mannering, F., Shankar, V., Haddock, J., 2012a. A study of factors affecting highway accident rates using the random-parameters tobit model. Accident Analysis and Prevention 45, 628–633.
Anastasopoulos, P., Shankar, V., Haddock, J., Mannering, F., 2012b. A multivariate tobit analysis of highway accident-injury-severity rates. Accident Analysis and Prevention 45, 110–119.
Barua, S., El-Basyouny, K., Islam, M., 2016. Multivariate random parameters collision count data models with spatial heterogeneity. Analytic Methods in Accident Research 9, 1–15.
Bhat, C., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research Part B 35 (7), 677–693.
Bolancé, C., Guillen, M., Nielsen, J., 2003. Kernel density estimation of actuarial loss functions. Insurance: Mathematics and Economics 32 (1), 19–36.
Bouchera, J., Santolino, M., 2010. Discrete distributions when modeling the disability severity score of motor victims. Accident Analysis and Prevention 42 (6), 2041–2049.
Boudreau, N., 2014. Traffic and Safety Engineering Project Selection Process. Available at 〈https://www.massdot.state.ma.us/Portals/17/docs/PSA/Traffic_Safety3–13-14_Final.pdf〉 (accessed 03.24.16).
Caird, J., Chisholm, S., Lockhart, J., 2008. Do in-vehicle advanced signs enhance older and younger drivers' intersection performance? Driving simulation and eye movement results. International Journal of Human-Computer Studies 66 (3), 132–144.
Campbell, J., Knapp, K., 2005. Alternative crash severity ranking measures and the implications on crash severity ranking procedures. In: Proceedings of the 2005 Mid-Continent Transportation Research Symposium, Ames, Iowa, August 2005.
Chiou, Y., Fu, C., 2013. Modeling crash frequency and severity using multinomial-generalized Poisson model with error components. Accident Analysis and Prevention 50, 73–82.
Cragg, J., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39 (5), 829–844.
Eling, M., 2012. Fitting insurance claims to skewed distributions: are the skew-normal and skew-student good models? Insurance: Mathematics and Economics 51 (2), 239–248.
Felsburg Holt, Ullevig, 2008. 2035 Regional Transportation Plan Update, Safety Analysis-procedural Overview. Prepared for Denver Regional Council of Governments, Centennial, Colorado.
Gregoriades, A., Mouskos, K., 2013. Black spots identification through a Bayesian Networks quantification of accident risk index. Transportation Research Part C 28, 28–43.
Harkey, D., 1999. Evaluation of truck crashes using a GIS-based crash referencing and analysis system. Transportation Research Record 1686, 13–21.
HRPDC, 2006. Regional Crash Database Supports Safety Planning and Drives State Crash Data Improvement. Available at 〈http://tsp.trb.org/assets/BP04_Hampton_Final.pdf〉 (accessed 03.24.16).
Hunter, W., Stewart, J., Eccles, K., Huang, H., Council, F., Harkey, D., 2001. Three-strand cable median barrier in North Carolina: in-service evaluation. Transportation Research Record 1743, 97–103.
Jones, A., 1989. A double-hurdle model of cigarette consumption. Journal of Applied Econometrics 4 (1), 23–39.
Kononov, J., Bailey, B., Allery, B., 2008. Exploratory analysis of relationship between the number of lanes and safety on urban freeways. In: Proceedings of the 87th Annual Meeting of the Transportation Research Board. Washington, D.C.
Lee, C., Abdel-Aty, M., Park, J., Wang, J., 2015. Development of crash modification factors for changing lane width on roadway segments using generalized nonlinear models. Accident Analysis and Prevention 76, 83–91.

Lehmann, E., Casella, G., 1998. Theory of Point Estimation. Springer, New York.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44 (5), 291–305.
Ma, J., Kockelman, K., 2006. Bayesian multivariate Poisson regression for models of injury count by severity. Transportation Research Record 1950, 24–34.
Ma, L., Yan, X., Weng, J., 2015. Modeling traffic crash rates of road segments through a lognormal hurdle framework with flexible scale parameter. Journal of Advanced Transportation 49 (8), 928–940.
Ma, L., Wang, G., Yan, X., Weng, J., 2016. A hybrid finite mixture model for exploring heterogeneous ordering patterns of driver injury severity. Accident Analysis and Prevention 89, 62–73.
Ma, L., Yan, X., Zhang, C., Weng, J., 2014. Traffic accident risk estimation based on the lognormal hurdle model with a flexible scale parameter. In: Proceedings of the 93rd Annual Meeting of the Transportation Research Board, Washington, D.C.
Maddala, G., 1983. Limited-dependent and Qualitative Variables in Econometrics. Cambridge University Press, Cambridge, United Kingdom.
Mannering, F., 2009. An empirical analysis of driver perceptions of the relationship between speed limits and safety. Transportation Research Part F 12 (2), 99–106.
Mannering, F., Bhat, C., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research 1, 1–22.
Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research 11, 1–16.
Martínez-Espiñeira, R., 2006. A box-cox double-hurdle model of wildlife valuation: the citizen's perspective. Ecological Economics 58 (1), 192–208.
Miller, P., Moffet, J., 1993. The Price of Mobility: Uncovering the Hidden Costs of Transportation. Natural Resources Defense Council, New York.
Montella, A., 2010. A comparative analysis of black spot identification methods. Accident Analysis and Prevention 42 (2), 571–581.
Oh, J., Washington, S., Lee, D., 2010. Property damage crash equivalency factors to solve crash frequency-severity dilemma: case study on South Korean rural roads. Transportation Research Record 2148, 83–92.
Ozbay, K., Bartin, B., Berechman, J., 2001. Estimation and evaluation of full marginal costs of highway transportation in New Jersey. Journal of Transportation and Statistics 4 (1), 81–103.
PIARC, World Road Association, Technical Committee on Road Safety C13, 2004. Road Safety Manual. Switzerland.
Rifaat, S., Tay, R., De Barros, A., 2010. Effect of street pattern on road safety: are policy recommendations sensitive to aggregations of crashes by severity? Transportation Research Record 2147, 58–65.
Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accident Analysis and Prevention 29 (6), 829–837.
Suhaila, J., Ching-Yee, K., Fadhilah, Y., Hui-Mean, F., 2011. Introducing the mixed distribution in fitting rainfall data. Opening Journal of Modern Hydrology 1 (2), 11–22.
Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26 (1), 24–36.
UMassSafe, 2011. Equivalent Property Damage Only Ranking of Commercial Vehicle Crashes. Available at ⟨http://cmvtoolkit.ecs.umass.edu/EPDO_2011_Report.pdf⟩ (accessed 03.24.16).
Vilar, J., Cao, R., Ausín, M., González-Fragueiro, C., 2009. Nonparametric analysis of aggregate loss models. Journal of Applied Statistics 36 (2), 149–166.
Washington, S., Haque, M., Oh, J., Lee, D., 2014. Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots. Accident Analysis and Prevention 66, 136–146.
Xiong, Y., Mannering, F., 2013. The heterogeneous effects of guardian supervision on adolescent driver-injury severities: a finite-mixture random-parameters approach. Transportation Research Part B 49, 39–54.
Zhu, X., Srinivasan, S., 2011. A comprehensive analysis of factors influencing the injury severity of large-truck crashes. Accident Analysis and Prevention 43 (1), 49–57.