

An Ordered Fractional Split Approach for Aggregate Injury Severity Modeling

Shamsunnahar Yasmin

Department of Civil, Environmental & Construction Engineering
University of Central Florida
Email: shamsunnahar.yasmin@mail.mcgill.ca

Naveen Eluru*

Associate Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: naveen.eluru@ucf.edu

Jaeyoung Lee

Department of Civil, Environmental & Construction Engineering
University of Central Florida
Email: jaeyoung.lee@ucf.edu

Mohamed A. Abdel-Aty

Professor
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: (407) 823-4535, Fax: 407-823-3315
Email: m.aty@ucf.edu

November 15, 2015

ABSTRACT

In crash frequency models, frequency by severity level are examined using multivariate count models. In these multivariate approaches the impact of exogenous variables is quantified through the propensity component of count models. The main interaction among variables across different severity levels is sought through unobserved effects i.e. there is no interaction of observed effects across the multiple count models. While this might not be a limitation per se, it might be beneficial to evaluate the impact of exogenous variables in a framework that directly relates a single exogenous variable to all severity count variables simultaneously. Towards this end, an alternative approach to examine crash frequency by severity is proposed. Specifically, as opposed to modeling the number of crashes, we adopt a fractional split modeling approach, to study the fraction of crashes by each severity level on a road segment. Given the ordered nature of injury severity, we employ an ordered probit fractional split model to study crash proportion by severity levels. The model is estimated for roadways segment data for single vehicle and multi vehicle crashes of Florida for the year 2009 through 2011. The model estimation results clearly highlight the importance of traffic volume, lane width, shoulder width, proportion of divided segments, and speed limit on crash proportion by severity. The model results are employed to predict hot spots for different crash types. The results clearly highlight how the ordered probit fractional split models can be employed for highway safety screening purposes.

Keywords: ordered probit fractional split model, crash frequency proportion by severity, single vehicle crashes and multiple vehicle crashes

INTRODUCTION

Road traffic crashes and their consequences such as injuries and fatalities are acknowledged to be a serious global health concern. In the United States (US), motor vehicle crashes are responsible for more than 90 deaths per day (1). Moreover, these crashes cost the society \$230.6 billion annually (2). There is a need for continued efforts to identify remedial measures to reduce crash occurrence and crash consequences. Traditionally, the transportation safety literature has evolved along two major streams: crash frequency analysis and crash severity analysis. Crash frequency or crash prediction analysis is focused on identifying attributes that result in traffic crashes and propose effective countermeasure to improve the roadway design and operational attributes (see (3) for a review of these studies). The crash frequency models study aggregate information; such as total number of crashes at an intersection or at a spatial aggregation level (zone or tract level). On the other hand, crash severity analysis is focused on examining crash events, identifying factors that impact the crash outcome and providing recommendations to reduce the consequences in the unfortunate event (injuries and fatalities) of a traffic crash (see (4-5) for a review). The crash severity models are quite disaggregate in nature because these consider every crash as a record for model development.

In safety planning, occasionally it might be useful to understand the proportion of crashes by severity type on roadway segments. A possible approach to achieve this is to develop count models for each severity type and then use the predicted values to identify proportion of severe crashes. While the approach is feasible, the count prediction by severity seems disjoint i.e. the observed components of the count variables do not interact. Moreover, any changes to observed variables does not directly affect proportion of crashes by severity. The impact of exogenous variables affects counts and these in turn will affect the proportion of crashes by severity. On the other hand, adopting disaggregate severity models for such analysis would require us to aggregate the findings to arrive at proportion of severe crashes. In this study, we propose an alternative approach based on modeling proportions directly as dependent variables. Hence, the impact of exogenous variables directly affects proportions by severity (not counts or severity in a single crash).

EARLIER RESEARCH AND CURRENT STUDY IN CONTEXT

Crash Frequency Literature

Researchers have predominantly examined total number of transportation related crash events either at the micro- (such as intersection and segment) or at the macro-level (such as zone, county, census tract) for different road user groups (vehicle, pedestrian and bicyclist) (for example see (6-8) for macro-level studies; and see (9-10) for micro-level study). However, crash count data are often compiled by injury severity outcomes (for example: no injury, minor injury, major injury, and fatal injury crashes). Researchers (11) have argued that it is important to examine crash frequency by severity levels as it would play significant role in model implications. To that extent, a number of studies have developed independent crash prediction models for different injury severity levels.

Among different crash severity outcomes, considerable research has been carried out for examining fatal crash counts (12-21). Several studies have also explored critical factors contributing to fatal/serious injury crash counts (22-28). Crash count events were also studied for injury crash outcome level by a number of researchers (12, 16, 21, 29-31). Moreover, serious injury (15, 20) and slight injury (20, 22) crash counts were studied in few studies, whilst property damage only/no injury crash counts (16) had been examined to a lesser extent.

In examining crash counts by severity levels, statistical approach has generally included the Negative binomial regression model (12-14, 16-17, 20, 22, 24). Among other statistical approaches, researchers have employed Generalized linear modeling techniques (28), Ordinary least square regression (29), Poisson-lognormal (32), Generalized Poisson regression (21), Negative multinomial regression (27), Random effect Negative binomial (12), Geographically weighted Poisson regression (19, 33), Geographically weighted Negative binomial regression (23), Bayesian Poisson Lognormal (7), Quasi induced exposure method (18) and Bayesian spatial regression model (30, 23).

Crash Severity Literature

A number of research efforts have examined crash injury severity to gain a comprehensive understanding of the factors that affect injury severity at a disaggregate crash or individual level. It is beyond the scope of the paper to review all the research on transportation crash severity analysis. For a detailed review of modeling frameworks employed in crash severity analysis, the reader is referred to earlier research (4-5, 34).

In general, many earlier studies have employed the logistic regression model (for example see 35-36) to identify the contributing factors of crash severity. In traffic crash reporting, injury severity is typically characterized as an ordered variable (for example: no injury, minor injury, serious injury and fatal injury). It is no surprise that the most commonly employed statistical framework in modeling crash injury severity is the ordered outcome models (ordered logit or probit) (37-40). Researchers have also employed unordered choice models to study injury severity due to additional flexibility offered by these frameworks. Specifically, the unordered systems allow for the estimation of alternative specific variable impacts while the ordered systems impose a uni-directional impact of the exogenous variable on injury severity alternatives. The most prevalent unordered outcome structure considered is the multinomial logit model (41-45). However, the unordered model does not recognize the inherent ordering of the crash severity outcome and therefore, it neglects vital information present in the data. More recently a generalized ordered framework that allows for alternative specific impacts within an ordered regime has been employed to study injury severity (5, 46-47). These studies have concluded that the generalized ordered variants (also referred to as partial proportional odds models) perform as well, if not better than the corresponding unordered models.

Bridging the Gap

More recently, the research in transportation safety has focused on bridging the gap between crash frequency models and crash severity models. Specifically, researchers are examining crash frequency by severity levels while recognizing that for the same observation record, crash

frequencies by different severity levels are likely to be dependent. Hence, as opposed to adopting the univariate crash frequency models as earlier, researchers have developed multivariate crash frequency models (7, 48-51). These studies have argued that crash counts across different crash severity levels share unobservable or omitted variables and are hence fundamentally multivariate in nature (49). Ignoring such correlations, if present, may result in biased parameter estimates and thus lead to inefficient policy implications (48).

In all of these joint approaches that study frequency and severity, the impact of exogenous variables is quantified through the propensity component of count models. The main interaction across different severity levels variables is sought through unobserved effects (studies discussed above) i.e. there is no interaction of observed effects across the multiple count models. While this might not be a limitation per se, it might be beneficial to evaluate the impact of exogenous variables in framework that directly relates a single exogenous variable to all severity count variables simultaneously i.e. a framework where the observed propensities of crashes are examined by severity level directly. In the traditional count modeling approaches this is not feasible.

Current Study

In this study, an alternative approach to examine crash frequency by severity is proposed. Specifically, as opposed to modeling the number of crashes, we adopt a fractional split modeling approach, to study the fraction of crashes by each severity level on a road segment. So for example, in a five severity count case (KABCO; fatal (K), incapacitating (A), non-incapacitating (B), possible injury (C), and property damage only (O)), the traditional approach would be to adopt a multivariate count model framework with five count equations. In the proposed approach, we adopt a fractional split model that examines the proportion of crashes (not frequency) by severity in a single probabilistic model system. In the case of five severity levels the dependent variable would be represented as proportions (number of specific crash level/total number of all crashes) as follows: (1) proportion of property damage only crashes, (2) proportion of minor injury crashes, (3) proportion of non-incapacitating injury crashes, (4) proportion of incapacitating injury crashes and (5) proportion of fatal crashes. For example, the dependent variable could take the following form – O: 0.45, C: 0.25, B: 0.15, A: 0.10 and K: 0.05.

The reader would note that the discretization of the variable as proportions does not lend itself to any traditional discrete modeling approaches because unlike the discrete modeling approaches where only one of the possible alternatives are chosen, in the crash proportion form we have possible non-zero values (ranging between 0 and 1 for each category) for multiple categories. In econometrics, Papke and Wooldridge (52) proposed a quasi-likelihood estimation method for binary probit model with a fractional dependent variable. The authors explored 401(K) plan participation rates in two portfolios using their proposed method. However, the approach is suitable only for two alternative proportions. The approach was extended to multinomial fractional model by Sivakumar and Bhat (53). The authors analyzed statewide interregional commodity-flow volumes in Texas using the proposed model. To be sure, the multinomial fractional approach has also been employed in safety literature. Milton et al. (54) developed a mixed multinomial fractional split model to study injury-severity distribution of crashes on highway segments by using highway-injury data from Washington State. The approach while allows for more than two alternatives

inherently ignores the relation between severity levels – the inherent ordering within severity level (from no injury to fatal).

A more appropriate polychotomous model would be an ordered extension of the Papke and Wooldridge (52) model. Eluru et al. (55) proposed a panel mixed ordered version that not only allows the analysis of proportion for variables with more than 2 alternatives but also recognizes the inherent ordering in the severity. Given the ordered nature of injury severity, we adopt this approach to study crash proportion by severity levels. Similar to the traditional ordered outcome model, a latent propensity is computed for each road segment with higher propensity indicating higher likelihood for the proportion of severe injury categories. Thus, in this model exogenous variables affect severity proportion through a single equation thus allowing us to obtain a parsimonious specification of exogenous variable impacts. The reader would note that if severity proportion were computed using count models, we needed to estimate as many equations as severity levels thus requiring us to estimate a large number of model parameters. To summarize, in this research we employ a road segment level ordered probit fractional split model to investigate the impact of exogenous factors on the proportion of crashes by severity in Florida. In the context of crash severity, we examine single vehicle crashes, and multivehicle crashes by crash type (head-on, rear-end, angular and sideswipe).

METHODOLOGY

The formulation for the Ordered Probit Fractional Split model (OPFS) for modeling the proportion of crashes by severity is presented in this section. The reader would note that conventional maximum likelihood approaches are not suited for fractional proportion models. Hence, we resort to a quasi-likelihood approach (proposed by 52-53, 56). The proposed approach is the ordered response extension of the binary probit model proposed by Papke and Woolridge (52).

Model Structure

Let q ($q = 1, 2, \dots, Q$) be an index to represent road segment, and let k ($k = 1, 2, 3, \dots, K$) be an index to represent severity category. The latent propensity equation for severity category at the q th site:

$$y_q^* = \alpha' z_q + \xi_q, \quad (1)$$

This latent propensity y_q^* is mapped to the actual severity category proportion y_{qk} by the ψ thresholds ($\psi_0 = -\infty$ and $\psi_K = \infty$). z_q is an ($L \times 1$) column vector of attributes (not including a constant) that influences the propensity associated with severity category. α is a corresponding ($L \times 1$)-column vector of mean effects. ξ_q is an idiosyncratic random error term assumed to be identically and independently standard normal distributed across segments q .

Model Estimation

The model cannot be estimated using conventional Maximum likelihood approaches. Hence we resort to quasi-likelihood based approach for our methodology. The parameters to be estimated in the Equation (1) are α , and ψ thresholds. To estimate the parameter vector, we assume that

$$E(y_{qk} | z_{qk}) = H_{qk}(\alpha, \psi), 0 \leq H_{qk} \leq 1, \sum_{k=1}^K H_{qk} = 1 \quad (2)$$

H_{qk} in our model takes the ordered probit probability (P_{qk}) form for severity category k defined as

$$P_{qk} = \{ G[\psi_k - \alpha'_q z_q] - G[\psi_{k-1} - \alpha'_q z_q] \} \quad (3)$$

The proposed model ensures that the proportion for each severity category is between 0 and 1 (including the limits). Then, the quasi-likelihood function (see (52) for a discussion on asymptotic properties of quasi-likelihood proposed), for a given value of δ_q vector may be written for site q as:

$$L_q(\alpha, \psi) = \prod_{k=1}^K \{ G[\psi_k - \alpha'_q z_q] - G[\psi_{k-1} - \alpha'_q z_q] \}^{d_{qk}} \quad (4)$$

where $G(\cdot)$ is the cumulative distribution of the standard normal distribution and d_{qk} is the proportion of crashes in severity category k . The model estimation is undertaken using routines programmed in Gauss matrix programming language.

After the model has been estimated, the model prediction can be undertaken based on the final convergence estimates. The approach is simpler than the approach required for the prediction of ordered outcome models. To elaborate, in the fractional split model, the probability computed is used as the proportion value directly for the severity category.

DATA PREPARATION AND DESCRIPTIVES

Crash, traffic, and roadway data used in this study were collected from multilane highway segments in Florida for the period 2009 through 2011. Crashes are classified by injury severity levels such as fatal (K), incapacitating (A), non-incapacitating (B), possible injury (C), and property damage only (O) crashes. The collected crash data are further classified by collision types. Crashes are firstly divided into single-vehicle (SV) and multiple-vehicle (MV) crashes. Then MV crashes are further classified as head-on, rear-end, angular and sideswipe collision type. The dependent variable proportions and sample size for each collision type are presented in Table 1. From the Table we can observe that head-on collision has the highest proportion of fatal crashes followed by SV and angular crashes. On the other hand, sideswipe collision has the highest proportion of no injury outcome.

The acquired traffic data include AADT (Annual average daily traffic), K-factor (i.e., the 30th highest hourly volume of the year expressed as a percentage of the AADT), D-factor (i.e., percentage of traffic moving in the peak travel direction), and T-factor (i.e., percentage of the

AADT volume generated by trucks or commercial vehicles). The collected roadway data consists of lane width, shoulder width, posted speed limit, and median division. The crash data were aggregated by segments and weighted average by segment length of candidate explanatory variables is computed. Table 2 provides a summary of explanatory variables used in the study.

TABLE 1 Severity Proportions

Crash Type	Property Damage Only	Minor Injury	Non-incapacitating Injury	Incapacitating injury	Fatal injury	Sample Size
Single Vehicle	0.406	0.208	0.228	0.135	0.023	124
Multi-Vehicle						
Head-on	0.261	0.292	0.197	0.173	0.076	59
Rear-end	0.427	0.322	0.205	0.046	0.001	126
Angular	0.521	0.254	0.157	0.065	0.004	114
Sideswipe	0.794	0.082	0.077	0.046	0.000	100

TABLE 2 Descriptive Statistics of the Processed Traffic and Roadway Data

Variable	Description	Mean	Std. dev	Minimum	Maximum
w_aadt	Average AADT weighted by segment length	22618.01	11379.61	2500	50000
w_kfctr	Average K-factor weighted by segment length	8.998	0.354	7.50	9.50
w_dfctr	Average D-factor weighted by segment length	58.565	8.940	50.80	99.90
w_tfctr	Average T-factor weighted by segment length	5.421	3.591	1.00	20.75
length	Segment length (sum of segment length)	5.756	6.471	0.143	33.585
w_lw	Average lane width weighted by segment length	11.857	0.433	10	13
w_sw	Average shoulder width weighted by segment length	4.101	1.811	1.5	10
p_div	Proportion of divided segment (opposed to undivided)	0.946	0.192	0.000	1.000
w_speed	Average speed limit width weighted by segment length	46.135	7.929	30	65

RESULTS

The effects of exogenous variables in model specifications are discussed in this section. To reiterate, we estimated five different OPFS models: one model for SV crashes and four different models for MV crashes (head-on, rear-end, angular and sideswipe collisions). In OPFS models,

the positive (negative) coefficient corresponds to increased (decreased) proportion for severe injury categories. The final specification of the model was based on removing the statistically insignificant variables in a systematic process based on statistical significance and intuitive coefficient effect. In some cases, parameters with a statistical significance up to 70% were retained given the small sample sizes in our data (ranging from 59 through 126). In estimating the models, several functional forms and variable specifications are explored. The functional form that provided the best result is used for the final model specifications.

SV Crash Model

The coefficients in Table 3 represent the estimation results of SV crash model. The threshold parameters identify the demarcation points between severity categories and have no substantial interpretation.

TABLE 3 Estimates for Single Vehicle Crash Proportion by Severity Model

Explanatory Variables	Coefficient (t-stat)
<i>Threshold Parameters</i>	
Threshold between property damage only and minor injury	-0.011 (-0.106)
Threshold between minor and non-incapacitating injury	0.533 (5.226)
Threshold between non-incapacitating and incapacitating injury	1.269 (11.367)
Threshold between incapacitating and fatal injury	2.299 (11.164)
<i>Weighted average AADT (Base: Weighted average AADT >10000)</i>	
Weighted average AADT <10000	-0.403 (-2.125)
<i>Weighted average lane width (Base: Weighted average lane width ≥ 12 ft)</i>	
Weighted average lane width less than 12 ft	0.207 (1.926)
<i>Weighted average shoulder width (Base: Weighted average shoulder width ≥ 3 ft)</i>	
Weighted average shoulder width <3 ft	0.164 (1.186)
<i>Weighted average speed (base: Weighted average speed ≤ 50)</i>	
Weighted average speed >50	0.468 (3.607)
<i>Interaction terms</i>	
Weighted average speed $\leq 40 \times$ Weighted average AADT <20000	-0.340 (-1.466)
Weighted average shoulder width greater than 5 ft $\times$ Weighted average AADT <20000	0.542 (2.268)

With respect to traffic volume, lower weighted average AADT (weighted average AADT <10000) categories are found to be associated with lower proportion of severe crash injury outcomes relative to higher weighted average AADT category (weighted average AADT $\geq$ 10000).

As is expected, for narrower lanes (weighted average lane width less than 12 ft), proportion of higher injury severity levels are found to be higher in SV crash events relative to SV crashes on wider lane (Weighted average lane width $\geq$ 12 ft). Narrow shoulder width is also found to be positively associated with higher proportion of SV crashes. We found that severe SV crashes are higher in the locations with narrow shoulder width (weighted average shoulder width <3 ft) compared to road sections with wider shoulder width. It is speculated that narrower lanes or shoulders may provide less space and less scope for error correction in the event of an impending crash, which in turn may result in more severe SV crashes (for instance run-off-road crashes) (57-58). As expected, the indicator variable representing higher speed limit (weighted average speed >50) increases the proportion of severe SV crashes compared to lower speed limit locations.

From the result of interaction terms, our model estimation results reveal that the effect of lower volume (weighted average AADT <20000) in a lower speed limit location (weighted average speed $\leq$ 40) increases the proportion of property damage only SV crashes. On the other hand, roadways with lower volume (weighted average AADT <20000) and wider shoulder width (weighted average shoulder width > 5 ft) increase the likelihood of more severe SV crashes. This is possible due to the sense of false safety that drivers feel under low traffic conditions.

MV Crash Model

The coefficients in Table 4 represent the estimation results of MV crash models. In terms of traffic volume, MV model results suggest that the impact of weighted average AADT variables vary across different MV collision types. Increase in weighted average AADT increases the likelihood of more severe rear-end crashes. On the other hand, lower AADT (weighted average AADT <20000) has a positive association with more severe angular crashes. The results related to rear-end collision is perhaps indicating lower headways in higher traffic volume. On the other hand, speeding driver behavior during low volume condition may result in more severe angular collision. Sideswipe collision results reveal lower proportion of severe crash outcomes for higher volume indicator (weighted average AADT 20000-30000 and weighted average AADT >30000) relative to lower volume (weighted average AADT <20000) condition. Result related to weighted average T-factor indicates an increase in severe crash proportions for rear-end collision.

TABLE 4 Estimates for Multi-vehicle Crash Proportion by Severity Model

Explanatory Variables	Collision Types			
	Head-on	Rear-end	Angular	Sideswipe
	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)	Coefficient (t-stat)
<i>Threshold Parameters</i>				
Threshold between property damage only and minor injury	1.503 (1.027)	2.260 (1.376)	0.228 (2.888)	1.565 (3.039)

Threshold between minor and non-incapacitating injury	2.315 (1.595)	3.122 (1.910)	0.944 (10.115)	1.915 (3.680)
Threshold between non-incapacitating and incapacitating injury	2.888 (1.963)	4.139 (2.500)	1.682 (15.367)	2.475 (4.501)
Threshold between incapacitating and fatal injury	3.697 (2.509)	5.500 (3.345)	2.915 (15.640)	6.384 (9.214)
<i>Weighted average AADT</i>				
<i>Weighted average AADT/1000</i>	—	0.098 (1.822)	—	—
Weighted average AADT <20000	—	—	0.239 (1.675)	—
Weighted average AADT 20000-30000	—	—	—	-0.456 (-1.902)
Weighted average AADT >30000	—	—	—	-0.491 (-2.486)
<i>Weighted average T-factor</i>	—	0.276 (1.035)	—	—
<i>Weighted average shoulder width</i>				
Weighted average shoulder width <3 ft	—	—	—	0.303 (1.208)
Weighted average shoulder width 3-5 ft	—	—	—	—
Weighted average shoulder width > 5 ft	—	—	0.344 (1.892)	—
<i>Weighted average lane width</i>	—	0.177 (1.255)	—	—
<i>Proportion of divided segments (opposed to undivided)</i>	1.904 (1.277)	—	—	1.001 (1.848)
<i>Weighted average speed</i>				
Weighted average speed ≤50	—	—	—	—
Weighted average speed >50	0.622 (2.379)	—	0.226 (1.810)	—
<i>Interaction terms</i>				
Weighted average speed ≤40×Weighted average AADT <30000	0.663 (1.779)	—	—	—
Weighted average shoulder width greater than 5 ft ×Weighted average AADT <20000	—	0.276 (1.035)	-1.078 (-3.722)	—

Weighted average shoulder width has no significant impact on head-on and rear-end crashes. From the model estimates, we found that in the presence of wider shoulder (weighted average shoulder width > 5 ft) on roadways, the possibility of more severe angular collision

increases. On the contrary, narrow shoulder width (weighted average shoulder width <3 ft) on roadway sections increase the possibility of more severe sideswipe crashes. In the presence of narrower shoulder drivers presumably show unsafe behaviour by shifting towards left most side of the lane (59) and thereby increases the possibility of more severe sideswipe crashes.

Weighted average lane width result has significant impact in rear-end collision model only. We found that, for rear-end collision, the likelihood of more severe crashes increases in the presence of wider lane width (see 60 for similar results). It is possible that drivers are less conscious of vehicles in the presence of wider lanes resulting in more rear-end crashes. Higher proportion of divided segment increases the possibility of more severe head-on and sideswipe collision, with greater impact in head-on collision followed by sideswipe collision. In terms of weighted average speed limit, higher speed limit (weighted average speed >50) indicator (relative to lower speed limit) has positive impact on the proportion of both head-on and angular collisions. It is interesting to note that speed limit does not influence rear-end and sideswipe collision proportions.

Among interaction terms, head-on collision model reveals a positive impact of weighted average speed ≤ 40 in a low traffic volume condition (weighted average AADT <30000) on severe crash proportion. Finally, the interaction term representing wider shoulder width (weighted average shoulder width > 5 ft) and weighted average AADT <20000 increases the probability of more severe rear-end crashes while reduces the possibility of more severe angular collision.

MODEL PREDICTION

The model results from the previous section clearly highlight the value of the proposed OPFS model for severity based crash proportion modeling. To further illustrate the model applicability, we employ the model results to undertake a prediction exercise.

The proposed measure is analogous to the Highway Safety Manual (HSM) (61) performance measure to identify crash hotspots. The HSM approach employs Excess Predicted Average Crash Frequency using Safety Performance Functions. The measure is calculated by subtracting the predicted crash frequency from the observed crash frequency. When the excess predicted average crash frequency is greater than zero, a segment experiences more traffic crashes than predicted. On the other hand, when the excess predicted average crash frequency is less than zero, a segment experiences fewer traffic crashes than predicted. Similar to this method, we propose the Excess Predicted Proportion (EPP) for a screening performance measure, which is the difference between the observed and predicted proportion of each severity for a segment.

$$EPP_{qk} = P(obs)_{qk} - P(prd)_{qk} \quad (5)$$

where, EPP_{qk} is the Excess Predicted Proportion of severity k for segment q . $P(obs)_{qk}$ is the observed proportion of crash severity; and $P(prd)_{qk}$ is the predicted crash proportion for severity k estimated from the OPFS model (Table 3 and 4). When EPP exceeds zero, the proportion for that segment is higher than predicted. In contrast, when EPP is smaller than zero, the proportion for that segment is lower than predicted. The EPP approach is slightly different from the earlier count based approaches. Because we deal with proportions that add up to 1 for a segment, a positive EPP for one severity automatically causes a negative EPP for at least another severity. Hence, directly identifying the segments with positive EPP as hotspots will not be appropriate. Because every

segment will be a hotspot for at least one severity (unless EPP is exactly 0 for all modes). Hence, after computing the measure across the entire sample for all severities, a severity specific hotspot (H) is identified based on the top 10 percentile ranking of segments for that mode. The other segments are labeled as normal (N). To be sure, the reader would note that 10 percentile ranking is for illustration purposes and any appropriate cut-off can be chosen by the analyst for analysis.

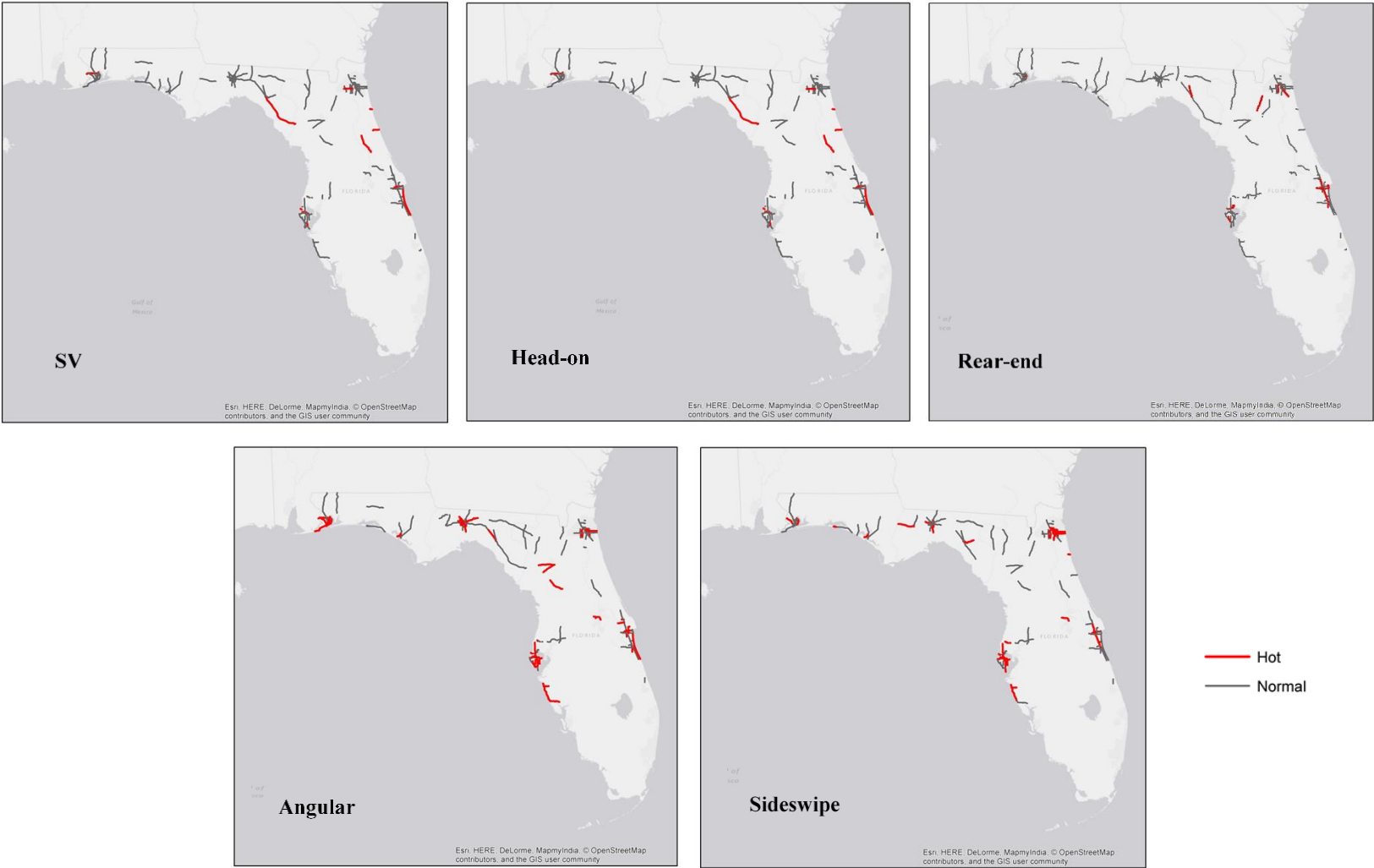
The EPP measure computation is undertaken for all severity categories in the model simultaneously. We presented the hotspots for fatal injury category in the Figure 1. From the spatial illustration we can see that hotspots for fatal crash are dispersed throughout the state. However, there are some clustering of hotspots for fatal crash categories specifically for angular and sideswipe collision.

CONCLUSIONS

Traditionally, the transportation safety literature has evolved along two major streams: crash frequency analysis and crash severity analysis. Crash frequency analysis is focused on identifying attributes that result in traffic crashes and propose effective countermeasure to improve the roadway design and operational attributes. On the other hand, crash severity analysis is focused on examining crash events, identifies factors that impact the crash outcome and provides solution to reduce the consequences in the unfortunate event (injuries and fatalities) of a traffic crash. More recently, the research in transportation safety has focused on bridging the gap between crash frequency models and crash severity models. Specifically, researchers are examining crash frequency levels by severity while recognizing that for the same observation record, crash frequencies by different levels of severity are likely to be dependent. Hence, as opposed to adopting the univariate crash frequency models as earlier, researchers have developed multivariate crash frequency models.

In multivariate approaches that study frequency and severity, the impact of exogenous variables is quantified through the propensity component of count models. The main interaction across different severity levels variables is sought through unobserved effects i.e. there is no interaction of observed effects across the multiple count models. While this might not be a limitation per se, it might be beneficial to evaluate the impact of exogenous variables in framework that directly relates a single exogenous variable to all severity count variables simultaneously i.e. a framework where the observed propensities of crashes by severity level are modeled directly, while also recognizing the inherent ordering of crash severity outcomes.

FIGURE 1 Spatial Representation of Hotspots for Fatal Injury Crashes (SV, Head-on, Rear-end, Angular and Sideswipe crashes)



Towards this end, we develop an alternative approach to examine crash frequency by severity. We employ a road segment level ordered probit fractional split model to investigate the impact of exogenous factors on the proportion of crashes by severity in Florida. In the context of crash severity, we examine single vehicle crashes, and multivehicle crashes by crash type (head-on, rear-end, angular and sideswipe). Crash, traffic, and roadway data used in this study were collected from multilane highway segments in Florida for the period 2009 through 2011. The model estimation results clearly highlight the importance of traffic volume, lane width, shoulder width, proportion of divided segments, and speed limit on crash proportion by severity. To further illustrate model applicability, we propose an Excess Predicted Proportion (EPP) metric for screening purposes. After computing the measure across the entire sample for all severities, a severity specific hotspot (H) is identified based on the top 10 percentile ranking of segments by severity.

The paper is not without limitations. In examining such aggregate crash event proportion across different crash outcome level, unobserved heterogeneity may play a significant role due to the omitted or unobserved variables. Due to the presence of such unobserved information, the effect of exogenous variables might not be the same across segments in the event of crash. However, in our analysis, we have not considered the effect of unobserved heterogeneity in examining the critical factors due to small sample sizes available for the analysis. It would be useful to consider such unobserved heterogeneity using mixed version of the ordered fractional split model in datasets with adequate number of observations. It would also be useful to compare the model performance from the proposed approach with crash frequency models (univariate and multivariate). Further, it might also be useful to study the influence of under-reporting for property damage only crashes on the proposed approach. In conclusion, the idea of the proposed approach is not to replace crash frequency methods but to augment the inventory of crash frequency models with an alternative approach.

ACKNOWLEDGMENTS

The authors wish to acknowledge the financial support of NCHRP 17-62 and the team represented by Drs. John Ivan, Bhagwant Persaud, and Raghavan Srinivasan. All opinions are those only of the authors. The authors also appreciate feedback from five anonymous reviewers on a previous version of the paper.

REFERENCES

1. NHTSA. *Traffic Safety Facts 2013: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*. Publication DOT HS 812 139. National Highway Traffic Safety Administration (NHTSA), U.S. Department of Transportation, 2015.
2. GHSA. *A 10-Point Plan for Reauthorizing the Federal Highway Safety Program*. Report No. 202-789-0942, Governors Highway Safety Association (GHSA), 2009.
3. Lord, D., and F. L. Mannering. The Statistical Analysis of Crash-frequency Data: A Review and Assessment of Methodological Alternatives. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 5, 2010, pp. 291-305.

- 1 4. Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. The Statistical Analysis of
2 Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives.
3 *Accident Analysis and Prevention*, Vol. 43, No. 5, 2011, pp. 1666-1676.
- 4 5. Yasmin. S., and N. Eluru. Evaluating Alternate Discrete Outcome Frameworks for Modeling
5 Crash Injury Severity. *Accident Analysis and Prevention*, Vol. 59, 2013, pp. 506-521.
- 6 6. Ukkusuri, S., L. F. Miranda-Moreno, G. Ramadurai, and J. Isa-Tavarez. The Role of Built
7 Environment on Pedestrian Crash Frequency. *Safety Science*, Vol. 50, No. 4, 2012, pp. 1141-
8 1151.
- 9 7. Lee, J., M. Abdel-Aty, and X. Jiang. Multivariate Crash Modeling for Motor Vehicle and
10 Non-Motorized Modes at the Macroscopic Level. *Accident Analysis and Prevention*, Vol. 78,
11 2015, pp. 146-154.
- 12 8. Wei, F., and G. Lovegrove. An Empirical Tool to Evaluate the Safety of Cyclists: Community
13 Based, Macro-Level Collision Prediction Models Using Negative Binomial Regression.
14 *Accident Analysis and Prevention*, Vol. 61, 2013, pp. 129-137.
- 15 9. Dong, C., D. B. Clarke, S. H. Richards, and B. Huang. Differences in Passenger Car and Large
16 Truck Involved Crash Frequencies at Urban Signalized Intersections: An Exploratory
17 Analysis. *Accident Analysis and Prevention*, Vol. 62, 2014, pp. 87-94.
- 18 10. Loo, B. P., and K. Tsui. Bicycle Crash Casualties in a Highly Motorized City. *Accident*
19 *Analysis and Prevention*, Vol. 42, No. 6, 2010, pp. 1902-1907.
- 20 11. Wang, C., M. A. Quddus, and S. G. Ison. Predicting Accident Frequency at Their Severity
21 Levels and Its Application in Site Ranking Using a Two-Stage Mixed Multivariate Model.
22 *Accident Analysis and Prevention*, Vol. 43, No. 6, 2011, pp. 1979-1990.
- 23 12. Noland, R.B., Traffic Fatalities and Injuries: the Effect of Changes in Infrastructure and Other
24 Trends. *Accident Analysis and Prevention*, Vol. 35, No. 4, 2003, pp 599-611.
- 25 13. Noland, R. B., and M. A. Quddus. A Spatially Disaggregate Analysis of Road Casualties in
26 England. *Accident Analysis and Prevention*, Vol. 36, No. 6, 2004, pp. 973-984.
- 27 14. Amoros, E., J. L. Martin, and B. Laumon. Comparison of Road Crashes Incidence and
28 Severity between Some French Counties. *Accident Analysis and Prevention*, Vol. 35, No. 4,
29 2003, pp. 537-547.
- 30 15. Quddus, M. A. Modelling Area-Wide Count Outcomes with Spatial Correlation and
31 Heterogeneity: An Analysis of London Crash Data. *Accident Analysis and Prevention*, Vol.
32 40, No. 4, 2008, pp. 1486-1497.
- 33 16. Naderan, A., and J. Shahi. Aggregate Crash Prediction Models: Introducing Crash Generation
34 Concept. *Accident Analysis and Prevention*, Vol. 42, No. 1, 2010, pp. 339-346.
- 35 17. Ng, K.-S., W.-T. Hung, and W.-G. Wong. An Algorithm for Assessing the Risk of Traffic
36 Accident. *Journal of Safety Research*, Vol, 33, No. 3, 2002, pp. 387-410.
- 37 18. Stamatiadis, N., and G. Puccini. Socioeconomic Descriptors of Fatal Crash Rates in the
38 Southeast USA. *Injury Control and Safety Promotion*, Vol. 7, No. 3, 2000, pp. 165-173.
- 39 19. Li, Z., W. Wang, P. Liu, J. M. Bigham, and D. R. Ragland. Using Geographically Weighted
40 Poisson Regression for County-Level Crash Modeling in California. *Safety Science*, Vol. 58,
41 2013, pp. 89-97.
- 42 20. Noland, R. B., and M. A. Quddus. A Spatially Disaggregate Analysis of Road Casualties in
43 England. *Accident Analysis and Prevention*, Vol. 36, No. 6, 2004, pp. 973-984.

21. Fridstrøm, L., J. Ifver, S. Ingebrigtsen, R. Kulmala, and L. K. Thomsen. Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts. *Accident Analysis and Prevention*, Vol. 27, No. 1, 1995, pp. 1-20.
22. Noland, R. B., and M. A. Quddus. Analysis of Pedestrian and Bicycle Casualties with Regional Panel Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1897, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 28-33.
23. Huang, H., M. Abdel-Aty, and A. Darwiche. County-Level Crash Risk Analysis in Florida. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 12148, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 27-37.
24. Abdel-Aty, M., C. Siddiqui, H. Huang, and X. Wang. Integrating Trip and Roadway Characteristics to Manage Safety in Traffic Analysis Zones. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2213, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 20-28.
25. Lee, J., M. Abdel-Aty, and X. Jiang. Development of Zone System for Macro-Level Traffic Safety Analysis. *Journal of Transport Geography*, Vol. 38, 2014, pp. 13-21.
26. Hadayeghi, A., A. Shalaby, and B. Persaud. Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1840, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 87-95.
27. Caliendo, C., M. Guida, and A. Parisi. A Crash-Prediction Model for Multilane Roads. *Accident Analysis and Prevention*, Vol. 39, No. 4, 2007, pp. 657-670.
28. Montella, A., L. Colantuoni, and R. Lamberti. Crash Prediction Models for Rural Motorways. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2083, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 180-189.
29. Wier, M., J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia. An Area-Level Model of Vehicle-Pedestrian Injury Collisions with Implications for Land Use and Transportation Planning. *Accident Analysis and Prevention*, Vol. 41, No. 1, 2009, pp. 137-145.
30. MacNab, Y. C. Bayesian Spatial and Ecological Models for Small-Area Accident and Injury Analysis. *Accident Analysis and Prevention*, Vol. 36, No. 6, 2004, pp. 1019-1028.
31. LaScala, E. A., D. Gerber, and P. J. Gruenewald. Demographic and Environmental Correlates of Pedestrian Injury Collisions: A Spatial Analysis. *Accident Analysis and Prevention*, Vol. 32, No. 5, 2000, pp. 651-658.
32. Abdel-Aty, M., J. Lee, C. Siddiqui, and K. Choi. Geographical Unit Based Analysis in the Context of Transportation Safety Planning. *Transportation Research Part A: Policy and Practice*, Vol. 49, 2013, pp. 62-75.
33. Hadayeghi, A., A. S. Shalaby, and B. N. Persaud. Development of Planning Level Transportation Safety Tools Using Geographically Weighted Poisson Regression. *Accident Analysis and Prevention*, Vol. 42, No. 2, 2010, pp. 676-688.
34. Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp. 1033-1054.

36. Conroy, C., G. T. Tominaga, S. Erwin, S. Pacyna, T. Velky, F. Kennedy, M. Sise, and R. Coimbra. The Influence of Vehicle Damage on Injury Severity of Drivers in Head-on Motor Vehicle Crashes. *Accident Analysis and Prevention*, Vol. 40, No. 4, 2008, pp. 1589-1594.
37. Fredette, M., L. S. Mambu, A. Chouinard, and F. Bellavance. Safety Impacts Due to the Incompatibility of Suvs, Minivans, and Pickup Trucks in Two-Vehicle Collisions. *Accident Analysis and Prevention*, Vol. 40, No. 6, 2008, pp. 1987-95.
38. Khattak, A. J., M. D. Pawlovich, R. R. Souleyrette, and S. L. Hallmark. Factors Related to More Severe Older Driver Traffic Crash Injuries. *Journal of Transportation Engineering*, Vol. 128, No. 3, 2002, pp. 243-249.
39. Kockelman, K. M., and Y. J. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis and Prevention*, Vol. 34, No. 3, 2002, pp. 313-321.
40. Kweon, Y. J., and K. M. Kockelman. Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. *Accident Analysis and Prevention*, Vol. 35, No. 4, 2003, pp. 441-450.
41. Gray, R.C., M.A. Quddus, and A. Evans, Injury Severity Analysis of Accidents Involving Young Male Drivers in Great Britain. *Journal of Safety Research*, Vol. 39, No. 5, 2008, pp. 483-495.
42. Khorashadi, A., D. V. N. Shankar, and F. Mannering. Differences in Rural and Urban Driver-Injury Severities in Accidents Involving Large-Trucks: An Exploratory Analysis. *Accident Analysis and Prevention*, Vol. 37, No. 5, 2005, pp. 910-921.
43. Islam, S., and F. L. Mannering. Driver Aging and Its Effect on Male and Female Single-Vehicle Accident Injuries: Some Additional Evidence. *Journal of Safety Research*, Vol. 37, No. 3, 2006, pp. 267-276.
44. Awadzi, K. D., S. Classen, A. Hall, R. P. Duncan, and C. W. Garvan. Predictors of Injury among Younger and Older Adults in Fatal Motor Vehicle Crashes. *Accident Analysis and Prevention*, Vol. 40, No.6, 2008, pp.1804-1810.
45. Schneider, W., P. Savolainen, and K. Zimmerman. Driver Injury Severity Resulting from Single-Vehicle Crashes Along Horizontal Curves on Rural Two-Lane Highways. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2102, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 85-92.
46. Ulfarsson, G. F., and F. L. Mannering. Differences in Male and Female Injury Severities in Sport-Utility Vehicle, Minivan, Pickup and Passenger Car Accidents. *Accident Analysis and Prevention*, Vol. 36, No. 2, 2004, pp. 135-47.
47. Eluru, N. Evaluating Alternate Discrete Choice Frameworks for Modeling Ordinal Discrete Variables. *Accident Analysis and Prevention*, Vol. 55, 2013, pp. 1-11.
48. Mooradian, J., J. N. Ivan, N. Ravishanker, and S. Hu. Analysis of Driver and Passenger Crash Injury Severity Using Partial Proportional Odds Models. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 53-58.
49. Ma, J., K.M. Kockelman, and P. Damien, A Multivariate Poisson-lognormal Regression Model for Prediction of Crash Counts by Severity Using Bayesian Methods. *Accident Analysis and Prevention*, Vol. 40, No. 3, 2008, pp. 964-975.
50. Aguero-Valverde, J. and P. Jovanis, Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking. In *Transportation Research Record: Journal of*

- 1 *the Transportation Research Board, No. 2136*, Transportation Research Board of the National
2 Academies, Washington, D.C., 2009, pp. 82-91.
- 3 51. Park, E. and D. Lord, Multivariate Poisson-Lognormal Models for Jointly Modeling Crash
4 Frequency by Severity. In *Transportation Research Record: Journal of the Transportation*
5 *Research Board, No. 2019*, Transportation Research Board of the National Academies,
6 Washington, D.C., 2007, pp. 1-6.
- 7 52. El-Basyouny, K. and T. Sayed, Collision Prediction Models Using Multivariate Poisson-
8 Lognormal Regression. *Accident Analysis and Prevention*, Vol. 41, No. 4, 2009, pp. 820-828.
- 9 53. Papke, L.E., and J.M. Wooldridge. Econometric Methods for Fractional Response Variables
10 with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics*, Vol.
11 11, 1996, pp. 619-632.
- 12 54. Sivakumar, A. and C.R. Bhat. A Fractional Split Distribution Model for Statewide
13 Commodity Flow Analysis, *Transportation Research Record*, Vol. 1790, 2002, pp. 80-88.
- 14 55. Milton, J. C., V. N. Shankar, and F. L. Mannering. Highway Accident Severities and the
15 Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis and Prevention*,
16 Vol. 40, No. 1, 2008, pp. 260-266.
- 17 56. Eluru N, V. Chakour, M. Chamberlain, and L. F. Miranda-Moreno. Modeling Vehicle
18 Operating Speed on Urban Roads in Montreal: A Panel Mixed Ordered Probit Fractional Split
19 Model. *Accident Analysis and Prevention*, Vol. 59, 2013, pp.125-34.
- 20 57. Gourieroux, C., A. Monfort, and A. Trognon, Pseudo Maximum Likelihood Methods: Theory.
21 *Econometrica: Journal of the Econometric Society*, 1984, pp. 681-700.
- 22 58. Stamatiadis N. *Impact of Shoulder Width and Median Width on Safety*. REPORT 633.
23 National Cooperative Highway Research Program (NCHRP), Transportation Research Board
24 of National Academies, 2009.
- 25 59. Dissanayake, S. Young Drivers and Run-Off-the-Road Crashes. *Proceedings of the 2003 Mid-*
26 *Continent Transportation Research Symposium*, Iowa, 2003.
- 27 60. Ben-Bassat, T. and D. Shinar, Effect of Shoulder Width, Guardrail and Roadway Geometry
28 on Driver Perception and Behavior. *Accident Analysis and Prevention*, Vol. 43, No. 6, 2011,
29 pp. 2142-2152.
- 30 61. Pant, P.D., A.S. Rajagopal, and Y. Cheng. Rational Schedule of Base Accident Rates for Rural
31 Highways in Ohio (Phase II). Ohio Department of Transportation and Federal Highway
32 Administration, Report No. FHWA/OH-2003/008, June, 2003.
- 33 62. AASHTO. *Highway Safety Manual*. American Association of State Highway and
34 Transportation Officials, Washington, DC, 2010.