

# URBAN FREIGHT CRASH ANALYSIS USING ORDINAL LOGIT AND ORDINAL PROBIT REGRESSION IN THE STATE OF ALABAMA

**Mehrnaz Doustmohammadi<sup>1</sup>**

<sup>1</sup>*Department of Civil and Environmental Engineering, University of Alabama in Huntsville, Huntsville, AL, 35899, USA*

*Received 28 November 2018; accepted 18 January 2019*

**Abstract:** Trucking is vital to the economic advancement of the nation as the majority of goods are transported on truck internally. This topic is critical to the state of Alabama since trucking is the most frequently selected travel mode for all goods shipped through the Port of Mobile and especially those necessary to support the automotive industry, one of the major sectors of the Alabama economy. Therefore, truck crashes have a significant impact on moving goods and delivery schedules vital to maintaining the operation of assembly plants. Truck crashes often cause severe delays and extra time is needed to move larger vehicles from roadways and transfer the carried good to other vehicles. Additionally, truck crashes are often severe due to the fact their increased size and weight tends to cause severe damage when crashes occur with passenger cars, tending to cause more injuries. This study examined urban, at-fault truck crashes where the driver operated using a commercial driver's license. The crashes were obtained from Alabama cities between 2012 and 2016. A statistical analysis of crashes using a logit regression model and a probit regression model was performed and the significant variables that influence truck crashes were determined and analyzed. From the analysis, the variables that most likely increased the severity of urban truck at-fault crashes were fatigue and speed. The analysis demonstrated the ability of the two models to determine influences and the models were in strong agreement of the variables, with distracted driving, driving too fast for conditions and fatigue being the leading issues that increased the severity of crashes.

**Keywords:** ordinal regression, urban freight, crash analysis.

## 1. Introduction

Urban and rural roadways have fundamentally different characteristics. The reason for the differences is the use of the roadways and users of the roadways. These differences materialize in factors such as density of the roadway network, travel patterns used, driver characteristics, and land use types. As the result the differences in roadways, the characteristics of motor

vehicle crashes differ as well. For instance, vehicle fatalities tend to be greater on high speed rural roads while bicycle and pedestrian fatalities are more frequent in urban areas (Lemp, 2017).

Trucks on the nation's highways carry over 70% of the freight in the U.S. (Kim, 2017). The safety of these vehicles on the roadways is necessary for economic vitality; maintain production schedules and delivering

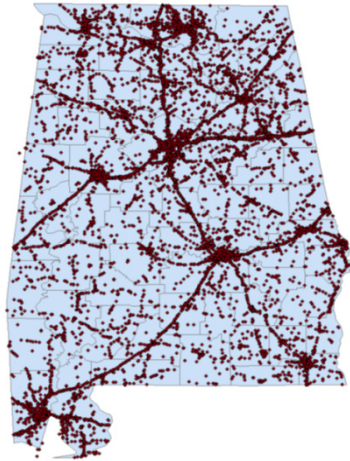
<sup>1</sup> Corresponding author: [md0033@uah.edu](mailto:md0033@uah.edu)

desired goods at the time and location for purchase. Therefore, trucks on the roadway will continue to be a common occurrence. However, these vehicles come with a price with regard to highway safety. Nationwide, in 2015, there were almost 3,600 fatal crashes and 83,000 injury crashes involving large trucks, and unfortunately the trend shows that the number of fatalities and injuries will continue to increase from a low in 2008 during the recession (“Large Truck and Bus Crash Facts”, 2015). The main issue associated with the severity of large truck crashes is the fact that trucks often weigh 20-30 times more than passenger cars and have higher ground clearance that has the potential to trap smaller vehicles under trailers (Lemp, 2017). Additionally, trucks tend to take longer to stop than passenger cars, with a greater discrepancy on wet pavement (Lemp, 2017). The safety of large trucks on the roadways can be summarized in the statistic that more people will be killed in crashes involving truck than were killed in all domestic airline crashes over the past 45 years (Abramson, 2015).

Vehicle crashes have a significant impact on the lives of those involved in the crash, as well as those living in the community due to delays from roadway closures and general rubbernecking. This is no different for urban truck freight crashes. From a perception view, truck crashes are

common occurrences and big business for trial lawyers, who tout “big money settlements for those involved in truck crashes” which is easy to determine through a quick google search of ‘truck crashes’. While truck crashes are very important to prevent, in reality, truck crashes are often a small percentage of the crashes because of the restrictions that have been implemented to regulate safety, including hours of operation limitations, strict DUI levels and pre-trip vehicle inspections. As previously mentioned, the size of the truck involved in the crash often has a serious impact on the level of severity of the crash.

This topic is especially important due to the heavy reliance on trucking in the state of Alabama. The automotive industry and Port of Mobile rely heavily on trucks for deliveries of parts and distribution of final products. While in Alabama, crashes in which the causal unit was a truck driven by a driver with a commercial driver license represent only 5 percent of the total crashes that occurred between 2012 and 2016, these crashes resulted in a higher percentage of fatalities and incapacitating injuries than crashes where a non-commercial driver was involved (CARE database, 2017). Figure 1 shows all the truck at-fault crashes in the state of Alabama between 2012 and 2016 with each dot on the map representing one or more crashes.



**Fig. 1.**  
*Truck-at-Fault Crashes in the State of Alabama*

This paper uses the Critical Analysis Reporting Environment (CARE) database developed from officer reports at crash scenes and available from the Center for Advanced Public Safety at the University of Alabama to perform a statistical evaluation of truck related crashes in Alabama. The paper presents a review of literature related to truck crashes, identifies the data and statistical models used in the analysis and presents factors that lead to increases in crash severity where the causal unit was a truck. The models used in the analysis are Ordered Logit and Ordered Probit. The results of the study indicate that driver distraction and fatigue are the primary issues that lead to severe crashes and that head-on crashes are the most severe, due to the difference between the weights of the vehicles. Overall, while steps have been implemented to improve safety there are still too many serious crashes involving trucks in Alabama in urban areas.

## 2. Literature Review

Because of the larger size and higher center of gravity, trucks are more likely to be influenced by wind and weather involved crashes. Goodwin has mentioned this issue in 2003 and as a result, several State Departments of Transportation alert truck drivers during severe weather (Goodwin, 2003). In the same year, (Khattak *et al.*, 2003) attempted to understand how truck driver behaviors, vehicle factors and crash events influence large-truck rollovers and occupant injuries in single-vehicle crashes using binary Probit models. (Khorashadi *et al.*, 2005) explored the differences between urban and rural driver injuries and large truck driver injuries in crashes that involved large trucks. This paper focused mainly on the geometric characteristics and a multinomial discrete probability model was used for this analysis (Khoradashi *et al.*, 2005). A study

of truck crashes in Nebraska was performed using multinomial and ordered mixed logit models to analyze the impact of weather on the severity of single-vehicle truck crashes (Naik *et al.*, 2016). The study found that air temperature, humidity, rain, and wind speed all had an effect on the severity of these types of crashes (Naik *et al.*, 2016).

Research on New York City traffic data (2008-2012) indicated that the influencing factors for single-vehicle and multi-vehicle truck crashes are different, and that there are statistically significant heterogeneous effects of truck weight on the crash severity of both models (Zou *et al.*, 2017). In this study, two-random parameter ordered probit models have been used to compare the differences between single-vehicle truck crashes and multi-vehicle truck crashes (Zou *et al.*, 2017). In 2016, (Osman *et al.*, 2016) conducted research employing several logit models in order to investigate causal factors contributing to the severity of injuries sustained in large truck crashes in work zones in Minnesota. The generalized ordered logit model was found to be superior in this study (Osman *et al.*, 2016). The study found that the most influential factors causing higher severity in these crashes were daytime crashes, no control of access, higher speed limits, and crashes occurring on rural principal arterials (Osman *et al.*, 2016). The study also found urban principal arterials, one-unit trucks with trailer, and single-vehicles to contribute to higher severity, as well, but the effects were less pronounced (Osman *et al.*, 2016). (Islam *et al.*, 2014) conducted research that used random parameter logit models to analyze the injury severity in large truck at-fault crashes in Alabama. The study found that some variables were influential in crash severity only for either single-vehicle crashes

or multi-vehicle crashes. Additionally, it was found that many variables were random and showed no influence, and that some variables were influential in single-vehicle, multi-vehicle, urban, and rural crashes (Islam *et al.*, 2014). (Zhu *et al.*, 2011) studied large truck crash data nationwide using an ordered-probit logit model in order to identify the factors influencing crash severity of these large-truck crashes. The study showed the need for increased accuracy in its findings; however, the results indicated the influence of driver behavior variables, presence of an intersection, head-on collisions, and presence of a barrier between oncoming traffic lanes on the severity of these large truck crashes (Zhu *et al.*, 2011).

(Lemp *et al.*, 2011) conducted a study that examined factors that affect crash severity in crashes involving heavy duty trucks. The study used data from the Large Truck Crash Causation Study (LTSCC) and employed standard ordered probit and heteroskedastic ordered probit models for the analysis. The results of both models were consistent with one another. Findings of the research include an increase in the likelihood of a low-severity crash if the crash occurs at a curve or roadway sag, an involved truck was overweight, a driver was under the influence, or a driver was exhibiting aggressive behavior (excluding speeding) (Lemp *et al.*, 2011). Additionally, the study found that non-bright lighting conditions, snowy/icy roadways, fog, and the number of involved occupants increase the probability of fatality in the crash (Lemp *et al.*, 2011). Finally, the study found that 1-trailer trucks are associated with low-severity crashes, while 0-trailer trucks are associated with the most severe truck crashes.

A study in New Zealand required truck drivers involved in crashes fill out

questionnaires related to their crashes (Gander *et al.*, 2006). The study found that fatigue should be given consideration as a possible contributing factor if the drivers had and combination of the following: less than six hours of sleep in the last twenty four hours, more than twelve hours awake, more than a week without two good nights of sleep in a row, or the crash occurred from midnight to eight a.m. (Gander *et al.*, 2006). The results of this study, however, were not reliable due to a low response rate to the questionnaire and a need for additional questions.

A research study from California employed multinomial logit models on four years of crash data in order to perform an analysis of the differences between urban and rural driver injuries in accidents involving large trucks (Greene, 2000). The study found that there were incredible differences between the factors influencing the driver-injury severity in rural areas in comparison with urban areas. Even variables that were influential in both scenarios had significantly different effects between the two cases. The results of the study found that driving under the influence of drugs or alcohol (as the primary collision factor) was the most influential factor in both severity models (Greene, 2000).

(Joshua and Garber, 1990) did a study that aimed to establish relationships between large truck accidents and traffic and roadway geometric variables through the use of multiple linear and Poisson Regression analyses. The results of the models indicated that slope change rate, average daily traffic, percent trucks, and difference in speed between trucks and non-trucks influence the occurrence of truck-involved accidents. (Miaou, 1994) evaluated the performance of Poisson and negative binomial regression

models in establishing the relationship between truck accidents and geometric design of road sections. Results from the study indicate that the NB regression model using the moment and regression-based methods should be used with caution. It also indicates that no particular model was better than the others for estimated relative frequencies of truck accident involvements across road sections (Miaou, 1994). Another study (Duncan *et al.*, 1998) examined the impact of various factors on injuries to passenger car occupants involved in truck-passenger car rear-end collisions and demonstrated the use of the ordered probit model in this scenario. The study found that darkness, high speed differentials, high speed limits, grades (especially when wet), being in a car struck to the rear (versus being the striking car), driving while drunk, and being female increase injury severity. On the other hand, snowy/icy roads, congestion, being in a station wagon (rather than a sedan), and using a child restraint decreases injury severity (Duncan *et al.*, 1998).

(Khattak *et al.*, 2002) focused on analyzing the effect of work zone duration mainly due to its policy-sensitivity. The study was performed by utilizing negative binomial modeling. The study found that longer work zone duration significantly increases both injury and non-injury crash frequencies. (Golob and Regan, 2003) did a study that examines the locations and conditions linked to truck-involved crashes using a binomial logit model to describe how probability that a crash involves a truck is influenced by percent trucks, time of day, day of week, weather conditions, mix of truck types, and absolute level of AADT. This model could be used to identify the locations with higher- or lower-than-expected truck crash rates. Additionally, multinomial logit

modeling was used to better understand patterns of truck-involved crashes by separating the crashes by type (Golob and Regan, 2003). (Hiselius, 2004) performed a study estimating the relationship between accident frequency and the traffic flow by using Poisson and Negative Binomial regression models. The study was performed by empirically treating the traffic flow as containing homogeneous vehicles and also as containing both cars and trucks. The study found that the estimated relationship between expected accident rate and traffic flow differs considerably depending on whether different types of traffic modes are considered or not implying that important information is lost if no consideration is taken to vehicle type (Hiselius, 2004). (Qin *et al.*, 2004) performed a study seeking to identify a linear relationship between crash counts and traffic volumes by disaggregating crashes into different types and modeling each separately using zero-inflated Poisson modeling. This study found both that the relationship between crashes and daily volume is non-linear and varies by crash type and that the relationship between these variables is significantly different from the relationship between crashes and segment length for all crash types (Qin *et al.*, 2004).

A study performed by (Caliendo *et al.*, 2007) sought to satisfy the need to quantify the safety effects on the expected number of crashes for multilane rural roads by using Poisson, Negative Binomial, and Negative Multinomial regression models to model frequency of accident occurrence as well as a Maximum Likelihood Method to estimate parameters. The models in the study were applied separately to curves and tangent section in order to perform the analyses. In the study, it was found that AADT was a significant variable in crash frequency on

both curves and tangent sections (Caliendo *et al.*, 2007). A study (Mulinazzi *et al.*, 2009) constructed a model to predict the likelihood of wind-induced truck crashes using a multivariate linear regression approach. The research found that wind speeds were not found to be a statistically significant part of any model based on available data. Another study (Jin *et al.*, 2010) developed ordered probability models to examine the factors significantly contributing to the severity of right-angle crashes given that a crash had occurred. Results of the study indicated that significant variables that affected average severity of an accident included whether person was ejected or not, alcohol/drug use, driver age, point of impact, and standardized yellow time for entering movement (Jin *et al.*, 2010).

(Lao *et al.*, 2011) adopted a diagonal inflated bivariate Poisson regression method to fit reported AVC and carcass removal data sets in order to model both reported AVC data and carcass removal data together and provide a better understanding of highway AVCs. Results found that an increase in speed limit, AADT, and shoulder width will increase the numbers of reported AVCs and carcass removals, while rolling hills and mountainous terrain decrease the number of reported AVCs (Lao *et al.*, 2011). Also (Pei *et al.*, 2011) performed a study that developed a joint probability model to integrate the predictions of both crash occurrence and crash severity into a single framework. Results of this study indicated that the proposed model demonstrates a good statistical fit and provides an appropriate analysis of the influences of explanatory factors. (Christoforou *et al.*, 2012) utilized univariate and ordered probit models in an effort to establish a framework for the integration of real-time traffic data in road



safety analysis in 2012. The results of the study found that increasing traffic volume has a consistently positive (decreasing) effect on injury severity. In addition, while in higher traffic volumes, higher traffic speeds aggravate severity outcomes, but in lower traffic volumes, speed does not significantly influence severity consistently (Christoforu *et al.*, 2012). (Aguero-Valverde, 2013) performed a study comparing random effects, zero inflated, and zero inflated random effects models for both goodness-of-fit and precision of posterior crash frequency estimates using a full Bayes hierarchical approach. The study found that fixing random effects over time significantly increases the fit of the models and that the Poisson gamma models perform better than Poisson lognormal models in terms of posterior deviance for this dataset (Aguero-Valverde, 2013). (Yang *et al.*, 2013) performed another study which identified risk factors in work zone safety evaluation through the estimation of a crash frequency model. It was found that the proposed improved approach outperformed the traditional approach in terms of goodness-of-fit statistics. Another study by (Amarasingha and Dissanayake, 2013) focused on understanding and evaluating the effects of both traffic conditions and site characteristics on the occurrence of large truck crashes through the application of Poisson and Negative Binomial Regression models. The study found that variables such as number of lanes, AADT, and percent trucks have a specific impact on large truck crashes. (Caliendo *et al.*, 2013) utilized Bivariate Negative Binomial regression, Random Effects Binomial regression, and Negative Multinomial regression models, along with the Maximum Likelihood Method in order to model frequency of accident occurrence and year effect on severe crashes in 2013.

The study found that the number of both non-severe and severe accidents occurring in tunnels increases with the tunnel length, the annual average daily traffic per lane, the percentage of trucks, and the number of lanes (Caliendo *et al.*, 2013). (Hamdar and Schorr, 2013), sought to develop a quantitative safety propensity index that captures overall propensity of a given surrounding environment to cause unsafe driving by using structural modeling techniques. It was found that shoulder width, number of lanes, grade, visibility, precipitation, speed, and AADT were significant parameters regarding traffic safety.

(Yu and Abdel-Aty, 2013) performed a study seeking to develop safety performance functions by analyzing a mountainous freeway in Colorado featuring mountainous terrain and adverse weather by using Bayesian random effects Poisson modeling. Their research found that AADT was not significant in the developed safety performance function. (Dong *et al.*, 2014) described a multivariate random-parameters zero-inflated negative binomial regression model that employed a full Bayesian method to model parameters that were used to jointly model crash counts. Results of the study show that the MRZINB model possesses most of the desirable statistical properties in terms of its ability to accommodate unobserved heterogeneity and excess zero counts in correlated data. (Mohammadi *et al.*, 2014) explored the effects of temporal correlation in crash frequency models at the highway segment level in 2014 by developing a negative binomial model using a generalized estimating equation procedure that incorporates the temporal correlations amongst yearly crash counts. The results of the study confirmed that the use of GEE is a good approach for addressing

the serial correlation in crash frequency data. (Sarker *et al.*, 2015) performed a study seeking to develop a procedure for identifying secondary crashes by using a static and a dynamic approach in a large-scale multimodal transportation system in order to develop a Secondary Crash Identification Algorithm. It was found that moderate traffic volumes tended to result in more secondary crashes than high traffic volumes. (Chen and Xie, 2016) aimed to use Generalized Additive Models in 2016 to model the complex relationship between traffic volumes and crash frequencies for signalized urban and suburban intersections and demonstrated how to use GAMs to identify appropriate safety performance functions for Generalized Linear Models and to model the joint safety effects of multiple covariates based on thin-plate splines. The study found that when the ratio of AADT of the minor approach to AADT of the major approach is between -1.5 and -1, there is a clear growing trend in expected crash frequency, but when this ratio continues to increase to 1 or higher, the expected number of crashes begins to fall. Also (Roshandeh *et al.*, 2016) investigated the effects of traffic, environmental, intersection geometric, and pavement-related characteristics on total crash frequencies at intersections using a random-parameter Poisson model. The results showed that evening peak period traffic volume, pavement condition, and unlighted intersections have the greatest on crash frequency of the identified factors. (Wang *et al.*, 2017) estimated crashes by both crash type and crash severity using the Integrated Nested Laplace Approximation Multivariate Poisson Lognormal model and identified the different effects of contributing factors on different crash

type and severity counts on rural two-lane highways. The comparison in the study shows that the MVPLN model outperforms the UPLN model in prediction accuracy (Wang *et al.*, 2017).

(Ma and Haque, 2017), utilized historical and archived crash data maintained by TDOT to closely analyze crash patterns in work zones while taking into account crash, roadway geometry, environmental, and several other characteristics in order to develop performance metrics for Tennessee work zones. The results showed that number of lanes, AADT, high speed limits, and length of roadway are all influential factors contributing to crashes at work zones (Ma and Haque, 2017). Another study by (Yan *et al.*, 2017) utilized Aggregated Crash Index and Time to Collision crash surrogate metrics in order to examine the relationship between traffic flow and potential conflict risks. The study found that potential conflict risk had a strong correlation with level of service. (Chiou *et al.*, 2017) aimed to identify key factors affecting crash frequencies under various times of the day in order to propose effective and time-specific countermeasures by using both clustering and multivariate approaches in 2017. The results found that “peak and off-peak” of crash counts in various segments markedly differ (Chiou *et al.*, 2017). This paper follows a similar design from previously identified literature by examining crashes where a large truck, driven by a commercial license holder, was the vehicle that caused the crash. This paper differs from previous research in focusing on specific truck crashes in urban areas and including various factors for contributing circumstance and manner of crash to see the influence on severity of the crash.

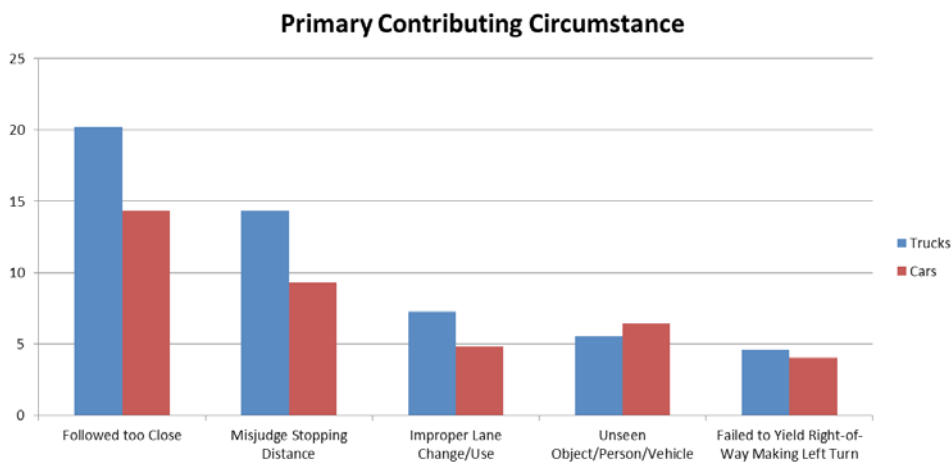


### 3. Preliminary Study: Descriptive Statistics

In this part of research, frequency analysis was performed on each category for more understanding of the crashes. The (CARE database, 2017) was used to examine the differences between truck at-fault and passenger car at-fault crashes in Alabama between 2012 and 2016. The two major frequency analyses are included.

The first analysis focuses on the primary contributing circumstance, which is

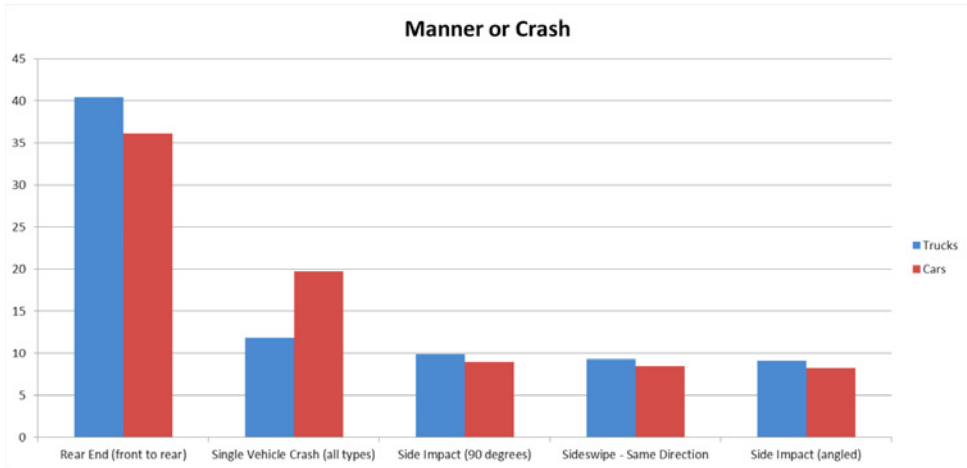
essentially the cause of the crash. In figure 2, the primary contributing circumstance for at-fault truck and passenger car crashes as shown by frequency. From the figure, there is definitely a trend that can be seen that truck at-fault crashes occur when the driver of the truck is following too closely and misjudging the stopping distance. These two contributing circumstances are related in the fact that truck take a longer time and distance to stop. Additionally, these factors are very important with to distracted driving, as any distraction will increase the time and distance to come to a stop.



**Fig. 1.**  
*Primary Contributing Circumstance*

The second analysis focused on the type of crash or “manner of crash” as recorded in the (CARE database, 2017). This category will allow us to have a better realization of what type of crash the at-fault vehicle was involved in. Figure 3 shows the graph that has the comparison of each subcategory in our study. From the figure, trucks have a

slightly higher likelihood of a rear-end crash, which is consistent with the contributing circumstance of following too close as this action will likely result in rear-end crashes. Interestingly, trucks have a much lower percentage of run of the road crashes which might indicate that truck drivers are less distracted or fatigued while driving.



**Fig. 2.**  
*Manner of Crash*

#### 4. Data

The data used in this analysis were all obtained from the Critical Analysis Reporting Environment (CARE) database that contains all crashes in Alabama between 2012 and 2016. The data extracted for the analysis included only crashes where a truck, driven by a commercial driver license holder, was at-fault. To focus the analysis on urban

trucks at-fault crashes, only crashes that occurred within the urbanized boundary of the metropolitan planning organizations in Alabama were included in the dataset. This limited the crashes to those that were within the largest 12 cities/areas in Alabama with a population greater than 50,000.

The independent variables for the study and the possible values are shown in table 1.

**Table 1**  
*Independent Variables and Values from the Dataset*

<b>Day of the Week</b>	Weekday
	Weekend
<b>Time of Day</b>	Evening Peak
	Morning Peak
	Off Peak
<b>Controlled Access</b>	Interchange
	Main Road
	Not a Controlled Access
	Off Ramp
	On Ramp

<b>Primary Contributing Circumstance</b>	Cargo fell/Defective Equipment
	Distraction
	Driving too Fast
	DUI
	Fatigued/Asleep
	Improper Driving
	Misjudge Stopping Distance
	Swerved to Avoid
	Vision Obstructed
<b>Manner of Crash</b>	Angle
	Causal Vehicle Backing
	Head on
	Non-Collision
	Rear End
	Side Impact
	Sideswipe
	Single Vehicle Crash
<b>Lighting Conditions</b>	Dark - Roadway Not Lighted
	Dark - Spot Illumination
	Dark-Roadway Lighted
	Daylight
	Dusk/Dawn
<b>Weather</b>	Clear
	Rain
<b>Driver Age</b>	15-20
	21-25
	26-30
	31-35
	36-40
	41-45
	46-50
	51-55
	56-60
	61-65
	66-70
	71 and older
<b>Speed Limit</b>	GT45
	LT45
<b>Number of Lanes</b>	Four Lanes
	Parking Lot
	Six Lanes
	Two Lanes

For the response variable, there were four levels of severity selected for this study. Property damage only (PDOs) are considered  $Y=1$ , which means it has the least severity. The next level of severity is possible injury crashes  $Y=2$ . The third category,  $Y=3$ , are non-incapacitating injuries where an individual from the crash is transported

to a hospital for medical care. The highest category involve fatalities and incapacitating injuries,  $Y=4$ , and include fatalities that occur within two days of the crash from injuries sustained during the crash. Two categories were consolidated due to low number of fatalities. The response variables for the tests are shown in Table 2.

**Table 2**  
*Response Variable Values*

PDO	1
Possible injury	2
Non incapacitating Injury	3
Fatalities, Incapacitating Injury	4

## 5. Data Preparation

Before conducting these statistical tests, the data were scrubbed in such a way that meets the requirements for the tests attempted.

Data were grouped into subcategories to verify that each subcategory had a sufficient number of data elements to conduct a fair analysis. An example of the data scrubbing is shown in Table 3.

**Table 3**  
*Example of Data Preparation*

Angle	Rear End
Angle (front to side) Opposite Direction	Rear End (front to rear)
Angle (front to side) Same Direction	<b>Side Impact</b>
Angle Oncoming (frontal)	Side Impact (90 degrees)
<b>Causal Vehicle Backing</b>	Side Impact (angled)
Causal Vehicle Backing: Rear to Rear	<b>Sideswipe</b>
Causal Vehicle Backing: Rear to Side	Sideswipe - Opposite Direction
Head on	Sideswipe - Same Direction
Head-On (front to front only)	<b>Single Vehicle Crash</b>
Non-Collision	Single Vehicle Crash (all types)
Non-Collision	

## 6. Model Selection

To perform the analysis, originally four statistical models were run selected to

perform the analysis. These models were taken from the reviewed literature as the desired models for studying crash data. The models were: binary probit regression, binary

logit regression, ordered probit regression and ordered logit regression. The difference between binary logit and probit regression and ordered probit and logit regression is the response variable. The response variable for binary regression is “severe” and “not severe”.

Fatalities and Incapacitating injuries are considered severe and non-incapacitating injuries, possible injuries and property damage only were grouped as not severe. The goodness of fit value for each model is shown in table 4.

**Table 4**  
*Goodness of Fit Values for Four Statistical Models*

Goodness of fit value	Binary probit	Binary Logit	Ordered probit	Ordered logit
Pearson chi-square	0.453	0.888	1.026	0.968

The Pearson Chi-Squared is used as the goodness of fit for the models for comparison purposes. The closer this value is to one the better fit the model is to the data. The goodness of fit value from table 4 indicates that the ordered probit and ordered logit are better models, and therefore were selected for inclusion in this study.

### 7. Model Methodology

Ordered probit and ordered logit models were selected for this study. The reason why these two models are practical in this paper is due to the nature of our dependent variables in this study. Dependent variable is categorical and it can certainly be arranged from the highest to the lowest using certain criteria. Each category needs to be associated with a number which shows the level of severity of the crash. In data preparation section the different levels of severity with their codes have been discussed. We associated PDO to number 1 in our coding since severity of this category is the lowest. Next category is possible injury which has the second level of severity. Numerical coding is used for software to be able to identify each categorical dependent variable correctly; however the difference between the categories may not

be the same. For example, the difference between PDOs and possible injury is less than the difference between incapacitating injuries and incapacitating injuries.

An index model for a single latent variable  $y^*$ :

$$y_i^* = x_i'\beta + u_i \tag{1}$$

Where  $y_i^*$  is unobservable meaning it is only detectable when it crosses thresholds.

$$y_i = j \quad \text{if} \quad \alpha_{j-1} < y_i^* \leq \alpha_j \tag{2}$$

In this case we would have latent continuous variable which would be formed into four groups with three thresholds. These are the cut-off points between the four categories.

The probability that observation i will select alternative j is:

$$p_{ij} = p(y_i = j) = p(\alpha_{j-1} < y_i^* \leq \alpha_j) = F(\alpha_j - x_i'\beta) - F(\alpha_{j-1} - x_i'\beta) \tag{3}$$

For ordered probit, F is the standard normal cdf,

For ordered logit model, F is the logistic cdf

$$F(z) = \frac{e^{z}}{(1+e^{z})}$$

The ordered logit/probit model with  $j$  alternatives will have one set of coefficients with  $(j-1)$  intercepts that represents  $\alpha$  as the cut-off points. Ordered choice models have multiple intercepts therefore the ordered models with  $j$  alternatives will have  $j$  sets of marginal effects.

Interpretation of coefficients in these models is that the sign of parameter indicates whether the latent variable  $\mathcal{V}^*$  increases with the regressor. The magnitude of coefficients will be different by a scale factor between the probit and logit models.

The marginal effect of an increase in a regressor  $x_r$  on the probability of selecting alternative  $j$  is

$$\partial p_{ij} / \partial x_{ri} = \{F'(\alpha_{j-1} - x_i'\beta) - F'(\alpha_j - x_i'\beta)\} \beta_r \quad (4)$$

The marginal effects of each variable on the different alternatives sums up to zero. Each unit increase in the dependent variable increase/decrease the probability of selecting alternative  $j$  by the marginal effect expressed as a percent.

## 8. Results

Before running the models, a few observations from the dataset were made to evaluate the nature of the urban truck at-fault crashes. Overall, these crashes have fewer fatalities compared to all drivers on Alabama roadways. This result was not surprising as professional drivers rely on their skill at driving to make a living, not as a mean to get to another profession. Additionally, there were very few DUI related crashes observed from the truck drivers. This is again due to the professionalism and the standards of blood alcohol level that professional drivers are subject to while operating a vehicle to retain their license.

Examining the models, from the independent variables only those that were statistically significant at a 0.05 level were retained for the model. Table 5 shows the parameter and standard error for the significant variables for the multinomial logit regression and table 6 shows the parameters and standard error for the significant variable for the multinomial probit regression.

**Table 5**  
*Multinomial Logit Regression Results*

Parameter	B	Std. Error
Primary Contributing Circumstance=Cargo fell/Defective Equipment	0.41	0.18
Primary Contributing Circumstance=Distraction	0.92	0.21
Primary Contributing Circumstance=Driving too Fast	1.55	0.19
Primary Contributing Circumstance=Fatigued/Asleep	1.58	0.32
Primary Contributing Circumstance=Improper Driving	0.48	0.12
Primary Contributing Circumstance=Misjudge Stopping Distance	0.65	0.14
Primary Contributing Circumstance=Swerved to Avoid	0.69	0.19
Manner of Crash=Causal Vehicle Backing	-1.16	0.47
Manner of Crash=Head on	1.24	0.38
Manner of Crash=Non-Collision	-0.63	0.22
Manner of Crash=Side Impact	0.55	0.12
Manner of Crash=Sideswipe	-0.67	0.12
Number of Lanes=Parking Lot	-0.69	0.31



**Table 6***Multinomial Probit Regression Results*

Parameter	B	Std. Error
Primary Contributing Circumstance=Distraction	0.52	0.12
Primary Contributing Circumstance=Driving too Fast	0.87	0.11
Primary Contributing Circumstance=Fatigued/Asleep	0.88	0.18
Primary Contributing Circumstance=Improper Driving	0.26	0.06
Primary Contributing Circumstance=Misjudge Stopping Distance	0.36	0.08
Primary Contributing Circumstance=Swerved to Avoid	0.38	0.11
Manner of Crash=Causal Vehicle Backing	-0.61	0.22
Manner of Crash=Head on	0.71	0.22
Manner of Crash=Non-Collision	-0.33	0.12
Manner of Crash=Side Impact	0.29	0.07
Manner of Crash=Sideswipe	-0.38	0.07
Number of Lanes=Four Lanes	0.09	0.04
Number of Lanes=Parking Lot	-0.32	0.15

**9. Discussion**

This study examined urban truck at-fault crashes between 2012 and 2016 in Alabama. The data were obtained from the (CARE database, 2017) of traffic crashes and the analysis performed included the development of statistical models to evaluate the variables that contributed to increases in crash severity. From the original input variables, several variables were removed from the results due to failure to meet the desired level of significance. Of the remaining variables, the following observations were made.

The values that are negative in Table 5 indicate situations that make the severity lower. For instance, causal vehicle backing up has a parameter value of -1.16. When the trucks are backing up, they are going slower therefore the severity would be lower. The second negative parameter is non-collision, like run off the road which is obviously less severe. Sideswipes generally don't have much impact on severity either. The last value from the negative parameters is parking lot crashes, which again is not contributing to

the severity because drivers generally are traveling at a low rate of speed.

Obviously, the positive parameter estimates on the variables indicate the situation is likely to increase the severity of a crash. The parameter estimate for category fatigue/asleep is 1.58 which is the most dangerous contributing circumstance. Driving fast for conditions is the second most important contributing factor for increasing the level of severity, with a parameter estimate of 1.55. After these two categories, parameter for distractions is equal to 0.92 which means driver distraction is a serious issue. Distracted driving seems to have a profound effect on motor vehicle crash severity. Analysis of manner of crash indicated head on collision has a high impact to increase severity which makes sense. Because the weight disparity between the truck and car is so great, it is not surprising that a truck-car crash in a head-on collision is likely to result in an injury or possible fatality. If a passenger car hits a truck head on, not only it is going to stop but it can cause the car to go backwards so the change in momentum

would go from forward to backwards. The change is going to make a rapid increase in severity. Fatigue also can be mixed well with head on collisions. There is a chance when a driver is tired, they drift to left and has a head on collision. They can also drift to the right and have a side impact or run off the road.

Examining the values in Table 6, almost all of the significant variables are the same with the last statistical model. This supports all of the analysis previously mentioned. Fatigued driving and driving too fast for conditions are the primary vehicle operating concerns that lead to higher severity crashes. Head-on collisions, likely to result from fatigued and distracted driving, tend to show large increases in severity.

In conclusion, this paper presents the statistical analysis of urban truck at-fault crashes analyzed using a logit and probit regression methodology. The analysis concluded that for truck drivers, driving too fast and driving while fatigued were the most likely actions that would lead to a greater severity of crash.

## References

- Abramson, H. 2015. *The Trucks Are Killing Us*. The New York Times. Available from internet: <<https://www.nytimes.com/2015/08/22/opinion/the-trucks-are-killing-us.html?mcubz=1>>.
- Aguero-Valverde, J. 2013. Full bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates, *Accident Analysis and Prevention* 50: 289-297.
- Amarasingha, N.; Dissanayake, S. 2013. Modeling frequency of truck crashes on limited-access highways, *Journal of the Transportation Research Forum* 52(3): 123-139.
- Caliendo, C.; De Guglielmo, M.; Guida, M. 2013. A crash-prediction model for road tunnels, *Accident Analysis and Prevention* 55: 107-115.
- Caliendo, C.; Guida, M.; Parisi, A. 2007. A crash-prediction model for multilane roads, *Accident Analysis and Prevention* 39(4): 657-670.
- CARE database. 2017. *Critical Analysis Reporting Environment (CARE)*. Center for Advanced Public Safety. The University of Alabama, Tuscaloosa, AL.
- Chen, C.; Xie, Y. 2016. Modeling the effects of AADT on predicting multiple-vehicle crashes at urban and suburban signalized intersections, *Accident Analysis and Prevention* 91: 72-83.
- Chiou, Y.; Sheng, Y.; Fu, C. 2017. Freeway crash frequency modeling under time-of-day distribution, *Transportation Research Procedia* 25: 664-676.
- Christoforou, Z.; Cohen, S.; Karlaftis, M. 2012. Integrating real-time traffic data in road safety analysis, *Procedia - Social and Behavioral Sciences* 48: 2454-2463.
- Dong, C.; Clarke, D.; Yan, X.; Khattak, A.; Huang, B. 2014. Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections, *Accident Analysis and Prevention* 70: 320-329.
- Duncan, C.; Khattak, A.; Forrest, C. 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions, *Transportation Research Record: Journal of the Transportation Research Board* 1635: 63-71.
- Gander, P. H.; Marshall, N. S.; James, I.; Le Quesne, L. 2006. Investigating driver fatigue in truck crashes: Trial of a systematic methodology, *Transportation Research Part F: Traffic Psychology and Behaviour* 9(1): 65-76. doi:10.1016/j.trf.2005.09.001.

- Golob, T.; Regan, A. 2003. Truck-involved crashes and traffic levels on urban freeways. *IDEAS Working Paper Series from RePEc*.
- Goodwin, L. C. 2003. Best practices for road weather management, version 2.0. FHWA road weather management program.
- Greene, W.H. 2000. *Econometric Analysis*. Prentice Hall, Englewood Cliffs, NJ.
- Hamdar, S.H.; Schorr, J. 2013. Interrupted versus uninterrupted flow: a safety propensity index for driver behavior, *Accident Analysis and Prevention* 55: 22-33.
- Hiselius, L. 2004. Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows, *Accident Analysis and Prevention* 36(6): 985-992.
- Jin, Y.; Wang, X.; Chen, X. 2010. Right-angle crash injury severity analysis using ordered probability models. In *Proceedings of the Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on*, 3: 206-209.
- Islam, S.; Jones, S. L.; Dye, D. 2014. Comprehensive analysis of single- and multi-vehicle large truck at-fault crashes on rural and urban roadways in Alabama, *Accident Analysis & Prevention* 67: 148-58. doi:10.1016/j.aap.2014.02.014.
- Joshua, S. C.; Garber, N. J. 1990. Estimating truck accident rate and involvements using linear and Poisson regression models, *Transportation Planning and Technology* 15(1): 41-58.
- Khattak, A.J.; Khattak, A.J.; Council, F. M. 2002. Effects of work zone presence on injury and non-injury crashes, *Accident Analysis and Prevention* 34(1): 19-29.
- Khattak, A. J.; Schneider, R.J.; Targa, F. 2003. Risk factors in large truck rollovers and injury severity: analysis of single-vehicle collisions. In *Proceedings of the 82nd Transportation Research Board annual meeting*, Washington, D.C.
- Khorashadi, A.; Niemeier, D.; Shankar, V.; Mannering, F. 2005. Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory analysis, *Accident Analysis & Prevention* 37: 910-921.
- Lao, Y.; Wu, Y.; Corey, J.; Wang, Y. 2011. Modeling animal-vehicle collisions using diagonal inflated bivariate Poisson regression, *Accident Analysis and Prevention* 43(1): 220-227.
- Large Truck and Bus Crash Facts. 2015. Federal Motor Carrier Safety Administration. Available from internet: <<https://www.fmcsa.dot.gov/safety/data-and-statistics/large-truck-and-bus-crash-facts-2015#A4>>.
- Lemp, J. D.; 2017. Large trucks. IIHS. Available from internet: <<http://www.iihs.org/iihs/topics/t/large-trucks/fatalityfacts/large-trucks>>.
- Lemp, J. D.; Kockelman, K. M.; Unnikrishnan, A. 2011. Analysis of large truck crash severity using heteroskedastic ordered probit models, *Accident Analysis & Prevention* 43(1): 370-380.
- Ma, T.; Haque, K. 2017. *Work zone crash performance data measurement*. Report for The Intermodal Freight Transportation Institute.
- Miaou, S. P. 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, *Accident Analysis and Prevention* 26(4): 471-482.

- Mohammadi, M.; Samaranayake, V.A.; Bham, G.H. 2014. Crash frequency modeling using negative binomial models: an application of generalized estimating equation to longitudinal data, *Analytic Methods in Accident Research* 2: 52-69.
- Mulinazzi, T.; Schrock, S.; Rescot, R.; Jasrotia, R.; Hovey, K.; Li, Y.; Pohlman, J. 2009. *Mitigating wind induced truck crashes*. Report for Mid-America Transportation Center (No. 25-1121-0001-263), University of Nebraska-Lincoln.
- Naik, B.; Tung, L.; Zhao, S.; Khattak, A. J. 2016. Weather impacts on single-vehicle truck crash injury severity, *Journal of Safety Research* 58: 57-65. doi:10.1016/j.jsr.2016.06.005.
- Osman, M.; Paleti, R.; Mishra, S.; Golias, M. M. 2016. Analysis of injury severity of large truck crashes in work zones, *Accident Analysis & Prevention* 97: 261-273. doi:10.1016/j.aap.2016.10.020.
- Pei, X.; Wong, S. C.; Sze, N. N. 2011. A joint-probability approach to crash prediction models, *Accident Analysis and Prevention* 43(3): 1160-1166.
- Qin, X.; Ivan, J.N.; Ravishanker, N. 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments, *Accident Analysis and Prevention* 36(2): 183-191.
- Roshandeh, A.M.; Agbelie, B.R.D.K.; Lee, Y. 2016. Statistical modeling of total crash frequency at highway intersections, *Journal of Traffic and Transportation Engineering (English Edition)* 3(2): 166-171.
- Sarker, A.A.; Naimi, A.; Mishra, S.; Golias, M.M.; Freeze, P.B. 2015. Development of a secondary crash identification algorithm and occurrence pattern determination in large scale multi-facility transportation network, *Transportation Research Part C* 60: 142-160.
- Wang, K.; Ivan, J.N.; Ravishanker, N.; Jackson, E. 2017. Multivariate Poisson lognormal modeling of crashes by type and severity on rural two lane highways, *Accident Analysis and Prevention* 99: 6-19.
- Yan, K.; Qu, X.; Yan, Y. 2017. Will higher traffic flow lead to more traffic conflicts? A crash surrogate metric based analysis, *PLoS ONE* 12(8): e0182458.
- Yang, H.; Ozbay, K.; Ozturk, O.; Yildirimoglu, M. 2013. Modeling work zone crash frequency by quantifying measurement errors in work zone length, *Accident Analysis and Prevention* 55: 192-201.
- Yu, R.; Abdel-Aty, M. 2013. Developing safety performance functions for a mountainous freeway. In *the Proceedings of the 16th International Conference Road Safety on Four Continents*, 15-17 May, Beijing, China.
- Zhu, X.; Srinivasan, S. 2011. A comprehensive analysis of factors influencing the injury severity of large-truck crashes, *Accident Analysis & Prevention* 43(1): 49-57. doi:10.1016/j.aap.2010.07.007.
- Zou, W.; Wang, X.; Zhang, D. 2017. Truck crash severity in New York city: An investigation of the spatial and the time of day effects, *Accident Analysis & Prevention* 99: 249-261. doi:10.1016/j.aap.2016.11.024.