Research Article

Analysis of Severe Injury Accident Rates on Interstate Highways Using a Random Parameter Tobit Model

Minho Park¹ and Dongmin Lee²

¹Highway & Transportation Research Division, Korea Institute of Civil Engineering and Building Technology, Gyeonggi-do, Republic of Korea

²Department of Transportation Engineering, University of Seoul, Seoul, Republic of Korea

Correspondence should be addressed to Dongmin Lee; dmllee@uos.ac.kr

Received 8 November 2016; Accepted 12 January 2017; Published 2 February 2017

Academic Editor: Inmaculada T. Castro

Copyright © 2017 Minho Park and Dongmin Lee. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this study, a random parameter Tobit regression model approach was used to account for the distinct censoring problem and unobserved heterogeneity in accident data. We used accident rate data (continuous data) instead of accident frequency data (discrete count data) to address the zero cell problems from data where roadway segments do not have any recorded accidents over the observed time period. The unobserved heterogeneity problem is also considered by using random parameters, which are parameter estimates that vary across observations instead of fixed parameters, which are parameter estimates that are fixed/constant over observations. Nine years (1999–2007) of panel data related to severe injury accidents in Washington State, USA, were used to develop the random parameter Tobit model. The results showed that the Tobit regression model with random parameters is a better approach to explore factors influencing severe injury accident rates on roadway segments under consideration of unobserved heterogeneity problems.

1. Introduction

Over the last decade, numerous studies have been conducted to explore the factors that cause accidents on roadway segments; various statistical modeling techniques have been employed, especially count models. Initially, simple linear regression models were employed. Although this is the simplest and easiest model, the data generally violate the basic assumption of homoscedasticity, which means that the variance increases as the variable increases. In addition, negative accident count values are predicted which should be greater than or equal to zero in reality [1]. To address these problems that the linear regression model had, previous investigators suggested a Poisson regression model wherein accident frequency is translated as a discrete random variable [2]. However, there was an important constraint in this model, namely, that the mean must be equal to the variance; the standard errors will be biased when this restriction is not valid. Meanwhile, real accident data were found to be

overdispersed, meaning that the variance is greater than the mean [3, 4]. As a result, the Poisson regression model incorrectly estimates the likelihood of accident frequencies and, to overcome this overdispersion problem, a negative binomial model was suggested, which relaxes the constraint that the mean is equal to the variance. Negative binomial models had been shown to be more appropriate than the Poisson model for describing the relationships between accident frequency and geometric elements [5–8]. Besides, a variety of attempts to analyze accident frequencies were made, and these resulted in a reduction in accidents and improved accident prevention (random effects models [9, 10], zero-inflated count models [11–13], and random parameters count models [14–18]).

Aside from methodologies that address accident frequency data, a Tobit regression method that uses accident rates as the dependent variable was also suggested [19]. This method employs continuous variables, that is, the number of accidents per vehicle-mile traveled, instead of accident frequency (discrete count data) data. In addition, the likelihood

that segments/spots will have no accident records during some period can be taken into account using data that is left-censored at zero. Particularly, accident frequencies that involve severe injury accidents (fatal and disabling injuries) are lower than those of other types of accidents such as EPDO, possible injuries, and evident injuries, that is, zero records in many cases. This has made it difficult to analyze severe injury accidents using the traditional frequency method, which is the reason why the Tobit model is a more appropriate approach to analyze the causes of severe injury accidents than the traditional frequency models.

In addition, there is another remaining problem, namely, that previous methods mentioned above do not account for heterogeneity. Since aggregated accident data for analysis do not have information related to the residual environmental effects and socioeconomic, driver, and vehicle characteristics, there is a possibility that unobserved heterogeneity may occur in accident data, which creates variation in the impact of the effect of observed variables on accident frequencies [20]. This problem cannot be addressed in the traditional count models in which the estimated parameters are fixed. Therefore, we must consider models that include random parameters, which allow some or all parameters to vary randomly across observations. Some relevant research has shown that the random parameters approach can account for variations in the variables [14–16, 18, 21].

Here, we developed a random parameter Tobit model that allowed us to address the unobserved heterogeneity in accident data for severe injuries and also compared the estimated results from a fixed parameter Tobit model on roadway segments except for interchange segments on interstates. To the best of the authors' knowledge, this is the first attempt to model severe injury accident case using random parameters Tobit model method. And this approach using accident rates could be applied in the process of select performance measures in HSM (Highway Safety Manual) whose framework and modeling architecture have been introduced in [22].

2. Methodology

The Tobit model is a regression model proposed by Tobin (1958) in which the dependent variable is either left- or right-censored. Here, left-censored means that the data are censored at a low threshold, while right-censored data are censored at a high threshold. In accident data, the data will be left-censored with clustering at a zero base since vehicle crash frequencies may not be observed on all/some segments during the observation period. Using this information, the Tobit model was constructed as follows:

$$Y_i^* = \beta'X_i + \varepsilon_i, \quad i = 1, 2, 3, \dots, N, \quad (1)$$

$$Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > 0 \\ 0 & \text{if } Y_i^* \leq 0. \end{cases}$$

Here, N is the number of observations, Y_i is the dependent variable (severe injury accident rate), β is a vector of estimable parameters, X_i is a vector of independent variables (e.g., traffic volumes and segment geometrics), and ε_i is a normally

and independently distributed error term with zero mean and constant variance σ^2 . Here, there is an implicit and stochastic index (latent variable) expressed as Y_i^* , which is observed only when the value of Y_i is greater than zero (positive). Hence, the likelihood function for the Tobit model over zero and positive observations is as follows:

$$L = \prod_0 \left[1 - \Phi \left(\frac{\beta X}{\sigma} \right) \right] \prod_1 \sigma^{-1} \Phi \left[\frac{(Y_i - \beta X)}{\sigma} \right]. \quad (2)$$

Here, Φ refers to the standard normal distribution function and ϕ is the standard normal density function.

The traditional Tobit (fixed parameter) model is described. However, it is difficult to account for heterogeneity (unobserved factors that may vary across observations) in this model. In order to account for heterogeneity using a random parameter, Greene [23] developed a simulated maximum likelihood estimation procedure, which has been shown to be an acceptable method [15, 16, 18, 21].

Estimable parameters that allow for random parameters are as follows:

$$\beta_i = \beta + \varphi_i, \quad i = 1, 2, 3, \dots, N. \quad (3)$$

Here, β indicates estimated parameters and φ_i is a randomly distributed term. Uniform, normal, lognormal, and other forms are considered to be potential density functions for random parameter estimation. The latent variable mentioned in (1) becomes $Y_i^* | \varphi_i = \beta X_i + \varepsilon_i$, and the likelihood function from (2) is as follows in log-likelihood form:

$$LL = \sum_{\forall i} \ln \int_{\varphi_i} g(\varphi_i) P(n_i | \varphi_i) d\varphi_i. \quad (4)$$

Here, g refers to the probability density function of φ_i .

To estimate the random parameters, a simulation-based maximum likelihood using Halton draws was employed which is an efficient distribution of draws for numerical integration [24, 25]. In summary, the random parameter Tobit model could account for unobserved factors and at the same time support the complete use of available data from left-censored severe injury traffic accident data.

3. Data

Vehicle crash accident data of roadway segments on interstates in Washington State (I-5, I-82, I-90, I-182, I-205, I-405, and I-705) had been collected over 9 years (1999 to 2007) to investigate the effects of geometrics and traffic flow conditions such as number of lanes, right and left shoulder width, number of horizontal and vertical curves, and traffic volumes on severe injury accident rates per 100-million vehicle-miles traveled (VMT).

Firstly, the collected data were divided into data on roadway segments and data on interchange segments of the interstate highways. In this study, only crash data on roadway segments were used because crashes on interchange might generally occur within various effects including traffic flow changes, weaving maneuvers, complex geometrics, driver behaviors by traffic signs, and other different conditions

TABLE 1: Descriptive statistics of variables.

Variable description	Mean	Std. dev.	Minimum	Maximum
Number of disabling injury accidents	0.161	0.439	0	5
Number of fatality injury accidents	0.058	0.246	0	2
Segment length (mi)	1.837	2.329	0.010	20.380
Average annual daily traffic volume	13,052	8,426	916	44,224
Number of lanes per direction	2.622	0.735	1	5
Left shoulder width (ft)	6.946	3.217	2	14
Right shoulder width (ft)	6.943	3.280	2	18
Number of horizontal curves	1.817	2.483	0	37
Number of vertical curves	3.183	3.248	0	30

TABLE 2: Model estimation results.

	Fixed parameter model		Random parameter model	
	Parameter estimate	<i>t</i> -ratio	Parameter estimate	<i>t</i> -ratio
Constant	−6.415	−10.819	−6.131	−11.568
Logarithm of segment length	1.015	20.651	1.033	17.528
Standard deviation of parameter distribution	NA	NA	0.989	28.438
Logarithm of annual average daily traffic volume	0.446	6.847	0.353	6.130
Standard deviation of parameter distribution	NA	NA	0.303	14.646
Number of lanes per direction	0.248	4.176	0.358	6.222
Standard deviation of parameter distribution	NA	NA	0.288	7.727
Left shoulder width	−0.023	−1.975	−0.020	−2.201
Standard deviation of parameter distribution	NA	NA	0.024	3.308
Right shoulder width	−0.011	−0.893	−0.024	−2.069
Standard deviation of parameter distribution	NA	NA	0.016	3.000
Number of horizontal curves	0.014	0.649	−0.025	−2.795
Standard deviation of parameter distribution	NA	NA	0.032	2.955
Number of vertical curves	−0.030	−2.044	−0.021	−2.487
Standard deviation of parameter distribution	NA	NA	0.019	2.962
Number of observations		589		
Log-likelihood function at zero		−3,485.29		
Log-likelihood function at convergence	−3,191.58		−3,169.62	

NA: not applicable.

from roadway segments. Consequentially, over a continuous period of nine years, the 589 roadway segments which were used for the analysis yielded a panel of 5,301.

Accident rate, the dependent variable, was calculated using the following equation:

$$\text{accident rates}_i = \frac{\sum_{y=1}^n \text{accidents}_{y,i}}{\left[\sum_{y=1}^n \text{AADT}_{y,i} \times L_i \times 365 \right] / 100,000,000} \quad (5)$$

Here, accident rates_i is the total number of severe accidents per 100-million VMT on segment i , y is the year of observed data, $\text{accidents}_{y,i}$ is the number of severe accidents on segment i in year y , $\text{AADT}_{y,i}$ is the average annual daily traffic volume on segment i in year y , and L_i is the length of segment i . Since we sought to determine the effects of geometrics on severe accidents, the dependent variable is defined as the summation of disabling and fatal injuries.

The descriptive statistic values for the primary variables are shown in Table 1. The average length of roadway segments was 1.837 miles, and 13,052 vehicles is the mean value of the average annual daily traffic volume on the objective segments during the study period. On average, 2.6 lanes per direction exist with a minimum of one lane and a maximum of five lanes. The mean value of the shoulder width is 6.9 ft on both the left and the right shoulder. In terms of curves, 1.8 horizontal curves and 3.2 vertical curves exist on the roadway segments.

4. Model Estimation Results

Two types of modeling methods were used to estimate whether parameters are fixed (fixed parameters, left side in Table 2) or they vary across observations (random parameters, right side in Table 2). For random parameter estimations, Halton draws were used, which has been shown to produce accurate parameter estimations [25]. The normal distribution

TABLE 3: Marginal effect and elasticity values of fixed and random parameters Tobit model.

	Fixed parameter		Random parameter	
	Marginal effect	Elasticity	Marginal effect	Elasticity
Logarithm of segment length	0.152	1.015	0.079	1.033
Logarithm of annual average daily traffic volume	0.067	0.446	0.022	0.389
Number of lanes per direction	0.037	0.649	0.023	0.938
Left shoulder width	−0.003	−0.159	−0.001	−0.139
Right shoulder width	0.002	0.074	0.002	0.166
Number of horizontal curves	0.002	0.025	−0.002	−0.046
Number of vertical curves	−0.004	−0.095	−0.001	−0.066

of density functional forms gave the best statistical results among the normal, uniform, and lognormal distributions mentioned in the Methodology.

The overall log-likelihood at convergence in the random parameter Tobit model (−3,169.62) showed relatively greater improvement than the fixed parameter Tobit model (−3,191.58). As described in Table 2, a total of seven variables with random parameters were derived to have an effect on the severe accident rates. These parameters are segment length, average annual daily traffic volume, number of lanes, left/right shoulder width, and number of horizontal/vertical curves. A random parameter was used when both the mean and the standard deviation of the parameter distribution were statistically significant ($\neq 0$). In this sense, a parameter in which standard deviation is not statistically significant ($= 0$) indicates that the effects are fixed across all segments. All derived variables with random parameters showed statistically significant mean and standard deviation values. On the other hand, some variables with fixed parameters were found to be statistically insignificant, which shows the flexibility of the random parameters in that the effect of the covariates must be constant/fixed across all observations [21].

The results of modeling, the marginal effect, and elasticity of the random parameters and fixed parameters models are presented in Tables 2 and 3, respectively. The logarithms of segment length, traffic volumes, and number of lanes were shown to have statistically significant fixed and random parameters with positive signs. This is consistent with the expectation of increased frequency of severe injury accidents with higher exposure (longer length, higher traffic volumes, and more lanes) on the roads.

A random parameter of the segment length that is normally distributed with a mean of 1.033 and standard deviation of 0.989 indicates that the effect of segment length decreases the severe injury accident frequency rate on 14.83% of the observed segments and increases the rate on 85.17% of the observed segments. In terms of elasticity, a 1% increase in length contributed to a 1.015% (fixed parameter) and 1.033% (random parameter) increase in severe injury accident rate; these are indications of elasticity.

We found that traffic volumes have a normally distributed random parameter with a mean of 0.353 and a standard deviation of 0.303. Given this distribution, the effect of traffic volume decreases the severe accident rate on 12.12% of segments and increases the severe accident rate on 87.88% of segments.

The number of lanes variable had a random parameter with a normal distribution with a mean of 0.358 and a standard deviation of 0.288. Given these distribution values, 10.69% of segments showed a decrease in the severe injury rate, and 89.31% of segments showed an increase in the severe injury rates.

With regard to shoulder width, a negative sign (severe injury rate decrease) was found for both fixed and random parameters. The left shoulder width had a normally distributed random parameter with a mean of −0.020 and a standard deviation of 0.024. These parameter values indicated that increasing the shoulder width decreases the severe injury rate in 79.86% of segments and increases the severe injury rate in 20.14% of segments. The right shoulder width variable shows a positive sign and is not statistically significant in the fixed parameter model. However, the right shoulder width variable was found to have a normally distributed random parameter with a mean of −0.024 and a standard deviation of 0.016 in the random parameter model. These distribution values mean that the right shoulder width increases severe injury rates in 7.04% of the main line segments and decreases severe injury rates in 92.96% of the main line segments. These results are consistent with previous studies [15, 17], which have shown that accident probability decreases as shoulder width increases.

We estimated that the number of horizontal curves variable had a random parameter with a mean of −0.025 and a standard deviation of 0.032. This variable decreases severe accident rates in 78.28% of the roadway segments and increases severe accident rates in 21.72% of the roadway segments. The effect of the number of horizontal curves on the severe injury accident rate in the fixed parameters model was positive for all interstate roadway segments considered, but no statistically significant influence was derived; this brings about additional support to the use of the random parameter. Finally, the variable for the number of vertical curves was shown to be statistically significant in both the fixed parameter model and the random parameter model under a normal distribution with a mean of −0.021 and a standard deviation of 0.019. This random parameter result indicates that the effect of the number of vertical curves decreases the likelihood of severe injury rates in 85.49% of all observed segments and increases the likelihood of severe injury rate in 14.51% of all observed segments. Results from these curve-related variables are similar to a previous study that showed

that some variations in roadway geometrics may improve driver alertness, resulting in more careful driving [26].

5. Conclusions

We used a Tobit regression model with fixed and random parameters to examine the geometric factors that influence the rate of accidents that result in severe injury (fatal and disabling injury). In the Tobit regression model, a dependent variable (severe injury accident rate) was applied as a continuous variable and was left-censored at zero, which is an alternative to the traditional discrete accident frequency approach. Unobserved heterogeneities were considered as well in the parameter estimation processes by employing a random parameter, which is difficult to do in traditional fixed parameter estimation. In other words, by using the random parameter Tobit method, heterogeneity from factors such as vehicle type, weather, individual character, and other unobserved factors which are not captured in the data collection was accounted and corrected for in finding significant factors on the severe injury accident rates on the interstates.

Nine years of severe injury accident and geometrics data from seven interstate main lines in Washington State, USA, were used to develop the models. Seven variables were found to have random parameters with statistically significant standard deviation values. The effects of these variables vary across observations: segment length, annual average daily traffic volume, number of lanes, left and right shoulder width, and numbers of horizontal and vertical curves.

While this study is exploratory in nature, random parameters Tobit regression model has the potential to provide a fuller understanding of the factors on severe accident rates, which showed the outstanding result related to fixed parameters model. Although the predictive power was improved (as shown in log-likelihood values), other possible variables that may affect the likelihood of severe accidents such as pavement conditions were not considered. In addition, interchange segments having more complex infrastructures and more diverse traffic flow types (merge, diverge, and weave) were not considered in this study. We recommend that future studies include those variables and various analyses for exploring additional causes of traffic accidents.

Competing Interests

The authors declare that there are no competing interests regarding the publication of this paper.

References

- [1] P. P. Jovanis and H. L. Chang, "Modeling the relationship of accidents to miles traveled," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1068, pp. 42–51, 1987.
- [2] S. C. Joshua and N. J. Garber, "Estimating truck accident rate and involvements using linear and poisson regression models," *Transportation Planning and Technology*, vol. 15, no. 1, pp. 41–58, 1990.
- [3] V. Shankar, F. Mannering, and W. Barfield, "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies," *Accident Analysis and Prevention*, vol. 27, no. 3, pp. 371–389, 1995.
- [4] A. Vogt and J. Bared, "Accident models for two-lane rural segments and intersections," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1635, pp. 18–29, 1998.
- [5] J. Engel, "Models for response data showing extra-poisson variation," *Statistica Neerlandica*, vol. 38, no. 3, pp. 159–167, 1984.
- [6] J. F. Lawless, "Negative binomial and mixed Poisson regression," *The Canadian Journal of Statistics*, vol. 15, no. 3, pp. 209–225, 1987.
- [7] M. J. Maher, "New bivariate negative binomial model for accident frequencies," *Traffic Engineering and Control*, vol. 32, no. 9, pp. 422–423, 1991.
- [8] S.-P. Miaou and H. Lum, "Modeling vehicle accidents and highway geometric design relationships," *Accident Analysis and Prevention*, vol. 25, no. 6, pp. 689–709, 1993.
- [9] V. N. Shankar, R. B. Albin, J. C. Milton, and F. L. Mannering, "Evaluating median crossover likelihoods with clustered accident counts an empirical inquiry using the random effects negative binomial model," *Transportation Research Record*, no. 1635, pp. 44–48, 1998.
- [10] H. C. Chin and M. A. Quddus, "Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections," *Accident Analysis and Prevention*, vol. 35, no. 2, pp. 253–259, 2003.
- [11] V. Shankar, J. Milton, and F. L. Mannering, "Modeling accident frequencies as zero-altered probability processes: an empirical inquiry," *Accident Analysis & Prevention*, vol. 29, no. 6, pp. 829–837, 1997.
- [12] J. Lee and F. Mannering, "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis," *Accident Analysis and Prevention*, vol. 34, no. 2, pp. 149–161, 2002.
- [13] N. Malyshikina and F. L. Mannering, "Zero-state Markov switching count-data models; an empirical assessment," *Accident Analysis and Prevention*, vol. 42, no. 1, pp. 122–130, 2010.
- [14] P. Ch. Anastasopoulos and F. L. Mannering, "A note on modeling vehicle-accident frequencies with random parameter count models," *Accident Analysis and Prevention*, vol. 41, no. 1, pp. 153–159, 2009.
- [15] N. S. Venkataraman, G. F. Ulfarsson, V. Shankar, J. Oh, and M. Park, "Model of relationship between interstate crash occurrence and geometrics: exploratory insights from random parameter negative binomial approach," *Transportation Research Record*, no. 2236, pp. 41–48, 2011.
- [16] N. Venkataraman, G. F. Ulfarsson, and V. N. Shankar, "Random parameter models of interstate crash frequencies by severity, number of vehicles involved, collision and location type," *Accident Analysis and Prevention*, vol. 59, pp. 309–318, 2013.
- [17] M. Park, "Relationship between interstate highway accidents and heterogeneous geometrics by random parameter negative binomial model—a case of interstate highway in Washington State, USA," *Journal of the Korean Society of Civil Engineers*, vol. 33, no. 6, pp. 2437–2445, 2013.
- [18] N. Venkataraman, V. Shankar, G. F. Ulfarsson, and D. Deptuch, "A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies," *Analytic Methods in Accident Research*, vol. 2, pp. 12–20, 2014.

- [19] P. C. Anastasopoulos, A. P. Tarko, and F. L. Mannering, "Tobit analysis of vehicle accident rates on interstate highways," *Accident Analysis and Prevention*, vol. 40, no. 2, pp. 768–775, 2008.
- [20] F. L. Mannering, V. Shankar, and C. R. Bhat, "Unobserved heterogeneity and the statistical analysis of highway accident data," *Analytic Methods in Accident Research*, vol. 11, pp. 1–16, 2016.
- [21] P. C. Anastasopoulos, F. L. Mannering, V. N. Shankar, and J. E. Haddock, "A study of factors affecting highway accident rates using the random-parameters tobit model," *Accident Analysis and Prevention*, vol. 45, pp. 628–633, 2012.
- [22] N. S. Venkataraman, G. F. Ulfarsson, and V. N. Shankar, "Extending the Highway Safety Manual (HSM) framework for traffic safety performance evaluation," *Safety Science*, vol. 64, pp. 146–154, 2014.
- [23] W. L. Greene, *Version 9.0*, Econometric Software Inc, Plainview, NY, USA, 2007.
- [24] J. H. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, vol. 2, no. 1, pp. 84–90, 1960.
- [25] C. R. Bhat, "Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences," *Transportation Research Part B: Methodological*, vol. 37, no. 9, pp. 837–855, 2003.
- [26] C. Winston, V. Maheshri, and F. Mannering, "An exploration of the offset hypothesis using disaggregate data: the case of airbags and antilock brakes," *Journal of Risk and Uncertainty*, vol. 32, no. 2, pp. 83–90, 2006.

Hindawi

Submit your manuscripts at
<https://www.hindawi.com>