This manuscript is the post-print of the following:

Anderson, J., Hernandez, S., 2017. Heavy Vehicle Crash Rate Analysis: A Comparison of Heterogeneity Methods Using Idaho Crash Data. Transportation Research Record: Journal of the Transportation Research Board, 2637, in press.

HEAVY VEHICLE CRASH RATE ANALYSIS: A COMPARISON OF HETEROGENEITY METHODS USING IDAHO CRASH DATA

Jason Anderson
PhD Student, School of Civil and Construction Engineering
Oregon State University, 211 Kearney Hall, Corvallis, OR 97331-3212
Tel: (541)-908-0921; Email: anderjas@oregonstate.edu

Salvador Hernandez, Corresponding Author
Assistant Professor, School of Civil and Construction Engineering
Oregon State University, 309 Owen Hall, Corvallis, OR 97331-3212
Tel: (541)-737-4740; Email: sal.hernandez@oregonstate.edu

TRR Paper number: 17-06641

ABSTRACT

Studies investigating crash rates by roadway classification are few and far between and even more so if extended to focus on heavy vehicles. This study explores and compares two advanced econometric methods, random-parameter Tobit regression and latent class Tobit regression, to determine contributing factors for heavy vehicle crashes per million-vehicle-miles-traveled while accounting for the unobserved heterogeneity present in crash data. The increasing crash rates in Idaho, crash proportion by roadway classification, and available data make for an ideal case study. Empirical results show that although the random-parameter Tobit regression model provides better insight into heavy vehicle crash rates than the fixed-parameter approach, the latent class Tobit regression model is the preferred methodology for the given dataset. Traffic volumes, roadway characteristics, and traffic control devices were among the variables found to be statistically significant. Results from this study provide an alternate framework to account for heterogeneity, while identifying key factors by roadway classification that influence heavy vehicle crash rates. The illustrated framework and analysis by roadway classification can provide guidance to transportation agencies and policy makers, and prompt future studies to include a latent class analysis and/or analysis by road classification.

Keywords: Heavy Vehicle Safety, Random-Parameter Tobit Model, Latent Class, Unobserved Heterogeneity, Crash Rate

INTRODUCTION

Heavy vehicle crashes have a substantial economic impact on commerce and society. In the United States, heavy vehicle crashes were about $87 billion in 2011 and costs due to delay and other consequences were roughly $28 billion *(1,2)*. These values will continue to increase as the economy continues to grow, as will the volume of heavy vehicles on the nation's freight infrastructure. For example, from 2010 to 2013 a 2.3% increase of heavy vehicles was experienced (about 6,000,000 vehicles) *(3)*. This number is expected to continue to grow and crashes associated with heavy vehicles will remain a concern for safety planners and safety related agencies. Although heavy vehicle crashes have decreased over the past two decades, the number of fatal crashes per 100-million vehicle-miles-traveled compared to passenger cars is higher (1.34 versus 1.08 in 2014)*(4, 5)*. In Idaho, the state experienced a 5.6% increase in heavy vehicle crashes and a 4.4% increase in heavy vehicle crashes per million-vehicle-miles-traveled (MVMT) from 2010 to 2013 *(6)*. 50% of these crashes occurred on local roads, 28% of injury crashes happened on interstates, and approximately 68% of fatalities happened on U.S. and State Highways *(6)*. These statistics illustrate the need for continued research in understanding the relationship between heavy vehicle crash rates and roadway classification.

There have been a number of studies that have addressed crash frequency through the application of count and spatial based models *(7–15)*. However, most of these studies have focused on data related to pedestrians, passenger cars, or all traffic mixes in a single modeling framework and do not address heavy vehicles explicitly. Although there have been several recent efforts to understand heavy vehicle injury severity factors *(16–18)*, heavy vehicle crash rate analyses are sparse. This is especially true for heavy vehicle crash rates by functional class of road. A possible reason for this deficiency in the literature may stem from the availability of sufficient data to capture the complex interactions of multiple crash rate factors under a single framework by functional class.

Recent studies have addressed the issue of insufficient data through the application of statistical and econometric methods that account for unobserved factors (unobserved heterogeneity), which are factors unknown to the analyst and that may vary across observations, see Mannering et al. for a complete review of these methods *(19)*. For instance, weather conditions that continually change over time, as well as driver response to the changing weather condition. These models allow the analyst to account for these variations and make more informed inferences regarding the effects of the contributing factors *(19)*.

With this in mind, the present study seeks to identify factors that impact heavy vehicle crashes per MVMT by road classification through the application and performance based comparison of two "heterogeneity" models, namely random-parameter- and latent class- Tobit regression. The Tobit modeling framework is selected due to the nature of crash rate data. Similar to frequency models, a crash rate analysis is likely to have several observations in which no crash has occurred, therefore a censoring method is recommended to account for the skewed nature of the response variable (crash rate). It has be shown that the Tobit regression framework can account for the skewed nature of crash rate data without omitting observations by censoring the analysis at a given value *(20)*. These models have been successfully applied to related transportation safety data, for example, Anastasopoulos et al. used the fixed-parameter Tobit model to investigate crash rates on interstates in Indiana and determine contributing factors *(21)*. To extend the Tobit framework, Anastasopoulos et al. utilized the random-parameter Tobit model to determine factors that influence crash rates per 100-million VMT on highways *(22)*. Islam and Hernandez investigated fatalities per million truck-miles-traveled and fatalities per ton-mile of freight for heavy vehicles through the application of a random-parameter Tobit regression model and Chen et al. utilized a random-effects Tobit model to analyze crash rates with refined-scale data *(23, 24)*. From a latent class Tobit regression application, there are no known applications to transportation safety data, however the method has been applied to social science studies (see *(25, 26)*).

Therefore, the present study will use the random-parameter Tobit method to identify significant contributing factors to crash rates by roadway functional class while accounting for heterogeneity. However, variables not found to be random in the random-parameter method may in fact have varying effects on heavy vehicle crash rates. Hence, the current study will be extended by investigating the results of the Tobit latent class approach by disaggregating the Tobit model into unobserved groups (or classes). To accomplish this, an extensive crash database collected and maintained by the Idaho Department of Transportation (IDT) is used. The findings of this study can provide insight that can aid safety planners and safety related agencies in identifying appropriate countermeasures to help reduce and mitigate heavy vehicle crashes. To the best of the authors' knowledge, these are first attempts at developing these types of models for heavy vehicle crash rate analysis.

SOURCE OF DATA

The current study uses 7 years of police-recorded crash data obtained from the state of Idaho (2007 to 2013). Each year was filtered to represent heavy vehicle crashes, then combined with traffic data from IDT utilizing segment codes and milepost markers that were present in both datasets. The segment codes and milepost numbers of the location of the crash were used to determine the intermediate segments within the milepost intervals in the traffic data—these segments are used for the modeling process. Using the complete dataset consisting of exposure variables (i.e., roadway geometrics, traffic control devices, number of lanes, etc.) and traffic volumes, several indicator variables were created to identify specific exposure conditions and traffic volumes that impact crash rates by road classification in Idaho. The final dataset for principal arterials, major collectors, and interstates had 1,560, 1,010, and 1,588 heavy vehicle crashes, respectively. (TABLE) displays the response variable and indicators found to be significant throughout the modeling process.

<Insert Table 1>

METHODOLOGY

Dependent Variable

To model heavy vehicle crash rates, a rate for each segment is calculated using the traffic data provided by Idaho *(27)*:

$$R_s = \frac{\sum_{y=1}^{n} N_{ys}}{\left[\sum_{y=1}^{n} AADT_{ys} \times L_s \times 365\right] / 1{,}000{,}000} \tag{1}$$

where R_s is the number of crashes per MVMT on segment s; y is the year (2007 to 2013); N_{ys} is the number of heavy vehicle crashes in year y on segment s; $AADT_{ys}$ is the average annual daily traffic for year y on segment s; and, L_s is the length of segment s in miles.

Given the datasets available for the current study, a specific methodology was employed to get corresponding traffic data for each crash, as formerly described. Being that the current study uses seven years of crash data, the likelihood of having one crash on each segment is higher when compared to a 3- to 5-year period. As such, many segments for each road classification were identified and analyzed, each of which had at least one crash. Referring to (Table 1), the minimum values of the response variables are approximately zero—if studying fewer years of crash data these values could have very likely been zero. (Figure 1) illustrates the crash rate distribution for each road classification, wherein the skewed distributions needed to be addressed during analysis.

<Insert Figure 1>

Random-Parameter Tobit Model

The distribution of crash rates illustrate the need to utilize a method that can account for the large lower bound cluster of observations while maintaining the linear assumptions required for regression of a continuous dependent variable (heavy vehicle crash rates by roadway classification). In regard to "other principal arterials", the data was too centered at zero and, even with censoring, produced erroneous estimates during analysis; therefore, this classification was omitted from the study. Taking this into consideration, the present study seeks to develop a statistical model that can be used to determine the contributing factors on heavy vehicle crash rates by roadway classification. This study will apply the Tobit regression modeling framework first introduced by James Tobin *(28)*. However, key variables not available within many crash datasets and variation across the available variables often results in unobserved heterogeneity, and if neglected, will lead to biased estimates and inaccurate inferences (see *(19)* for further discussion). To account for such *(22, 23, 29–31)*, the current study will apply the random-parameter approach to the traditional Tobit regression framework. As mentioned earlier, Anastasopoulos et al., Islam and Hernandez, and Chen et al. all apply successful applications of the random-parameter Tobit regression model to transportation safety data *(22–24)*. Therefore, for this work, the standard Tobit model is expressed as:

$$\begin{aligned} &\mathrm{Y}_s^* = \boldsymbol{\beta}'\mathbf{X_s} + \varepsilon_s \quad \text{with} \quad \varepsilon_s \sim \mathrm{N}[0, \sigma^2] \text{ and } s = 1, 2, \ldots \mathrm{N} \\ &\mathrm{Y}_s = \mathrm{Y}_s^* \ \text{ if } \ \mathrm{Y}_s^* > \mathrm{L} \\ &\mathrm{Y}_s = 0 \ \text{ if } \ \mathrm{Y}_s^* \le \mathrm{L} \end{aligned} \tag{2}$$

where Y_s is the number of crashes per MVMT; L is the value the model is left-censored at; $\mathbf{X_s}$ is the vector of explanatory variables (AADT, roadway geometrics, etc.); $\boldsymbol{\beta}'$ is the vector of estimated parameters; and, ε_s is the normally and independently distributed error term with a mean of zero and constant variance, σ^2. To determine the likelihood for the Tobit model over zero observations (e.g., the value the Tobit model is left-censored at) and positive observations (1), the following function applies *(21, 26)*:

$$\mathrm{L} = \prod_0 \left[1 - \Psi\left(\frac{\boldsymbol{\beta}'\mathbf{X_s}}{\sigma}\right)\right] \prod_1 \left(\frac{1}{\sigma}\right) \psi\left(\frac{\mathrm{Y_s} - \boldsymbol{\beta}'\mathbf{X_s}}{\sigma}\right) \tag{3}$$

where $\Psi\left(\frac{\boldsymbol{\beta}'\mathbf{X_s}}{\sigma}\right)$ is the standard normal distribution function and $\psi\left(\frac{\mathrm{Y_s}-\boldsymbol{\beta}'\mathbf{X_s}}{\sigma}\right)$ is the standard normal density function.

In an attempt to capture the unobserved heterogeneity, the random-parameter approach is now applied to the Tobit framework and estimated parameters can be written as *(32)*:

$$\beta_s = \beta + \phi_s \tag{4}$$

where the equivalent log-likelihood function is *(26)*:

$$\mathrm{LL} = \sum_{\forall s} \ln \int_{\phi_s} g(\phi_s) P(Y_s^* | \phi_s) d\phi_s \tag{5}$$

where $g(\phi_s)$ is the probability density function of ϕ_s and $P(Y_s^*|\phi_s)$ is the probability of the Tobit model (i.e., probability that the value is uncensored). As stated in previous studies *(22, 23)*, the maximum likelihood estimations encounter computing issues due to its complexity. To address this issue, a common approach developed by Halton is used to simulate the maximum likelihood by utilizing Halton draws to solve the complex integral seen in Eq. (5) and has been shown to be preferable over merely random draws *(33–35)*.

Latent Class Tobit Model

Although the random-parameter method accounts for unobserved heterogeneity, there is a possible disadvantage due to the assumption that the parameters vary in a predefined distribution and that parameters vary only across singular observations (see *(19)* for further discussion). Taking that into consideration, the latent class approach attempts to capture unobserved heterogeneity by allowing estimable parameters to vary with an underlying discrete distribution across unobserved groups of observations (or classes). The heterogeneity is accounted for by defining a finite number of points and measuring the mass probability of the intervals between points. Applying this to the Tobit regression structure results in:

$$\begin{aligned} &Y_s^*|(\mathrm{Class} = \mathrm{C}) = \boldsymbol{\beta'}_{\mathbf{c}}\mathbf{X_s} + \varepsilon_{s|c} \quad \text{with} \quad \varepsilon_{s|c} \sim N[0, {\sigma^2}_c] \text{ and } s = 1, 2, \ldots N \\ &Y_s = Y_s^* \ \text{ if } \ Y_s^* > L \\ &Y_s = 0 \ \text{ if } \ Y_s^* \leq L \end{aligned} \tag{6}$$

where $\boldsymbol{\beta'}_{\mathbf{c}}$ is a vector of estimated parameters belonging to class C and $Y_s^*|(\mathrm{Class} = \mathrm{C})$ is the number of crashes per MVMT of segment s in class C. The corresponding log-likelihood function is now *(26)*:

$$\mathrm{LL} = \sum_{s=1}^{N} \log\left[\sum_{c=1}^{C} P_{sc}(\delta_c, \omega_s)\left[f\left(Y_s | \mathrm{Class} = \mathrm{C}, \mathbf{X_s}, \boldsymbol{\beta'}_{\mathbf{c}}, \sigma_c\right)\right]\right] \tag{7}$$

where $P_{sc}(\delta_c, \omega_s)$ is the prior to model estimation logit probability of being in class C and represented by the multinomial logit form *(26)*:

$$P_{sc}(\delta_c, \omega_s) = \frac{e^{(\omega_s \delta_c)}}{\sum_{c=1}^{C} e^{(\omega_s \delta_c)}} \quad \text{with } c = 1, 2, \ldots C \text{ and } \delta_C = 0 \text{ for normalization} \tag{8}$$

Lastly, after the parameters have been estimated a second estimation is conducted to determine the posterior probabilities of crash rate Y_s belonging to class C *(36)*. The posterior probability that a heavy vehicle crash belongs to class C is determined post-estimation. That is, the posterior probability utilizes the estimated parameters to determine a class probability based on the observed crash data *(26, 37)*:

$$P(\text{Class} = C|\text{Crash Rate } Y_s) = \frac{f(\text{Crash Rate } Y_s|\text{Class} = C)P(\text{Class } C)}{\sum_{c=1}^{C} f(\text{Crash Rate } Y_s|\text{Class} = C)\, P(\text{Class } C)} \tag{9}$$

As mentioned previously, the application of the latent class modeling structure to the Tobit regression modeling framework in a safety context is a first.

MODEL ESTIMATION RESULTS

Random-Parameter Tobit Model

As shown in (TABLE and Table 3), no parameters were found to be random for principal arterials and interstates. However, two parameters were found to be random for major collectors and are displayed in (Table 4). To statistically asses the more significant log-likelihood for major collectors, fixed- or random-parameter, the ensuing log-likelihood ratio test was conducted:

$$\chi^2 = -2[LL(\beta_{FP}) - LL(\beta_{RP})] \tag{10}$$

where $LL(\beta_{FP})$ is the log-likelihood at convergence for the fixed-parameter model; $LL(\beta_{RP})$ is the log-likelihood at convergence for the random-parameter model; and, χ^2 is a chi-square statistic with degrees of freedom equal to the number of random parameters. One more goodness of fit measure was applied, the Maddala Pseudo R^2 value *(38)*:

$$R^2 = 1 - e^{\left(\frac{-2[LL(\beta) - LL(0)]}{N}\right)} \tag{11}$$

where $LL(\beta)$ is the log-likelihood at convergence for the best fit model; $LL(0)$ is the log-likelihood at zero; and, N is the number of observations.

With regard to the principal arterial model, a chi-square statistic of 83.11 and 6 degrees of freedom indicated with 99.99% confidence that the fixed-parameter model is preferred over the model with simply the constant. For interstates, a chi-square statistic of 61.25 and 6 degrees of freedom showed with 99.99% confidence that the fixed-parameter model is of more significance than the model with no estimated parameters. In the case of major collectors, where variables were found to be random, a chi-square statistic of 80.81 and 2 degrees of freedom demonstrated with 99.99% confidence that the random-parameter model is statistically preferred.

<Insert Table 2, Table 3, and Table 4>

Latent Class Tobit Model

Latent class regression models for each road classification are shown in (Table 5, Table 6, and Table 7). In line with previous studies, the number of latent classes for each model were selected using the Bayesian Information Criterion (BIC)—the number of latent classes that produced the smallest BIC were used *(39, 40)*. However, Louviere et al. suggest that the smallest Akaike Information Criterion (AIC) be used to determine the best fit number of classes and this was the case for the major collector model *(41)*—Yang found similar results in terms of the number of latent classes based on AIC *(42)*.

The class-split for each classification is highly significant and the best fit number of classes are different for each model. Principal arterials have a best fit model with three latent classes as is shown in (Table 5), major collectors with four latent classes as is shown in (Table 6), and interstates with two latent classes as is shown in (Table 7).

<Insert Table 5, Table 6, and Table 7>

DISCUSSION

Tobit Model

High passenger vehicle AADT (PAADT) decreases crash rates for each road classification and has a significant impact on crash rates based on partial effects. Partial effects refer to a one-unit increase in an exposure variable, while holding all others constant, and the outcome it has on heavy vehicle crash rates. Partial effects show that PAADT greater than 10,500 on principal arterials decreases the number of heavy vehicle crashes per MVMT by 0.084. Similarly, PAADT greater than 2,500 on major collectors reduces the number of heavy vehicle crashes per MVMT by 0.143 and PAADT greater than 15,000 on interstates reduces heavy vehicle crashes by 0.013 per MVMT. Conversely, low total AADT (passenger vehicles and heavy vehicles) increases crash rates. For instance, partial effects indicate that AADT less than 500 on major collectors increases the number of heavy vehicle crashes by 0.279 per MVMT. Likewise, AADT less than 6,500 on interstates results in an increase of 0.016 crashes per MVMT. These findings are in line with previous work *(21, 22, 30, 43)*, in which lower AADT increases crash rates while higher AADT decreases crash rates. The same literature finds that the presence of heavy vehicle traffic decreases crash rates, yet the present study finds that the presence of heavy vehicle traffic increases crash rates for principal arterials and interstates. A possible explanation could be that such a finding is exclusive to the state of Idaho.

Two parameters were found to be random on major collectors based on the statistical significance of the mean and standard deviation. The estimated parameter for a speed limit less than or equal to 40 miles per hour was found to be random and normally distributed with a mean of 0.42 and standard deviation of 0.73. This suggests that for 28.2% of heavy vehicles the estimated parameter mean is less than zero and greater than zero for 71.8%. In other words, lower speed limits on major collectors decreases crash rates for 28.2% of heavy vehicles and increases crash rates for 71.8%. Chen et al., however, found that lower speed limits increase crash rates for all observations using the random-effects Tobit model, possibly indicating that the random-effects approach is not accounting for all the heterogeneity in their dataset *(24)*. On the other hand, high speed limits decrease crash rates on principal arterials and interstates. Speed limits of 65 miles per hour decrease crash rates on principal arterials and partial effects show a reduction of 0.041 crashes per MVMT. Interstates with a speed limit equal to 75 miles per hour see a decrease in crash rates and partial effects indicate a decrease of 0.008 heavy vehicle crashes per MVMT. Although higher speed limits are prone to more severe crashes, they have been shown to reduce crash rates (see *(44)* for a thorough literature review regarding implications of high speed limits).

As for road configuration, 2-way major collectors with a double-yellow painted divider was found to have a normally distributed random parameter. With a mean of 0.36 and standard deviation of 0.99, this configuration decreases crash rates for 35.8% of heavy vehicles and increases crashes rates for 64.2%. 2-way interstates with a raised/depressed divider experience a reduction in crash rates and have a partial effect of -0.014. Road configuration, however, on principal arterials increases crash rates—partial effects suggest that 2-way principal arterials with a 2-way left-turn lane result in an increase of 0.041 heavy vehicle crashes per MVMT.

With regard to horizontal geometrics, straight and curved conditions increase crash rates for major collectors and interstates. Major collectors experience an increase in crash rates due to straight horizontal geometrics and partial effects show an increase of 0.210 heavy vehicle crashes per MVMT. Horizontal curves increase crash rates on interstates, though there is just a 0.010 increase. Curved geometrics were found to increase crash risk by Yu et al., while the degree of curvature was found to increase crash rate by Chen et al. *(24, 29)*.

Other notable contributing crash rate factors are traffic control devices and surface defects. No traffic control devices on principal arterials decrease crash rates and partial effects indicate a reduction of 0.047 heavy vehicle crashes per MVMT. On the contrary, stop signs on major collectors increase heavy vehicle crashes by 0.116 per MVMT. Interstates with no surface defects decrease crash rates. This variable has the largest effect on interstate crash rates, as partial effects suggest a decrease of 0.023.

Latent Class Tobit Model

The presence of latent classes suggest that various explanatory variables are heterogeneous. For example, 2-way roads with a 2-way left turn lane on principal arterials is positively significant in latent class 2, but negative and not significant in latent classes 1 and 3. These results indicate the presence of heterogeneity and that such road configurations can have a negative and positive impact on crash rates *(45)*. Similar findings are presented in each latent class specification and exist for each variable.

With regard to class probability, the prior probabilities for principal arterials indicate that the probability of a crash belonging to latent class 3 is the highest at 0.716. This is seen in the posterior probabilities, as 80.7% of the heavy vehicle crashes belong to latent class 3 while 17.9% and 1.4% belong to latent class 1 and 2, respectively. For major collectors, there is a 0.577 prior probability that crashes belong to latent class 2. Posterior probabilities suggest this is so, as 72.4% belong to latent class 2 with 4.6%, 12.4%, and 10.7% of crashes belonging to latent class 1, 3 and 4. Prior class probabilities for interstates indicate a 0.361 probability of a heavy vehicle crash belonging to latent class 1 and 0.639 probability of belonging to latent class 2. Posterior probabilities agree, being that 10.8% of heavy vehicle crashes belong to latent class 1 and 89.2% belong to latent class 2.

Looking at principal arterials, the partial effects of the Tobit model are significantly greater than those of the latent class model. The partial effect for PAADT greater than 10,500 using the Tobit model was -0.084, but according to the latent class model, this PAADT decreases the number of heavy vehicle crashes by 0.007 per MVMT. Overall, the partial effects for the latent class model were much less than the Tobit model.

Moving to major collectors, latent class partial effects were substantially less when compared to the Tobit model. For example, PAADT greater than 2,500 has a partial effect of -0.143 for the Tobit model, yet the same variable based on the latent class model results in a reduction of 0.059 heavy vehicle crashes per MVMT.

Interstates, however, experienced a decrease in partial effects for some variables and an increase in others, even a change in signs for one variable. For instance, 2-way interstates with a raised/depressed divider has a partial effect of -0.014 for the Tobit model while the latent class model has the opposite effect and results in an increase of 0.002 heavy vehicle crashes per MVMT. The partial effect of the Tobit model for PAADT greater than 15,000 is -0.013, but increases to -0.008 for the latent class model. Interstates with no surface defects decrease the number of heavy vehicle crashes per MVMT by 0.023 according to the Tobit model and increases the reduction to 0.108 according to the latent class estimations.

Model Comparison

To determine the best fit model for the Idaho crash data, three metrics were evaluated: overall model fit, partial effect inferences, and the rate of prediction of actual crash rate values. To illustrate, the latent class approach for each road classification had a better overall model fit. It should be noted that log-likelihoods are typically negative, however, it is possible to see positive values for regression of a continuous dependent variable. In such a case, the greater the value (if positive), the better the fit of the model. In terms of partial effects, the latent class framework identified different high impact variables and partial effects were much less, as were the partial effects for the random-parameter model. (Table 8) quantitatively shows the fit of the actual crash rates versus the predicted crash rates for both regression estimates, while the corresponding plots are presented in (Figure 4). The plots visually illustrate that the Tobit model substantially under-predicted the crash rates for each road classification and that the latent class model outperformed the Tobit model significantly.

<Insert Table 8, Figure 2>

SUMMARY & INSIGHTS

This study utilized two specific econometric frameworks, namely random-parameter Tobit regression and latent class Tobit regression, to determine factors that contribute to the number of heavy vehicle crashes per MVMT by roadway classification while identifying a preferred method to account for unobserved heterogeneity. Police-reported crash data is often missing key variables (e.g., not on data collection forms) and has variation across existing variables, therefore utilizing the random-parameter Tobit method allows the analyst to account for heterogeneity by defining a distribution. The latent class approach also accounts for heterogeneity, but no distribution is defined and the parameters are permitted to vary across a specified number of classes. Using goodness-of-fit measures and the rate of prediction, the estimates of the two approaches were examined.

The Idaho case study provides new insights into crash rates by roadway classification. Different road configurations, horizontal geometrics, and traffic control devices were found to be significant for each road classification. A specific road configuration was found to decrease crash rates for major collectors and interstates, but increase crash rates on principal arterials. Curved horizontal geometrics increase crash rates on interstates and straight horizontal geometrics increase crash rates on major collectors. Stop signs on major collectors increase crash rates, yet no traffic control devices on principal arterials decrease crash rates. High speed limits decrease crash crates on principal arterials and interstates, and lower speed limits increase crash rates for the majority of heavy vehicles on major collectors. The most common insight from this study is that high traffic volumes decrease crash rates and low volumes increase crash rates. With that in mind, unlike previous works, this study finds that the presence of heavy vehicles has the potential to increase crash rates.

To assess the accuracy of the two frameworks, the actual crash rates and predicted crash rates were plotted and the Pearson product moment correlation coefficient was provided for each. The latent class approach outperformed the traditional Tobit method for each road classification and, as a result, should be considered in future crash rate analyses. In addition, the sample size may indicate what information criterion (AIC or BIC) to use when selecting the correct number of latent classes. It needs to be noted, however, that results indicating a better fit for the latent class approach are entirely data-specific. Although the present study has found that the latent class model better describes the Idaho crash data, there is the potential that the latent class model may not be better suited in crash datasets of other states. This strongly suggests that further work be conducted in the comparison of these two heterogeneity methods. Unfortunately, this is an inherent limitation of the latent class modeling framework. Although the model, in this case, captures more heterogeneity and provides a better fit for the data, this is not always the case.

In summary, this study exhibits two distinct methodologies to model crash rates while accounting for heterogeneity. Factors that contribute to crash rates differ dependent on road classification and in future work should be analyzed separately. Such findings can assist with safety measures in Idaho by providing transportation agencies, engineers, planners, and policy makers with contributing crash rate factors coupled with more precision. For example, road configuration was found to impact crash rates by road classification and restriping to reconfigure configurations can be an economically viable solution to reduce heavy vehicle crash rates in Idaho. In addition, stop signs were found to increase crash rates on major collectors and a possible explanation could be ineffective stop sign location; hence, relocating stop signs is yet another economically viable solution to reduce heavy vehicle crash rates in Idaho. The presented framework, censored latent class regression, should strongly be considered when conducting future crash rate analyses, as well as analysis by roadway classification in other geographic regions.

ACKNOWLEDGMENT

We would like to acknowledge the Idhao Department of Transportation for providing the crash data and for their assistance in understanding the data.

REFERENCES

1. Federal Motor Carrier Safety Administration. Commercial Motor Vehicle Facts. March. 13–14. https://www.fmcsa.dot.gov/safety/data-and-statistics/commercial-motor-vehicle-facts-–-march-2013. Accessed Feb. 11, 2015.

2. Blincoe, L., T. R. Miller, E. Zaloshnja, and B. A. Lawrence. *The Economical and Societal Impact of Motor Vehicle Crashes, 2010 (Revised May 2015)*. 2015.

3. Bureau of Transportation Statistics. *Freight Facts and Figures*. United States Department of Transportation, Washington, D.C., 2015.

4. NHTSA. NCSA DATA RESOURCE WEBSITE.

5. Federal Motor Carrier Safety Administration. *Large Truck and Bus Crash Facts 2014*. 2016.

6. Idaho Office of Highway Safety. *Idaho Traffic Crashes*. 2014.

7. Bhat, C. R., K. Born, R. Sidharthan, and P. C. Bhat. A Count Data Model with Endogenous Covariates: Formulation and Application to Roadway Crash Frequency at Intersections. *Analytic Methods in Accident Research*, Vol. 1, 2014, pp. 53–71.

8. Castro, M., R. Paleti, and C. R. Bhat. A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections. *Transportation Research Part B: Methodological*, Vol. 46, No. 1, 2012, pp. 253–272.

9. Aguero-Valverde, J., and P. P. Jovanis. Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2061, No. 1, 2008, pp. 55–63.

10. Xie, K., X. Wang, K. Ozbay, and H. Yang. Crash Frequency Modeling for Signalized Intersections in a High-Density Urban Road Network. *Analytic Methods in Accident Research*, Vol. 2, 2014, pp. 39–51.

11. Xu, P., and H. Huang. Modeling Crash Spatial Heterogeneity: Random Parameters Versus Geographically Weighting. *Accident Analysis and Prevention*, Vol. 75, 2015, pp. 16–25.

12. Park, E. S., and D. Lord. Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2019, 2007, pp. 1–6.

13. Xie, Y., and Y. Zhang. Crash Frequency Analysis with Generalized Additive Models. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2061, 2008, p. pp 39-45.

14. Ukkusuri, S., S. Hasan, and H. M. A. Aziz. Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2237, 2011, pp. 98–106.

15. Shea, M. S., T. Q. Le, and R. J. Porter. Combined Crash Frequency–Crash Severity Evaluation of Geometric Design Decisions. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2521, 2015, pp. 54–63.

16. Islam, S., S. L. Jones, and D. Dye. Comprehensive Analysis of Single and Multi-Vehicle Large Truck At-Fault Crashes on Rural and Urban Roadways In Alabama. *Accident Analysis & Prevention*, Vol. 67, 2014, pp. 148–158.

17. Moore, D. N., W. H. Schneider IV, P. T. Savolainen, and M. Farzaneh. Mixed Logit Analysis of Bicyclist Injury Severity Resulting From Motor Vehicle Crashes at Intersetion and Non-Intersection Locations. *Accident Analysis and Prevention*, Vol. 43, No. 3, 2011, pp. 621–630.

18. Pahukula, J., S. Hernandez, and A. Unnikrishnan. A Time of Day Analysis of Crashes Involving Large Trucks in Urban Areas. *Accident Analysis and Prevention,* Vol. 75, 2015, pp. 155–163.

19. Mannering, F. L., V. Shankar, and C. R. Bhat. Unobserved Heterogeneity and the Statistical Analysis of Highway Accident Data. *Analytic Methods in Accident Research*, Vol. 11, 2016, pp. 1–16.

20. Washington, S. P., M. G. Karlaftis, and F. L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC, 2011.

21. Anastasopoulos, P. C., A. P. Tarko, and F. L. Mannering. Tobit Analysis of Vehicle Accident Rates on Interstate Highways. *Accident Analysis & Prevention*, Vol. 40, No. 2, 2008, pp. 768–775.

22. Anastasopoulos, P. C., 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, 2012, pp. 628–633.

23. Islam, M. Bin, and S. Hernandez. Fatality Rates for Crashes Involving Heavy Vehicles on Highways: A Random Parameter Tobit Regression Approach. *Journal of Transportation Safety & Security*, Vol. 8, No. 3, 2015, pp. 247–265.

24. Chen, F., X. Ma, and S. Chen. Refined-Scale Panel Data Crash Rate Analysis Using Random-Effects Tobit Model. *Accident Analysis and Prevention*, Vol. 73, 2014, pp. 323–332.

25. Jedidi, K., V. Ramaswamy, and W. S. Desarbo. A Maximum Likelihood Method for Latent Class Regression Involving a Censored Dependent Variable. *Psychometrika*, Vol. 58, No. 3, 1993, pp. 375–394.

26. Brown, S., W. H. Greene, M. N. Harris, and K. Taylor. An Inverse Hyperbolic Sine Heteroskedastic Latent Class Panel Tobit Model: An Application to Modelling Charitable Donations. *Economic Modelling*, Vol. 50, 2015, pp. 228–236.

27. Golembiewski, G. A., and B. Chandler. *Roadway Departure Safety: A Manual for Local Rural Road Owners*. Washington, D.C., 2011.

28. Tobin, J. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, Vol. 26, No. 1, 1958, pp. 24–36.

29. Yu, R., Y. Xiong, and M. Abdel-Aty. A Correlated Random Parameter Approach to Investigate the Effects of Weather Conditions on Crash Risk for a Mountainous Freeway. *Transportation Research Part C: Emerging Technologies*, Vol. 50, 2015, pp. 68–77.

30. Anastasopoulos, P. C., and F. L. Mannering. A Note on Modeling Vehicle Accident Frequencies With Random-Parameters Count Models. *Accident Analysis and Prevention*, Vol. 41, No. 1, 2009, pp. 153–159.

31. Carlsson, F., and P. Martinsson. Willingness to Pay Among Swedish Households to Avoid Power Outages: A Random Parameter Tobit Model Approach. *Energy Journal*, Vol. 28, No. 1, 2007, pp. 75–89.

32. Greene, W. H. *Econometric Analysis*. Pearson Education, New York, 2012.

33. Bhat, C. R. Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences. *Transportation Research Part B: Methodological*, Vol. 37, No. 9, 2003, pp. 837–855.

34. Train, K. Halton Sequences for Mixed Logit. *Working Papers*, 1999, pp. 1–18.

35. Halton, J. H. On The Efficiency Of Certain Quasi-Random Sequences of Points in Evaluating Multi-Dimension Integrals. *Numerische Mathematik*, Vol. 2, 1960, pp. 84–90.

36. Greene, W. H. *Econometric Modeling Guide*. Econometric Software, Inc., Plainview, NY, 2012.

37. Greene, W. H. *LIMDEP Version 10 Reference Guide*. Econometric Software, Inc., New York, 2012.

38. Veall, M. R., and K. F. Zimmerman. Pseudo-R2 Measures for Some Common Limited Dependent Variable Models. *Journal of Economic Surveys*, Vol. 10, No. 3, 1996.

39. Magidson, J., and J. K. Vermunt. Latent Class Models. No. 617, 2004, p. 11.

40. Hagenaars, J. A., and A. L. McCutcheon, Eds. *Applied Latent Class Analysis*. Cambridge University Press, New York, 2002.

41. Louviere, J. J., D. A. Hensher, and J. D. Swait. *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, New York, 2000.

42. Yang, C. C. Evaluating Latent Class Analysis Models in Qualitative Phenotype Identification. *Computational Statistics and Data Analysis*, Vol. 50, No. 4, 2004, pp. 1090–1104.

43. Anastasopoulos, P. C., V. N. Shankar, J. E. Haddock, and F. L. Mannering. A Multivariate Tobit Analysis of Highway Accident-Injury-Severity Rates. *Accident Analysis and Prevention*, Vol. 45, 2012, pp. 110–119.

44. Dutta, A., and D. A. Noyce. *Impact of Raising Speed Limits on Traffic Safety*. 2005.

45. Shaheed, M. S., and K. Gkritza. A Latent Class Analysis of Single-Vehicle Motorcycle Crash Severity Outcomes. *Analytic Methods in Accident Research*, Vol. 2, 2014, pp. 30–38.

LIST OF TABLES

LIST OF FIGURES

TABLE 1 Descriptive Statistics for Significant Variables by Road Classification

Classification	Variable	Mean	Standard Deviation	Min	Max
Principal Arterials	Crashes per Million-Vehicle Miles-Traveled (Response Variable)	0.127	0.208	0.005	3.488
	Speed Limit (1 if 65MPH, 0 Otherwise)	0.443	0.497	-	-
	Traffic Control Device (1 if No Device, 0 Otherwise)	0.702	0.458	-	-
	Road Configuration (1 if 2-Way & 2-Way Left-Turn Lane	0.126	0.333	-	-
	Heavy Vehicle AADT (1 if Less Than or Equal to 300, 0 Otherwise)	0.158	0.365	-	-
	Passenger Vehicle AADT (1 if Greater Than 10,500, 0 Otherwise)	0.108	0.310	-	-
	Total AADT (1 if Between 5,000 and 7,000, 0 Otherwise)	0.167	0.373	-	-
Major Collectors	Crashes per Million-Vehicle Miles-Traveled (Response Variable)	0.620	1.204	0.017	13.105
	Speed Limit (1 if Less Than or Equal to 40 MPH, 0 Otherwise)	0.241	0.428	-	-
	Traffic Control Device (1 if Stop Sign, 0 Otherwise)	0.159	0.366	-	-
	Horizontal Geometrics (1 if Straight, 0 Otherwise)	0.789	0.408	-	-
	Road Configuration (2-Way & Double-Yellow Painted Divider, 0 Otherwise)	0.129	0.335	-	-
	Total AADT (1 if Less Than 500, 0 Otherwise)	0.228	0.420	-	-
	Passenger Vehicle AADT (1 if Greater Than 2,500, 0 Otherwise)	0.227	0.419	-	-
Interstates	Crashes per Million-Vehicle Miles-Traveled (Response Variable)	0.034	0.047	0.003	0.726
	Speed Limit (1 if 75 MPH, 0 Otherwise)	0.675	0.469	-	-
	Total AADT (1 When Less Than 6,500, 0 Otherwise)	0.164	0.370	-	-
	Passenger Vehicle AADT (1 if Greater Than 15,000, 0 Otherwise)	0.116	0.321	-	-
	Heavy Vehicle AADT (1 if Between 2,000 and 3,000, 0 Otherwise)	0.256	0.437	-	-
	Horizontal Geometrics (1 if Curved, 0 Otherwise)	0.230	0.421	-	-
	Road Configuration (1 if 2-Way and Raised/Depressed Divider)	0.929	0.258	-	-
	Surface Defects (1 if No Surface Defects, 0 Otherwise)	0.960	0.195	-	-

Note: AADT = Average Annual Daily Traffic

TABLE 2 Best Fit Fixed-Parameter Tobit Regression Estimates for Principal Arterials

Variable	Coefficient	*t*-stat	Partial Effect
Constant	0.12	4.48	
Speed Limit (1 if 65MPH, 0 Otherwise)	-0.11	-3.95	-4.09
Traffic Control Device (1 if No Device, 0 Otherwise)	-0.12	-4.33	-4.66
Road Configuration (1 if 2-Way & 2-Way Left-Turn Lane)	0.10	2.81	4.06
Heavy Vehicle AADT (1 if Less Than or Equal to 300, 0 Otherwise)	0.10	2.88	3.76
Passenger Vehicle AADT (1 if Greater Than 10,500, 0 Otherwise)	-0.22	-4.70	-8.44
Total AADT (1 if Between 5,000 and 7,000, 0 Otherwise)	-0.06	-1.71	-2.26
Sigma, σ	0.31	26.85	
Number of Observations		862	
Log-Likelihood at Zero		-396.68	
Log-Likelihood at Convergence		-355.12	
χ^2		83.11	
Maddala Pseudo R^2		0.092	

TABLE 3 Best Fit Fixed-Parameter Tobit Regression Estimates for Interstates

Variable	Coefficient	*t*-stat	Partial Effect
Constant	0.09	4.54	
Speed Limit (1 if 75 MPH, 0 Otherwise)	-0.02	-2.20	-0.80
Total AADT (1 When Less Than 6,500, 0 Otherwise)	0.03	3.88	1.64
Passenger Vehicle AADT (1 if Greater Than 15,000, 0 Otherwise)	-0.03	-2.28	-1.30
Heavy Vehicle AADT (1 if Between 2,000 and 3,000, 0 Otherwise)	0.02	2.06	0.78
Horizontal Geometrics (1 if Curved, 0 Otherwise)	0.02	2.36	0.95
Road Configuration (1 if 2-Way and Raised/Depressed Divider)	-0.03	-2.20	-1.37
Surface Defects (1 if No Surface Defects, 0 Otherwise)	-0.05	-2.81	-2.30
Sigma, σ	0.06	20.74	
Number of Observations		379	
Log-Likelihood at Zero		211.91	
Log-Likelihood at Convergence		242.54	
χ^2		61.25	
Maddala Pseudo R^2		0.149	

TABLE 4 Best Fit Random-Parameter Tobit Regression Estimates for Major Collectors

Variable	Fixed-Parameter Tobit			Random-Parameter Tobit		
	Coefficient	***t*-stat**	**Partial Effect**	**Coefficient**	***t*-stat**	**Partial Effect**
Constant	-1.31	-5.94		-1.21	-5.40	
Speed Limit (1 if Less Than or Equal to 40 MPH, 0 Otherwise)	0.52	2.87	0.20	0.42	2.30	12.02
Standard Deviation of Parameter, Normally Distributed	-	-	-	*0.73*	*6.67*	-
Traffic Control Device (1 if Stop Sign, 0 Otherwise)	0.45	2.22	0.18	0.40	2.01	11.59
Horizontal Geometrics (1 if Straight, 0 Otherwise)	0.91	4.38	0.36	0.73	3.67	20.96
Road Configuration (1 if 2-Way & Double-Yellow Painted Divider, 0 Otherwise)	0.70	3.10	0.27	0.36	1.53	10.41
Standard Deviation of Parameter, Normally Distributed	-	-	-	*0.99*	*5.45*	-
Total AADT (1 if Less Than 500, 0 Otherwise)	1.03	5.61	0.41	0.97	4.88	27.90
Passenger Vehicle AADT (1 if Greater Than 2,500, 0 Otherwise)	-0.60	-2.90	-0.24	-0.50	-2.22	-14.31
Sigma, σ	1.82	25.91		1.59	56.33	
Number of Observations		768			768	
Log-Likelihood at Zero		-1003.14			-1003.14	
Log-Likelihood at Convergence		-967.07			-926.67	
χ^2		72.13			80.81	
Maddala Pseudo R^2		0.090			0.181	

TABLE 5 Best Fit Latent Class Tobit Regression Estimates for Principal Arterials

Variable	Latent Class 1		Latent Class 2		Latent Class 3		Partial Effect
	Coefficient	*t*-stat	Coefficient	*t*-stat	Coefficient	*t*-stat	
Constant	1.51	2.59	0.26	7.90	0.04	11.83	
Speed Limit (1 if 65MPH, 0 Otherwise)	0.05	0.06	-0.12	-3.25	0.00	-1.07	-0.45
Traffic Control Device (1 if No Device, 0 Otherwise)	-1.06	-2.06	-0.11	-3.15	-0.01	-1.93	-0.78
Road Configuration (1 if 2-Way & 2-Way Left-Turn Lane)	-0.94	-0.74	0.09	2.31	0.00	-0.12	0.04
Heavy Vehicle AADT (1 if Less Than or Equal to 300, 0 Otherwise)	0.93	1.86	0.05	1.34	0.00	0.69	0.49
Passenger Vehicle AADT (1 if Greater Than 10,500, 0 Otherwise)	-0.53	-0.10	-0.12	-1.58	-0.01	-1.56	-0.71
Total AADT (1 if Between 5,000 and 7,000, 0 Otherwise)	1.15	0.65	0.00	-0.05	0.00	-0.09	0.34
Sigma, σ	0.61	2.31	0.15	10.52	0.02	15.46	
Class Probability (*t*-statistic)	0.022 (2.77)		0.262 (9.06)		0.716 (25.24)		
Number of Observations	862						
Log-Likelihood at Zero	166.81						
Log-Likelihood at Convergence	195.60						
Akaike Information Criterion	-339.20						
Bayesian Information Criterion	-215.50						

TABLE 6 Best Fit Latent Class Tobit Regression Estimates for Major Collectors

Variable	Latent Class 1		Latent Class 2		Latent Class 3		Latent Class 4		Partial Effect
	Coefficient	*t*-stat	Coefficient	*t*-stat	Coefficient	*t*-stat	Coefficient	*t*-stat	
Constant	2.11	0.71	0.08	3.83	0.07	0.93	0.43	1.69	
Speed Limit (1 if Less Than or Equal to 40 MPH, 0 Otherwise)	2.66	1.40	0.02	1.16	0.05	0.88	0.43	2.67	9.94
Traffic Control Device (1 if Stop Sign, 0 Otherwise)	1.94	1.05	0.01	0.39	0.06	1.19	0.55	3.08	1.18
Horizontal Geometrics (1 if Straight, 0 Otherwise)	-1.28	-0.61	0.02	1.00	0.13	2.17	0.02	0.07	-5.24
Road Configuration (2-Way & Double-Yellow Painted Divider, 0 Otherwise)	2.46	1.07	0.02	0.81	0.13	1.85	-0.09	-0.44	6.56
Total AADT (1 if Less Than 500, 0 Otherwise)	2.28	1.17	0.01	0.33	0.29	5.52	0.74	4.26	7.00
Passenger Vehicle AADT (1 if Greater Than 2,500, 0 Otherwise)	-3.71	-1.04	-0.03	-1.39	-0.06	-0.90	0.35	1.87	-5.91
Sigma, σ	3.10	3.83	0.07	9.64	0.16	5.44	0.44	5.38	
Class Probability (*t*-statistic)	0.076 (3.93)		0.577 (13.22)		0.227 (4.97)		0.120 (4.98)		
Number of Observations	768								
Log-Likelihood at Zero	-437.00								
Log-Likelihood at Convergence	-414.63								
Akaike Information Criterion	899.30								
Bayesian Information Criterion	1061.80								

TABLE 7 Best Fit Latent Class Tobit Regression Estimates for Interstates

Variable	Latent Class 1		Latent Class 2		Partial Effect
	Coefficient	***t*-stat**	**Coefficient**	***t*-stat**	
Constant	0.72	5.05	-0.01	-0.20	
Speed Limit (1 if 75 MPH, 0 Otherwise)	0.00	0.02	-0.01	-1.49	-0.37
Total AADT (1 When Less Than 6,500, 0 Otherwise)	0.00	-0.10	0.03	2.23	0.69
Passenger Vehicle AADT (1 if Greater Than 15,000, 0 Otherwise)	-0.01	-0.29	-0.02	-0.61	-0.77
Heavy Vehicle AADT (1 if Between 2,000 and 3,000, 0 Otherwise)	0.00	-0.12	0.02	2.32	0.65
Horizontal Geometrics (1 if Curved, 0 Otherwise)	-0.01	-0.55	0.02	1.63	0.24
Road Configuration (1 if 2-Way and Raised/Depressed Divider)	-0.01	-0.24	0.01	0.48	0.23
Surface Defects (1 if No Surface Defects, 0 Otherwise)	-0.70	-4.81	0.01	0.36	-10.78
Sigma, σ	0.03	4.21	0.04	11.84	
Class Probability (*t*-statistic)	0.361 (2.30)		0.639 (4.06)		
Number of Observations	379				
Log-Likelihood at Zero	363.56				
Log-Likelihood at Convergence	385.48				
Akaike Information Criterion	-733.00				
Bayesian Information Criterion	-658.10				

TABLE 8 Fit of Predicted Crash Rates

Classification	Tobit R^2	Latent Class R^2
Principal Arterials	0.09	0.76
Major Collectors	0.29	0.88
Interstates	0.22	0.72

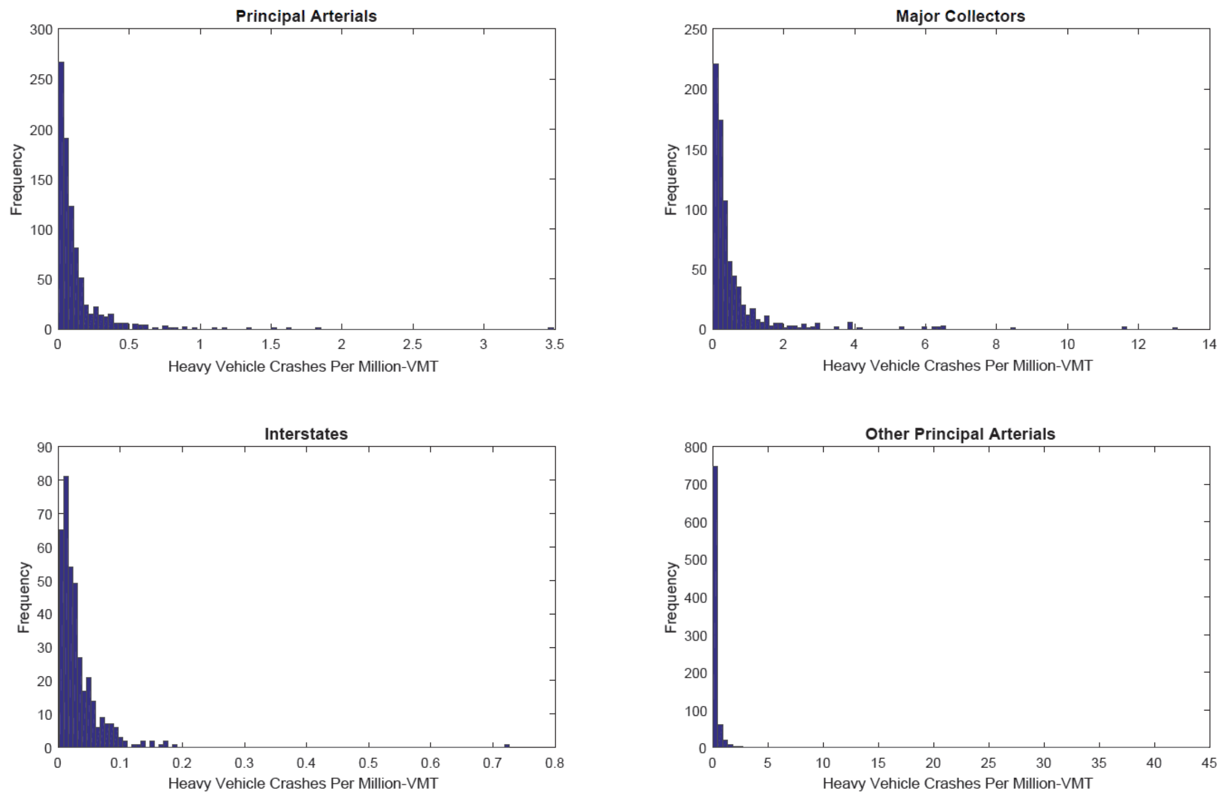

FIGURE 3 Heavy vehicle crash rate distribution by roadway classification.

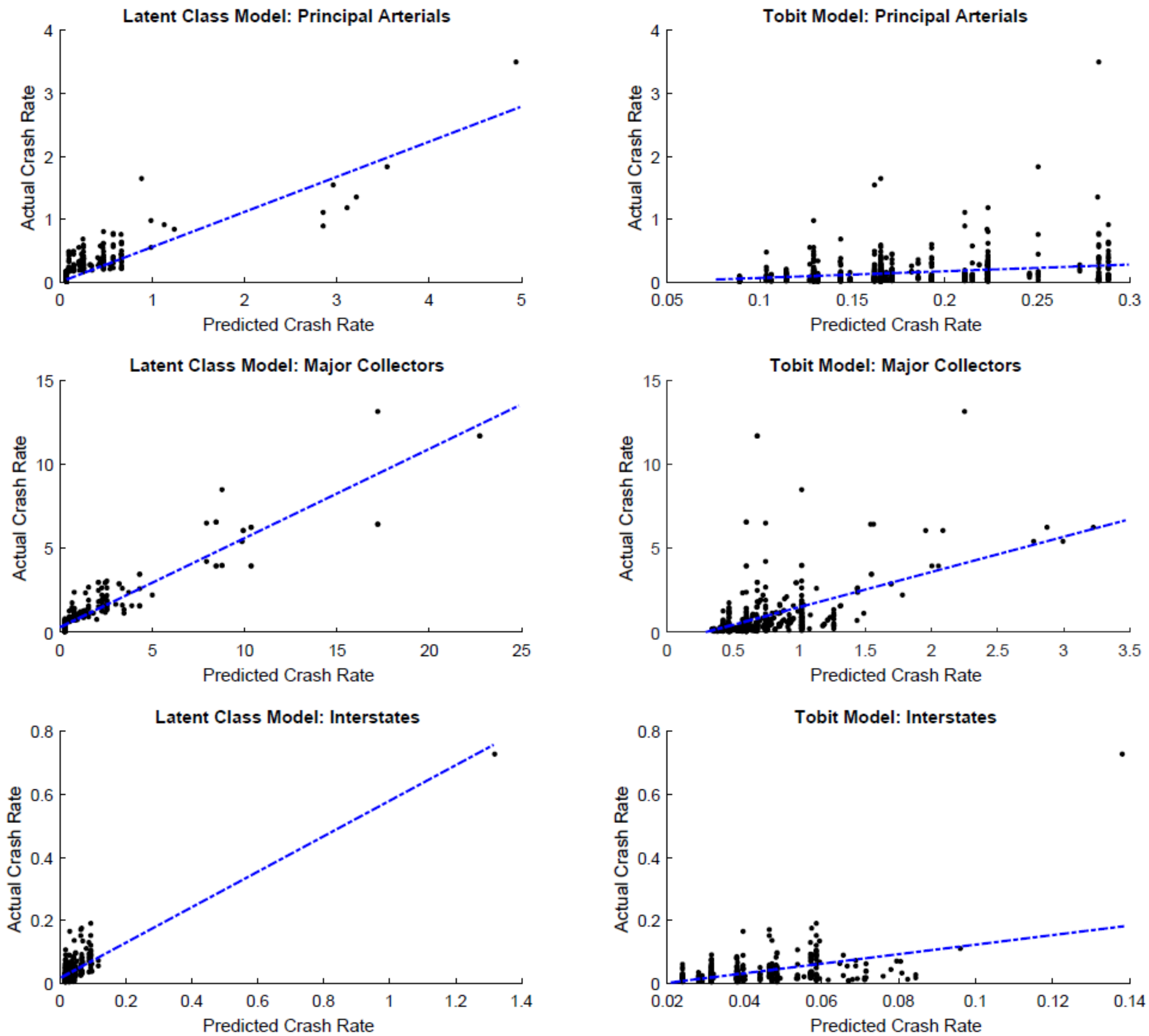

FIGURE 4 Actual crash rates versus predicted crash rates.