Accident Analysis and Prevention 41 (2009) 153–159

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A note on modeling vehicle accident frequencies with random-parameters count models

Panagiotis Ch. Anastasopoulos, Fred L. Mannering *

School of Civil Engineering, 550 Stadium Mall Drive, Purdue University, West Lafayette, IN 47907-2051, United States

ARTICLE INFO

Article history:
Received 29 April 2008
Received in revised form
29 September 2008
Accepted 9 October 2008

Keywords:
Accident frequency
Count data
Random-parameters Poisson and negative binomial models

ABSTRACT

In recent years there have been numerous studies that have sought to understand the factors that determine the frequency of accidents on roadway segments over some period of time, using count data models and their variants (negative binomial and zero-inflated models). This study seeks to explore the use of random-parameters count models as another methodological alternative in analyzing accident frequencies. The empirical results show that random-parameters count models have the potential to provide a fuller understanding of the factors determining accident frequencies.

1. Introduction

Despite ongoing improvements in highway design and vehicle safety, the toll of traffic accidents in terms of personal injury and lost productivity remains unacceptably high. While countless research efforts have been conducted to better understand the factors that influence the frequency and severity of accidents to provide more effective safety-related countermeasures, there is still much work to be done. The intent of this paper is to demonstrate an approach that can be used to better understand the factors that influence the frequency of accidents.

Accident frequencies (the number of accident occurrences on roadway segments or at intersections over a specified period of time) have been extensively studied in recent years using a variety of methodological approaches. The Poisson and negative binomial models have been widely used to predict accident occurrences, with respect to the count data modeling for highway safety (Jones et al., 1991; Shankar et al., 1995; Hadi et al., 1995; Poch and Mannering, 1996; Milton and Mannering, 1998; Abdel-Aty and Radwan, 2000; Savolainen and Tarko, 2005; Lord, 2006; Lord and Park, 2008). In addition, some variations of count models have also been developed, such as zero-inflated negative binomial models (Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002), negative binomial with random effects models (Shankar et al., 1998), and Conway–Maxwell–Poisson generalized linear models (Lord et al., 2008). With respect to accident occurrences and safety in general, the Transport Research Laboratory has performed an exhaustive amount of research that set the standard, in terms of a rigorous methodology, for later work in this area (see Satterthwaite, 1981; Maycock and Hall, 1984; Pickering et al., 1986; Hall, 1986; Wright and Barnett, 1991). Interestingly, Maher and Summersgill (1996) described various technical issues relating to generalized linear models such as the low-mean value problem, overdispersion, data disaggregation over time, allowing the presence of a trend over time in accident risk, random errors in the flow estimates, the estimation of prediction uncertainty, correlations between predictions for different accident types, and combination of model predictions with site observations.

While past studies that have applied count-data models have added to our knowledge of factors affecting accident frequencies, they have assumed that parameters are fixed across observations (roadway segments or intersections). However, the possibility of accounting for heterogeneity across observations by allowing some or all parameters to vary (across observations) has considerable potential. In the case of discrete outcome models (multinomial and ordered logit models), relatively recent research conducted by Train (1997), Revelt and Train (1997, 1999), Brownstone and Train (1999), McFadden and Train (2000), Bhat (2001), and Eluru et al. (2008), has demonstrated the applicability of a random-parameters approach to explicitly account for the variations of the effect of variables across observations (for an application in safety analysis see Milton et al., 2008 and Gkritza and Mannering, 2008). Constraining the parameters to be constant when they actually vary across

* Corresponding author. Tel.: +1 765 496 7913.
E-mail addresses: panast@purdue.edu (P.Ch. Anastasopoulos), flm@ecn.purdue.edu (F.L. Mannering).

0001-4575/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2008.10.005

observations could lead to inconsistent and bias parameter estimates (Washington et al., 2003). Given the potential heterogeneity in accident-frequency data, random-parameters approaches may be appropriate in some cases.

2. Methodology

Count-data modeling techniques are commonly used for accident frequency analysis because the number of accidents on roadway segment per unit of time is a non-negative integer. As mentioned above, count data are generally modeled with a Poisson regression or its derivatives which include the negative binomial and zero-inflated models (see Washington et al., 2003). For the basic Poisson model, the probability $P(n_i)$ of road segment i having n_i accidents is

$$P(n_i) = \frac{EXP(-\lambda_i)\lambda_i^{n_i}}{n_i!} \tag{1}$$

where λ_i is the Poisson parameter for road segment i, which is road segment i's expected number of accidents, $E[n_i]$. Poisson regression specifies the Poisson parameter λ_i (the expected number of accidents) as a function of explanatory variables by typically using a log-linear function:

$$\lambda_i = EXP(\boldsymbol{\beta}\mathbf{X}_i) \tag{2}$$

where $\mathbf{X}_i$ is a vector of explanatory variables and $\boldsymbol{\beta}$ is a vector of estimable parameters (Washington et al., 2003).

Depending on the data, a Poisson model may not always be appropriate because the Poisson distribution restricts the mean and variance to be equal ($E[n_i] = VAR[n_i]$). If this equality does not hold, the data are said to be underdispersed ($E[n_i] > VAR[n_i]$) or overdispersed ($E[n_i] < VAR[n_i]$), and the standard errors of the estimated parameter vector will be incorrect and incorrect inferences could be drawn. To account for this possibility, the negative binomial model is derived by rewriting

$$\lambda_i = EXP(\boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i) \tag{3}$$

where $EXP(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance α. The addition of this term allows the variance to differ from the mean as $VAR[n_i] = E[n_i][1 + \alpha E[n_i]] = E[n_i] + \alpha E[n_i]^2$. The negative binomial probability density function has the form:

$$P(n_i) = \left(\frac{1/\alpha}{(1/\alpha)+\lambda_i}\right)^{1/\alpha} \frac{\Gamma[(1/\alpha)+n_i]}{\Gamma(1/\alpha)n_i!} \left(\frac{\lambda_i}{(1/\alpha)+\lambda_i}\right)^{n_i} \tag{4}$$

where $\Gamma(\cdot)$ is a gamma function. The Poisson regression is a limiting model of the negative binomial regression as α approaches 0. Thus, if α is significantly different from 0, the negative binomial is appropriate and if it is not, the Poisson model is appropriate (Washington et al., 2003).

To account for heterogeneity (unobserved factors that may vary across observations) with random parameters, Greene (2007) has developed estimation procedures (using simulated maximum likelihood estimation) for incorporating random parameters in Poisson and negative binomial count-data models.[1] To allow for such random parameters in count-data models, estimable parameters can be written as

$$\beta_i = \beta + \varphi_i \tag{5}$$

where φ_i is a randomly distributed term (for example a normally distributed term with mean 0 and variance σ^2). With this equation, the Poisson parameter becomes $\lambda_i|\varphi_i = EXP(\boldsymbol{\beta}\mathbf{X}_i)$ in the Poisson model and $\lambda_i|\varphi_i = EXP(\boldsymbol{\beta}\mathbf{X}_i + \varepsilon_i)$ in the negative binomial with the corresponding probabilities for Poisson or negative binomial now $P(n_i|\varphi_i)$ (see Eq. (1)). With this, the log-likelihood can be written as

$$\mathrm{LL} = \sum_{\forall i} \ln \int_{\varphi_i} g(\varphi_i) P(n_i|\varphi_i)\, \mathrm{d}\varphi_i \tag{6}$$

where $g(\cdot)$ is the probability density function of the φ_i. Because maximum likelihood estimation of the random-parameters Poisson and negative binomial models is computationally cumbersome (due to the required numerical integration of the Poisson or negative binomial function over the distribution of the random parameters), a simulation-based maximum likelihood method is used (the estimated parameters are those that maximize the simulated log-likelihood function; and to allow for heterogeneity, $\sigma \neq 0$). The most popular simulation approach uses Halton draws,[2] which has been shown to provide a more efficient distribution of draws for numerical integration than purely random draws (see Bhat, 2003; Train, 1999).

Note that this random-parameters formulation is equivalent to a random effects model (see Shankar et al., 1998 for an accident frequency application) if only the constant term is a random parameter. Please see Greene (2007) for further details on random-parameters count models.

To describe the relative magnitude between the dependent and independent variables based on parameter estimates from the estimations of the Poisson and negative binomial models and their variations, marginal effects can be estimated. In the case of accident frequencies, marginal effects give the change in the number of accidents given a unit change in any independent variable, x, and are simply calculated as the partial derivative, $\partial\lambda_i/\partial x$, where λ_i is defined as in Eqs. (2), (3), or (5) depending on the model being considered (Poisson, negative binomial with fixed parameters, or random-parameters models, respectively). Although marginal effects are generated for each road segment i, in the forthcoming estimations, averages over the road-segment population will be presented.

3. Data

To illustrate the application of the random-parameters model, data previously used to estimate a tobit model of accident rates (as described in Anastasopoulos et al., 2008) are used. These data consist of accident data from rural interstate highways in Indiana (I-64, I-65, I-70, I-74, and I-164) collected over a 5-year period (1 January 1995 to 31 December 1999). The data consist of 322 roadway segments, with each segment having homogenous characteristics (segments begin and end when characteristics change).[3] The number of police-reported vehicle accidents (fatalities, injuries, and property-damage-only) occurring on each segment over the 5-year period was obtained from the Indiana State Patrol accident-data files. Detailed roadway-segment information was acquired from

[1] An alternative to a random-parameters approach in the negative-binomial case would be to allow α to vary as a function of the mean, λ (see Cameron and Trivedi, 1986). While this would simplify estimation, it also would be more restrictive in terms of its ability to account for heterogeneity across roadway segments.

[2] Halton draws are sequences used to generate deterministically constructed, nearly uniformly distributed points in the interval [0, 1], that appear to be random (Halton, 1960). Based on prime p ($\geq$2), the one-dimensional Halton sequence divides the 0–1 space into p segments, and systematically fills in the empty spaces using cycles of p-length that place one draw in each segment (see Hess et al., 2003).

[3] 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.

Table 1
Summary statistics of key pavement, geometric and traffic variables (for the 322 roadway segments).

Variable	Mean	S.D.	Min	Max
Dependent variable: 5-year number of motor vehicle accidents	17.561	36.643	0	329
Pavement characteristics				
5-year minimum friction readings	30.558	6.631	15.9	48.2
5-year maximum international roughness index (IRI) readings	101.400	32.816	50	264
Average pavement condition rating (PCR) indicator variable (1 if greater than 95, 0 otherwise)	0.542	0.499	0	1
Good rutting indicator variable (1 if all 5-year average rutting readings are below 0.2 in., 0 otherwise)	0.870	0.337	0	1
Excellent rutting indicator variable (1 if all 5-year rutting readings are below 0.12 in., 0 otherwise)	0.182	0.386	0	1
Geometric characteristics				
Road segment length (in miles)	0.900	1.491	0.1	11.53
Median width indicator variable (1 if greater than 74 ft, 0 otherwise)	0.318	0.466	0	1
Median barrier presence indicator variable (1 if present, 0 otherwise)	0.158	0.365	0	1
Inside shoulder width indicator variable (1 if 5 ft or greater, 0 otherwise)	0.306	0.462	0	1
Outside shoulder width (in feet)	11.258	1.735	6.2	21.8
Average horizontal curve degree curvature per mile	3.036	5.398	0	37.980
Number of vertical curves per mile	1.200	2.520	0	14.493
Ratio of the vertical curve length over the road segment length (in tenths of miles)	2.215	3.539	0	10
Number of bridges per mile	0.872	2.175	0	9.479
Number of ramps in the driving direction per lane-mile	0.141	0.410	0	3.268
Traffic characteristics				
Average annual daily traffic (AADT) of passenger cars (in thousands of vehicles per day)	22.998	26.885	5.988	128.034
Average daily percent of combination trucks in traffic stream	0.233	0.145	0.034	0.439

Indiana highway and pavement databases. The resulting database is very rich in terms of its segment-level detail. Due to the very large amount of data available, only the summary statistics of variables that were found to be significant in the subsequent model estimations are presented in this paper, as shown in Table 1.

4. Estimation results

The random-parameters negative binomial model resulted in the best statistical fit (relative to the random- and fixed-parameters Poisson models) and was estimated by specifying a functional form of the parameter density function and using simulation-based maximum likelihood with 200 Halton draws (this number of draws has been empirically shown to produce accurate parameter estimates: see Milton et al., 2008, and Bhat, 2003). Halton draws were used instead of random draws (200) because it has been shown (see for example Train, 2003) that fewer draws are needed to achieve convergence and they are more efficient in general.[4]

For the functional form of the random-parameters density functions, consideration is given to normal, lognormal (which restricts the impact of the estimated parameter to be strictly positive or negative), uniform and triangular distributions. For all parameters found to be random, the normal distribution was found to provide the best statistical fit.

Tables 2 and 3 present the model estimation results and marginal effects for the standard and random-parameters negative binomial models, respectively. Table 2 shows that the random-parameters negative binomial results in a significantly better log-likelihood at convergence and better overall fit with the ρ_C^2 statistic improving from 0.177 in the fixed-parameters case to 0.223 in the random-parameters case.[5] Table 3 shows that the average marginal effects generated by the standard negative binomial and random-parameters negative binomial models can be quite different.[6] The likelihood ratio test comparing the fixed- and random-parameters models indicates that we are more than 99.99 percent confident that the random-parameters model is statistically superior.[7] Also, the mean-predicted over the actual values for the two models shown in Fig. 1, indicate that the random-parameters model provides better overall fit relative to the standard negative binomial model. These findings are an outgrowth of the standard negative binomial model's restriction that the estimated parameters are the same for all observations.

A random parameter is used when the standard deviation of the parameter density is statistically significant. If its estimated standard deviation is not statistically different from 0, the parameter is fixed across the population of roadway segments. The estimation results shown in Table 2 indicate that the roadway segment's international roughness index (IRI), the rutting indicator variable (if rutting readings are below 0.12 in.), the road segment length, the median barrier indicator variable (if present), the interior shoulder width indicator variable (if 5 ft or wider), the horizontal curve's degree of curvature per mile, and the average annual daily traffic (AADT) of passenger cars were found to produce statistically significant random parameters (the standard deviation of the parameter's distribution was significantly different from 0).

With regard to the parameters found to be random, the maximum 5-year IRI reading (over the 5 years data accident

[4] Note that the data were also used as cross-sectional time series, considering annual accident occurrences instead of the 5-year accident frequencies. In doing so, each roadway segment generates five observations. To account for the correlation that will be set up with such repeated observations, we estimated negative binomial and Poisson models with fixed and random effects. Our finding with this approach was that both fixed and random effects were not significant. Also, the 5-year data provided better estimation results with regard to overall model fit and the significance of individual fixed and random parameters. In addition, note that higher order terms and a very large number of variable interactions were used for model estimation, and that only the statistically significant variables are shown herein.

[5] The ρ_C^2 statistic is calculated as $1 - \mathrm{LL}(\beta)/\mathrm{LL}(C)$ where $\mathrm{LL}(\beta)$ is the log-likelihood at convergence and $\mathrm{LL}(C)$ is the log-likelihood with only the constant term included in the model (see Washington et al., 2003).

[6] Note that the marginal effects are presented as the averages over the roadway segment population. In fact, the marginal effects and their magnitudes vary considerably across observations.

[7] The test statistic is $\chi^2 = -2[\mathrm{LL}(\beta_R) - \mathrm{LL}(\beta_U)]$, where $\mathrm{LL}(\beta_R)$ is the log likelihood at convergence of the 'restricted' (fixed-parameters) model, and $\mathrm{LL}(\beta_U)$ is the log likelihood at convergence of the 'unrestricted' (random-parameters) model (Washington et al., 2003). This statistic is χ^2 distributed, with degrees of freedom equal to the difference in the numbers of parameters between the restricted and unrestricted models. The resulting χ^2 value was 105.5 with 7 degrees of freedom which gives a p-value of near 0.

Table 2
Model estimation results for random- and fixed-parameters negative binomial models (*t*-statistics in parentheses).

Variable	Negative binomial parameter estimates			
	Random-parameters model		Fixed-parameters model	
Constant	3.158	(9.04)	5.112	(6.734)
Pavement characteristics				
5-year minimum friction readings	−0.026	(−4.779)	−0.030	(−2.54)
5-year maximum international roughness index (IRI) readings	0.004	(3.267)	0.004	(1.292)
Standard deviation of parameter distribution	*0.001*	*(2.336)*		
Average pavement condition rating (PCR) indicator variable (1 if greater than 95, 0 otherwise)	0.132	(1.707)	0.186	(0.939)
Good rutting indicator variable (1 if all 5-year average rutting readings are below 0.2 in., 0 otherwise)	−0.244	(−2.728)	−0.406	(−1.846)
Excellent rutting indicator variable (1 if all 5-year rutting readings are below 0.12 in., 0 otherwise)	−0.473	(−5.078)	−0.469	(−2.183)
Standard deviation of parameter distribution	*0.552*	*(6.795)*		
Geometric characteristics				
Road segment length (in miles)	0.647	(25.592)	0.460	(9.088)
Standard deviation of parameter distribution	*0.171*	*(11.773)*		
Median width indicator variable (1 if greater than 74 ft, 0 otherwise)	−0.344	(−4.302)	−0.554	(−2.981)
Median barrier presence indicator variable (1 if present, 0 otherwise)	−23.875	(−3.079)	−1.866	(−6.538)
Standard deviation of parameter distribution	*11.545*	*(3.061)*		
Inside shoulder width indicator variable (1 if 5 ft or greater, 0 otherwise)	−0.727	(−7.199)	−0.531	(−2.586)
Standard deviation of parameter distribution	*0.581*	*(6.791)*		
Outside shoulder width (in feet)	−0.090	(−3.698)	−0.078	(−1.744)
Average horizontal curve degree curvature per mile	−0.079	(−5.609)	−0.106	(−3.952)
Standard deviation of parameter distribution	*0.061*	*(5.201)*		
Number of vertical curves per mile	−0.237	(−5.859)	−0.237	(−3.089)
Ratio of the vertical curve length over the road segment length (in tenths of miles)	0.104	(5.814)	0.098	(2.252)
Number of bridges per mile	−0.106	(−4.392)	−0.124	(−3.005)
Number of ramps in the driving direction per lane-mile	0.421	(6.177)	0.370	(2.42)
Traffic characteristics				
Average annual daily traffic (AADT) of passenger cars (in thousands of vehicles per day)	0.033	(8.354)	−0.007	(−1.398)
Standard deviation of parameter distribution	*0.026*	*(14.372)*		
Average daily percent of combination trucks in traffic stream	−0.0065	(−1.802)	−0.036	(−4.389)
Dispersion parameter for negative binomial distribution				
Dispersion parameter	0.850	(6.086)	0.878	(9.088)
Number of observations	322		322	
Log-likelihood with constant only	−1140.73		−1140.73	
Log-likelihood at convergence	−886.46		−939.21	
ρ_c^2	0.223		0.177	

data were collected) results in a random parameter that is normally distributed, with a mean 0.004 and standard deviation 0.001.[8] This indicates that increasing IRI (implying rougher pavement) nearly always increases the accident frequency (less than 0.1 percent of the distribution would have a negative value) but with varying magnitude across the population of roadway segments. With regard to marginal effects, Table 3 shows that a unit increase in the maximum IRI (which has average 101.4 in the sample) results in an average 0.007 increase in the number of accidents. This contrasts with an average marginal effect of 0.005 in the fixed-parameter model.

An indicator variable for excellent rutting conditions (1 if all 5-year average rutting readings are below 0.12 in., 0 otherwise) resulted in a random parameter that is normally distributed, with a mean −0.473 and standard deviation 0.552. Given these distributional parameters 80.42 percent of the distribution is less than 0 and 19.58 percent is greater than 0. This implies that the majority of the road segments result in a decrease in accident occurrences for this excellent rutting indicator, and a few of the road segments result in an increase. The average marginal effect of this variable is −0.440.

The road segment length also results in a random parameter that is normally distributed, with a mean 0.647 and standard deviation 0.171. Given these distributional parameters, the vast majority of the road segments result in an accident increase but with varying magnitude. The average marginal effect for this variable (see Table 3) shows that a mile increase in the road segment length results in an average 1.514 increase in the number of accidents (and only a 1.027 increase in the number of accidents in the fixed-parameter model, again underscoring the varying impact across roadway segments). It is also important to note that, although the effect of segment length varies across observations, the mean-parameter for segment length in the random-parameters model is not equal to 1.[9] This likely is the result of the segment-boundary effects on accident frequencies (recall segment lengths are defined by changing roadway geometrics) in that changing geometrics (number of lanes, shoulder widths, etc.) may be associated with accident clustering.[10]

The presence of a median barrier (1 if a median barrier is present and 0 otherwise) resulted in a normally distributed random parameter, with a mean −23.875 and standard deviation 11.545

[8] The international roughness index (IRI) measures pavement roughness with a filtered ratio (referred to as the average rectified slope) of a standard vehicle's accumulated suspension motion (inches) divided by the distance traveled by the vehicle during the measurement (miles).

[9] A parameter value of 1.0 indicates that if roadway segments were cut to produce smaller segments, the same overall accident frequency would result.

[10] On and off ramps from interstate interchanges do not constitute a geometric change as we have defined segments. Thus accidents from these sources are typically within the roadway segment and accounted for with the number-of-ramps variable included in the model.

Table 3
Average marginal effects for random- and fixed-parameters negative binomial models.

Variables	Random-parameters model	Fixed-parameters model
Pavement characteristics		
5-year minimum friction readings	−0.011	−0.012
5-year maximum international roughness index (IRI) readings	0.007	0.005
Average pavement condition rating (PCR) indicator variable (1 if greater than 95, 0 otherwise)	0.142	0.207
Good rutting indicator variable (1 if all 5-year average rutting readings are below 0.2 in., 0 otherwise)	−0.198	−0.288
Excellent rutting indicator variable (1 if all 5-year rutting readings are below 0.12 in., 0 otherwise)	−0.440	−0.436
Geometric characteristics		
Road segment length (in miles)	1.514	1.027
Median width indicator variable (1 if greater than 74 ft, 0 otherwise)	−0.312	−0.478
Median barrier presence indicator variable (1 if present, 0 otherwise)	−5.160	−6.093
Inside shoulder width indicator variable (1 if 5 ft or greater, 0 otherwise)	−0.611	−0.463
Outside shoulder width (in feet)	−0.033	−0.032
Average horizontal curve degree curvature per mile	−0.066	−0.085
Number of vertical curves per mile	−0.199	−0.199
Ratio of the vertical curve length over the road segment length (in tenths of miles)	0.142	0.131
Number of bridges per mile	−0.099	−0.115
Number of ramps in the driving direction per lane-mile	0.455	0.395
Traffic characteristics		
Average annual daily traffic (AADT) of passenger cars (in thousands of vehicles per day)	0.153	−0.006
Average daily percent of combination trucks in traffic stream	−0.006	−0.018

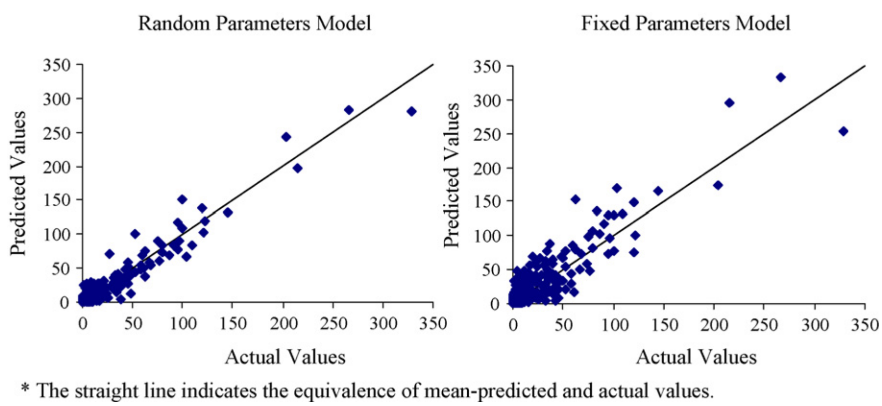

Fig. 1. Mean-predicted over actual number of accidents of random- and fixed-parameters negative binomial models (the straight line indicates the equivalence of mean-predicted and actual values).

(98.06 percent of the distribution is less than 0 and 1.94 percent is greater than 0), implying that median barriers generally reduce the frequency of accidents but their effect varies across the roadway-segment population. One possible explanation for this finding is that the presence of median barriers may result in accidents of lower severity, and because the data used consist only of police-reported accidents, this finding could reflect, among other factors, under-reporting of minor accidents. Table 3 shows that road segments with median barriers decrease the number of accidents by an average of 5.160 which is close to the fixed-parameter model's average of 6.093.[11]

Inside shoulders that were 5 ft or greater resulted in a random parameter that is normally distributed, with a mean −0.727 and standard deviation 0.581 (giving 89.46 percent of the distribution less than 0 and 10.54 percent greater than 0) again reflecting heterogeneity across roadway segments. Marginal effects show that this inside-shoulder indicator variable resulted in an average 0.611 decrease in the number of accidents (the average marginal effect for the fixed-parameters model indicates a 0.463 decrease in the number of accidents).

The degree of horizontal curvature (which is 5731 divided by the radius of the curve in feet; see Mannering et al., 2008) was included in the model by taking the average degree of curvature of all curves in the roadway segment and dividing by the roadway-segment length. Table 3 shows that this variable results in a random parameter that is normally distributed, with a mean −0.079 and standard deviation 0.061 (90.23 percent of the distribution is less than 0 and 9.77 percent is greater than 0)—suggesting considerable variability in the effect of horizontal curvature over the roadway-segment population. The parameter estimates translate into a unit increase in the average degree-of-curve value of the horizontal curves per mile decreasing the number of accidents by an average of 0.142 (and average of 0.131 was found for the fixed-parameters model)—which may be an outgrowth of increased driver alertness in relatively sharper curves. To explain this finding, please keep in mind that all roadway segments are at interstate standards, so the variation in the degree of curvature is limited.

Finally, the annual average daily travel (AADT) results in a random parameter that is normally distributed, with a mean 0.033 and standard deviation 0.026 (both of which were highly signifi-

[11] It is noteworthy that the average estimated parameter in the random-parameters model is large compared to the fixed-parameters model (see Table 2). However, the average marginal effects of the fixed- and random-parameters models are quite close (−5.160 for the random-parameters model and −6.093 for the fixed-parameters model as shown in Table 3). This reflects the effect of the standard deviation of the random parameter when calculating average marginal effects.

cant as indicated by their t-statistics). Given these distributional parameters 89.78 percent is greater than 0 and 10.22 percent of the distribution is less than 0. This implies that the vast majority of the road segments result in an accident increase as AADT increases, and a small proportion of road segments result in a decrease. Note that in the fixed-parameters model, the AADT variable was found not to be significantly different from 0 (as indicated by the t-statistic). This provides additional support for the use of the random-parameters model in that the effect of increasing AADT has a significant impact that generally results in an increase in the number of accidents. The marginal effects for the random-parameters model (Table 3) show that a 1000 vehicle per day increase in AADT increased the number of accidents by an average of 0.153 (for the fixed-parameters model the result was a 0.006 average decrease in the number of accidents). This AADT finding is likely picking up a complex interaction among traffic volume, driver behavior and accident frequency. Because the roadway segments are scattered throughout Indiana, the finding that the effect of AADT increases the number of accidents on some segments and decreases it on others may be capturing, among other factors, the response and adaptation of local drivers to various levels of traffic volume.

5. Summary and conclusions

This paper provides a demonstration of a random-parameters negative binomial regression as a methodological approach to gain new insights into the ways that factors significantly influence accident frequencies. The random-parameters negative binomial regression is an important approach because it allows one to account and correct for heterogeneity that can arise from a number of factors relating to road geometrics, pavement and traffic characteristics, driver behavior, vehicle types, socioeconomic factors, variations in police recording accidents, time and other unobserved factors.

Using 5 years of vehicle accident data from Indiana, the estimation results provide some interesting findings. For example, a variety of factors relating to pavement condition and quality were found to significantly influence vehicle accident occurrences including the effects of friction, the international roughness index (IRI), pavement rutting and the pavement's condition rating (PCR). In terms of geometric factors and their effect on vehicle accident frequencies, road segment length, median types and width, number of ramps and bridges, horizontal and vertical curves and shoulder widths were all found to be statistically significant. Also, the traffic variables of annual average daily travel (AADT) and the percent of combination trucks in the traffic stream were both found to have a significant impact on accident frequencies. However, because a number of these factors (IRI, rutting, road segment length, median barrier presence, inside shoulder width, horizontal curvature, AADT) resulted in random parameters, their effect on accident frequencies was found to vary significantly across roadway segments.

While this study is exploratory in nature, it does suggest the considerable potential for random-parameters count models in analyzing accident frequency. The findings show that ignoring the possibility of random parameters when estimating count-data models can result in substantially different marginal effects and subsequent inferences relating to the magnitude of the effect of factors affecting accident frequencies.

References

Abdel-Aty, M.A., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accident Analysis and Prevention 32 (5), 633–642.

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

Bhat, C., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research Part B 17 (1), 677–693.

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

Brownstone, D., Train, K., 1999. Forecasting new product penetration with flexible substitution patterns. Journal of Econometrics 89 (1–2), 109–129.

Cameron, A., Trivedi, P., 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. Journal of Applied Econometrics 1 (1), 29–53.

Carson, J., Mannering, F., 2001. The effect of ice warning signs on accident frequencies and severities. Accident Analysis and Prevention 33 (1), 99–109.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention 40 (3), 1033–1054.

Gkritza, K., Mannering, F., 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.

Hadi, M.A., Aruldhas, J., Chow, L.-F., Wattleworth, J.A., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. Transportation Research Record 1500, 169–177.

Hall, R.D., 1986. Accidents at Four-arm Single Carriageway Urban Traffic Signals. Transport Research Laboratory. Contractor Report CR65. Crowthorne, U.K.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2, 84–90.

Hess, S., Polak, J.W., Daly, A., 2003. On the performance of shuffled Halton sequences in the estimation of discrete choice models. Paper presented at the European Transport Conference, Strasbourg.

Jones, B., Janssen, L., Mannering, F., 1991. Analysis of the frequency and duration of freeway accidents in Seattle. Accident Analysis and Prevention 23 (2), 239–255.

Lee, J., Mannering, F., 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.

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., Park, Y.-J., 2008. Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. Accident Analysis and Prevention 40 (4), 1441–1457.

Lord, D., Guikema, S.D., 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.

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., Washburn, S., Kilareski, W., 2008. Principles of Highway Engineering and Traffic Analysis, 4th ed. John Wiley and Sons, New York, NY.

Maycock, G., Hall, R.D., 1984. Accidents at 4-Arm Roundabouts. Laboratory Report LR1120. Transport Research Laboratory, Crowthorne, Berks, U.K.

McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15, 447–470.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor vehicle accident frequencies. Transportation 25 (4), 395–413.

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

Pickering, D., Hall, R.D., Grimmer, M., 1986. Accidents at Rural T-junctions. Research Report RR65. Transport Research Laboratory, Crowthorne, Berks, U.K.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering 122 (2), 105–113.

Revelt, D., Train, K., 1997. Mixed logit with repeated choices: households' choice of appliance efficiency level. Review of Economics and Statistics 80 (4), 647–657.

Revelt, D., Train, K., 1999. Customer Specific Taste Parameters and Mixed Logit. Working Paper. University of California, Department of Economics, Berkeley.

Satterthwaite, S.P., 1981. A Survey of Research into Relationships Between Traffic Accidents and Traffic Volumes. Supplementary Report 692. Transport Research Laboratory, Crowthorne, Berks, U.K.

Savolainen, P.T., Tarko, A.P., 2005. Safety impacts at intersections on curved segments. Transportation Research Record 1908, 130–140.

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

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

Train, K., 1997. Mixed logit models for recreation demand. In: Kling, C., Herriges, J. (Eds.), Valuing the Environment Using Recreation Demand Models. Elgar Press, New York.
Train, K., 1999. Halton Sequences for Mixed Logit. Working Paper. University of California, Department of Economics, Berkley.
Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, Cambridge, UK.
Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC.
Wright, D.E., Barnett, V., 1991. Fitting Predictive Accident Models in GLIM with Uncertainty in the Flow Estimates. Contractor Report CR286. Transport Research Laboratory, Crowthorne, Berks, U.K.