A Tobit Model of Vehicle Accident Rates

by

Panagiotis Ch. Anastasopoulos
Research Assistant - School of Civil Engineering
550 Stadium Mall Drive
Purdue University
West Lafayette, IN 47907-2051
panast@purdue.edu

Andrew P. Tarko
Professor of Civil Engineering
550 Stadium Mall Drive
Purdue University
West Lafayette, IN 47907-2051
tarko@ecn.purdue.edu

and

Fred L. Mannering
Professor of Civil Engineering
550 Stadium Mall Drive
Purdue University
West Lafayette, IN 47907-2051
(765) 496-7913
flm@ecn.purdue.edu

July 27, 2007
Revised September 9, 2007
Prepared for *Accident Analysis and Prevention*

Abstract

In recent years, there has been an abundance of research that has used Poisson models and its variants (negative binomial and zero-inflated models) to improve our understanding of the factors that affect accident frequencies on roadway segments. This study explores the application of an alternate method, tobit regression, by viewing vehicle accident rates directly (instead of frequencies) as a continuous variable that is left-censored at zero. Using data from vehicle accidents on Indiana interstates, the estimation results show that many factors relating to pavement condition, roadway geometrics and traffic characteristics significantly affect vehicle accident rates.

Keywords: Accident rates; Tobit regression; Pavement condition; Roadway geometrics

Introduction

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. Because such accident-frequency data are non-negative count data, Poisson and negative binomial models have been widely used (Shankar et al., 1995; Hadi et al., 1995; Poch and Mannering, 1996; Abdel-Aty and Radwan, 2000; Savolainen and Tarko, 2005). In addition, some variations of these modeling approaches have also been applied, such as zero-inflated Poisson and zero-inflated negative binomial (Shankar et al., 1997; Carson and Mannering, 2001; Lee and Mannering, 2002) and negative binomial with random effects (Shankar et al., 1998). To be sure, this body of literature has provided numerous insights into the factors affecting accident frequencies on roadway segments. However, using exposure-based accident rates instead of traditional accident counts as the dependent variable has considerable appeal because accident rates are widely used in accident reporting. For example, the number of accidents per 100 million miles driven is a commonly reported standard in the literature. The use of accidents per vehicle-miles traveled has an intuitive appeal in highway safety – providing a standardized measure of the relative safety of roadway segments that is more easily interpreted than the number of accidents per some time period.

From the perspective of statistical modeling, accidents per vehicle-miles traveled is a continuous variable instead of the non-negative integer used when studying the number of accidents on a roadway segment over some time period. However, because accident rates on specific highway segments are assessed over some finite time period, there is the likelihood that many highway segments will have no accidents reported during the analysis period. Thus, modeling accident rates by standard ordinary least squares would result in biased and inconsistent parameter estimates (Washington et al., 2003). The solution to this is to consider accident rates as a censored dependent variable (censored at zero) and apply a tobit model.

While having not been previously applied to the study of highway accident rates (to the authors' knowledge), tobit models have been widely used in economics (Tobin, 1958; Nelson, 1977; Keeley et al., 1978), social sciences (Fair, 1978; Roncek, 1992; Smith and Brame, 2003), and medicine (Baser et al., 2003; Ekstrand and Carpenter, 1998). In the transportation arena in general, tobit models are not commonly used and only a hand full of applications can be found. As examples, Talley (1995) estimated a tobit model to predict the determinants of accident passenger-vessel damage cost; Nolan (2002) used a tobit regression to model household petrol, bus and taxi fare expenditure household decisions in Ireland; and Weiss (1992) used the approach to asses helmet use on motorcycle head-injury accidents.

The intent of the current paper is to demonstrate the potential of tobit regression as a tool to study accident rates on roadway segments. This will be done by applying the model to vehicle accident rates on interstate roadway segments in Indiana.

Methodology

The tobit model was first proposed by James Tobin (1958), and refers to regression models in which the range of the dependent variable is censored in some way. Censoring refers to a data limitation that can result in a data clustering at a lower threshold (left censored), an upper threshold (right censored), or both. Censored data differ from truncated data in that in truncated data only non-limited values are available (censored data provide information on limited data as well). For the accident-rate data that we will consider, the data will be 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.

The tobit model is expressed (for roadway segment *i*) using a limit of zero (which will be the case for our analysis) as:

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

where N is the number of observations, Y_i is the dependent variable (accidents per 100-million vehicle miles traveled), $\boldsymbol{X}_i$ is a vector of independent variables (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. It is assumed 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

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

where Φ is the standard normal distribution function and ϕ is the standard normal density function.

With this model, the expected value of the dependent variable for all cases, $E[Y]$, is (removing the subscript *i* to simplify the exposition):

$$E[Y] = \boldsymbol{\beta X}F\left(z\right) + \sigma f\left(z\right) \tag{3}$$

where $z = \boldsymbol{\beta X}/\sigma$ 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 the zero, $f(z)$ the unit normal density (value of the derivative of the normal curve at a particular point), and σ the standard deviation of the error term. And, the relationship among the expected value of all observations, $E[Y]$; the expected value for cases above zero, $E[Y']$; and the probability of being above zero, $F(z)$ is:

$$E[Y] = F(z)E[Y'] \tag{4}$$

where Y' denotes observations above zero in our accident-rate case.

To determine the effect of an independent variable on the expected value, the first-order partial derivative of equation 3 is used and, as shown by Amemiya (1973, 1985), the expected value of Y for observations above zero, is $\boldsymbol{\beta X}$ plus the expected value of the truncated normal error term:

$$\begin{aligned} E[Y'] &= E[Y \mid Y > 0] \\ &= E[Y \mid \varepsilon > -\boldsymbol{\beta X}] \quad . \\ &= \boldsymbol{\beta X} + \sigma f(z)/F(z) \end{aligned} \tag{5}$$

McDonald and Moffitt (1980) find for the first-order partial derivative for a specific independent variable X_k of equation 3 to be:

$$\frac{\partial E[Y]}{\partial X_k} = F(z)\left(\frac{\partial E[Y']}{\partial X_k}\right) + E[Y']\left(\frac{\partial F(z)}{\partial X_k}\right), \tag{6}$$

where $\partial E[Y]/\partial X_k$ is the change in the overall expected value, $\partial E[Y']/\partial X_k$ is the change in the expected value for cases above zero (weighted by the probability of being above zero), and $\partial F(z)/\partial X_k$ is the change in the cumulative probability of being above zero (weighted by the expected value of Y if above the limit), associated with an independent variable (see Roncek, 1992). Assuming that either $\partial E[Y']/\partial X_k$ or $\partial F(z)/\partial X_k$ has estimates of β_k and σ, each of the terms in equation 6 can be evaluated at some value of $\beta_k X_k$, usually at the mean of X_k. The value of $E[Y']$ can be calculated from equation 5, and the value of $F(z)$ can be obtained directly (from statistical tables, etc.). Roncek's decomposition method (Roncek, 1992) provides the change in the cumulative probability of being zero as,

$$\frac{\partial F(z)}{\partial X_k} = \beta_k \frac{f(z)}{\sigma}, \qquad (7)$$

where β_k is the estimated tobit parameter for a particular independent variable k, and from equation 5, the change in the expected value for cases above zero is,

$$\frac{\partial E[Y']}{\partial X_k} = \beta_k \left[1 - z\frac{f(z)}{F(z)} - \frac{f(z)^2}{F(z)^2} \right], \qquad (8)$$

where z is the z-score associated with the area under the normal curve. A common error in literature is to assume that the estimated tobit estimated parameters measure the correct regression parameters for observation above the limit. As indicated McDonald and Moffitt (1980) presented, this is only true when $X = \infty$, which would mean that $F(z) = 1$ and $f(z) = 0$.

Empirical Setting

Vehicle accident data from interstate highways in Indiana (I-64, I-65, I-70, I-74, and I-164) were collected for a 5-year period (1 January 1995, to 31 December 1999) to investigate the effect of pavement characteristics, highway geometries, and traffic characteristics on accident rates per 100-million vehicle miles traveled (VMT). To study this relationship, the roadway data were collected from the Indiana Department of Transportation and 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 337 roadway segments were defined and the number of police-reported vehicle accidents occurring on each segment over the five-year period was obtained from the Indiana State Patrol accident-data files.

For model estimation, our data included the aggregated number of accidents on each roadway segment over a five year period: from 1 January 1995, to 31 December 1999. The accident rate (number of accidents per 100-million VMT) was calculated as:

$$Accident\ Rate_i = \frac{\sum_{Year=1}^{5} Accidents_{Year,i}}{\left[\sum_{Year=1}^{5} AADT_{Year,i} \times L_i \times 365\right] \big/ 100{,}000{,}000}, \qquad (8)$$

where $Accident\ Rate_i$ is the number of accidents per 100-million VMT on roadway segment i, $Year$ denotes the year (from 1 to 5 representing 1995 to 1999), $Accidents_{Year,i}$ is the number of accidents, $AADT_{Year,i}$ the average annual daily traffic, L_i the length of roadway segment i.

Estimation results

Tables 1 and 2 present summary statistics and tobit estimation results.[1] To assess the influence of specific variables, Table 3 presents the computed values for the change in the expected value of accident rates for roadway segments (vehicle accidents per 100-million VMT) and the change in the probability of having a 100-million VMT accident rate above zero for the accident-observation period.[2] The estimation results provided in Table 2 show that the estimated parameters are significant and of plausible sign based on the sample of 325 roadway segments that had complete information on all variables used. Also, the overall statistical fit of the model was quite good as indicated by the Maddala psuedo-R^2 of 0.948.[3]

Turning to specific estimation results, Table 3 shows that Interstates I-70 and I-164 tend to have a lower accident rate 26.38 and a 22.48% lower probability of having a 100-million VMT accident rate above zero. While it is impossible to know for sure, there are a number of speculative explanations for this finding. One possibility is that, because both of these interstates are located to the southern part of Indiana, this finding could be reflecting more favorable weather conditions. There is also the possibility that different driving behavior and/or temporal distributions of traffic on these interstates may be contributing to the heterogeneity that this indicator variable seems to be capturing.

---

[1] There are a number of econometric software packages available to estimate tobit models. The estimations presented in this paper were estimated using Limdep (Greene, 2007).

[2] We also used the data as cross-sectional time-series, considering annual accident rates instead of the 5-year accident rate. In doing so, each roadway segment generates five observations. To account for the correlation that will be set up with such repeat observations, we estimated a tobit model with random effects. Our finding with this approach was that the 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 variables.

[3] With regard to overall goodness of fit, Veall and Zimmermann (1996) extensively discuss this topic in the context of the tobit model. They conclude that the Maddala pseudo R^2 is a valid measurement (Maddala, 1983). The Maddala R^2 is computed as $1\text{-}EXP[2(LL(\beta)\text{-}LL(0))/N]$ were $LL(\beta)$ is the log-likelihood at convergence, $LL(0)$ is the log-likelihood at zero and N is the number of observations.

With regard to pavement characteristics, roadway friction is typically measured on 0 to 100 friction scale, with friction considered to be good if its value is 40 or above. Given this, we constructed an indicator variable that is equal to one if the roadway segment had a minimum friction number of more than 40 for the entire 5-year analysis period. Table 3 shows that the estimation findings for this high-friction indicator variable (1 if all 5-year friction readings are 40 or above, 0 otherwise) suggests a 29.82 decrease of the number of accidents per 100-million VMT and a 25.52% lower probability of having a 100-million VMT accident rate above zero. This result supports the earlier finding of Noyce et al. (2005) that showed that higher pavement friction lowered the risk of accident.

Another important measure of pavement condition is the International Roughness Index (IRI) which measures irregularities that can result from rutting, potholes, patching, and other factors. The IRI is used to define a characteristic of the longitudinal profile of a traveled wheel track and its units are inches per mile or meters per kilometer. The IRI is based on a filtered ratio (refereed to as the average rectified slope) of a standard vehicle's accumulated suspension motion (usually in meters or inches) divided by the distance traveled by the vehicle during the measurement (usually km or mi). In Indiana, the IRI is measured in inches/mile, with lower values indicating a smoother pavement (see Noyce et al., 2005 and Shafizadeh and Mannering, 2003, 2006). A pavement is considered as smooth (good) when the IRI is less than 95; hence, a maximum IRI (over a 5 year period) of less than 75 (the median of the average IRI is 75 in our sample) would indicate a very smooth pavement – presumably resulting in lower driver fatigue and thus lower accident rates. Table 3 shows the smooth-pavement indicator variable (1 if 5-year IRI readings are below 75, 0 otherwise) resulted in a 27.06 decrease in the number of accidents per 100-million VMT and a 23.80% lower probability of having a 100-million VMT accident rate above zero.

Because excessive rutting could contribute to vehicle tracking and loss of control during maneuvering, one would expect pavements with minimal rutting to have lower accident rates. Our model results support this – we found that pavements that had rutting below 0.12 inches over the entire 5-year observation period had a 23.55 decrease in the number of accidents per 100-million VMT and a 20.19% lower probability of having a 100-million VMT accident rate above zero. We also found that Good overall rutting conditions (1 if all 5-year average rutting readings are below 0.2 inches, 0 otherwise) gave an 11.77 decrease in the number of accidents per 100-million VMT and an 11.91% lower probability of having a 100-million VMT accident rate above zero.

Pavement Condition Ratings (PCRs) range from 0 (completely deteriorated) to 100 (excellent pavement condition). Values above 95 are considered to be very good. In our model, roadway segments that had average PCR values above 95 over the 5-year period had a 19.92 increase in the number of accidents per 100-million VMT and a 13.21% higher probability of having a 100-million VMT accident rate above zero. This finding is somewhat surprising and could be picking up a variety of unobserved factors. For example, one possible explanation could relate to driver behavior. Mannering (2007) found that drivers, who believe pavement quality in Indiana interstates is good, tend to drive at higher speeds. The potential reactions of drivers to excellent pavement conditions could be one explanation for this finding.[4]

Turning to roadway geometrics, Table 3 shows that wide medians (greater than 74 feet) result in a 13.70 decrease in the number of accidents per 100-million VMT and an 11.78% lower probability of having a 100-million VMT accident rate above zero. This is expected because wide medians have been shown to reduce median-related accidents and thus reduce accident rates (see for example Shankar et al. 1998).

The presence of a median barrier resulted in a 61.20 decrease in the number of accidents per 100-million VMT and a 60.88% lower probability of having a 100-million VMT accident rate above zero. One possible explanation for this finding is that the presence of median barriers may result in accidents of lower severity. And, because our data consist only of police-reported accidents, this finding could reflect, among other factors, under-reporting of minor accidents.

With regard to other median variables, it is noteworthy that inside (median) shoulders that were 5 feet or greater resulted in a 26.73 decrease in the number of accidents per 100-million VMT and a 25.34% lower probability of having a 100-million VMT accident rate above zero. It is speculated that inside shoulders of 5 feet or more are may be providing adequate recovery space for some driver errors.

---

[4] As a final point relating to pavement attributes, we did try an indicator variable to test concrete versus asphalt because simple correlation suggested that concrete pavements had lower accident rates. Of the 325 roadway segments in our sample, 259 were asphalt and 78 were concrete. The concrete segments had higher 5-year average friction (43.71 vs. 36.07) and lower 5-year average rut depth (0.10 vs. 0.14 inches) and asphalt segments had a better average 5-year average International Roughness Index (74.87 vs. 106.20 in/mi) and a slightly better 5-year average Pavement Condition Rating (95.04 vs. 92.33). When these detailed pavement condition variables are included in the model, the simple concrete/asphalt indicator variable became statistically insignificant. This suggests that the safety characteristics of the pavement are readily defined by measurable attributes and not fundamental differences in pavement types that are difficult to measure.

Table 3 also shows that outside shoulders play an important role in vehicle accident rates. We found that a one foot increase in the outside shoulder width decreased the number of accidents per 100-million VMT by 6.93 and lowered the probability of having a 100-million VMT accident rate above zero by 4.72%. This finding is consistent with the earlier work of Shankar et al. (1997) that found that shoulder widths less than 5 feet significantly increased accident frequency.

With regard to the effects of a number of roadway geometric characteristics, at least two factors must be considered. First, there is an entire body of literature that suggests some variation in roadway geometrics may improve drivers' alertness. For example, Wertheim (1978), Cerezuela et. al., (2004) and others have explored what has been termed "highway hypnosis" which may result from constant roadway geometrics and other factors. Given this possibility, some visual and/or physical variation in roadway characteristics may increase driver alertness and thus decrease the likelihood of an accident. Second, there is a body of literature that has indicated that improvements in roadway design do not always reduce accidents – the idea being that drivers may compensate by driving more carefully in conditions which they perceive as dangerous and less carefully in conditions they do not (see the risk compensation literature including Assum et. al., 1999; Dulisse, B., 1997; Winston et. al., 2006). These two factors play into a number of the estimation findings. For example, Table 3 shows that each additional bridge in a roadway segment was found to decrease the number of vehicle accidents per 100-million VMT by 6.57 and lowered the probability of having a 100-million VMT accident rate above zero by 5.41%. Although the effect is not large, it is possible that bridges may be playing a role in keeping drivers alert. Similarly, shoulder rumble strips, which are designed to alert drivers by producing noise and vibration to warn of leaving the roadway, were found to be effective in reducing accident rates with the presence of rumble strips on both inside and outside shoulders resulting in a 16.16 decrease in the number of accidents per 100-million VMT and a 17.32% lower probability of having a 100-million VMT accident rate above zero.

With regard to vertical curves, a unit increase in the number of vertical curves per mile resulted in a 4.43 decrease in the number of accidents per 100-million VMT and a 3.88% lower probability of having a 100-million VMT accident rate above zero. This finding is consistent with previous research on accident frequencies and severities. For example, Shankar et al. (1997) found that flat roadway segments had lower accident frequencies and Milton et al. (2007) found that the number of changes in vertical profile per mile had a significant negative effect on the likelihood of injury accidents. A possible explanation for this is that drivers may adjust their driving behavior in response to the geometric changes of the roadway (vertical curves) – reducing their speed, and becoming more cautious (risk compensation as discussed above). However, Table 3 also shows that a unit (10%) increase in the ratio of the vertical curve length over the total roadway segment length was found to increase the number of accidents per 100-million VMT by 3.13 and increase the probability of having a 100-million VMT accident rate above zero by 2.24%. This, combined with the number of vertical curves per mile finding, suggest that the vertical curve effect on accident rates is complex – with the presence of some vertical curves reducing vehicle accident rates but the presence of many elevation changes (measured as a proportion of the total segment length) countering this effect.

Consistent with the findings of Shankar et al. (1997) we find that a higher degree of curvature of horizontal curves has a negative effect on accident rates (the higher the degree and thus the sharper the curve, the fewer the accidents). Table 3 shows that if the average degree-of-curve value of the horizon curves in the roadway segment is greater than 2.1 degrees (which is the average of all horizontal curves in our sample plus one standard deviation), the number of accidents per 100-million VMT decreased by 17.98 and the probability of having a 100-million VMT accident rate above zero decreased by 14.79%. One must keep in mind that even these "sharper" curves are still designed to interstate standards with 70-mph design speeds. As was the finding with the number of vertical curves, it is possible that these sharper curves are making drivers more alert and/or they are compensating for the increase in danger by driving more cautiously – thus reducing accident rates.

A unit increase in the number of ramps in the roadway segment (per lane mile) was found to increase the number of accidents per 100-million VMT by 22.36 and increase the probability of having a 100-million VMT accident rate above zero by 16.87%. This is consistent with other findings that have shown that vehicle maneuvers (turning, overtaking, merging, changing lanes, etc.) have a significant effect on accident frequency (Li and Kim, 2000; Islam and Mannering, 2006).

Turning to traffic-related variables, we find that as the annual average daily travel (AADT) increased (in thousands of vehicles per day), the number of accidents per 100-million VMT decreased by 0.54 and the probability of having a 100-million VMT accident rate above zero by decreased by 0.61%. This finding is consistent with a number of previous studies (Qi et al., 2007; Dickerson et al., 2000; Zhou and Sisiopiku,

1997) that have shown that accident rates are higher on low-traffic volume roads, and significantly decrease with increasing volumes.

Finally, a one percent increase in the average daily percentage of combination trucks was found to decrease the number of accidents per 100-million VMT by 0.69 and decrease the probability of having a 100-million VMT accident rate above zero by 0.79%. This finding is consistent with previous work (for example, Shankar et al., 1997) and suggest that higher combination-truck percentages may be reflecting greater driver experience (truck drivers) and may also be having a calming effect on traffic.

As a final comment on model estimation, it should be mentioned that, because some of our roadway segments were adjacent to one another, we considered tobit models with both fixed and random effects to account for possible error term correlation among the adjacent segments. All tests indicated that fixed and random effects were not statistically significant.[5]

Summary and Conclusions

This paper provides a demonstration of tobit regression as a methodological approach to gain new insights into the factors that significantly influence accident rates. The ability of the tobit model to consider roadway segments with and without observed

---

[5] We also estimated a traditional negative binomial model of the number of accidents over the five-year observation period occurring on specific roadway segments (the significance of the negative binomial dispersion parameter allowed us to reject the Poisson model, see Washington et. al., 2003). The results were broadly similar with four variables differing: 1) road section length was included in the negative binomial model to account for the fact that the number of accidents will increase with longer sections (this variable was not statistically significant in the tobit model because section length would not affect accident rates); 2) the interstate indicator variable for I-70 and I-164 was modified and now found to be significant for I-64 and I-164; 3) number of bridges per mile was used instead of number of bridges per roadway segment; and 4) the square of AADT of passenger cars was used instead of just the AADT of passenger cars. The overall fit of the negative binomial using McFadden's ρ^2 was 0.869 (see Washington et. al., 2003).

accidents provides for complete use of available data. Using five years of vehicle accident data from Indiana, our estimation results provide some interesting findings. For example, a variety of factors relating to pavement condition and quality were found to significantly influence vehicle accidents rates including the effects of friction, the International Roughness Index, pavement rutting and the pavement's condition rating. In terms of geometric factors and their effect on vehicle accident rates, median types and width, shoulder widths, number of ramps and bridges, horizontal and vertical curves and rumble strips were all found to be statistically significant. And, the traffic variables of annual average daily travel and the percent of combination truck in the traffic stream were both found to have a significant impact on accident rates.

While this study is exploratory in nature, it does suggest the considerable potential that tobit regression has in analyzing accidents rates. Applying the approach to other geographic areas and to non-interstate road sections would potentially provide more information on the effect that pavement, geometric, and traffic characteristics have on accident rates.

Acknowledgements

The authors are grateful to Mr. Bill Flora for providing the pavement characteristics datasets and for his helpful suggestions and comments.

References

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

Amemiya, T., 1973. Regression-analysis when dependent variable is truncated normal. *Econometrica* **41**(6), 997-1016.

Amemiya, T., 1985. Advanced econometrics. Harvard University Press. Cambridge, MA.

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

Baser, O., Bradley, C. J., Gardiner, J. C., Given C., 2003. Testing and correcting for non-random selection bias due to censoring: An application to medical costs. *Health Services and Outcomes Research Methodology* **4**(2), 93-107.

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

Cerezuela, P., Tejero, P., Chóliz, M., Chisvert, M., Monteagudo, M. J., 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

Dickerson, A., Peirson, J., Vickerman, R., 2000. Road accidents and traffic flows: An econometric investigation. *Economica* **67**, 101-121.

Dulisse, B., 1997. Methodological issues in testing the hypothesis of risk compensation. *Accident Analysis and Prevention*, **39**(3), 285-292.

Ekstrand, C., Carpenter, T. E., 1998. Using a tobit regression model to analyze risk factors for foot-pad dermatitis in commercially grown broilers. *Preventive Veterinary Medicine* **37**(1-4), 219-228.

Fair, R. C., 1978. A theory of extramarital affairs. *The Journal of Political Economy* **86**(1), 45-61.

Greene, W., 2007. Limdep Version 9.0. Econometric Software, Inc., Plainview, NY.

Hadi, M. A., Aruldhas, J., Lee-Fang Chow, 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.

Islam, S., Mannering, F., 2006. Driver aging and its effect on male and female single vehicle accident injuries: Some additional evidence. *Journal of Safety Research* **37**(3), 267-276.

Keeley, M., Robins, P., Spiegelman, R., West, R., 1978. The labor supply effects and costs of alternative negative income tax programs. *Journal of Human Resources* **13**, 3-36.

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.

Li, L., Kim, K., 2000. Estimating driver crash risks based on the extended Bradley-Terry model: an induced exposure method. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **163**(2), 227–240.

Maddala, G. S., 1983. Limited-dependent and qualitative variables in econometrics. *Cambridge University Press*. New York.

Mannering, F. L., 2007. Effects of Interstate Speed Limits on Driving Speeds: Some New Evidence. Compendium of papers CD-ROM, Transportation Research Board 86th Annual Meeting, Washington DC.

McDonald, J. F., Moffitt, R. A., 1980. The uses of tobit analysis. *The Review of Economics and Statistics* **62**(2), 318-321.

Milton, J. C., Shankar, V. N., Mannering, F. L., 2007. Highway accident severities and the mixed logit model: An exploratory empirical analysis. Forthcoming in *Accident Analysis and Prevention*.

Nelson, F. D., 1977. Censored regression models with unobserved, stochastic censoring thresholds. *Journal of Econometrics* **6**, 309-327.

Nolan, A., 2002. The determinants of urban households' transport decision: A microeconomic study using Irish data. *Presentation to the VIIth Spring Meeting of Young Economists, Paris-France*.

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

Poch, M., Mannering, F., 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 (ASCE)* **133**, 149-156.

Roncek, D. W., 1992. Learning more from tobit coefficients: Extending a comparative analysis of political protest. *American Sociological Review* **57**(4), 503-507.

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

Shafizadeh, K., Mannering, F., 2003. Acceptability of pavement roughness on urban highways by driving public. *Transportation Research Record 1860*, 187-193.

Shafizadeh, K., Mannering, F., 2006. Statistical modeling of user perceptions of infrastructure condition: An application to the case of highway roughness. *Journal of Transportation Engineering* 132(2), 133-140.

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.

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.

Smith, D. A., Brame, R., 2003. Tobit models in social science research: Some limitations and a more general alternative. *Sociological Methods and Research* **31**(3), 364-388.

Talley W. K., 1995. Safety investments and operating conditions: Determinants of accident passenger-vessel damage cost. *Southern Economic Journal* **61**, 819-829.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24-36.

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

Washington, S. P., Karlaftis, M. G., Mannering, F. L., 2003. Statistical and Econometric methods for transportation data analysis. *Chapman & Hall/CRC*.

Weiss, A. A., 1992. The effects of helmet use on the severity of head injuries in motorcycle accidents. *Journal of the American Statistical Association* **87**(417), 48-56.

Wertheim, A. H., 1978. Explaining highway hypnosis: Experimental evidence for the role of eye movements. *Accident Analysis and Prevention*, **10**(2), 111-129.

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.

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

Table 1. Summary statistics of key pavement, geometric and traffic variables.

Variable	Mean	St. Dev.	Min.	Max.
Dependent Variable: Accidents per 100-million VMT	41.29	47.50	0	350.40
Interstate indicator variable (1 if I-70 or I-164, 0 otherwise)	0.15	0.36	0	1
High Friction indicator variable (1 if all 5-year friction readings are 40 or higher, 0 otherwise)	0.14	0.35	0	1
Smooth pavement indicator variable (1 if 5-year International Roughness Index (IRI) readings are below 75, 0 otherwise)	0.20	0.40	0	1
Excellent rutting indicator variable (1 if all 5-year rut readings are below 0.12 inches, 0 otherwise)	0.18	0.39	0	1
Good rutting indicator variable (1 if all 5-year average rutting readings are below 0.2 inches, 0 otherwise)	0.87	0.33	0	1
Average Pavement Condition Rating (PCR) indicator variable (1 if PCR greater than 95, 0 otherwise)	0.55	0.50	0	1
Median width indicator variable (1 if greater than 74 feet, 0 otherwise)	0.32	0.47	0	1
Median barrier presence indicator variable (1 if present, 0 otherwise)	0.16	0.37	0	1
Inside shoulder width indicator variable (1 if 5 feet or greater, 0 otherwise)	0.31	0.46	0	1
Outside shoulder width (in feet)	11.28	1.74	6.2	21.8
Number of bridges (per road section)	0.34	0.85	0	5
Number of ramps in the driving direction per lane-mile	0.14	0.41	0	3.27
Number of vertical curves per mile	1.18	2.50	0	14.49
Ratio of the vertical curve length over the road section length	2.23	3.56	0	10
Horizontal curve's degree curvature indicator variable (1 if average degrees per road section is greater than 2.1, 0 otherwise)	0.12	0.33	0	1
Rumble strips indicator variable (1 if both inside and outside rumble strips are present, 0 otherwise)	0.72	0.45	0	1
AADT of passenger cars (in 1,000 vehicles per day)	23.16	26.83	5.99	128.03
Average daily percent of combination trucks	0.23	0.14	0.03	0.44

Table 2. Tobit regression estimation of accidents per 100-million VMT.

Variable	Parameter estimate	*t*-ratio
Constant	173.414	5.91
Interstate indicator variable (1 if I-70 or I-164, 0 otherwise)	-29.518	-3.45
Pavement Characteristics		
High-friction indicator variable (1 if all 5-year friction readings are 40 or higher, 0 otherwise)	-33.521	-3.52
Smooth pavement indicator variable (1 if 5-year IRI readings are below 75, 0 otherwise)	-31.262	-3.71
Excellent rutting indicator variable (1 if all 5-year rutting readings are below 0.12 inches, 0 otherwise)	-26.516	-2.92
Good rutting indicator variable (1 if all 5-year average rutting readings are below 0.2 inches, 0 otherwise)	-15.639	-1.88
Average PCR indicator variable (1 if greater than 95, 0 otherwise)	17.349	2.64
Geometric Characteristics		
Median width indicator variable (1 if greater than 74 feet, 0 otherwise)	-15.466	-2.41
Median barrier presence indicator variable (1 if present, 0 otherwise)	-79.955	-4.80
Inside shoulder width indicator variable (1 if 5 feet or greater, 0 otherwise)	-33.281	-4.26
Outside shoulder width (in feet)	-6.204	-2.91
Number of bridges (per road section)	-7.101	-1.80
Rumble strips indicator variable (1 if both inside and outside rumble strips are present, 0 otherwise)	-22.746	-2.47
Number of vertical curves per mile	-5.093	-2.73
Ratio of the vertical curve length over the road section length (in tenths)	2.945	2.28
Horizontal curve's degree curvature indicator variable (1 if average degrees per road section is greater than 2.1, 0 otherwise)	-19.422	-2.02
Number of ramps in the driving direction per lane-mile	22.156	3.36
Traffic Characteristics		
AADT of passenger cars (in 1,000 vehicles per day)	-0.800	-2.82
Average daily percent of combination trucks	-1.041	-3.41
Number of observations	325	
Log-likelihood at zero	-1724.20	
Log-likelihood at convergence	-1242.99	
Madalla Psuedo R^2	0.948	

Table 3. Sensitivity of estimated tobit regression parameters.

Variable	Overall sensitivity[a]	Zero sensitivity[b]
Interstate indicator variable (1 if I-70 or I-164, 0 otherwise)	-26.38	-22.48%
Pavement Characteristics		
High-friction indicator variable (1 if all 5-year friction readings are 40 or higher, 0 otherwise)	-29.82	-25.52%
Smooth pavement indicator variable (1 if 5-year IRI readings are below 75, 0 otherwise)	-27.06	-23.80%
Excellent rutting indicator variable (1 if all 5-year rutting readings are below 0.12 inches, 0 otherwise)	-23.55	-20.19%
Good rutting indicator variable (1 if all 5-year average rutting readings are below 0.2 inches, 0 otherwise)	-11.77	-11.91%
Average PCR indicator variable (1 if greater than 95, 0 otherwise)	19.22	13.21%
Geometric Characteristics		
Median width indicator variable (1 if greater than 74 feet, 0 otherwise)	-13.70	-11.78%
Median barrier presence indicator variable (1 if present, 0 otherwise)	-61.20	-60.88%
Inside shoulder width indicator variable (1 if 5 feet or greater, 0 otherwise)	-26.73	-25.34%
Outside shoulder width (in feet)	-6.93	-4.72%
Number of bridges (per road section)	-6.57	-5.41%
Rumble strips indicator variable (1 if both inside and outside rumble strips are present, 0 otherwise)	-16.16	-17.32%
Number of vertical curves per mile	-4.43	-3.88%
Ratio of the vertical curve length over the roadway segment length (in tenths of miles)	3.13	2.24%
Horizontal curve's degree curvature indicator variable (1 if average degrees per road section is greater than 2.1, 0 otherwise)	-17.98	-14.79%
Number of ramps in the driving direction per lane-mile	22.36	16.87%
Traffic Characteristics		
AADT of passenger cars (in 1,000 vehicles per day)	-0.54	-0.61%
Average daily percent of combination trucks	-0.69	-0.79%

[a] Change in the overall expected value $\partial E\left[Y\right]/\partial X_k$ (see equation 6).

[b] Percent change in the cumulative probability of being above zero (using $\partial F\left(z\right)/\partial X_k$ in equation 7 and converting to percent).