Accident Analysis and Prevention 45 (2012) 110–119

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

A multivariate tobit analysis of highway accident-injury-severity rates

Panagiotis Ch. Anastasopoulos[a,*], Venky N. Shankar[b,1], John E. Haddock[c,2], Fred L. Mannering[d,3]

[a] School of Civil Engineering, Purdue University, Indiana Local Technical Assistance Program, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, United States
[b] Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, PA 16802, United States
[c] Department of Civil Engineering, Purdue University, and Indiana Local Technical Assistance Program, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, United States
[d] Department of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, United States

ARTICLE INFO

Article history:
Received 23 August 2011
Received in revised form 14 October 2011
Accepted 9 November 2011

Keywords:
Accident-injury-severity rates
Censoring
Contemporaneous error correlation
Multivariate tobit regression
Roadway geometrics
Interstate highways

ABSTRACT

Relatively recent research has illustrated the potential that tobit regression has in studying factors that affect vehicle accident rates (accidents per distance traveled) on specific roadway segments. Tobit regression has been used because accident rates on specific roadway segments are continuous data that are left-censored at zero (they are censored because accidents may not be observed on all roadway segments during the period over which data are collected). This censoring may arise from a number of sources, one of which being the possibility that less severe crashes may be under-reported and thus may be less likely to appear in crash databases. Traditional tobit-regression analyses have dealt with the overall accident rate (all crashes regardless of injury severity), so the issue of censoring by the severity of crashes has not been addressed. However, a tobit-regression approach that considers accident rates by injury-severity level, such as the rate of no-injury, possible injury and injury accidents per distance traveled (as opposed to all accidents regardless of injury-severity), can potentially provide new insights, and address the possibility that censoring may vary by crash-injury severity. Using five-year data from highways in Washington State, this paper estimates a multivariate tobit model of accident-injury-severity rates that addresses the possibility of differential censoring across injury-severity levels, while also accounting for the possible contemporaneous error correlation resulting from commonly shared unobserved characteristics across roadway segments. The empirical results show that the multivariate tobit model outperforms its univariate counterpart, is practically equivalent to the multivariate negative binomial model, and has the potential to provide a fuller understanding of the factors determining accident-injury-severity rates on specific roadway segments.

Published by Elsevier Ltd.

1. Introduction

The preponderance of past research has studied the occurrence of accidents by considering their frequency and applying count-data modeling techniques to study factors that affect the frequency of accidents over some time period on specific roadway segments. This body of literature has applied a wide variety of modeling approaches such as Poisson and negative binomial models, Poisson-lognormal models, zero-inflated count models, Conway–Maxwell–Poisson models, negative multinomial models, Gamma models, generalized estimating equation models, generalized additive models, random effects and random parameters count models, and finite mixture and Markov switching models (for a complete review of this literature see Lord and Mannering, 2010).

While traditional accident-frequency approaches have undeniably improved our understanding of factors affecting accident occurrence, some recent research has suggested the tobit regression as an alternative (Anastasopoulos et al., 2008, 2012). The tobit-regression approach considers accident rates (such as the number of accidents per vehicle-miles traveled) on roadway segments as opposed to accident frequencies. This results in data that are continuous (instead of the discrete count data in the traditional frequency approaches) and in data that are left-censored at zero because accidents may not be reported on some roadway segments during the time period over which data are collected. This censoring may occur for a number of reasons ranging from the simple possibility that no accidents occurred on the roadway segment over the study period, to the possibility that accidents not involving injury may not be reported if their property damage does not exceed a specified threshold (this threshold is open to the interpretation of the officer at the accident scene and thresholds may vary from one jurisdiction to the next).

* Corresponding author. Tel.: +1 765 494 7760.
E-mail addresses: panast@purdue.edu (P.Ch. Anastasopoulos), shankarv@engr.psu.edu (V.N. Shankar), jhaddock@purdue.edu (J.E. Haddock), flm@purdue.edu (F.L. Mannering).
[1] Tel.: +1 814 865 9434.
[2] Tel.: +1 765 496 3996.
[3] Tel.: +1 765 496 7913.

0001-4575/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.aap.2011.11.006

In the traditional tobit framework, which models overall accident rates irrespective of their injury severity, the equivalent effect of these censoring sources is assumed to be the same across all accident-injury-severity levels – which can be problematic if no-injury accidents, for example, are less likely to be reported. A solution to this would be to model the accident rates by injury-severity level. However, if accident rates by injury-severity level are modeled independently, significant estimation error could be introduced because unobserved effects at the roadway-segment level are likely to be shared across severities. This general problem of shared unobserved effects[4] has been previously encountered in traditional accident-frequency modeling where, instead of modeling the total number of accidents, separate count-data models are estimated for each accident-injury type. This has led to the application of bivariate models, where two accident types are considered (Subrahmaniam and Subrahmaniam, 1973; Maher and Summersgill, 1996; N'Guessan et al., 2006; Geedipally and Lord, 2010; N'Guessan, 2010), and multivariate models, where three or more accident types are considered (Winkelmann, 2008; Bijleveld, 2005; Ma and Kockelman, 2006; Song et al., 2006; Park and Lord, 2007; Ma et al., 2008; El-Basyouny and Sayed, 2009; Park et al., 2010). The intent of this paper is to address both issues – censoring masking, and contemporaneous error correlation of the latent variables due to commonly shared unobserved characteristics – simultaneously, by demonstrating the multivariate tobit model as an approach of dealing with correlation among injury-severity models when the data are continuous and censored (accident rates on roadway segments) as opposed to the more traditional count-frequency data.

2. Methodology

Tobit regression (Tobin, 1958) is used with left-censored (censored at a low threshold) or right-censored (censored at a high threshold) dependent variables. For the accident rate of specific injury-severity levels (defined by the degree of injury of the most severely injured occupant in the accident), the data can be left-censored with a clustering at zero (zero accidents of a specified injury-severity per 100-million vehicle miles traveled) because, as discussed earlier, accidents of a specific injury severity may not be observed simply because none have occurred or due to the non-availability of data (accidents that do not involve injury, typically are only reported if the property value damage exceeds a pre-specified value threshold). In tobit regression, the equivalent effect is assumed to be the same, clustered-at-zero observations.

To address the possibility of multiple tobit regressions by injury-severity level (no-injury crashes per 100-million vehicle miles traveled, possible-injury per 100-million vehicle miles traveled, and injury crashes per 100-million vehicle miles traveled), past research that has dealt with estimation techniques to model multiple tobit equations with contemporaneous (cross-equation) error correlation (Huang et al., 1987; Huang, 1999; Trivedi and Zimmer, 2005) can be used to develop a multivariate tobit model for accident rates on specific roadway segments[5] considering accidents whose most severely injured vehicle occupant was either not injured, possibly injured, or injured.[6] Using a left-censored limit of zero, the multivariate tobit model with three dependent variables is expressed as:

$$\begin{array}{lll} Y_{ik}^* = \mathbf{X}_{ik}'\boldsymbol{\beta}_k + \varepsilon_{ik}, \; i = 1, 2, \ldots, N, & k = 1, 2, 3 & \\ \quad Y_{ik} = Y_{ik}^* & \text{if } \; Y_{ik}^* > 0 & , \\ \quad \;\; = 0 & \text{if } \; Y_{ik}^* \leq 0, & \end{array} \tag{1}$$

where N is the number of observations, Y_{ik} is the dependent variable for the kth accident-injury-severity rate (1–3 for no-injury, possible injury, and injury accidents per 100-million vehicle miles traveled in roadway segment i, respectively) for the ith segment, $\mathbf{X}_{ik}'$ is a vector of independent variables (pavement, traffic, weather and roadway segment characteristics), $\boldsymbol{\beta}_k$ is a vector of estimable parameters, and ε_{ik} are multivariate normally and independently distributed error terms with zero mean, variance σ^2, correlation ρ, and covariance matrix:

$$\Sigma_{\varepsilon_k} = \begin{pmatrix} \sigma_{\varepsilon_1}^2 & \rho_{\varepsilon_2\varepsilon_1}\sigma_{\varepsilon_2}\sigma_{\varepsilon_1} & \rho_{\varepsilon_3\varepsilon_1}\sigma_{\varepsilon_3}\sigma_{\varepsilon_1} \\ \rho_{\varepsilon_1\varepsilon_2}\sigma_{\varepsilon_1}\sigma_{\varepsilon_2} & \sigma_{\varepsilon_2}^2 & \rho_{\varepsilon_3\varepsilon_2}\sigma_{\varepsilon_3}\sigma_{\varepsilon_2} \\ \rho_{\varepsilon_1\varepsilon_2}\sigma_{\varepsilon_1}\sigma_{\varepsilon_3} & \rho_{\varepsilon_2\varepsilon_3}\sigma_{\varepsilon_2}\sigma_{\varepsilon_3} & \sigma_{\varepsilon_3}^2 \end{pmatrix}. \tag{2}$$

And given these error terms, the density function of Y_{ik} is (Trivedi and Zimmer, 2005):

$$fk(Y_{ik}|\mathbf{X}_{ik}\boldsymbol{\beta}_k') = \prod_{Y_{ik}=0}\left[1 - \Phi\left(\frac{\mathbf{X}_{ik}'\boldsymbol{\beta}_k}{\sigma_k}\right)\right]\prod_{Y_{ik}>0}\phi\left(\frac{Y_{ik} - \mathbf{X}_{ik}'\boldsymbol{\beta}_k}{\sigma_k}\right), \tag{3}$$

where Φ is the multivariate normal distribution function and ϕ is the multivariate normal density function. Eq. (1) shows that there is an implicit, stochastic index (latent variable) equal to Y_{ik}^* which is observed only when positive.

The multivariate (trivariate) tobit distribution is:

$$F(Y_1, Y_2, Y_3) = C[F_1(Y_{i1}|\mathbf{X}_{i1}'\boldsymbol{\beta}_1), F_2(Y_{i2}|\mathbf{X}_{i2}'\boldsymbol{\beta}_2), F_3(Y_{i3}|\mathbf{X}_{i3}'\boldsymbol{\beta}_3); \theta], \tag{4}$$

and the corresponding log-likelihood function for the multivariate (trivariate) tobit model is:

$$\begin{aligned} & L_N[(Y_1|\mathbf{X}_1; \boldsymbol{\beta}_1), (Y_2|\mathbf{X}_2; \boldsymbol{\beta}_2), (Y_3|\mathbf{X}_3; \boldsymbol{\beta}_3); \theta] \\ & \quad = \sum_{i=1}^{N}\sum_{k=1}^{3} \ln f_{ik}\left(Y_{ik}|\mathbf{X}_{ik}; \boldsymbol{\beta}_k\right) + \sum_{i=1}^{N} C_{123}\left[F_1\left(Y_{i1}|\mathbf{X}_{i1}; \boldsymbol{\beta}_{i1}\right),\right. \\ & \quad \left. F_2\left(Y_{i2}|\mathbf{X}_{i2}; \boldsymbol{\beta}_{i2}\right), F_3\left(Y_{i3}|\mathbf{X}_{i3}; \boldsymbol{\beta}_{i3}\right); \theta\right], \end{aligned} \tag{5}$$

where $C_{123}(\cdot)$ is the cross partial derivative for the copula (the function linking marginal variables into the multivariate distribution; see Trivedi and Zimmer (2005) and Prokhorov and Schmidt (2009) for specifics on the copula estimation), θ the dependence

[4] Such unobserved effects may be capturing driver-specific information (age, gender, marital status, socioeconomic status, risk taking, driving experience, driving behavior, driving adjustment in situational responses, etc.), or information on vehicle characteristics, such as vehicle type, safety features (airbags, anti-lock brakes, etc.), and horsepower (see Janssen, 1994; Dee, 1998; Winston et al., 2006). Such information is typically available only after an accident has occurred and thus cannot be used to predict accident rates.

[5] Note that modeling accident rates using tobit regression is an appropriate solution, in order to avoid biased and inconsistent parameter estimates that would otherwise be the result of the ordinary least squares (OLS) estimation which does not account for the censoring in the data. The same limitations exist when modeling censored data as a system of equations (such as seemingly unrelated regression equations, two-stage least squares, or three-stage least squares), which implicitly assume that the zero observations are generated by the same process that generates the positive observations (see Washington et al., 2011). The univariate and multivariate tobit models by definition assume different generation processes for the zero and non-zero observations (see Eq. (1)), which makes them suitable for the analysis and modeling of left-censored, continuous data, as the accident rates.

[6] Evident injury rates include injury severity levels classified as evident injury, disabling injury and fatal injury. The disabling and fatal injury rates individually resulted in a sparse column of positive rates, while disabling and fatal injuries combined did not significantly improve the column sparseness of non-zero values. It is acknowledged that this is a limiting assumption in that correlations are constrained to be the same within evident, disabling and fatal injury categories due to the composite definition.

parameter assumed to be a scalar, and the rest of the terms as previously defined. The copula density is given as:

$$c[F_1(\cdot), F_2(\cdot), F_3(\cdot)] = C_{123}[(F_1(\cdot), F_2(\cdot), F_3(\cdot))f_1(\cdot)f_2(\cdot)f_3(\cdot)], \tag{6}$$

where

$$C_{123}[(F_1|\mathbf{X}_1; \boldsymbol{\beta}_1), (F_2|\mathbf{X}_2; \boldsymbol{\beta}_2), (F_3|\mathbf{X}_3; \boldsymbol{\beta}_3); \theta] = \frac{\partial C[(F_1|\mathbf{X}_1; \boldsymbol{\beta}_1), (F_2|\mathbf{X}_2; \boldsymbol{\beta}_2), (F_3|\mathbf{X}_3; \boldsymbol{\beta}_3); \theta]}{\partial F_1 \partial F_2 \partial F_3}. \tag{7}$$

And for model estimation, the use of copula methods allows for the use of standard maximum likelihood method, where the estimates are obtained by solving the nonlinear equations:

$$\frac{\partial L_N}{\partial \Omega} = 0, \quad \text{where} \quad \Omega = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \theta). \tag{8}$$

Assuming that the solution, $\hat{\Omega}$, is consistent for the true parameter vector Ω_0, its asymptotic distribution is (see Trivedi and Zimmer, 2005; Prokhorov and Schmidt, 2009):

$$\sqrt{N}(\hat{\Omega} - \Omega_0) \xrightarrow{d} \mathrm{N}\left[0, -\left(p \lim\left(\frac{1}{N}\right)\left(\frac{\partial^2 L_N(\Omega)}{\partial \Omega \partial \Omega'}\right)\right)^{-1}_{\Omega_0}\right]. \tag{9}$$

3. Data

To illustrate the application of the multivariate tobit model, the data previously presented in Milton et al. (2008) are used. These data consist of historical motor vehicle accident data from multilane divided highways in Washington over a 5-year period (1990–1994). A total of 274 roadway segments of varying lengths (the mean segment length was 2.4 miles, and the standard deviation 2.7 miles) were defined by dividing the data into homogenous roadway segments (defined by roadway geometrics)[7] and the police-reported vehicle accidents (injuries, possible injuries, and property damage only)[8] occurring on each segment over the five-year period were obtained from the Washington State Department of Transportation. The accident data were combined with weather data from the Western Regional Climate Center which included total precipitation (all forms) and snowfall precipitation. Detailed roadway-segment, pavement, and traffic information was acquired from the Washington State Department of Transportation databases. The resulting database is very rich in terms of its roadway segment-level detail. Due to the very large amount of data available, only the summary statistics of key variables are presented in this paper, as shown in Table 1.

For model estimation, the data included the aggregated number of no-injury, possible injury, and injury accidents on each roadway segment, i, over the five-year period of analysis. The accident rate (number of accidents per 100-million VMT) by injury-severity level, k, was calculated as:

$$Accident \quad Rate_i^k = \frac{\sum_{t=1}^{n} Accident_{t,i}^k}{\left[\sum_{t=1}^{n} AADT_{t,i} \times L_i \times 365\right] / 100,000,000}, \tag{10}$$

where $Accident\ Rate_i^k$ is the number of accidents per 100-million VMT in the three injury-severity categories on roadway segment i,

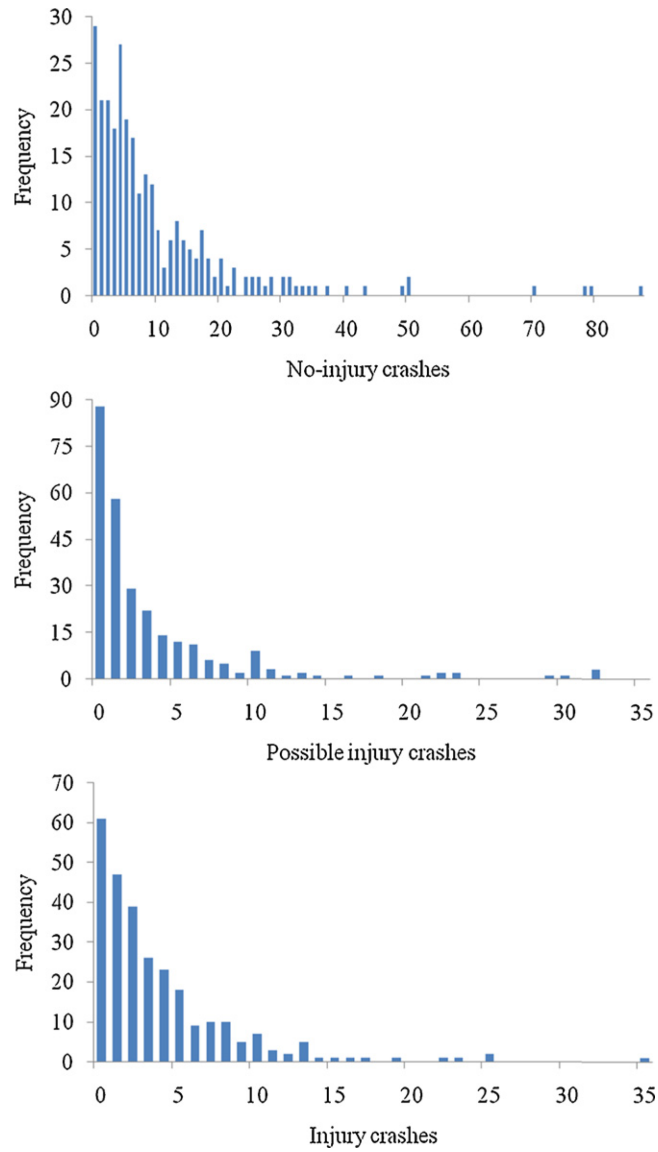

Fig. 1. Histograms of no-injury, possible injury, and injury crashes.

k is the injury-severity level defined as no-injury, possible injury, and injury, t denotes the year (from 1 to n), $Accident_{t,i}^k$ is the number of injury-severity level k accidents in year t on segment i, $AADT_{t,i}$ the average annual daily traffic in year t on segment i, and L_i the length of roadway segment i in miles.

Of the 274 road segments, 29 did not have any no-injury accidents (245 had at least one), 88 did not have any possible injury accidents (186 had at least one), and 61 did not have any injury accidents (213 had at least one) over the five-year analysis period. The frequency of the accident rates by injury-severity level is presented in Fig. 1.

4. Model estimation results

Table 2 presents the univariate (three separate tobit models – one for each injury-severity level – estimated without cross equation correlation) and multivariate tobit estimation results, and Table 3 the models' goodness-of-fit. To assess the influence of specific variables, Table 4 presents the computed values for the change in the expected values of no-injury, possible injury, and injury accident rates for roadway segments and the change in the

[7] The segment-defining information included primarily median treatments (safety barriers, cables or landform barriers). The beginning and ending points of the roadway segments were matched with the beginning and ending points of a barrier, respectively.

[8] As previously mentioned, due to the limited number of accidents that resulted in disabling injury and fatality, the injury category included evident injury, disabling injury and fatality. With this categorization, 12,612 individual accidents resulted in no injury, 4866 in possible injury, and 5090 in injury as the most severe outcome of the accident.

Table 1
Descriptive statistics of key variables.

	Mean	Std Dev	Minimum	Maximum
Total number of no-injury crashes	9.749	12.587	0	87
Total number of possible injury crashes	3.415	5.623	0	32
Total number of injury crashes	3.771	4.762	0	35
Traffic characteristics				
Average annual daily traffic (in 10,000)	3.735	3.703	0.335	17.256
Average annual daily traffic of passenger cars (in 10,000)	3.319	3.484	0.280	16.404
Average annual daily traffic of single unit trucks (in 10,000)	0.140	0.115	0.017	0.538
Daily percentage of trucks (single unit, tractor, trailer, and two-trailer trucks)	0.142	0.067	0.032	0.32
Road segments with access control	1.45%		0	1
Road segments without access control	98.55%		0	1
Road geometrics				
Segment length in miles	2.430	2.694	0.5	19.3
Total number of lanes on the road segment	4.604	1.133	2	8
Road segments with 5 or more lanes in total	26.18%		0	1
Number of interchanges in the segment	0.847	0.836	0	4
Road segments with interchanges	61.09%		0	1
Road segments without interchanges	38.91%		0	1
Road segments with median width less than 30 ft	4.73%		0	1
Road segments with median width between 30 and 40 ft	27.64%		0	1
Road segments with median width between 40 and 50 ft	11.64%		0	1
Road segments with median width between 50 and 60 ft	5.82%		0	1
Road segments with median width greater than 60 ft	50.18%		0	1
Minimum median shoulder in feet	4.484	1.686	0	10
Maximum median shoulder in feet	5.305	2.496	0	18
Number of grade breaks in the segment	3.865	4.095	0	28
Road segments with 4 or less grade breaks	68.73%		0	1
Road segments with horizontal curves	90.91%		0	1
Road segments without horizontal curves	9.09%		0	1
Maximum grade in the segment (in percent)	−0.218	3.072	−5.5	6.72
Road segments with maximum grade between −1% and 1%	9.45%		0	1
Minimum radius curve on roadway segment (in 1000 ft)	4.267	4.882	0	38.4
Tangent length (miles) in the segment per mile	0.630	1.425	0	18.941
Road segments with flat slope	64.36%		0	1
Road segments with slight slope	19.66%		0	1
Road segments with medium slope	12.75%		0	1
Road segments with high slope	3.23%		0	1
Pavement condition				
Pavement friction (scaled 0–100)	49.654	4.738	38.8	61.5
Weather				
Average annual precipitation in inches	37.771	28.874	4.56	131.76
Average annual snowfall in inches	20.319	60.061	0	651.96

probability of having a 100-million VMT accident rate above zero for the accident-observation period. The estimation results provided in Table 2 show that the estimated parameters are significant and of plausible sign.

It should be mentioned here that the comparison of the estimated multivariate tobit model with its univariate counterparts clearly indicated the statistical superiority of the first (in terms of goodness-of-fit and prediction accuracy).[9] This is further evidenced in Table 2 from the cross-equation correlation coefficients (ρ) of the error terms of the three latent variables, all of which are statistically significant at 99.9% level of confidence, and show positive association between no-injury, possible injury, and injury accident rates.[10]

Moreover, Table 2 shows that a number of variables (the segment slope in the no-injury and injury models, the pavement friction in the no-injury model, and the average annual snowfall on segments with horizontal curves and on straight segments on the possible injury model) are not statistically significant in the univariate tobit models, but are found to be significant in the multivariate model. This indicates that the multivariate tobit model is capturing underlying effects (possibly those of the variables that are statistically insignificant in the univariate tobit models) that may be masked in the error terms of the univariate models, hence explaining more of the variance in the data.

Turning to the specific estimation results of the multivariate tobit model, Table 4 shows that a number of traffic characteristics affect the accident-injury-severity rates. For example, it is found that as the annual average daily travel (AADT) of passenger cars (in 10,000 s of passenger cars per day) and single unit trucks

[9] With regard to overall goodness-of-fit, Veall and Zimmermann (1996) extensively discuss this topic in the context of the tobit model. They conclude that the Maddala pseudo R^2 is a valid measurement (Maddala, 1983). The Maddala pseudo-R^2 is computed as $1 - \exp[-2(LL(\beta) - LL(0))/N]$ were $LL(\beta)$ is the log-likelihood at convergence, $LL(0)$ is the log-likelihood at zero and N is the number of observations. The overall statistical fit of the multivariate tobit model was better than its univariate counterparts, as indicated by the Maddala pseudo-R^2 of 0.496 for the multivariate model, and 0.129, 0.247, and 0.22 for the univariate no-injury, possible injury, and injury models, respectively. And with respect to the predicting accuracy, it is found that the forecasts of the multivariate tobit model under or overestimate the observed values by roughly 17%, whereas the forecasts of the univariate tobit models under or overestimate the observed values by roughly 22%.

[10] Note that the statistically significant cross-equation correlation coefficients (ρ) of the error terms of the three latent variables indicate that there are commonly shared unobserved factors among the three levels of injury-severity rates. The sharing of these unobserved factors appears to be stronger between the no injury and possible injury levels, as compared to the no injury and injury, and possible injury and injury levels, with the later being the least strong. The signs of the ρ terms are in all three cases positive, indicating the increasing indirect effect of the commonly shared unobserved factors on the injury-severity rates. Thus, a road segment with a high injury rate may also have a high possible injury rate due to commonly shared unobserved characteristics, which in turn indicate that the specific road segment may have potential safety issues that are not captured by the available information.

Table 2
Univariate and multivariate tobit model estimation results.

	Univariate tobit models			Multivariate tobit model		
	Coefficient	*t*-Ratio	*P*-Value	Coefficient	*t*-Ratio	P-Value
Constant [*NI*]	150.739	4.12	0.000	143.309	4.35	0.000
Constant [*PI*]	138.844	2.07	0.038	146.630	2.44	0.015
Traffic characteristics						
Average annual daily traffic of passenger cars (in 10,000) [*NI*]	−2.848	−2.21	0.027	−2.635	−2.42	0.016
Average annual daily traffic of single unit trucks (in 10,000) [*I*]	−57.465	−4.28	0.000	−58.038	−4.49	0.000
Daily percentage of trucks (all categories) [*NI*]	−159.402	−2.95	0.003	−106.807	−2.23	0.026
Daily percentage of trucks (all categories) [*PI*]	−134.007	−4.94	0.000	−115.518	−4.85	0.000
Access control on segments with horizontal curves (1 if no access control, 0 otherwise) [*PI*]	31.068	2.31	0.021	25.671	2.32	0.020
Access control on segments with horizontal curves (1 if no access control, 0 otherwise) [*I*]	27.834	2.04	0.042	24.658	1.99	0.047
Road geometrics						
Median width (1 if less than 40 ft; 0 otherwise) [*NI*]	−57.628	−3.40	0.001	−36.756	−2.43	0.015
Minimum median shoulder in feet [*NI*]	−11.229	−4.26	0.000	−8.491	−3.99	0.000
Maximum median shoulder in feet [*NI*]	5.473	3.21	0.001	4.568	3.08	0.002
Maximum median shoulder in feet [*PI*]	1.895	2.37	0.018	1.363	1.93	0.054
Tangent length (miles) in the segment per mile [*I*]	−5.204	−2.47	0.013	−5.482	−2.94	0.003
Minimum radius curve on roadway segment (in 1000 ft) [*PI*]	0.583	1.69	0.090	0.529	2.01	0.045
Segment slope (1 if flat, 0 otherwise) [*I*]	−5.140	−1.59*	0.112	−5.248	−1.82	0.068
Segment slope (1 if medium or high, 0 otherwise) [*NI*]	−23.532	−1.24*	0.214	−28.081	−1.79	0.073
Segment slope (1 if medium or high, 0 otherwise) [*I*]	18.167	2.06	0.039	15.883	1.92	0.055
Maximum grade in the segment (1 if between −1% and 1%, 0 otherwise) [*I*]	−14.648	−2.64	0.008	−14.463	−2.93	0.003
Number of grade breaks in the segment (1 if 4 or less, 0 otherwise) [*PI*]	−8.531	−2.39	0.017	−6.789	−2.45	0.015
Total number of lanes on the road segment (1 if 5 or more, 0 otherwise) [*PI*]	−18.480	−3.60	0.000	−13.649	−3.25	0.001
Number of interchanges in the segment (1 if none, 0 otherwise) [*PI*]	−8.047	−2.34	0.019	−5.962	−2.21	0.027
Pavement condition						
Pavement friction (scaled 0–100) [*NI*]	−1.041	−1.57*	0.117	−1.223	−2.04	0.042
Natural logarithm of pavement friction (scaled 0–100) [*PI*]	−29.989	−1.75	0.080	−32.067	−2.09	0.036
Natural logarithm of pavement friction (scaled 0–100) [*I*]	7.434	8.14	0.000	7.589	9.00	0.000
Weather						
Average annual precipitation (in inches) on segments with horizontal curves [*PI*]	0.158	2.29	0.022	0.129	2.36	0.018
Average annual precipitation (in inches) on straight segments [*PI*]	0.370	2.26	0.024	0.314	2.43	0.015
Average annual precipitation (in inches) on straight segments [*I*]	0.251	1.74	0.083	0.236	1.83	0.067
Average annual snowfall (in inches) on segments with horizontal curves [*PI*]	−0.049	−1.28*	0.201	−0.055	−1.79	0.073
Average annual snowfall (in inches) on segments with horizontal curves [*I*]	0.089	3.65	0.000	0.082	3.81	0.000
Average annual snowfall (in inches) on straight segments [*PI*]	−1.925	−1.49*	0.136	−2.112	−1.95	0.051
Average annual snowfall (in inches) on straight segments [*I*]	−1.455	−1.82	0.069	−1.299	−1.72	0.086
σ [*NI*]	48.762	21.79	0.000	49.184	21.56	0.000
σ [*PI*]	23.942	18.55	0.000	21.969	19.62	0.000
σ [*I*]	23.638	19.90	0.000	24.435	20.22	0.000
Correlation (ρ) between endogenous variables						
No injury vs. possible injury				0.595	13.66	0.000
No injury vs. injury				0.458	9.00	0.000
Possible injury vs. injury				0.337	5.44	0.000

NI, no-injury crash rates; *PI*, possible injury crash rates; *I*, injury crash rates.
* The variable is statistically insignificant at 0.90 level of confidence.

(in 10,000 s of single unit trucks per day) increases by 10,000 units, the number of no-injury and injury accidents per 100-million VMT is expected to decrease by 2.079 and 4.282, respectively, and the probability of having a 100-million VMT no-injury and injury accident rate above zero decreases by 0.979% and 7.1%, respectively. Also, a 10% increase in the daily percentage of trucks is found to decrease the number of no-injury and possible injury accidents per 100-million VMT by 8.455 and

Table 3
Goodness-of-fit measures of the univariate and multivariate tobit models.

	Univariate tobit models			Multivariate tobit model
	NI	*PI*	*I*	
LL(0)	−1351	−953	−1063	−3289
LL(β)	−1332	−914	−1029	−3195
Maddala pseudo R^2	0.129	0.247	0.220	0.496
N	274	274	274	274

NI, no-injury crash rates; *PI*, possible injury crash rates; *I*, injury crash rates.

Table 4
Sensitivity of estimated multivariate tobit regression parameters.

	Multivariate tobit	
	Overall sensitivity[a]	Zero sensitivity[b] (%)
Traffic characteristics		
Average annual daily traffic of passenger cars (in 10,000) [*NI*]	−2.079	−0.979
Average annual daily traffic of single unit trucks (in 10,000) [*I*]	−4.282	−7.996
10% increase in truck traffic (all categories) [*NI*]	−8.455	−3.968
10% increase in truck traffic (all categories) [*PI*]	−6.796	−18.872
Access control on segments with horizontal curves (1 if no access control, 0 otherwise) [*PI*]	15.173	41.937
Access control on segments with horizontal curves (1 if no access control, 0 otherwise) [*I*]	18.187	30.163
Road geometrics		
Median width (1 if less than 40 ft, 0 otherwise) [*NI*]	−28.891	−13.654
Minimum median shoulder in feet [*NI*]	−6.689	−3.154
Maximum median shoulder in feet [*NI*]	3.598	1.697
Maximum median shoulder in feet [*PI*]	0.802	2.226
Tangent length (miles) in the segment per mile [*I*]	−4.036	−6.706
Minimum radius curve on roadway segment (in 1000 ft) [*PI*]	0.313	0.864
Segment slope (1 if flat, 0 otherwise) [*I*]	−3.889	−6.420
Segment slope (1 if medium or high, 0 otherwise) [*NI*]	−22.082	−10.431
Segment slope (1 if medium or high, 0 otherwise) [*I*]	11.726	19.429
Maximum grade in the segment (1 if between −1% and 1%, 0 otherwise) [*I*]	−10.686	−17.693
Number of grade breaks in the segment (1 if 4 or less, 0 otherwise) [*PI*]	−4.007	−11.091
Total number of lanes on the road segment (1 if 5 or more, 0 otherwise) [*PI*]	−8.017	−22.298
Number of interchanges in the segment (1 if none, 0 otherwise) [*PI*]	−3.502	−9.741
Pavement condition		
Pavement friction (scaled 0–100) [*NI*]	−0.965	−0.454
Natural logarithm of pavement friction (scaled 0–100) [*PI*]	−18.887	−52.388
Natural logarithm of pavement friction (scaled 0–100) [*I*]	5.599	9.283
Weather		
Average annual precipitation (in inches) on segments with horizontal curves [*PI*]	0.076	0.211
Average annual precipitation (in inches) on straight segments [*PI*]	0.185	0.514
Average annual precipitation (in inches) on straight segments [*I*]	0.175	0.289
Average annual snowfall (in inches) on segments with horizontal curves [*PI*]	−0.033	−0.090
Average annual snowfall (in inches) on segments with horizontal curves [*I*]	0.060	0.100
Average annual snowfall (in inches) on straight segments [*PI*]	−1.243	−3.450
Average annual snowfall (in inches) on straight segments [*I*]	−0.957	−1.589

NI, no-injury crash rates; *PI*, possible injury crash rates; *I*, injury crash rates.

[a] Change in the overall expected value $\partial E[Y]/\partial X$, where $E[Y]$ (removing the subscripts to simplify the exposition) is the expected value of the dependent variable for all cases (see Anastasopoulos et al., 2008).

[b] Percent change in the cumulative probability of being above zero, using $\partial F[z]/\partial X = \beta_k f(z)/\sigma$ (removing the subscripts to simplify the exposition) and converting to percent, where $F(z)$ is the cumulative normal distribution function associated with the proportion of cases above the zero, $f(z)$ the unit normal density (value of the derivative of the normal curve at a particular point), and σ the standard deviation of the error term (see Anastasopoulos et al., 2008).

6.796, respectively, and found to decrease the probability of having a 100-million VMT no-injury and possible injury accident rate above zero by 3.968% and 18.872%, respectively. These findings are consistent with a number of previous studies that have shown that accident rates are higher on low-traffic volume roads, and significantly decrease with increasing volumes (Anastasopoulos and Mannering, 2009; Qi et al., 2007; Dickerson et al., 2000; Zhou and Sisiopiku, 1997), and in line with previous work that has suggested that higher truck percentages may be reflecting greater driver experience (truck drivers) and may also be having a calming effect on traffic (for example, Shankar et al., 1997; Anastasopoulos et al., 2008).

Segments with horizontal curves and no access control are found to increase the number of possible injury and injury accidents per 100-million VMT by 15.173 and 18.187, respectively, and increase the probability of having a 100-million VMT possible injury and injury accident rate above zero by 41.937% and 30.163%, respectively. Given the strong effect of this variable on accident-injury-severity rates, these results indicate that locations with horizontal curves and no access control roadway segments should be viewed as safety “hot spots” that require some treatment (access control, warning signs in advance of the curves, etc.).

The roadway-geometric variables included in the multivariate model (median and shoulder width, tangent length, curve radius, segment slope, grade, grade breaks in the segment, number of lanes, and number of interchanges) were found to have variable effects across the three accident-injury-severity rates. For example, narrow medians (less than 40 ft) tend to have a 28.891 lower no-injury accident rate, and a 13.654% lower probability of having a 100-million VMT no-injury accident rate above zero. Typically, narrow (less than 40 ft) medians are treated with barriers, and because median slopes become flatter as the median width increases, the median width variable may be capturing the effect of segments with slopes that are flat enough to prevent severe overturning accidents but do not have adequate median widths to prevent severe median crossover accidents (Shankar et al., 1998).

It is also found that a 1 ft increase in the minimum median shoulder width decreases the number of no-injury accidents per 100-million VMT by 6.689, and decreases the probability of having a 100-million VMT no-injury accident rate above zero by 3.154%. Whereas, a 1 ft increase in the maximum median shoulder width increases the number of no-injury and possible injury accidents per 100-million VMT by 3.598 and 0.802, respectively, and increases the probability of having a 100-million VMT no-injury and possible injury accident rate above zero by 1.697% and 2.226%, respectively. These results are in line with past studies (for example Knuiman et al., 1993; Shankar et al., 1997, 1998; Anastasopoulos and Mannering, 2009, 2011) that have shown that median shoulder widths may have variable effects on accident occurrences and severities.

Curve characteristics are also found to affect accident-injury-severity rates. A unit increase in the tangent length (in miles) in the roadway segment per mile resulted in a 4.036 decrease in the number of injury accidents per 100-million VMT and a 6.706%

lower probability of having a 100-million VMT injury accident rate above zero. Whereas, a 1000 ft increase of the minimum curve radius resulted in a 0.313 increase in the number of possible injury accidents per 100-million VMT and a 0.864% higher probability of having a 100-million VMT possible injury accident rate above zero. Flat slopes (0–2% grades) on average have a 3.889 lower injury accident rate, and a 6.42% lower probability of having a 100-million VMT injury accident rate above zero; whereas, medium (greater than 2% to 4% grades) or high slopes (greater than 4% grades) are found to have a 22.082 lower no-injury accident rate on average and a 10.431% lower probability of having a 100-million VMT no-injury accident rate above zero, and 11.726 higher injury accident rate and a 19.429% higher probability of having a 100-million VMT injury accident rate above zero. Furthermore, level segments (with a maximum grade in the segment ranging from −1% to 1%), tend to have a 10.686 lower injury accident rate, and a 17.693% lower probability of having a 100-million VMT injury accident rate above zero. Finally, segments with four or less grade breaks (defined by the presence of a vertical curve or a vertical point of inflection for grade changes that would result in vertical curves below minimum curve lengths), segments with more than 5 lanes (in both directions), and segments with no interchanges, are found to have a 4.007, 8.017, and 3.502 lower possible injury accident rate, respectively, and a 11.091%, 22.298%, and 9.741% lower probability, respectively, of having a 100-million VMT injury accident rate above zero.

With regard to pavement characteristics, pavement friction is typically measured on 0–100 friction scale, with friction considered to be good if its value is 40 or above. Given this, the results indicate that a one unit increase in pavement friction results in a 0.965 decrease of the number of no-injury per 100-million VMT, and in a 0.454% lower probability of having a 100-million VMT no-injury and possible injury accident rate above zero, respectively. This result supports the earlier finding of Shankar et al. (1995), Noyce and Bahia (2005), Milton et al. (2008), and Anastasopoulos and Mannering (2009, 2011), that showed that higher pavement friction lowered the risk of accident. Interestingly, an increase in pavement friction results in an increase in the number of injury accidents per 100-million VMT, and in a higher probability of having a 100-million VMT injury accident rate above zero. This finding is somewhat surprising and could be capturing unobserved heterogeneity, possibly relating to, among other factors, the potential reactions of drivers with respect to good pavement quality (for example, Mannering, 2009, found that drivers tend to drive faster than usual on roads with good pavement condition).

The last findings relate to the effect of average annual precipitation and snowfall in conjunction with curve presence in the segment. Table 4 shows that a 1-in. increase in the annual precipitation on segments with horizontal curves results in a 0.076 increase on the number of possible injury accidents per 100-million VMT, and in a 0.211% higher probability of having a 100-million VMT possible injury accident rate above zero. A 1-in. increase in the annual precipitation on straight segments (without horizontal curves) results in a 0.185 and 0.175 increase of the number of possible injury and injury accidents per 100-million VMT, respectively, and in a 0.514% and 0.289% higher probability of having a 100-million VMT possible injury and injury accident rate above zero, respectively.

With respect to snowfall, it is found that a 1-in. increase in the average annual snowfall on segments with horizontal curves results in a 0.033 decrease in the number of possible injury accidents per 100-million VMT, and in a 0.09% lower probability of having a 100-million VMT possible injury accident rate above zero; whereas, a 1-in. increase in the average annual snowfall on segments with horizontal curves results in a 0.06 increase in the number of injury accidents per 100-million VMT, and in a 0.1% higher probability of having a 100-million VMT possible injury accident rate above zero. And, a 1-in. increase in the average annual snowfall on straight segments (without horizontal curves) results in a 1.243 and 0.957 decrease of the number of possible injury and injury accidents per 100-million VMT, respectively, and in a 3.45% and 1.589% lower probability of having a 100-million VMT possible injury and injury accident rate above zero, respectively.

These results are in line with Milton et al. (2008), and are likely picking up geographical differences in driving behavior in response to rain and snow conditions. It appears that drivers generally respond to higher snowfalls by driving more carefully, thus decreasing the likelihood of more severe accidents (possible injuries and injuries). However, this compensating behavior (driving slower) may not be sufficient to overcome the reduced friction and other adverse factors associated with rain and snow. Therefore, increasing precipitation (regardless of the presence of a horizontal curve) increases the possible injury and injury accident rates (and their respective probabilities of having an accident-injury-severity rate above zero), and increasing snowfall on segments with horizontal curves also increases the injury accident rate (and its respective probability of having an accident-injury-severity rate above zero).

5. Model evaluation

To further evaluate the multivariate tobit model, a multivariate negative binomial model was also estimated (the overdispersion parameters, α, indicated the superiority of the multivariate negative binomial model over its Poisson counterpart).[11] As described in Winkelmann (2008) and Zhao and Kockelman (2002), with the expected number of accidents:

$$\lambda_{ik} = \exp(\mathbf{X}'_{ik}\boldsymbol{\beta}_k + \varepsilon_{ik}), \quad i = 1, 2, \ldots, N, \quad k = 1, 2, 3, \tag{11}$$

where $\exp(\varepsilon_{ik})$ is a multivariate gamma-distributed error term with mean 1 and variance a^{-1}, the multivariate negative binomial model has a joint probability function:

$$f(Y_{1k}, \ldots, Y_{ik}) = \sum_{m=0}^{S_k} f_{NB}(m) \prod_{i=1}^{N} f_{NB}(Y_{ik} - m), \tag{12}$$

where $s_k = \min(Y_{1k}, \ldots, Y_{ik})$ and for $Z_{ik} = Y_{ik} - m$,

$$f_{NB}(Z_{ik}) = \frac{\Gamma(\lambda_{ik}/\alpha + Z_{ik})}{\Gamma(\lambda_{ik}/\alpha)\Gamma(Z_{ik}+1)} \left(\frac{1}{1+\alpha}\right)^{\lambda_{ik}/\alpha} \left(\frac{\alpha}{1+\alpha}\right)^{Z_{ik}}, \tag{13}$$

where $\Gamma(\cdot)$ is the gamma function, and the dispersion parameter $\alpha = \lambda_{ik}/a$ allows for overdispersion as long as $\alpha > 0$. The model parameters can be estimated by maximizing the log-likelihood function:

$$L = \sum_{i,k} \left\{ \begin{array}{c} \log\left[\Gamma(\alpha^{-1} + Y_{ik})\right] - \log\left[\Gamma(\alpha^{-1})\right] - \log(Y_{ik}!) - \\ \alpha^{-1}\ln[1 + \alpha \exp(\mathbf{X}_i\boldsymbol{\beta}'_k)] - Y_{ik}\log[1 + \alpha^{-1}\exp(\mathbf{X}_i\boldsymbol{\beta}'_k)] \end{array} \right\}. \tag{14}$$

The model estimation results of the two models are presented side-by-side in Table 5, and corresponding goodness-of-fit measures are provided in Table 6. Table 5 shows that the two models are similar, with the multivariate negative binomial model giving 27 statistically significant independent parameters, and the multivariate tobit 31 (to save space, we do not provide a detailed comparison of the individual parameter estimates of these two models). The

[11] A further extension of the multivariate tobit would be to include random parameters as in Anastasopoulos et al. (2012). Unfortunately, allowing for the possibility of random parameters in the context of the multivariate tobit would present a formidable computational problem that is beyond the scope of this paper. This may be a fruitful direction for future research.

Table 5
Multivariate tobit and multivariate negative binomial model estimation results.

	Multivariate negative binomial model			Multivariate tobit model		
	Coefficient	*t*-Ratio	*P*-Value	Coefficient	*t*-Ratio	*P*-Value
Constant [*NI*]	1.083	4.91	0.000	143.309	4.35	0.000
Constant [*PI*]				146.630	2.44	0.015
Traffic characteristics						
AADT of passenger cars (in 10,000) [*NI*]	−0.103	−2.54	0.011	−2.635	−2.42	0.016
AADT of single unit trucks (in 10,000) [*NI*]	−2.748	−2.08	0.038			
AADT of single unit trucks (in 10,000) [*PI*]	−3.988	−5.67	0.000			
AADT of single unit trucks (in 10,000) [*I*]	−1.873	−3.17	0.002	−58.038	−4.49	0.000
Daily percentage of trucks (any category) [*NI*]				−106.807	−2.23	0.026
Daily percentage of trucks (any category) [*PI*]	−8.516	−7.57	0.000	−115.518	−4.85	0.000
Daily percentage of trucks (any category) [*I*]	−2.974	−3.31	0.001			
Access control on segments with horizontal curves (1 if no access control, 0 otherwise) [*PI*]				25.671	2.32	0.020
Access control on segments with horizontal curves (1 if no access control, 0 otherwise) [*I*]				24.658	1.99	0.047
Road geometrics						
Segment length in miles [*NI*]	0.169	6.91	0.000			
Segment length in miles [*PI*]	0.106	3.52	0.000			
Segment length in miles [*I*]	0.143	5.66	0.000			
Median width (1 if less than 40 ft, 0 otherwise) [*NI*]	−1.179	−3.55	0.000	−36.756	−2.43	0.015
Minimum median shoulder in feet [*NI*]	−0.128	−2.78	0.005	−8.491	−3.99	0.000
Maximum median shoulder in feet [*NI*]	0.102	3.10	0.002	4.568	3.08	0.002
Maximum median shoulder in feet [*PI*]	0.072	2.44	0.015	1.363	1.93	0.054
Maximum median shoulder in feet [*I*]	0.055	2.10	0.036			
Tangent length in the segment per mile [*I*]	−0.171	−1.93	0.054	−5.482	−2.94	0.003
Minimum radius curve on roadway segment (in 1000 ft) [*PI*]				0.529	2.01	0.045
Segment slope (1 if flat, 0 otherwise) [*I*]				−5.248	−1.82	0.068
Segment slope (1 if medium or high, 0 otherwise) [*NI*]				−28.081	−1.79	0.073
Segment slope (1 if medium or high, 0 otherwise) [*I*]				15.883	1.92	0.055
Maximum grade in the segment (1 if between −1% and 1%, 0 otherwise) [*I*]	−0.554	−2.43	0.015	−14.463	−2.93	0.003
Number of grade breaks in the segment (1 if 4 or less, 0 otherwise) [*PI*]	−0.744	−4.41	0.000	−6.789	−2.45	0.015
Number of grade breaks in the segment (1 if 4 or less, 0 otherwise) [*I*]	−0.410	−2.85	0.004			
Total number of lanes on the road segment (1 if 5 or more, 0 otherwise) [*NI*]	−0.442	−2.26	0.024			
Total number of lanes on the road segment (1 if 5 or more, 0 otherwise) [*PI*]				−13.649	−3.25	0.001
Number of interchanges in the segment (1 if none, 0 otherwise) [*PI*]	−0.675	−4.58	0.000	−5.962	−2.21	0.027
Number of interchanges in the segment (1 if none, 0 otherwise) [*I*]	−0.356	−2.88	0.004			
Pavement condition						
Pavement friction (scaled 0–100) [*NI*]				−1.223	−2.04	0.042
Natural logarithm of pavement friction (scaled 0–100) [*PI*]	−1.372	−5.36	0.000	−32.067	−2.09	0.036
Natural logarithm of pavement friction (scaled 0–100) [*I*]	0.297	4.83	0.000	7.589	9.00	0.000
Weather						
Average annual precipitation (in inches) on segments with horizontal curves [*PI*]				0.129	2.36	0.018
Average annual precipitation (in inches) on straight segments [*PI*]	0.014	2.13	0.033	0.314	2.43	0.015
Average annual precipitation (in inches) on straight segments [*I*]				0.236	1.83	0.067
Average annual snowfall (in inches) on segments with horizontal curves [*PI*]				−0.055	−1.79	0.073
Average annual snowfall (in inches) on segments with horizontal curves [*I*]				0.082	3.81	0.000
Average annual snowfall (in inches) on straight segments [*PI*]	−0.089	−1.80	0.072	−2.112	−1.95	0.051
Average annual snowfall (in inches) on straight segments [*I*]	−0.043	−1.69	0.091	−1.299	−1.72	0.086
σ [*NI*]				49.184	21.56	0.000
σ [*PI*]				21.969	19.62	0.000
σ [*I*]				24.435	20.22	0.000
α [*NI*]	0.614	9.26	0.000			
α [*PI*]	0.604	6.42	0.000			
α [*I*]	0.456	6.24	0.000			
Correlation between endogenous variables						
Property damage only vs. possible injury	0.535	9.70	0.000	0.595	13.66	0.000
Property damage only vs. injury	0.473	8.27	0.000	0.458	9.00	0.000
Possible injury vs. injury	0.343	5.22	0.000	0.337	5.44	0.000

NI, no-injury crash rates or frequency; *PI*, possible injury crash rates or frequency; *I*, injury crash rates or frequency.

correlation between the endogenous variables is, nonetheless, very similar between the two models.

To further compare these two modeling approaches, the mean absolute deviation (*MAD*) and the mean square error (*MSE*) measures were estimated for the two models, and are illustrated in Table 6. These were estimated as follows (Washington et al., 2011):

$$MAD = \frac{\sum_{i=1}^{N} |\varepsilon_i|}{N}, \tag{15}$$

Table 6
Goodness-of-fit of the multivariate tobit and multivariate negative binomial models.

	Multivariate negative binomial model	Multivariate tobit model tobit
$LL(0)$	−2319	−3289
$LL(\beta)$	−1968	−3195
Maddala pseudo R^2		0.496
McFadden pseudo ρ^2	0.151	
Mean absolute deviation	7.235	7.103
Mean square error	317.008	309.233
N	274	274

and

$$MSE = \frac{\sum_{i=1}^{N} \varepsilon_i^2}{N}. \quad (16)$$

The two models are found to share similar predictive accuracy, even though the multivariate tobit model appears to produce more precise parameter estimates with additional variables found to be statistically significant.

6. Summary and concluding remarks

This paper studies accident rates categorized by injury severities (the most severely injured vehicle occupant in an accident) by estimating a multivariate tobit model of accident injury-severity rates, which accounts for both the underlying masking of the censoring sources, and the contemporaneous error correlation of the latent variables due to commonly shared unobserved characteristics. Using five-year of vehicle accident data from Washington, the model estimation results showed that the effects of the explanatory variables, involving traffic characteristics (average annual daily traffic of passenger cars and single unit trucks, daily percentage of trucks, and access control on segments with curves), road geometrics (median and shoulder width, tangent length, curve radius, slope, grade, grade breaks, number of lanes, and number of interchanges), pavement condition (pavement friction), and weather characteristics (average annual precipitation and snowfall on segments with and without curves), are not the same within different accident-injury-severity categories, something that is implicitly assumed in a standard tobit model approach where the accident rates are modeled irrespective of their injury severity.

Furthermore, a number of variables (the segment slope in the no-injury and injury models, the pavement friction in the no-injury model, and the average annual snowfall on segments with horizontal curves and on straight segments on the possible injury model) were found to be statistically insignificant in the univariate tobit models (models that do not account for the interrelation among injury-severity levels), when they are clearly significant in the multivariate model. This illustrates the potential that the multivariate tobit model has in capturing underlying effects (possibly those, among others, of the variables that are statistically insignificant in the univariate tobit models) that may be masked in the error terms of the univariate models, thus allowing additional inferences to be drawn. In addition, the comparison of the univariate tobit models with their multivariate counterpart clearly indicated the statistical superiority of the latter.

The multivariate tobit model was further evaluated relative to its multivariate negative binomial counterpart. The comparison was based on the predictive accuracy of the two models, which showed that the models are practically equivalent in terms of forecasting. Overall, the empirical findings in this paper suggest that the multivariate tobit model may have considerable potential as an alternative to traditional accident-frequency models.

References

Anastasopoulos, P.Ch., Mannering, F.L., 2009. A note on modeling vehicle-accident frequencies with random parameter count models. Accident Analysis and Prevention 41 (1), 153–159.

Anastasopoulos, P.Ch., Mannering, F.L., 2011. An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. Accident Analysis and Prevention 43 (3), 1140–1147.

Anastasopoulos, P.Ch., Mannering, F.L., Shankar, V.N., Haddock, J.E., 2012. A study of factors affecting highway accident rates using the random-parameters tobit model. Accident Analysis and Prevention, doi:10.1016/j.aap.2011.09.015.

Anastasopoulos, P.Ch., Tarko, A.P., Mannering, F.L., 2008. Tobit analysis of vehicle accident rates on interstate highways. Accident Analysis and Prevention 40 (2), 768–775.

Bijleveld, F.D., 2005. The covariance between the number of accidents and the number of victims in multivariate analysis of accident related outcomes. Accident Analysis and Prevention 37 (4), 591–600.

Dee, T.S., 1998. Reconsidering the effects of seat belt laws and their enforcement status. Accident Analysis and Prevention 30 (1), 1–10.

Dickerson, A., Peirson, J., Vickerman, R., 2000. Road accidents and traffic flows: an econometric investigation. Economica 67 (265), 101–121.

El-Basyouny, K., Sayed, T., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accident Analysis and Prevention 41 (4), 820–828.

Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson-gamma models. Accident Analysis and Prevention 42 (4), 1273–1282.

Huang, H.-C., 1999. Estimation of the SUR Tobit model via the MCECM algorithm. Economics Letters 64 (1), 25–30.

Huang, H.-C., Sloan, F.A., Adamache, K.W., 1987. Estimation of seemingly unrelated Tobit regressions via the EM algorithm. Journal of Business and Economic Statistics 5 (4), 425–430.

Janssen, W., 1994. Seat-belt wearing and driver behavior: an instrumented vehicle study. Accident Analysis and Prevention 26 (2), 249–261.

Knuiman, M.W., Council, F.M., Reinfurt, D.W., 1993. Association of median width and highway accident rates. Transportation Research Record 1401, 70–82.

Lord, D., Mannering, F.L., 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.M., 2006. Bayesian multivariate Poisson regression for models of injury count by severity. Transportation Research Record 1950, 24–34.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accident Analysis and Prevention 40 (3), 964–975.

Maddala, G.S., 1983. Limited-Dependent and Qualitative Variables in Econometrics. Cambridge University Press, New York.

Maher, M.J., Summersgill, I., 1996. A comprehensive methodology for the fitting of predictive accident models. Accident Analysis and Prevention 28 (3), 281–296.

Mannering, F.L., 2009. An empirical analysis of driver perceptions of the relationship between speed limits and safety. Transportation Research Part F 12 (2), 99–106.

Milton, J., Shankar, V., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40 (1), 260–266.

N'Guessan, A., 2010. Analytical existence of solutions to a system of nonlinear equations with application. Journal of Computational and Applied Mathematics 234 (1), 297–304.

N'Guessan, A., Essai, A., N'Zi, M., 2006. An estimation method of the average effect and the different accident risks when modeling a road safety measure: a simulation study. Computational Statistics and Data Analysis 51 (2), 1260–1277.

Noyce, D.A., Bahia, H.U., 2005. Incorporating road safety into pavement management: maximizing surface friction for road safety improvements. Technical Report 04-04, University of Wisconsin, Madison.

Park, E.-S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transportation Research Record 2019, 1–6.

Park, E.-S., Park, J., Lomax, T.J., 2010. A fully Bayesian multivariate approach to before–after safety evaluation. Accident Analysis and Prevention 42 (4), 1118–1127.

Prokhorov, A., Schmidt, P., 2009. Likelihood-based estimation in a panel setting: robustness, redundancy and validity of copulas. Journal of Econometrics 153 (1), 93–104.

Qi, Y., Smith, B.L., Guo, J., 2007. Freeway accident likelihood prediction using a panel data analysis approach. Journal of Transportation Engineering 133 (3), 149–156.

Shankar, V., Albin, R., Milton, J., Mannering, F.L., 1998. Evaluating median crossover likelihoods with clustered accident counts: an empirical inquiry using the random effects negative binomial model. Transportation Research Record 1635, 44–48.

Shankar, V., Mannering, F.L., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural accident frequencies. Accident Analysis and Prevention 27 (3), 371–389.

Shankar, V., Milton, J., Mannering, F.L., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accident Analysis and Prevention 29 (6), 829–837.

Song, J.J., Ghosh, M., Miaou, S., Mallick, B., 2006. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis 97 (1), 246–273.

Subrahmaniam, K., Subrahmaniam, K., 1973. On the estimation of the parameters in the bivariate negative binomial distribution. Journal of the Royal Statistical Society Series B 35 (1), 131–146.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26 (1), 24–36.

Trivedi, P.K., Zimmer, D.M., 2005. Copula modeling: an introduction for practitioners. Foundations and Trends in Econometrics 1 (1), 1–111.

Veall, M.R., Zimmermann, K.F., 1996. Pseudo-R^2 for some common limited dependent variable models. Journal of Economic Surveys 10 (3), 241–259.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Statistical and Econometric Methods for Transportation Data Analysis, Second edition. Chapman and Hall/CRC, Boca Raton, FL.

Winkelmann, R., 2008. Econometric Analysis of Count Data, Fifth edition. Springer, Germany.

Winston, C., Maheshri, V., Mannering, F.L., 2006. An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes. Journal of Risk and Uncertainty 32 (2), 83–99.

Zhao, Y., Kockelman, K., 2002. Household Vehicle Ownership by Vehicle Type: Application of a Multivariate Negative Binomial Model. Compendium of the Transportation Research Board's 81st Annual Meeting , Washington, DC, January 2002.

Zhou, M., Sisiopiku, V.P., 1997. Relationship between volume-to-capacity ratios and accident rates. Transportation Research Record 1581, 47–52.