Accident Analysis and Prevention 45 (2012) 628–633

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

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

A study of factors affecting highway accident rates using the random-parameters tobit model

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

[a] Visiting Assistant Professor, School of Civil Engineering, Indiana Local Technical Assistance Program, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, USA
[b] Charles Pankow Professor of Civil Engineering, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, USA
[c] Associate Professor of Civil and Environmental Engineering, Pennsylvania State University, University Park, PA 16802, USA
[d] Associate Professor of Civil Engineering, Director, Indiana Local Technical Assistance Program, Purdue University, 550 Stadium Mall Drive, West Lafayette, IN 47907-2051, USA

ARTICLE INFO

Article history:
Received 27 June 2011
Received in revised form 27 August 2011
Accepted 15 September 2011

Keywords:
Accident rates
Random parameters
Tobit regression
Pavement condition
Roadway geometrics
Interstate highways

ABSTRACT

A large body of previous literature has used a variety of count-data modeling techniques to study factors that affect the frequency of highway accidents over some time period on roadway segments of a specified length. An alternative approach to this problem views vehicle accident rates (accidents per mile driven) directly instead of their frequencies. Viewing the problem as continuous data instead of count data creates a problem in that roadway segments that do not have any observed accidents over the identified time period create continuous data that are left-censored at zero. Past research has appropriately applied a tobit regression model to address this censoring problem, but this research has been limited in accounting for unobserved heterogeneity because it has been assumed that the parameter estimates are fixed over roadway-segment observations. Using 9-year data from urban interstates in Indiana, this paper employs a random-parameters tobit regression to account for unobserved heterogeneity in the study of motor-vehicle accident rates. The empirical results show that the random-parameters tobit model outperforms its fixed-parameters counterpart and has the potential to provide a fuller understanding of the factors determining accident rates on specific roadway segments.

Published by Elsevier Ltd.

1. Introduction

The analysis of the frequency of accidents on roadway segments over a specified period of time has been the most common approach to study factors that affect the likelihood of accident occurrence. Models that look at the frequency of vehicle accidents are abundant in the literature and encompass a wide variety of modeling approaches including (see Lord and Mannering (2010) for a complete review of this literature): Poisson models (Jovanis and Chang, 1989; Jones et al., 1991; Miaou and Lum, 1993); negative binomial models (Miaou, 1994; Shankar et al., 1995; Poch and Mannering, 1996; El-Basyouny and Sayed, 2006; Lord, 2006; Kim and Washington, 2006; Malyshkina and Mannering, 2010a); Poisson-lognormal models (Lord and Miranda-Moreno, 2008); zero-inflated count models (Miaou, 1994; Shankar et al., 1997; Lee and Mannering, 2002; Lord et al., 2005, 2007; Malyshkina and Mannering, 2010b); Conway–Maxwell–Poisson models (Lord et al., 2008; Sellers and Shmueli, 2010); Gamma models (Oh et al., 2006; Daniels et al., 2010); generalized estimating equation models (Wang and Abdel-Aty, 1996; Lord and Mahlawat, 2009); generalized additive models (Xie and Zhang, 2008; Li et al., 2011); random effects models (Shankar et al., 1998; Quddus, 2008; Sittikariya and Shankar, 2009; Guo et al., 2010); negative multinomial models (Ulfarsson and Shankar, 2003; Caliendo et al., 2007); random parameters count models (Anastasopoulos and Mannering, 2009; El-Basyouny and Sayed, 2009); and finite mixture and Markov switching models (Malyshkina et al., 2009; Park and Lord, 2009; Malyshkina and Mannering, 2010a; Park et al., 2010).

An alternative to this traditional accident-frequency approach is to consider an accident rate (such as the number of accidents per 100-million vehicle miles traveled), which has long been a standardized measure of roadway safety. Such accident-rate data are continuous, but because some highway segments will have no accidents reported during the analysis period over which accidents are observed, the data will be left-censored at zero. Anastasopoulos et al. (2008) have identified the tobit regression as an appropriate approach to this censoring problem. However, their study of accident rates used a statistical approach that constrained

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

0001-4575/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.aap.2011.09.015

the estimated parameters to be fixed across observations. In the presence of unobserved heterogeneity across observations, such a fixed-parameter approach could potentially result in biased parameter estimates and incorrect inferences (Washington et al., 2011).[4] Given the potential heterogeneity in accident-rate data, a random (as opposed to a fixed) parameter approach may be appropriate. In fact, there is a growing body of accident research that has dealt with possible heterogeneity across observations by allowing some or all parameters to vary randomly across observations. For example, relatively recent research conducted by Milton et al. (2008), Gkritza and Mannering (2008), Anastasopoulos and Mannering (2009, 2011), El-Basyouny and Sayed (2010), and Dinu and Veeraragavan (2011) have all empirically demonstrated the applicability of the random-parameters approach to explicitly account for the variations in the effect of variables across observations. The intent of the current paper is to address the possibility of unobserved heterogeneity by estimating a random-parameters tobit model and comparing the estimation results with a traditional fixed-parameters tobit model.

2. Methodology

The tobit model is a regression model with a dependent variable that is either left-censored (censored at a low threshold) or right-censored (censored at a high threshold) (Tobin, 1958). For the accident rates, the data are left-censored with a clustering at zero (zero accidents per 100-million vehicle miles traveled) because accidents may not be observed on all roadway segments during the period of observation.[5] Using a left-censored limit of zero, the tobit model is expressed as (see Tobin, 1958; Anastasopoulos et al., 2008):

$$\begin{aligned} &Y_i^* = \boldsymbol{\beta}'\mathbf{X}_i + \varepsilon_i, \quad i = 1, 2, \ldots, N, \\ &Y_i = Y_i^* \quad \textit{if} \quad Y_i^* > 0 \\ &Y_i = 0 \quad \textit{if} \quad Y_i^* \le 0, \end{aligned} \tag{1}$$

where N is the number of observations, Y_i is the dependent variable (accidents per 100-million vehicle miles traveled in roadway segment i), $\mathbf{X}_i$ is a vector of independent variables (pavement, traffic and roadway segment characteristics), $\boldsymbol{\beta}$ is a vector of estimable parameters, and ε_i is a normally and independently distributed error term with zero mean and constant variance σ^2. Eq. (1) shows that there is an implicit, stochastic index (latent variable) equal to Y_i^* which is observed only when positive. The corresponding likelihood function for the tobit model is:

$$L = \prod_0 [1 - \boldsymbol{\Phi}(\boldsymbol{\beta}\mathbf{X}/\sigma)] \prod_1 \sigma^{-1} \boldsymbol{\Phi}[(Y_i - \boldsymbol{\beta}\mathbf{X})/\sigma] \tag{2}$$

where $\boldsymbol{\Phi}$ is the standard normal distribution function and $\boldsymbol{\Phi}$ is the standard normal density function.[6]

To account for heterogeneity (unobserved factors that may vary across observations, in our case individual road segments) with random parameters, Greene (2007) has developed estimation procedures (using simulated maximum likelihood estimation) for incorporating random parameters in tobit regression. This approach views estimable parameters as,

$$\beta_i = \beta + \varphi_i, \quad i = 1, 2, \ldots, N, \tag{3}$$

where φ_i is a randomly distributed term (for example a normally distributed term with mean zero and variance σ^2). With this equation, the latent variable becomes $Y_i^*|\varphi_i = \beta X_i + \varepsilon_i$, and the likelihood function from Eq. (2) can be written as the log-likelihood,

$$\mathrm{LL} = \sum_{\forall i} \ln \int_{\varphi_i} g(\varphi_i) P(Y_i^*|\varphi_i) d\varphi_i \tag{4}$$

where $g(.)$ is the probability density function of the φ_i. Because maximum likelihood estimation of the random parameters tobit model is computationally cumbersome, a simulation-based maximum likelihood method is employed using Halton draws, which has been shown to provide an efficient distribution of draws for numerical integration (see Halton, 1960; Train, 1999; Bhat, 2003).[7]

3. Data

Motor vehicle accident data from urban interstate roads in Indiana (I-465, I-64, I-65, I-69, I-70, I-74, I-90, and I-164) were collected over a 9-year period (1 January 1999 to 31 December 2007) to investigate the effect of pavement characteristics (pavement roughness, rutting, surface deflection, and pavement condition rating), road geometrics (number of lanes, presence and characteristics of horizontal/vertical curves, junctions, barriers, gores, medians, shoulders, etc.), and traffic characteristics (speed limits, passenger car and truck traffic volumes) on accident rates per 100-million vehicle miles traveled (VMT). The pavement condition data were collected from Indiana's Department of Transportation pavement management databases and from INDIPAVE (a database consisting of data on pavement condition, weather, pavement structure, traffic, maintenance, and other information of over 10,000 1-mile (1.61 km) pavement segments in the State of Indiana). For additional information on the data specifics, the reader is referred to Anastasopoulos (2009).

The data were divided into homogeneous roadway segments (defined by roadway geometrics and pavement type). The segment-defining information included shoulder characteristics (inside and outside shoulder presence and width and rumble strips), pavement characteristics (pavement type), median characteristics (median width, type, condition, barrier presence and location), number of lanes, and speed limit. A total of 200 roadway segments were defined and the number of police-reported motor vehicle accidents occurring on each segment over the 9-year period was obtained from the Indiana State Patrol accident-data files.

[4] Because accident-frequency/rate data do not contain residual environmental effects, and socio-economic and behavioral characteristics of drivers or vehicle-specific information (such data elements are available only after an accident has occurred and thus cannot be used to predict the likelihood of an accident or an accident rate), significant unobserved heterogeneity across observations is likely to be present.

[5] Accidents may not be observed simply because none have occurred. However, another reason for not observing an accident may be due to the availability of data. For example, in most accident databases, accidents that do not involve injury are only reported if the property value damage exceeds a pre-specified value threshold. There may also be under-reporting even when damage thresholds are exceeded due to a variety of reasons relating to data acquisition and entry. From an empirical standpoint however, the equivalent effect is the same with the tobit model treating these as the same, clustered-at-zero observations.

[6] The expected value of the dependent variable for all cases, $E[Y]$, is (removing the subscript i for simplification) $E[Y] = \boldsymbol{\beta}\mathbf{X}F(z) + \sigma f(z)$, where $z = \boldsymbol{\beta}\mathbf{X}/\sigma$ is the z-score for an area under the normal curve, $F(z)$ is the cumulative normal distribution function associated with the proportion of cases above zero, $f(z)$ the unit normal density, and σ the standard deviation of the error term. To determine the effect of an independent variable on the expected value, the first-order partial derivative of $E[Y]$ is used, and the expected value of Y for observations above zero is $\boldsymbol{\beta}\mathbf{X}$ plus the expected value of the truncated normal error term, $E[Y'] = \boldsymbol{\beta}\mathbf{X} + \sigma f(z)/F(z)$. Note that the estimated tobit parameters do not measure the correct regression parameters for observations above the limit. In fact, this is only true when $X = \infty$, which would mean that $F(z) = 1$ and $f(z) = 0$. For more information on the tobit model formulation and interpretation in safety, see Anastasopoulos et al. (2008).

[7] Note that this random parameters model formulation is equivalent to a random effects model if only the constant term is a random parameter (see Anastasopoulos and Mannering, 2009).

Table 1
Descriptive statistics of selected variables.

	Mean	Std. dev.	Minimum	Maximum
Total number of motor vehicle accidents over the 9-year analysis period	10.57	17.1167	0	61
Average traffic volume of passenger cars over the 9-year analysis period (in 1000 s per day)	11.399	20.897	6.042	153.345
Average combination truck volume over the 9-year analysis period (in 1000 s per day)	1.904	4.281	0	40.648
Average international roughness index (IRI) over the 9-year analysis period (in./mi)	102.589	32.416	51.167	311.667
Standard deviation of the international roughness index – IRI – over the 9-year analysis period (in./mi)	24.346	15.894	1.155	111.776
Average pavement condition rating (PCR) over the 9-year analysis period (scale 0–100)	88.818	3.876	74.167	97
Average rut depth over the 9-year analysis period (in.)	0.149	0.060	0.052	0.536
Average surface deflection over the 9-year analysis period (thousandths of inches)	8.514	4.491	2.721	25.799
Presence of horizontal curve	0.495		0	1
Presence of vertical curve	0.117		0	1
Presence of median barrier	0.056		0	1
Junction presence	0.157		0	1

For model estimation, the data included the aggregated number of accidents on each roadway segment over the 9-year period of analysis, and the accident rate (number of accidents per 100-million VMT) was calculated as:

$$\text{Accident rate}_i = \frac{\sum_{\text{yr}=1}^{n} \text{Accidents}_{\text{yr},i}}{[\sum_{\text{yr}=1}^{n} \text{AADT}_{\text{yr},i} \times L_i \times 365]/100,000,000} \quad (5)$$

where Accident rate$_i$ is the number of accidents per 100-million VMT on roadway segment i, yr denotes the year (from 1 to n), Accidents$_{\text{yr},i}$ is the number of accidents in year yr (year from 1 to n) on segment i, AADT$_{\text{yr},i}$ the average annual daily traffic in year yr on segment i, and L_i the length of roadway segment i in miles.

Of the 200 road segments, 65 had no accidents over the 9-year analysis period, and 135 had at least one accident. Table 1 presents summary statistics of selected variables.

4. Model estimation results

The random parameters tobit model is estimated by specifying a functional form of the parameter density function and using simulation-based maximum likelihood with 200 Halton draws – a number of draws that has been empirically shown by Bhat (2003) to produce accurate parameter estimates. For the functional form of the random parameters density functions, consideration is given to normal, lognormal, Weibull, uniform and triangular distributions. For all random parameters, the normal distribution was found to provide the best statistical fit.

Table 2 shows that the random-parameters tobit model has a better log-likelihood at convergence compared to the fixed-parameters tobit model. To statistically compare the two models, a likelihood ratio test is conducted. The test statistic is:

$$\chi^2 = -2[\text{LL}(\beta_{\text{F}}) - \text{LL}(\beta_{\text{RP}})], \quad (6)$$

where LL(β_{F}) is the log likelihood at convergence of the fixed-parameters tobit model, and LL(β_{RP}) is the log likelihood of the random-parameters tobit model (Washington et al., 2011). This statistic is χ^2 distributed, with degrees of freedom equal to the difference in the numbers of parameters between the fixed- and random-parameter models. The resulting likelihood ratio test gives a χ^2 value of 83.8 with 6 degrees of freedom indicating that we are more than 99.99% confident that the random-parameters model is statistically superior.[8]

With regard to specific parameter estimates, for the random-parameters model formulation, a random parameter is used when both the mean and standard deviation of the parameter density are statistically significant. If an individual parameter's estimated standard deviation is not statistically different from zero, the parameter is fixed across the population of roadway segments. The estimation results shown in Table 2 indicate that the average annual daily traffic (AADT) of passenger cars, the roadway segment's international roughness index (IRI) and rut depth, the horizontal and vertical curvature indicator variables (if present), and the median barrier indicator variable (if present), were all found to produce statistically significant random parameters (the standard deviation of the parameter's distribution was significantly different from zero).[9]

Looking first at the traffic-related variables, for the random-parameters model the annual average daily travel of passenger cars results in a random parameter that is normally distributed, with a mean −0.43 and standard deviation 0.44. Given these distributional parameters, for 16.42% of the observations the effect of AADT increases accident rates and for 83.58% of the observations the effect of AADT decreases accident rates. This implies that for the vast majority of the roadway segments the accident rate will decrease with increasing AADT – a finding that is consistent with a number of previous studies (Zhou and Sisiopiku, 1997; Dickerson et al., 2000; Qi et al., 2007; Anastasopoulos and Mannering, 2009) that have shown that the likelihood of an accident is higher on low-traffic volume roads, and significantly decreases with increasing volumes. However, for a small percentage of the roadway segments the opposite is true. For the fixed-parameters model the effect of AADT on the crash rate is negative for all roadway segments, but it is statistically insignificant.

Moving to the effect of combination truck volumes on accident rates, in the random-parameters model it was found that an

[8] Note that a number of variables in the fixed parameters model were found to be statistically insignificant, whereas they were significant in the random parameters model. This is attributed to the flexibility of the latter that relaxes the restriction (of the fixed parameters model) that the effect of the covariates must be constant across the observations.

[9] It should be noted that a number of variables reflecting the interaction among road geometrics, pavement condition and traffic (trucks, passenger cars, etc.) were tested in the model. These interaction-variables yielded fixed and random parameters that were barely statistically insignificant (at 0.90 level of confidence). Therefore, in this paper the models that provided the best overall statistical fit and yielded parameters that were statistically significant (at 0.90 level of confidence) are only presented.

Table 2
Tobit model estimation results.

	Fixed parameters model		Random parameters model	
	Parameter estimate	t-ratio	Parameter estimate	t-ratio
Average traffic volume of passenger cars over the 9-year analysis period (in 1000 s per day)	−0.04	−0.10	−0.43	−9.85**
Standard deviation of parameter distribution			0.44	4.89**
Average traffic volume of combination trucks over the 9-year analysis period (in 1000 s per day)	−0.04	−0.02	−1.31	−4.67**
Average international roughness index – IRI – over the 9-year analysis period (in./mi)	1.44	4.85**	1.41	4.69**
Standard deviation of parameter distribution			0.82	9.75**
Standard deviation of the international roughness index (IRI) over the 9-year analysis period (in./mi)	0.61	2.42*	0.71	6.47**
Average pavement condition rating (PCR) over the 9-year analysis period (scale 0–100)	−3.20	−15.42**	−2.63	−10.08**
Average rut depth over the 9-year analysis period (in.)	571.87	4.16**	222.33	16.27**
Standard deviation of parameter distribution			140.92	10.22**
Average surface deflection over the 9-year analysis period (thousandths of inches)	5.51	2.13*	5.50	3.57**
Curvature indicator variable (1 if horizontal curve is present, 0 otherwise)	−31.23	−3.29**	−33.78	−6.60**
Standard deviation of parameter distribution			14.58	11.51**
Grade indicator variable (1 if vertical grade is present, 0 otherwise)	5.01	0.47	4.63	3.14*
Standard deviation of parameter distribution			15.73	10.67**
Median barrier indicator variable (1 if median barrier is present, 0 otherwise)	5.89	0.44	11.98	7.94**
Standard deviation of parameter distribution			17.69	11.03**
Junction indicator variable (1 if present, 0 otherwise)	2.36	0.25	15.49	5.09**
$LL(0)$	−1228.4		−1228.4	
$LL(\beta)$	−711.5		−669.6	
Number of observations	200		200	

* Significant at 0.95 level of confidence.
** Significant at 0.99 level of confidence.

increase in the number of combination trucks is associated with lower accident rates. This finding may be reflecting the calming effect that combination trucks can have on traffic flow by lowering speeds, and so on (see also similar findings by Shankar et al., 1997; Anastasopoulos et al., 2008). In the fixed-parameters model, the combination truck volume was found to be statistically insignificant (less than 90% level of confidence).

With regard to pavement characteristics, important measures of pavement condition considered in this study were the international roughness index, pavement condition rating (PCR), rutting depth, and surface deflection. In the fixed- and random-parameter models all of these factors were found to have a statistically significant effect on accident rates as found in previous literature (Anastasopoulos et al., 2008; Anastasopoulos, 2009; Anastasopoulos and Mannering, 2009, 2011).[10] As shown in Table 2, in the random-parameters model, the average IRI readings produce a random parameter that is normally distributed, with a mean 1.41 and standard deviation 0.82. This indicates that increasing IRI (implying rougher pavement) nearly always increases the accident rate (with only 4.28% of the roadway segments having a negative value). In the fixed-parameters tobit model, increasing IRI increases accidents rates for all roadway segments.

In a similar context, roadway segments with a high IRI standard deviation over the sample period (implying a large performance change in the pavement condition over the 9-year period due to either some potential maintenance of the pavement or rapid pavement deterioration reflected in a high increase in the IRI over the years) have higher accident rates, and this results in a fixed parameter in both models.

Higher pavement condition ratings (PCRs), which range from zero (completely deteriorated) to 100 (excellent pavement condition), decrease accident rates as one would suspect. This variable results in fixed parameters for both fixed- and random-parameter tobit models.

It is found that pavements with high rutting (a deformation of the pavement in the wheel path, measured in inches) generally result in higher accident rates as expected, because excessive rutting increases chances of hydroplaning when moisture is present and therefore could contribute to vehicle tracking and loss of control during lane changes and other maneuvering. For the fixed-parameters model, high rut depth increases accident rates on all roadway segments. For the random-parameters model, the rutting variable is normally distributed such that increases in rut depth increase accident rates in 94.27% of the roadway segments and decrease accident rates on 5.73% of the roadway segments. It appears that on at least some roadway segments the vehicle tracking caused by rutting may be beneficial, perhaps on segments were lane-changing is infrequent. In any event, the effect of rutting on accidents rates is not uniform across roadway segments.

Pavement surface deflection is used to evaluate the load carrying ability of both flexible and rigid pavements. It is an important pavement evaluation method because the magnitude and shape of pavement deflection is a function of the pavement structural segment, traffic, temperature, and moisture affecting the pavement structure. Surface deflection is measured as a pavement surface's vertical deflection distance as a result of an applied static or dynamic load. The units for the surface deflection used in the analysis are measured in thousandths of inches of deflection (at the center of the load) resulting from a falling weight deflectometer load, corrected to a 9000 lb load applied on a 11.8-in. diameter plate and adjusting for temperature (using 65 °F as the standard). For

[10] Note that Anastasopoulos (2009) and Anastasopoulos et al. (2011) used the data presented in this paper (in their disaggregate form) to develop fixed and random parameters seemingly unrelated regression equations of pavement performance, and showed that the pavement condition deteriorates over time linearly (other non-linear forms of deterioration were also tested, with the linear form providing far better pavement performance predictors). Hence, using the average pavement condition (and standard deviation) is appropriate, as the possibility that the aggregated data may be masking some of the pavement condition effects is slim.

both fixed- and random-parameters models, high average surface deflection (implying poor pavement load carrying ability) increases accident rates for all roadway segments.

The roadway-geometric variables included in the model (horizontal curve, vertical grade, and median barrier presence on the roadway segment) were found to have mixed effects on the accident rates. In the random-parameters model, all of these variables produced statistically significant random parameters, indicating that their effect varies across roadway segments. When horizontal curves were present, the random-parameters model estimation results indicate that accident rates decreased on 98.97% of the roadway segments and increased on the other 1.03% (for the fixed-parameters model the accident rate decreased on all roadway segments). This general finding (that horizontal curves generally decrease accident rates) could be the result of two factors. The first is highway hypnosis, which suggests that some variation in roadway geometrics may improve drivers' alertness and hence decrease the likelihood of an accident (Wertheim, 1978; Cerezuela et al., 2004; Anastasopoulos et al., 2008). The second is risk compensation, which suggests that drivers may compensate by driving more carefully in conditions which they perceive as dangerous and less carefully in conditions they do not (Assum et al., 1999; Dulisse, 1997; Winston et al., 2006).

The other roadway-geometric variables included grade and median-barrier indicator variables. Neither of these variables produced statistically significant parameters in the fixed-parameters model. In the random-parameters model, 61.58% of the roadway segments with non-zero absolute vertical grade presence, and 75.09% of the roadway segments with median barriers present had higher accident rates (with 38.42% and 24.91% with lower accident rates respectively).[11] Thus, the effect of grades and median barriers on crash rates varies significantly across roadway segments.

Finally, the presence of a junction (ramp or interchange presence along the road segment) was statistically insignificant in the fixed-parameters model, but was found to be a fixed parameter that increased accident rates in the random-parameters model.[12]

5. Summary and conclusions

This paper explores the use of random parameters in tobit regression to examine factors that significantly influence accident rates expressed as the number of accidents per 100-million vehicle miles traveled. In addition to the ability of the tobit model to consider roadway segments with and without observed accidents (by viewing accident rates as a continuous variable that is left censored at zero), which allows for a complete use of all available data, the underlying unobserved heterogeneity can potentially be accounted for through the use of the random parameters.

Using 9-years of motor vehicle accident data from urban interstate roads in Indiana, our estimation results show that the random-parameters tobit model provides a superior statistical fit – with eleven variables producing statistically significant parameters as opposed to only six in the case of the traditional fixed-parameters tobit. The random-parameters model shows that a variety of factors relating to pavement condition and quality are found to significantly influence motor vehicle accident rates including the average and standard deviation (over the 9-year analysis period) of the international roughness index, and the average rut depth, pavement condition rating, and surface deflection. In terms of geometric factors and their effect on the accident rates, presence of horizontal curve, vertical grade, median barrier, and junction, are all found to be statistically significant. And, the annual average daily traffic of passenger cars and combination trucks are both found to have a significant impact on accident rates. Of these variables, the AADT of passenger cars, the average IRI and rut depth, the horizontal and vertical curvature indicator variables, and the median barrier indicator variable, all produce statistically significant random parameters indicating that their impact on accident rates varies across roadway segments.

While this study is exploratory in nature, it does suggest the considerable potential that random parameters have in analyzing accident rates using tobit regression. The empirical results show that the random parameters tobit model outperforms its fixed parameters tobit counterpart, having thus the potential to provide a fuller understanding of the factors determining accident rates. Applying the approach to other geographic areas and to non-interstate road segments would potentially provide more information on the effect that pavement, geometric, and traffic characteristics have on accident rates.

Acknowledgements

The authors would like to acknowledge the contribution of Bill Flora (Indiana Department of Transportation) and Lakhwinder Singh (US Department of Transportation) for their assistance in data acquisition. The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of any agency, nor do the contents constitute a standard, specification, or regulation.

References

Anastasopoulos, P.Ch., 2009. Infrastructure asset management: a case study on pavement rehabilitation. Ph.D. Dissertation, Purdue University, West Lafayette, Indiana (available electronically from http://upload.cos.com/etdadmin/files/43/13462_pdf_A7F9D4C8-2E08-11DE-B14E-0869F0E6BF1D.pdf).

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., Haddock, J.E., 2011. A random parameters seemingly unrelated equations approach to the post-rehabilitation performance of pavements. Working paper.

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.

Assum, T., Bjornskau, T., Fosser, S., Sagberg, F., 1999. Risk compensation—the case of road lighting. Accident Analysis and Prevention 31 (5), 545–553.

Bhat, C., 2003. Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. Transportation Research Part B 37 (1), 837–855.

Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis and Prevention 39 (4), 657–670.

Cerezuela, G., Tajero, P., Choliz, M., Chivert, M., Monteagudo, M., 2004. Wertheim's hypothesis on 'highway hypnosis': empirical evidence from a study on motorway and conventional road driving. Accident Analysis and Prevention 36 (6), 1045–1054.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accident Analysis and Prevention 42 (2), 393–402.

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

Dinu, R.R., Veeraragavan, A., 2011. Random parameter models for accident prediction on two-lane undivided highways in India. Journal of Safety Research 42 (1), 39–42.

[11] Regarding the median barrier variable, it is important to note that there were roughly 11 roadway segments with median barriers and 189 without. Because the estimated models did not have statistically significant constants, it may be presumed that this variable may be capturing an intercept shift, or simply the random effects associated with non-median barrier roadway segments, rather than a true median barrier effect.

[12] A traditional accident-frequency negative binomial model was also estimated (the Poisson model was rejected due to the significance of the dispersion parameter, α), and the results were broadly similar to the fixed parameters tobit model. The overall fit of the negative binomial model using McFadden's pseudo-ρ^2 was 0.393 (see Washington et al., 2011).

Dulisse, B., 1997. Methodological issues in testing the hypothesis of risk compensation. Accident Analysis and Prevention 29 (3), 285–292.
El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transportation Research Record 1950, 9–16.
El-Basyouny, K., Sayed, T., 2009. Accident prediction models with random corridor parameters. Accident Analysis and Prevention 41 (5), 1118–1123.
El-Basyouny, K., Sayed, T., 2010. Accident prediction models with random corridor parameters. Accident Analysis and Prevention 42 (1), 131–139.
Gkritza, K., Mannering, F.L., 2008. Mixed logit analysis of safety-belt use in single- and multi-occupant vehicles. Accident Analysis and Prevention 40 (2), 443–451.
Greene, W., 2007. Limdep, Version 9.0. Econometric Software, Inc., Plainview, NY.
Guo, F., Wang, X., Abdel-Aty, M., 2010. Modeling signalized intersection safety with corridor spatial correlations. Accident Analysis and Prevention 42 (1), 84–92.
Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2, 84–90.
Jones, B., Janssen, L., Mannering, F.L., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention 23 (2), 239–255.
Jovanis, P., Chang, H., 1989. Disaggregate model of highway accident occurrence using survival theory. Accident Analysis and Prevention 21 (5), 445–458.
Kim, D., Washington, S.P., 2006. The significance of endogeneity problems in crash models: an examination of left-turn lanes in intersection crash models. Accident Analysis and Prevention 38 (6), 1094–1100.
Lee, J., Mannering, F.L., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis and Prevention 34 (2), 149–161.
Li, X., Lord, D., Zhang, Y., 2011. Development of accident modification factors for rural frontage road segments in Texas using results from generalized additive models. Journal of Transportation Engineering 137 (1), 74–83.
Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention 38 (4), 751–766.
Lord, D., Guikema, S., Geedipally, S.R., 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. Accident Analysis and Prevention 40 (3), 1123–1134.
Lord, D., Mahlawat, M., 2009. Examining the application of aggregated and disaggregated Poisson-gamma models subjected to low sample mean bias. Transportation Research Record 2136, 1–10.
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.
Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Science 46 (5), 751–770.
Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accident Analysis and Prevention 37 (1), 35–46.
Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero inflated models in highway safety. Accident Analysis and Prevention 39 (1), 53–57.
Malyshkina, N., Mannering, F.L., 2010a. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accident Analysis and Prevention 42 (1), 131–139.
Malyshkina, N., Mannering, F.L., 2010b. Zero-state Markov switching count-data models: an empirical assessment. Accident Analysis and Prevention 42 (1), 122–130.
Malyshkina, N., Mannering, F.L., Tarko, A.P., 2009. Markov switching negative binomial models: an application to vehicle accident frequencies. Accident Analysis and Prevention 41 (2), 217–226.
Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident Analysis and Prevention 26 (4), 471–482.
Miaou, S.-P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention 25 (6), 689–709.
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.
Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway-highway interfaces. Accident Analysis and Prevention 38 (2), 346–356.
Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accident Analysis and Prevention 41 (4), 683–691.
Park, B.J., Lord, D., Hart, J.D., 2010. Bias properties of Bayesian statistics in finite mixture of negative regression models for crash data analysis. Accident Analysis and Prevention 42 (2), 741–749.
Poch, M., Mannering, F.L., 1996. Negative Binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering 122 (2), 105–113.
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.
Quddus, M.A., 2008. Time series count data models: an empirical application to traffic accidents. Accident Analysis and Prevention 40 (5), 1732–1741.
Sellers, K.F., Shmueli, G., 2010. A flexible regression model for count data. Annals of Applied Statistics 4 (2), 943–961.
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.
Sittikariya, S., Shankar, V., 2009. Modeling Heterogeneity: Traffic Accidents. VDM-Verlag, 80 pp.
Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26 (1), 24–36.
Train, K., 1999. Halton sequences for mixed logit. Working Paper, University of California, Department of Economics, Berkley.
Ulfarsson, G., Shankar, V., 2003. An accident count model based on multi-year cross-sectional roadway data with serial correlation. Transportation Research Record 1840, 193–197.
Wang, X., Abdel-Aty, M., 1996. Temporal and spatial analyses of rear-end crashes at signalized intersections. Accident Analysis and Prevention 38 (6), 1137–1150.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2011. Statistical and Econometric methods for transportation data analysis, 2nd ed. Chapman & Hall/CRC.
Winston, C., Maheshri, V., Mannering, F., 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.
Wertheim, A.H., 1978. Explaining highway hypnosis: experimental evidence for the role of eye movements. Accident Analysis and Prevention 10 (2), 111–129.
Xie, Y., Zhang, Y., 2008. Crash frequency analysis with generalized additive models. Transportation Research Record 2061, 39–45.
Zhou, M., Sisiopiku, V.P., 1997. Relationship between volume-to-capacity ratios and accident rates. Transportation Research Record 1581, 47–52.