A Multilevel Generalized Ordered Probit Fractional Split Model For Analyzing Vehicle Speed

Tanmoy Bhowmik
Doctoral Student
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 1-407-927-6574; Fax: 1-407-823-3315
Email: tanmoy78@knights.ucf.edu

Shamsunnahar Yasmin
Postdoctoral Associate
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-823-4815, Fax: 407-823-3315
Email: shamsunnahar.yasmin@ucf.edu

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

*Corresponding author

ABSTRACT

Vehicle operating speed plays a significant role in many fields of transportation engineering including safety, operation, design and management. The current research effort contributes to literature on examining vehicle speed on arterial roads methodologically and empirically. Specifically, we propose and estimate a panel mixed generalized ordered probit fractional split (PMGOPFS) model to examine critical factors contributing to vehicle operating speed on roadways. The proposed modeling framework allows for the exogenous variable impacts to vary across the alternatives. Further, the model is formulated to allow for the impact of common unobserved factors across multiple levels (roadway, segment, direction, day and time period). To the best of the authors' knowledge, this is the first time such an econometric model is proposed and estimated in any literature (not just in transportation). The proposed model is estimated employing a maximum simulated quasi-likelihood based objective function. Vehicular speed data obtained from 8 arterial roads in Orlando for the year 2016 is used for estimating the model. The data is obtained for weekday morning and evening peak and off-peak hours for one randomly chosen week for each roadway throughout the year. The exogenous variables that are considered in the current empirical study include geometry, roadway, traffic, land use and environmental attributes. The model estimation results are further augmented by conducting elasticity analysis to highlight the important factors affecting the vehicular speed profile.

1 INTRODUCTION

Vehicle operating speed plays a significant role in many fields of transportation engineering. In transportation safety literature, vehicle speed is one of the most common factors identified as a contributing factor for crash occurrence and its consequences. Crashes involving high operating speeds often result in severe injury or fatality (Eluru and Bhat, 2007). As evident, in 2015, of the 32,166 fatal crashes in the US, 27% are attributed to high speed (National Center for Statistics and Analysis, Speeding: 2015 data). Vehicle speed is also considered as a critical factor in other domains of transportation research. For example, vehicle speed is employed as a performance measure in the evaluation process of geometric design consistencies. In traffic flow modeling, vehicle operating speed is an essential parameter to build or validate simulation frameworks. Recently, with the emergence of global climate change associated challenges, accurately estimating transport emissions relies on models that provide vehicle speed estimates.

Thus, it is not surprising that a number of studies explored the relationship between vehicular speed on roadway facilities (including roads, curves and tangents) and various exogenous factors including geometric attributes, traffic characteristics and driver characteristics. A majority of these studies developed vehicle speed prediction models based on the 85th percentile speed as opposed to employing the full speed distribution (Krammes et al., 1995; Gattis and Watts, 1999; Fambro et al., 2000; McFadden et al., 2001). The analysis neglecting the available full speed distribution of vehicles is likely to result in biased parameter estimations. Studies have attempted to address this by considering the full vehicle operating speed distribution (Tarris et al., 1996; Figueroa Medina and Tarko, 2005;). The methodological approaches considered in these studies involve disaggregate level linear regression and ordered fractional split approaches. The ordered fractional split model proposed by Eluru et al., 2013 evaluated the proportion of vehicles traveling in each speed interval for the segment of roadway by using the data from the city of Montreal. Further, to accommodate for repeated vehicle speed proportion observations, the authors developed a panel mixed version of the ordered fractional split model.

While ordered fractional split approach was successfully employed for vehicle speed proportions and injury severity proportions in existing transportation literature (Yasmin et al., 2016; Yasmin and Eluru, 2018a), it is associated with restrictive assumptions. Specifically, the ordered fractional split model assumes that the thresholds in the model structure remain the same for the entire population. The limitation is analogous to the limitation of the traditional ordered response model (Eluru et al., 2008; Yasmin et al., 2014; Yasmin et al., 2016; Fountas and Anastasopoulos, 2017). To address this limitation, we propose a generalized version of the ordered fractional split model by allowing for the exogenous variable impacts to vary across the alternatives. Further, in the presence of adequate repeated proportion measures for the same roadway segment multiple levels of unobserved factors influence the dependent variable (see Mannering and Bhat, 2014 and Mannering et al., 2016 for a discussion of the impact of ignoring unobserved heterogeneity on model estimates). The ordered fractional split model is enhanced to allow for the impact of common unobserved factors[1] across multiple levels (roadway, segment, direction, day and time period). Specifically, we propose to estimate a panel mixed generalized ordered probit fractional split (PMGOPFS) model to examine critical factors contributing to vehicle operating speed on roadways. The proposed model is estimated for vehicular speed data obtained from 8 arterial roads in Orlando for the year 2016. The data is obtained for weekday peak (morning and evening) and off-peak hours for one randomly chosen week for each roadway throughout the year. The exogenous variables that are considered in the

---

[1] A number of studies in transportation literature adopted various modeling framework to show that dependent variable in multiple dimension (vehicular speed or crash) share unobservable factors and hence multivariate in nature. The reader would refer to (Russo et al., 2014); (Alarifi et al., 2017); (Bogue et al., 2017); (Xin et al., 2017); (Cai et al., 2018); (Fountas et al., 2018); (Yasmin and Eluru, 2018b)

current empirical study include geometry, roadway, traffic, land use and environmental attributes. To the extent of the authors' knowledge, this is the first attempt to employ such an econometric framework for ordered fraction/proportion variables in extant literature (not just in transportation).

2 BACKGROUND AND CURRENT STUDY IN CONTEXT

2.1 Earlier Research

Several research efforts in existing literature have explored the factors affecting vehicle speed. A detailed review of earlier literature is available in Eluru et al., 2013. Table 1 provides a brief summary of the literature on vehicle speed modeling. The table provides information on the study, facility location, and roadway functional classification, application area, dependent variable considered and modeling methodology employed. The studies presented in Table 1 are categorized along two streams: (1) studies examining vehicle speed profile considering partial distribution and (2) studies examining vehicle speed profile considering full distribution.

Several observations can be made from Table 1. First, a large share of earlier research on vehicle speed modeling considered partial operating speed profile (20 studies out of 25). Second, among studies considering partial vehicle speed profile, the dependent variable representation was based on 85th percentile or mean vehicular speed. Studies that considered full operating speed profile, the exact representations of vehicular speed include proportions of vehicle speed categories and mean vehicular speed over a specific time interval. Third, the analysis of vehicle speed has extensively examined both rural and urban facilities with higher number of studies in the rural category. In terms of roadway functional classification, earlier research has explored vehicle speed profiles for two lane highways, multi-lane highways, freeways, local, collector and arterial roads. Fourth, the major focus in examining vehicle speed profile were road safety, geometric design, and/or speed limit compliance while few studies also focused on exploring several modeling frameworks in this aspect. Finally, the methodologies considered in these studies include simple approaches such as graphical plots, linear regression (or ordinary least squares) approaches, and Analysis of Variance (ANOVA). Among advanced econometric modeling approaches, researchers have employed mixed effect models (to recognize presence of repeated measures), three stage least square estimation, panel mixed ordered fractional split model and Artificial Neural Network methods.

Based on earlier literature, the various factors identified as influencing vehicle speed distribution profile include – radius and length of the curve, curvature change ratio, preceding tangent length, grade, roadway geometry, uninterrupted travel distance, width and type of street, posted speed limit, lane width, roadside hazard, available sight distance, heavy vehicle proportion, access density, heavy rainfall and high wind speed.

2.2 Current Study in Context

Based on the literature review, we can see that only a handful of studies have considered modeling the entire vehicle speed distribution. The current study contributes to literature in this stream by proposing, formulating and developing a new econometric model structure. Specifically, we build on earlier research of Eluru et al., 2013 by proposing a panel mixed generalized ordered probit fractional split (PMGOPFS) model. In this approach, we recognize that the threshold parameters are not constant across the entire sample population. The framework is formulated to allow these parameters to vary in response to observational attributes.

The reader would note that a maximum likelihood approach used for discrete choice/outcome variables is not valid for fractional dependent variables. For traditional discrete outcome models, the maximum likelihood approach is employed for maximizing the likelihood

4

of a single observed outcome. However, the dependent variable in the current analysis represents the proportion of vehicles in various speed category. Thus, the presence of the multiple observed outcomes with fractional values that sum up to 1 does not lend itself to maximum likelihood approaches. An alternative approach referred to as a quasi-likelihood approach has been proposed and widely implemented in econometrics literature for modeling such fractional dependent variables (see for example Papke and Wooldridge, 1996, Sivakumar and Bhat, 2002). Hence, the proposed model is estimated employing a maximum simulated quasi-likelihood based objective function.

The data for the analysis is drawn from 8 arterial roads with a total of 368 segments in the state of Florida. The hourly speed proportions on weekdays (Monday to Friday) from 6am – 9pm for a random week in the year are considered for each roadway. Further, within each time period - morning peak (6am-9am), off-peak (9am-4pm and 7pm-9pm) and evening peak (4pm-7pm) - 2 hourly data records are randomly drawn for analysis. Based on the data considered, several levels of hierarchy exist including arterial, segment, week, day and hour. In our proposed model, we allow for unobserved factors at these various levels. Finally, we compute aggregate level elasticity measures to provide a clear picture of attribute impact on vehicular speed profile. The effects are computed for the proposed framework (PMGOPFS) as well as the model counterparts (OPFS and GOPFS) across all speed levels (for different roads) for comparison purpose. The proposed study is the first attempt at modeling multiple levels of unobserved factors in fractional split models.

3   METHODOLOGY

The focus of our study is to model proportions of vehicle speed categories by employing PMGOPFS modeling approach. The econometric framework for the PMGOPFS model is presented in this section.

3.1   Model Structure

Let q ($q = 1, 2, …, Q$) be an index to represent observation unit (hour of weekdays for roadway segment), p (p = 1, 2, …, P) be an index for different level of repetition measures (roadway, roadway-day, roadway-hour, roadway-day-direction) at observation unit q and let k ($k = 1, 2, 3, …, K$) be an index to represent vehicle speed category. The latent propensity equation for vehicle speed at the qth unit and pth interval can be written as:

$$y_q^* = \left(\alpha + \gamma_q + \delta_{qp}\right)Z_q + \xi_q, \tag{1}$$

This latent propensity y_q^* is mapped to the actual vehicle speed category proportion y_{qk} by the ψ thresholds ($\psi_0 = -\infty$ and $\psi_k = \infty$). Z_q is a vector of attributes that influences the propensity associated with vehicle speed. α is a corresponding vector of mean effects, and γ_q is another vector of unobserved factors moderating the influence of attributes in Z_q on the vehicle speed propensity for analysis unit q. δ_{qp} is a vector of unobserved effects specific to repetition level p. ξ_q is an idiosyncratic random error term assumed to be identically and independently standard normal distributed across individuals q.

The PMGOPFS model relaxes the constant threshold across observation to provide a flexible form of the OPFS model. The basic idea of the PMGOPFS is to represent the threshold parameters as a linear function of exogenous variables. Thus, the thresholds are expressed as:

$$\psi_k = fn(x_{qk}) \tag{2}$$

where, x_{qk} is a set of exogenous variables (including a constant) associated with k th threshold. Further, to ensure the accepted ordering of observed vehicle speed proportion ($-\infty < \psi_1 < \psi_2 < \ldots\ldots\ldots < \psi_{K-1} < +\infty$), we employ the following parametric form as employed by Eluru et al.(Eluru et al., 2008):

$$\psi_k = \psi_{k-1} + exp((\beta_k + \theta_{kq})x_{qk}) \tag{3}$$

where, β_k is a vector of parameters to be estimated. θ_{kq} is another vector of unobserved factors moderating the influence of attributes in x_{qk} on the vehicle speed propensity for analysis unit q and vehicle speed category k.

3.2   Model Estimation

The model cannot be estimated using conventional Maximum likelihood approaches. Hence, we resort to quasi-likelihood based approach for our methodology (see (Papke and Wooldridge, 1996; Sivakumar and Bhat, 2002 for detailed discussion). To estimate the parameter vector, we assume that

$$E(y_{qk}|Z_{qk}) = H_{qk}(\alpha, \psi, \delta_q, \theta_{kq}), 0 \le H_{qk} \le 1, \sum_{k=1}^{k} H_{qk} = 1 \tag{4}$$

H_{qk} in our model is the generalized ordered probit probability for vehicle speed category k defined as

$$P_{qpk} = \{G\left[(\psi_{k-1} + exp((\beta_k + \theta_{kq})x_{qk})) - \{(\alpha' + \delta'_{qp})Z_q\}\right] \quad - \quad G\left[(\psi_{k-2} + exp(\beta_k + \theta_{kq})x_{qk}) - \{(\alpha + \gamma_q + \delta_{qp})Z_q\}\right]\} \tag{5}$$

The proposed model ensures that the proportion for each vehicle speed category is between 0 and 1 (including the limits). In estimating the PMGOPFS model, it is necessary to specify the structure for the unobserved vectors $\delta, \gamma \ and \ \theta$ represented by Ω. In this paper, it is assumed that these elements are drawn from independent realization from normal population: $\Omega \sim N(0, (\pi^2, \sigma_p^2, \upsilon_k^2))$. Thus, conditional on Ω, the quasi likelihood function (see (Papke and Wooldridge, 1996)) for a discussion on asymptotic properties of quasi likelihood proposed) may be written for unit q for various repetitive measures as:

$$L_q(\alpha, \psi| \Omega) = \prod_{p=1}^{P}\prod_{k=1}^{K} = \{P_{qpk}\}^{d_{qk}} \tag{6}$$

where d_{qk} is the proportion of vehicles in speed category k. Finally, the unconditional likelihood function can be computed for site unit q can be written as:

$$L_q(\alpha, \psi, \Omega) = \int_{\Omega} L_q(\alpha, \psi| \Omega) \ dF(\Omega) \tag{7}$$

where F is the multidimensional cumulative normal distribution. The quasi log-likelihood function is:

$$L(\Omega) = \sum_q L_q (\alpha, \psi, \Omega) \tag{8}$$

The log-likelihood function in Equation (8) involves the evaluation of a multi-dimensional integral of size equal to the number of rows in δ_{qp}.

In the current study context, we estimate δ_{qp} for different levels of repetition measures (p). Specifically, we evaluate unobserved effects at roadway, roadway-day, roadway-hour, roadway-day-direction levels in addition to the simpler hourly level. The flexibility offered by testing for unobserved heterogeneity enhances the model development exercise. For example, at a roadway level, factors such as design and pavement quality differences across roadways can influence speed distribution. On some facilities, based on prior crash experience, a number of speed bumps might be installed at multiple locations. The data on the presence of such speed bumps (or other such micro-treatments) are rarely available for analysis. At a roadway-day level, repaving or maintenance activities on a particular day could influence the speed distribution across the roadway. At a roadway-hour level, the specific design characteristics of the roadway and its interaction with traffic volumes could result in a specific speed profile. While some design characteristics and interactions with traffic volumes are considered in modeling exercise, it is not possible to consider all possible combinations. At a roadway-day-direction level, a major accident occurring on a particular day on a roadway facility could affect speed distribution. The reader would note that the multiple levels identified here apply to any random parameters estimated such as lane drop i.e. the influence of unobserved factors associated with an observed attribute can also be accommodated at multiple levels identified above. Thus, the proposed framework by allowing for additional flexibility allows the analyst to avoid conflation of unobserved effects on speed profiles.

In accommodating unobserved effects at different levels, random numbers are assigned to the appropriate observations of the repetition measures. For example, at roadway level, we have 8 arterials. Thus, in evaluating unobserved effect at the roadway level, 8 sets of different random numbers are generated specific to 8 roads and assigned to the data records based on their roadway ID. The random numbers are assigned for other repetition levels following the same analogy in estimating the model. In the current paper, we use a quasi-Monte Carlo (QMC) method proposed by Bhat (Bhat, 2001) for discrete choice models. Within the broad framework of QMC sequences, we specifically use the Halton sequence in the current analysis. In our analysis, a rigorous examination of the influence of the number of Halton draws was conducted. The model estimation routine is coded in GAUSS Matrix Programming software (Aptech, 2015)

4   DATA AND DEPENDENT VARIABLE

The data for the empirical study is obtained for 8 major arterials in state of Florida including 368 segments for the year 2016. Figure 1 represents the location of the arterial roads considered in the study. The number of segments for each road range from 6 to 89. In our data, we have a total of 368 segments for the current analysis and we extract the vehicular speed for 8 arterial roads for weekdays only from 6am to 9pm. For these 8 arterial roads, one random week in 2016 is chosen for data retrieval. For our study, weekday data (randomly a week for each roadway) from 6am-9pm were retrieved. Specifically, a total of 6 hours with 2 records from each time period (morning peak (6am-9am), off-peak (9am-4pm and 7pm-9pm) and evening peak (4pm-7pm)) were considered for model analysis. The reader would note that the 2 records chosen are selected randomly i.e. they vary across days and arterials. The vehicle speed data is compiled from Regional Integrated Transportation Information System (RITIS) for the year 2016. The RITIS database is an automated data sharing system which includes real time data feeds. After

processing the records, we obtained a total of 27600 observations. These records were split into two datasets: 1) model estimation sample with 11040 observations and 2) holdout sample with 16560 records set aside for validation analysis. The sample size for estimation was guided by run times for estimating multiple levels of unobserved heterogeneity.

The vehicle speed data gathered from RITIS includes information on the travelling speed of vehicles on a second by second level for each roadway segment. The second by second data were aggregated to obtain number of vehicles under various speed categories for each hour. For the analysis purposes, the following speed categories are used: 1) ≤ 20 mph; 2) 20-25 mph; 3) 25-30 mph; 4) 30-35 mph; 5) 35-40 mph and 6) ≥ 40 mph. The dependent variable for fractional split models can be represented as a proportion (number of vehicles with speed category in a segment during an hour/total number of vehicles in a segment during an hour) as follows: (1) Proportion of vehicle speed ≤ 20 mph, (2) Proportion of vehicle speed greater than 20 mph and ≤ 25 mph, (3) Proportion of vehicle speed greater than 25 mph and ≤ 30 mph, (4) Proportion of vehicle speed greater than 30 mph and ≤ 35 mph, (5) Proportion of vehicle speed greater than 35 mph and ≤ 40 mph and (6) Proportion of vehicle speed greater than 40 mph. Table 2 provides the summary statistics of the vehicle speed proportion variables at segment level. From Table 2, we can observe that majority of the traffic flow on arterial roads is in the speed range of 25-30 mph.

4.1   Exogenous Variables

For examining the critical factors contributing to vehicle speed profile of arterials, we compiled additional information on a host of exogenous variables including roadway characteristics, land use attributes, built environment, traffic characteristics and temporal variables. These variables are generated for each segment across all roadways considered. Roadway characteristics are obtained from the Florida Department of Transportation (FDOT). Transportation Statistics division. Traffic, land use and built environment attributes are obtained from the US Census Bureau and FDOT. Information about the temporal attributes are collected from Florida Automated Weather Network (FAWN).

Roadway attributes considered include segment lengths, width and type of median, maximum and minimum number of lane, lane drop, total number of intersections in the segment, average posted speed limit, average width of the sidewalk, inside and outside shoulder mean width and length of bike lane. Two types of median are considered based on the shoulder type: soft or hard. Lane drop variable is defined as the difference between maximum and minimum number of lane in the segment. Land use attributes included area of urban, residential, industrial, institutional, recreational, office, and land use mix within a 1-mile buffer around the roadway. For traffic characteristic, average annual daily traffic (AADT), average annual daily truck traffic (truck AADT) and proportion of heavy traffic are considered. In case of built environment, the study includes total number of financial, commercial, business, educational, recreational and parking facilities within a 1-mile buffer. Finally, weather data such as information on temperature, average precipitation, wind speed, relative humidity and dew point temperature at an hourly level are gathered for environmental characteristics.

Table 3 summarizes sample characteristics of the explanatory variables with the appropriate definition considered for final model estimation along with the minimum, maximum and mean values at a segment level. Several functional forms and specification for different variables are explored and those are used which provide the best result. The final specification of the model development was based on removing the statistically insignificant variables in a systematic process based on 90% significance level.

8

5   EMPIRICAL ANALYSIS

5.1   Model Specification and Overall Measure of Fit

The empirical analysis involves estimation of three different models: 1) Ordered Probit Fractional Split model (OPFS), 2) Generalized Ordered Probit Fractional Split Model (GOPFS) and 3) Panel Mixed Generalized Ordered Probit Fractional Split Model (PMGOPFS) model. The quasi log-likelihood values at convergence for the different models are as follows: 1) OPFS (18 parameters) is -17518.38, 2) GOPFS (26 parameters) is -17437.79 and 3) PMGOPFS (30 parameters) is -17388.07. The corresponding log-likelihood at constants is -18467.60. Prior to discussing the estimation results, we compare the performance of these models in this section. We employ log-likelihood ratio test for comparing these models. The log-likelihood test statistic is computed as $2[LL_U - LL_R]$, where LL_U and LL_R are the log-likelihood of the unrestricted and the restricted models, respectively. The computed value of the LR test is compared with the $\aleph^2$ value for the corresponding degrees of freedom (*dof*). The resulting LR test values for the comparison of OPFS/GOPFS, OPFS/PMGOPFE and GOPFS/PMGOPFS models are 161.18 (8 *dof*), 260.62 (12 *dof*) and 99.44 (4 *dof*), respectively. The log-likelihood ratio test values indicate that PMGOPFS model outperform both OPFS and GOPS at any level of statistical significance. The reader would note that Akaike Information Criterion and Bayesian Information Criterion offer the same conclusions in favor of the PMGOPFS model.

5.2   Model Estimation Results

In discussing the estimation results, to conserve on space, we will restrict ourselves to the discussion of PMGOPFS model results (see Appendix for the results of OPFS and GOPFS models). Table 4 represents the model estimation results of the PMGOPFS model. In Table 4, column 2 presents the estimates for propensity and columns 3 through 6 represent the threshold parameters. In the propensity, a positive (negative) coefficient increases (decreases) the likelihood of higher speed categories. When the threshold parameter is positive (negative), the result implies that the threshold is bound to increase (decrease). For the ease of presentation, the estimation results are discussed by variable groups

5.2.1   Roadway Characteristics

The results associated with segment length indicate that longer segments have a higher likelihood of observing higher vehicle speeds (see (Dinh et al., 2013) for similar result). As the stretch of roads between intersections increases (as is the case in our study for segments) speed proportions are likely to skew rightward. Lane drop variable shows a negative association with speed proportion propensity. Lane drop variable implies a reduction in number of lanes thus resulting in drivers slowing down. Further, we found that lane drop has significant variability across observations (at an hourly level) as indicated by the significant standard deviation parameter. An increase in width of average inside shoulder is positively associated with speed propensity. Wider inside shoulder is likely to provide additional safety margin for drivers encouraging them to drive at a higher vehicular speed.

Average sidewalk width shows a negative effect on vehicular speed proportion propensity. With wider sidewalk, there is likely to be higher pedestrian activity that potentially results in vehicle operation at a lower speed (for a similar result see Wang, 2006). A negative relationship is found between presence of an intersection and vehicular speed proportion propensity. An increase in average bike lane length results in reduced vehicular speed propensity. The impact on the threshold value for the variable indicates a lower likelihood of speed proportion of above 40mph with increased average length of bike lane.

5.2.2 Traffic Characteristics

As expected, vehicular speed is also influenced by the volume of traffic on the road (Ericsson, 2000). The estimated result for AADT implies a negative impact on vehicular speed proportion propensity. With higher volume of traffic, vehicle density will increase which will result in low vehicular speed, a result also observed in (Islam and El-Basyouny, 2015).

5.2.3 Land Use Characteristics

The results associated with industrial area indicate that higher proportion of industrial area is negatively associated with the vehicular speed proportion. With higher proportion of industrial area, there is likely to be more traffic activity which results in vehicle operation at lower speeds. The variable also has a positive impact on the threshold between 30-35mph and 35-40mph i.e. further increasing the probability of lower speed alternatives.

Among different built environment characteristics explored in the study, only number of shopping centers within 1-mile of the road segment is found to have a significant impact on the vehicular speed proportion. The variable does not have any effect on the speed propensity but demonstrates a lower likelihood of speed proportion above 30-35mph (as the thresholds move to the right starting with the threshold between 20-25 mph and 30-35 mph). This is intuitive because, with high number of shopping centers, there is likely to be higher traffic and pedestrian activity that potentially results in low vehicular speed.

5.2.4 Environmental Characteristics

Various types of environmental attributes were considered in the model estimation process including average temperature, average precipitation, wind speed and humidity. However, only, average precipitation was found to have significant impact on vehicle speed profile in our study context. The estimated result shows that drivers have a disinclination towards higher vehicular speed in the presence of high precipitation, perhaps because of the reduced visibility (see (Feng, 2001) for similar result). The effect of rain on the threshold also indicates the lower likelihood of speed proportion above 35-40mph.

5.2.5 Road Specific Characteristics

Given the availability of repeated measures for each roadway, we also considered road specific measures in our analysis. Among the road specific indicator variables, Lake Underhill, SR 15, SR 426, SR 551 and University Blvd variables were found to have a significant impact on speed propensity (SR 434, SR 436 and SR 50 serve as the base for these variables). The negative sign on latent propensity associated with Lake Underhill, SR 15 and SR 426 indicator variable suggests that the vehicular speed on these roads are usually lower compared to SR 434, SR 436 and SR 50. Further, the positive effect of Lake Underhill SR 426 and SR 15 on the fourth and second threshold indicates a lower likelihood of higher vehicular operating speed. On the other hand, we observe an increase in speed propensity on University Blvd relative to SR 434, SR 436 and SR 50 indicating a rightward shift in the speed proportion. The result associated with SR 551 indicates that the variable does not have any effect on the speed propensity but demonstrates a higher likelihood of speed proportion above 20-25 mph.

5.2.6 Unobserved Effects

In our proposed model, we estimated unobserved effects at multiple levels: roadway, roadway-day, roadway-day-direction and roadway-hour. Among different considered levels, we found that the roadway, roadway-day and roadway-day-direction level effects have significant influence on vehicular speed profile. The estimation results of these standard deviations are presented in last row panel of Table 4. The significant standard deviation parameters at different repetition measures provide evidence toward supporting our hypothesis that it is

necessary to incorporate these unobserved effects in examining vehicle speed. These variables indicate that the vehicle speed profile may vary for different roadway based on the unobserved effects specific to different levels.

6   MODEL PREDICTION EXERCISE

In order to demonstrate the applicability of the proposed framework for modelling vehicular speed proportion on arterial roads in Orlando, a prediction exercise was undertaken using the model parameter estimates. The predicted proportion are compared with the actual observed proportion in the dataset. We perform two prediction exercises: 1) In-sample prediction: for the records used in model analysis and 2) holdout sample prediction: for the records that have been set aside for validation analysis. Figure 2 represents the heat map of absolute differences between observed and predicted proportion for the in-sample (2a) and holdout-sample (2b) vehicular speed data.

From Figure 2a, we can observe that, for the in-sample prediction comparison, the result is very good. For around 71% of the categories (37 out of 48), we have less than 0.03 difference between observed and predicted vehicular speed proportions. For SR 436, SR 434, SR 436 and University Blvd, the prediction is no different than the actual observed proportion. For SR 15, the result is slightly worse because in the observed data we have a unusually high percentage in the lower speed category. The corresponding results presented in Figure 2b, for the holdout sample prediction exercise is slightly inferior but acceptable. For about 58% of the categories (28 out of 48), the predicted proportion differs by less than 0.03. Overall trend of the heat map is quite similar to the in-sample data as prediction is good for SR 436, SR 434, SR 436 and University Blvd while we observe a higher difference between observed and predicted proportion in the lower speed categories for SR 15. From both figures, we can conclude that, the model performance for the in-sample and holdout-sample predictions are indicative of capturing the major speed trends.

7   ELASTICITY EFFECTS

The estimated results from Table 4 do not directly provide the exact magnitude of the effects of variables on the probability of each speed level. To quantify the impact of factors more clearly, we compute aggregate level elasticity effects for a subset of independent variables including length of the segment, intersection density, average bike lane length, AADT and proportion of industrial area. In our study, we investigate the effect as percentage change in the expected proportions of vehicular speed in response to the increase of the explanatory variable by 10% (see Eluru and Bhat, 2007 for a discussion on the methodology for computing elasticities). The effects are computed for the proposed framework (PMGOPFS) as well as the model counterparts (OPFS and GOPFS) across all speed levels. However, for the sake of brevity, we present the results for the top three speed categories PMGOPFS and GOPFS models (Figure 3). A detailed table documenting all the elasticity effects across speed categories and models is included in the appendix (Table A.3). The numbers in Figure 3 and Table A.3 can be interpreted as the percentage change in the expected vehicular speed proportions (increase for positive sign and decrease for negative sign) due to the change in the exogenous variable. For instance, the value of elasticity corresponding to segment length (from Figure 3, for speed >40mph) indicates that an increase in segment length of Lake road by 10% results in a 10.65% increase in the speed proportion of above 40mph.

The reader would note that the elasticity effects need to be described across multiple dimensions including overall effects from the PMGOPFS model, variations in elasticity across different roads and differences between PMGOPFS and GOPFS models. Based on the elasticity effects presented in Figure 3, following observations can be made. First, among the variables considered for elasticity computation, segment length is the most significant contributor

associated with higher vehicular speed. On the other hand, AADT, intersection density and proportion of industrial area are the important factors responsible for reduction in speed probability. Second, the average bike lane length variable does not have a significant influence on speed profile except for SR 426's highest speed level (>40mph). Third, with regards to road specific analysis, we found substantial and significant differences in elasticities across different roads. For instance, increased segment length is likely to lead to higher vehicular speeds for SR 426 relative to the other roads. On the other hand, AADT and intersection density variables result in a higher reduction in speed proportion for SR 15 compared to the other roads. Finally, in terms of comparison across different frameworks adopted in the study, we found substantial differences in elasticities. For example, in case of Lake road, for the industrial area variable, the PMGOPFS model predicts an increase in speed proportion between >20-25mph (Table A.3) while the other two models predict a reduction for the same speed category. Thus, it is evident that allowing for a flexible specification of observed and unobserved factors provides more accurate representation of variable impacts.

8   CONCLUSIONS

Vehicle operating speed plays a significant role in many fields of transportation engineering including transportation safety, traffic flow modeling, geometric design, vehicle emission and road user route decisions. Thus, it is not surprising that a number of studies explored the relationship between vehicular speed on roadway facilities (including roads, curves and tangents) and various exogenous factors; including geometric attributes, traffic characteristics and driver characteristics. A majority of these studies developed vehicle speed prediction models based on the 85th percentile speed as opposed to employing the full speed distribution. The current study contributes to literature in this stream by proposing, formulating and developing a new econometric model structure considering the full vehicle operating speed distribution. Specifically, we estimated a panel mixed generalized ordered probit fractional split (PMGOPFS) model to examine critical factors contributing to vehicle operating speed on roadways. The study was conducted by using vehicular speed data obtained from 8 arterial roads in the state of Florida for the year 2016. The data is obtained for weekday 6 hours (2 from each time period) for one randomly chosen week for each roadway throughout the year. A host of exogenous variables are considered including geometry, roadway, traffic, land use and environmental attributes. The proposed model is estimated employing a quasi-loglikelihood based objective function because maximum likelihood approach is not valid for fractional outcomes. A comparison of the proposed model with traditional OPFS and GOPFS was conducted using a loglikelihood ratio test. The test values clearly highlighted the superiority of the PMGOPFS model over other two approaches.

The model is developed to allow for the impact of parameter heterogeneity and common unobserved factors across multiple levels (roadway, segment, direction, day and time period). In terms of random effects, we found that lane drop has significant variability for the vehicular speed proportion. Again, we estimated unobserved effects for different repetition measures including: roadway, roadway-day, roadway-day-direction and roadway-hours. Of these parameters, roadway, roadway-day and roadway-day-direction level effects significantly influenced vehicular speed profile. The findings suggest that vehicle speed profile may vary for different roadway based on the unobserved effects specific to different repetition measures and thus it is necessary to incorporate these unobserved effects in examining vehicle speed. Finally, we undertake a prediction exercise to evaluate the performance and applicability of the proposed framework and the results indicate that the model performs quite adequately for both in-sample and holdout-sample datasets.

In our research, to further quantify the impact of various exogenous factors, we compute aggregate level elasticity effects for a subset of independent variables across different roads for

all speed levels. The effects are computed for the proposed framework (PMGOPFS) as well as the model counterparts (OPFS and GOPFS) across all speed levels for comparison purpose. The elasticity analysis indicated that segment length is the most significant contributor associated with higher vehicular speed whereas in terms of reduction in speed, AADT, intersection density and proportion of industrial area are the important factors. Moreover, the effects indicated that there were substantial and significant differences in elasticities across different roads. Finally, we found substantial differences in elasticities across the three model structures. For instance, the proposed model predicts an increase in speed proportion between >20-25mph for industrial area while the other two models show a reduction for the same category. Such differences could be attributed to the fact that the proposed approach allows for a more flexible framework by accommodating for observed and unobserved heterogeneity.

However, the study is not without limitations. In our analysis, we used 2 hours of speed data from each time slot. It might be interesting to estimate a similar model using all hourly records. Further, the records considered in our analysis are entirely for conditions with potential traffic congestion. A future research effort can be conducted by considering the night hours to capture the significant differences across the different time periods.

REFERENCES
Alarifi, S.A., Abdel-Aty, M.A., Lee, J. and Park, J., 2017. Crash modeling for intersections and segments along corridors: a Bayesian multilevel joint model with random parameters. Analytic Methods in Accident Research, 16, 48-59.

Anastasopoulos, P.C. and Mannering, F.L., 2016. The effect of speed limits on drivers' choice of speed: a random parameters seemingly unrelated equations approach. Analytic Methods in Accident Research, 10, 1-11.

Aptech 2015, Aptech Systems Inc, URL http://www.aptech.com/ (accessed 9.19.15).

Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research Part B, 35 (7), 677–693.

Bogue, S., Paleti, R. and Balan, L., 2017. A Modified Rank Ordered Logit model to analyze injury severity of occupants in multivehicle crashes. Analytic Methods in Accident Research, 14, 22-40.

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L. and Wang, X., 2018. Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. Analytic Methods in Accident Research, 19, 1-15.

Dinh, D.D., Kojima, A., Kubota, H., 2013. Modeling Operating Speeds on Residential Streets with a 30 km/h Speed Limit: Regression versus Neural Networks Approach. Journal of the Eastern Asia Society for Transportation Studies, 10, 1650–1669.

Eluru, N., and Bhat, C.R., 2007. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention, 39 (5), 1037–1049.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis and Prevention, 40 (3), 1033–1054.

Eluru, N., Chakour, V., Chamberlain, M., Miranda-Moreno, L.F., 2013. Modeling vehicle operating speed on urban roads in Montreal: A Panel Mixed Ordered Probit Fractional Split Model. Accident Analysis and Prevention, 59, 125–134.

Ericsson, E., 2000. Variability in urban driving patterns. Transportation Research Part D, 5 (5), 337–354.

Fambro, D., Fitzpatrick, K., Russell, C., 2000. Operating speed on crest vertical curves with limited stopping sight distance. Transportation Research Record: Journal of the Transportation Research Board, 1701, 25–31.

Feng, C. (2001). Synthesis of studies on speed and safety. Transportation Research Record: Journal of the Transportation Research Board, 1779, 86-92.

Figueroa Medina, A., and Tarko, A., 2005. Speed factors on two-lane rural highways in free-flow conditions. Transportation Research Record: Journal of the Transportation Research Board, 1912, 39–46.

Fitzpatrick, K., Carlson, P., Brewer, M., Wooldridge, M., 2001. Design factors that affect driver speed on suburban streets. Transportation Research Record: Journal of the Transportation Research Board, 1751, 18–25.

Fitzpatrick, K., Carlson, P., Brewer, M., Wooldridge, M.D., 2003. Design speed, operating speed, and posted speed limit practices. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, DC.

Fountas, G., and Anastasopoulos, P. C. (2017). A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. Analytic Methods in Accident Research, 15, 1-16.

Fountas, G., Anastasopoulos, P.C. and Abdel-Aty, M., 2018. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. Analytic Methods in Accident Research, 18, 57-68.

Garber, N. J. and Gadiraju, R. (1988). Speed variance and its influence on accidents. AAA Foundation for Traffic Safety, Washington, DC.

Gattis, J.L., and Watts, A., 1999. Urban street speed related to width and functional class. Journal of Transportation Engineering, 125 (3), 193–200.

Gong, H., and Stamatiadis, N., 2008. Operating speed prediction models for horizontal curves on rural four-lane highways. Transportation Research Record: Journal of the Transportation Research Board, 2075, 1–7.

Himes, S.C., and Donnell, E.T., 2010. Speed prediction models for multilane highways: simultaneous equations approach. Journal of Transportation Engineering, 136 (10), 855–862.

Islam, M.T. and El-Basyouny, K., 2015. Multilevel models to analyze before and after speed data. Analytic Methods in Accident Research, 8, 33-44.

Jacob, A., Dhanya, R., Anjaneyulu, M. V. L. R., 2013. Geometric design consistency of multiple horizontal curves on two-lane rural highways. Procedia - Social and Behavioral Sciences, 104, 1068–1077.

Kanellaidis, G., Golias, J., Efstathiadis, S., 1990. Drivers' speed behaviour on rural road curves. Traffic Engineering and Control, 31 (7-8), 414–415.

Ko, J., and Guensler, R.L., 2005. Characterization of congestion based on speed distribution: a statistical approach using Gaussian mixture model. Presented at 84th Annual Meeting of the Transportation Research Board, Washington, DC.

Krammes, R.A., Brackett, R.Q., Shafer, M.A., Ottesen, J.L., Anderson, I.B., Fink, K.L., Collins, K.M., Pendleton, O.J., Messer, C.J., 1995. Horizontal alignment design consistency for rural two-lane highways. Rep. No. FHWA-RD-94-034. Federal Highway Administration, U.S. Department of Transportation.

Lamm, R., Choueiri, E.M. and Mailaender, T., 1990. Comparison of operating speeds on dry and wet pavements of two-lane rural highways. Transportation Research Record: Journal of the Transportation Research Board, 1280 (8), 199-207.

Liang, W., Kyte, M., Kitchener, F., Shannon, P., 1998. Effect of environmental factors on driver speed: A case study. Transportation Research Record: Journal of the Transportation Research Board, 1635, 155–161.

Mannering, F.L. and Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research, 1, 1-22.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11, 1–16.

McFadden, J., Yang, W.T., Durrans, S., 2001. Application of artificial neural networks to predict speeds on two-lane rural highways. Transportation Research Record: Journal of the Transportation Research Board, 1751, 9–17.

National Center for Statistics and Analysis, Speeding: 2015 data (Traffic Safety Facts. DOT HS 812 409). Washington, DC: National Highway Traffic Safety Administration, 2017.

Papke, L.E., and Wooldridge, J.M., 1996. Econometric methods for fractional response variables with an application to 401 (k) plan participation rates. Journal of Applied Econometrics, 11 (6), 619–632.

Park, B.-J., Zhang, Y., Lord, D., 2010. Bayesian mixture modeling approach to account for heterogeneity in speed data. Transportation Research Part B, 44 (5), 662–673.

Perco, P., 2008. Influence of the general character of horizontal alignment on operating speed of two-lane rural roads. Transportation Research Record: Journal of the Transportation Research Board, 2075, 16–23.

Poe, C., and Mason Jr, J., 2000. Analyzing influence of geometric design on operating speeds along low-speed urban streets: mixed-model approach. Transportation Research Record: Journal of the Transportation Research Board, 1737, 18–25.

Poe, C.M., Tarris, J.P., Mason, J.M., 1996. Influence of Access and Land Use on Vehicle Operating Speeds Along Low-Speed Urban Streets. Presented at In Second National Access Management Conference, Transportation Research Board; Federal Highway Administration and Colorado Department of Transportation.

Polus, A., Fitzpatrick, K., Fambro, D., 2000. Predicting operating speeds on tangent sections of two-lane rural highways. Transportation Research Record: Journal of the Transportation Research Board, 1737, 50–57.

Russo, B.J., Savolainen, P.T., Schneider IV, W.H. and Anastasopoulos, P.C., 2014. Comparison of factors affecting injury severity in angle collisions by fault status using a random parameters bivariate ordered probit model. Analytic Methods in Accident Research, 2, 21-29.

Russo, F., Fric, S., Biancardo, S.A., Gavran, D., 2015. Driver speed behavior on circular curves of undivided two-lane rural roads: Serbian and Italian case studies. Transportation Research Record: Journal of the Transportation Research Board, 2472, 117–128.

Semeida, A.M., 2014. Application of artificial neural networks for operating speed prediction at horizontal curves: a case study in Egypt. Journal of Modern Transportation, 22 (1), 20–29.

Sivakumar, A. and Bhat, C., 2002. Fractional split-distribution model for statewide commodity-flow analysis. Transportation Research Record: Journal of the Transportation Research Board, 1790, 80-88.

Tarris, J., Poe, C., Mason Jr, J., Goulias, K., 1996. Predicting operating speeds on low-speed urban streets: Regression and panel analysis approaches. Transportation Research Record: Journal of the Transportation Research Board, 1523, 46–54.

Wang, J., 2006. Operating speed models for low speed urban environment based on in-vehicle GPS data. Doctoral dissertation, Georgia Institute of Technology.

Wang, B., Hallmark, S., Savolainen, P. and Dong, J., 2018. Examining vehicle operating speeds on rural two-lane curves using naturalistic driving data. Accident Analysis and Prevention, 118, 236-243.

Xin, C., Guo, R., Wang, Z., Lu, Q. and Lin, P.S., 2017. The effects of neighborhood characteristics and the built environment on pedestrian injury severity: a random parameters generalized ordered probability model with heterogeneity in means and variances. Analytic Methods in Accident Research, 16, 117-132.

Yasmin. S., N. Eluru, C. R. Bhat and R. Tay (2014). A latent segmentation generalized ordered logit model to examine factors influencing driver injury severity. Analytic Methods in Accident Research, 1, 23-38.

Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M., 2016. Ordered fractional split approach for aggregate injury severity modeling. Transportation Research Record: Journal of the Transportation Research Board 2583, 119–126.

Yasmin, S. and Eluru, N., 2018a. A joint econometric framework for modeling crash counts by severity. Transportmetrica A: Transport Science, 14 (3), 230-255.

Yasmin, S. and Eluru, N., 2018b. A mixed grouped response ordered logit count model framework. Analytic Methods in Accident Research, 19, 49-61.

LIST OF FIGURES

FIGURE 1 Location of the arterial roads.

FIGURE 2a In-sample Prediction Comparison for Arterial Roads

FIGURE 2b Out-sample Prediction Comparison for Arterial Roads

FIGURE 3 Elasticity effect analysis at road specific level

LIST OF TABLES

TABLE 1: Brief Summary of Earlier Research

TABLE 2: Descriptive Statistics of Vehicle Speed Related Dependent Variables

TABLE 3: Summary Characteristics for Independent Variables

TABLE 4: PMGOPFS Model Results
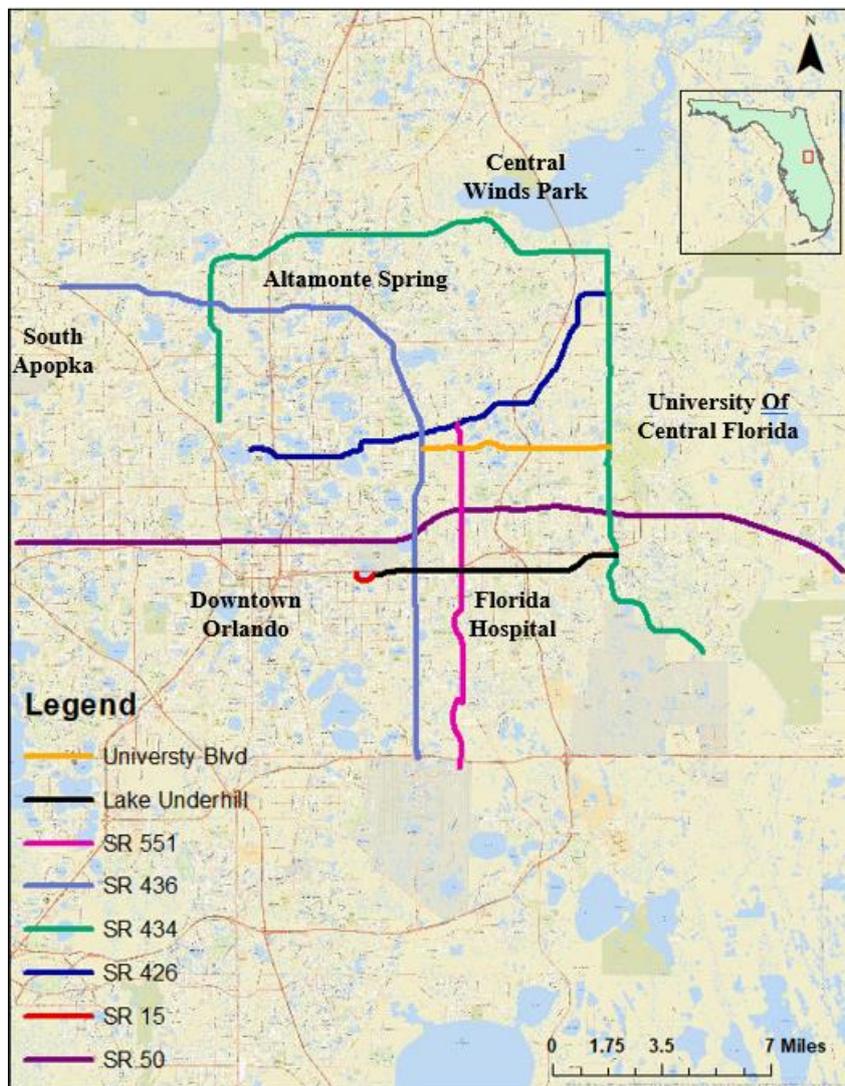
FIGURE 1 Location of the arterial roads.
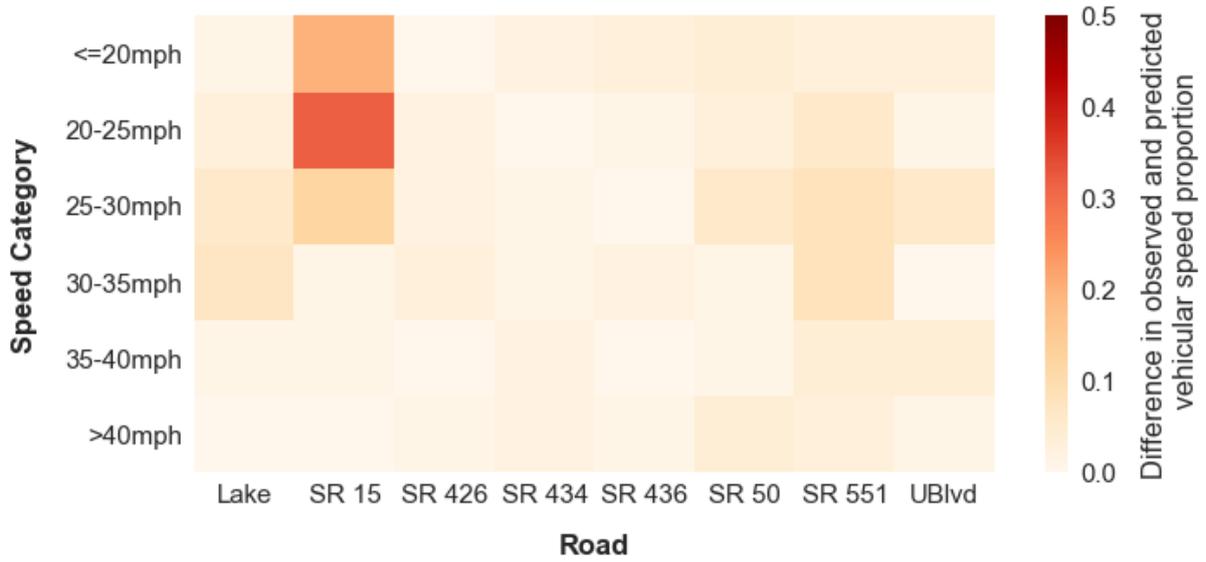
FIGURE 2a In-sample Prediction Comparison for Arterial Roads (sample size = 11,040)
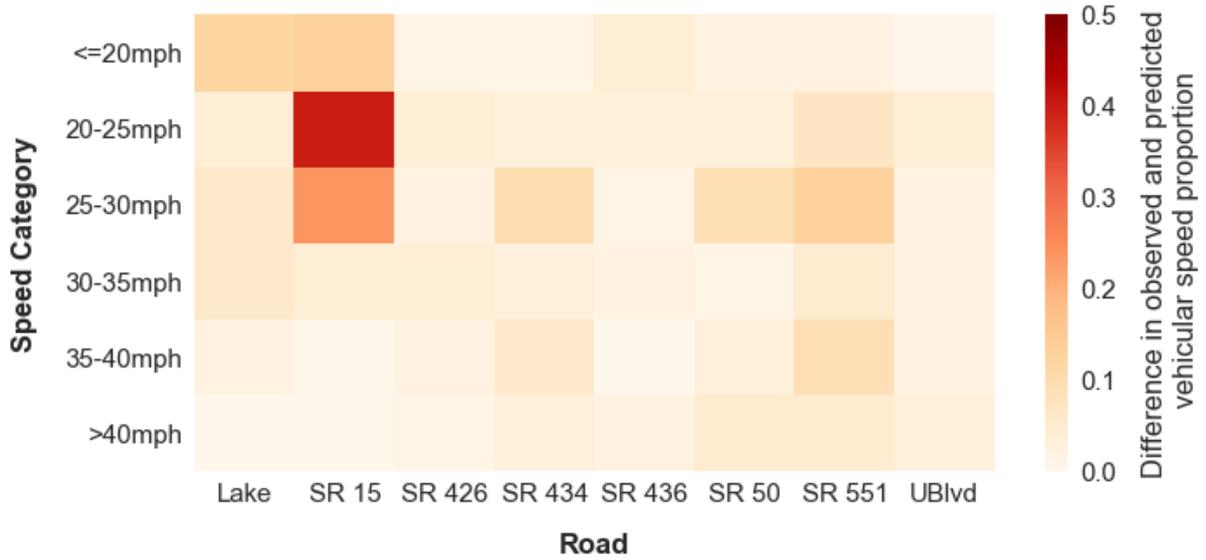
FIGURE 2b Out-sample Prediction Comparison for Arterial Roads (sample size = 16,560)
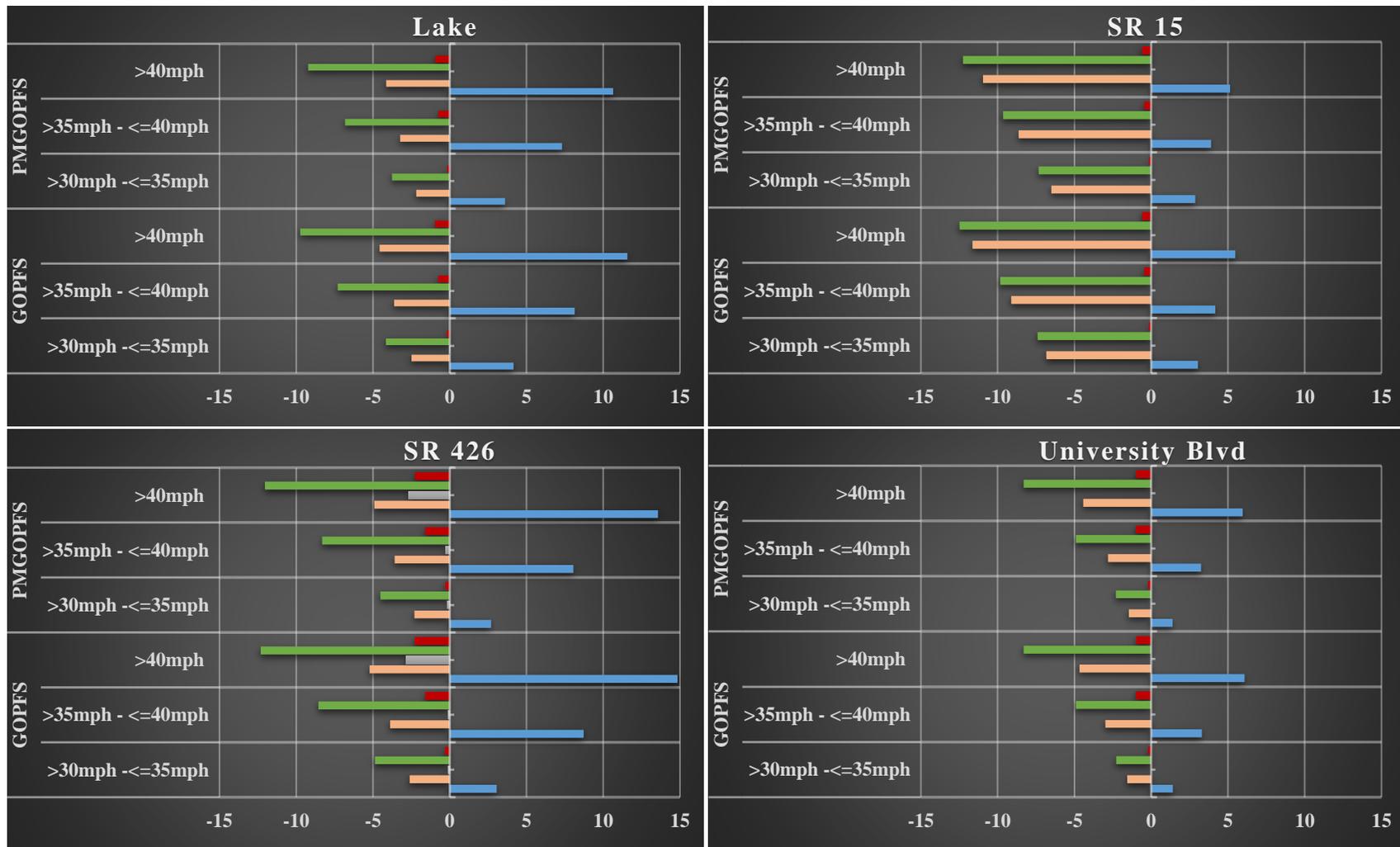
FIGURE 3 Elasticity effect analysis at road specific level[2]

---

[2] ■ Industrial area   ■ AADT   ■ Average Bike Lane   ■ Intersection Density   ■ Length

GOPFS= Generalized ordered probit fractional split model, PMGOPFS= Panel mixed generalized ordered probit fractional split mode

TABLE 1 Brief Summary of Earlier Research

Study	Facility Type	Roadway type	Application areas	Dependent Variable	Model Structure
Studies Considering Partial Vehicle Speed Profile					
(Garber and Gadirau, 1988)	Rural	Interstate, arterial and major collector roads	Road safety	85th and 50th percentile speed	Regression and Analysis of Variance (ANOVA) analysis
(Kanellaidis et al., 1990)	Rural	Two lane roadways	Promoting road design consistency	85th percentile speed	Linear regression analysis
(Lamm et al., 1990)	Rural	Two lane roadways	Geometric design consistency evaluation	85th percentile speed	Ordinary linear regression analysis
(Krammes et al., 1995)	Rural	Two lane roadways	Geometric design consistency evaluation	85th percentile speed	Ordinary least square linear regression model (OLS)
(Poe et al., 1996)	Urban	Low-speed collector street	Geometric design	85th percentile speed	Single regression equation
(Liang et al., 1998)	Rural	Interstate highway	Road safety	Mean speed	Regression model
(Gattis and Watts, 1999)	Urban	Two lane street	Road safety	85th percentile speed	Graphical comparison, linear regression analysis
(Fambro et al., 2000)	Rural	Multilane and two-lane roadway	Geometric design and safety	Mean speed	A comparison study and regression analysis
(Poe and Mason Jr, 2000)	Urban	Low-speed collector street	Geometric design	85th percentile speed	A mixed effect model
(Polus et al., 2000)	Rural	Two lane highways	Geometric design	85th percentile speed	Linear regression analysis
(Fitzpatrick et al., 2001)	Urban and Suburban	Arterial road	Design consistency evaluation	85th percentile speed	Multiple regression analysis
(McFadden et al., 2001)	Rural	Two lane roadways	Methodological	85th percentile speed	Artificial neural network (ANN) models
(Fitzpatrick et al., 2003)	Urban and Suburban	Arterial, collector and local	Policy on geometric design	85th percentile speed	Linear regression model

(Gong and Stamatiadis, 2008)	Rural	Four lane highways	Geometric design consistency evaluation	85th percentile speed	Ordinary least square linear regression model (OLS)
(Perco, 2008)	Rural	Two lane roadways	Geometric design consistency evaluation	85th percentile speed	Regression equations
(Himes and Donnell, 2010)	Rural and urban	Four lane highways	Evaluate speed consistency	Mean speed	A simultaneous equations framework, three stage least square estimator (3SLS)
(Dinh et al., 2013)	Urban	Residential street	Road Safety	Mean and 85th percentile speed	Single equation regression (SER), Simultaneous Equation Approach (SEA) and
(Jacob et al., 2013)	Rural	Two lane roadways	Geometric design consistency evaluation	85th percentile speed	Regression equations and Analysis of Variance (ANOVA) analysis
(Semeida, 2014)	Rural	Multi-lane highways	Road safety	85th percentile speed	Multiple linear regression model and Artificial neural network (ANN) models
(Russo et al., 2015)	Rural	Two lane roadways	Road Safety	85th percentile speed	Comparison study, graphical plotting
(Islam and El-Basyouny, 2015)	Urban	Collector and local street	Policy on geometric design	Mean speed (free flow)	Varying intercept multilevel model
(Anastasopoulos and Mannering, 2016)	Rural and urban	Interstate highways	Road Safety	55mph, 65mph and 70mph	Fixed and random parameter seemingly unrelated regression equation
(Wang et al., 2018)	Rural	Two lane highways (curve)	Road Safety	45-55mph speed	Generalized estimation equation (GEE)
Studies Considering Full Vehicle Speed Profile					
(Tarris et al., 1996)	Urban	Low-speed collector street	Geometric design	Entire speed profile – Individual vehicle operating speed	Regression and Panel analysis
(Figueroa Medina and Tarko, 2005)	Rural	Two-lane highways	Road safety	Entire speed profile Mean speed and speed dispersion, any percentile speed	Ordinary least square regression with panel data (OLS-PD)
(Ko and Guensler, 2005)	Urban	Freeway segment	Traffic flow modeling	15 minutes mean speed	Gaussian Mixture Model

(Park et al., 2010)	Urban	Interstate highway	Traffic flow modeling	24 hours speed data, Morning, evening and off-peak	Bayesian Mixture Model
(Eluru et al., 2013)	Urban	Local and arterial roads	Methodological and speed compliance	Entire speed profile grouped as proportions	Panel mixed ordered probit fractional split model

TABLE 2 Descriptive Statistics of Vehicle Speed Related Dependent Variables

Variable Names (N=8160)	Definition	Minimum	Maximum	Average	Standard deviation
Speed ≤ 20 mph	Proportion of vehicle speed ≤ 20 mph	0.000	1.000	0.163	0.300
Speed 20-25 mph	Proportion of speed greater than 20 mph and ≤ 25 mph	0.000	1.000	0.214	0.260
Speed 25-30 mph	Proportion of speed greater than 25 mph and ≤ 30 mph	0.000	1.000	0.268	0.270
Speed 30-35 mph	Proportion of speed greater than 30 mph and ≤ 35 mph	0.000	1.000	0.209	0.261
Speed 35-40 mph	Proportion of speed greater than 35 mph and ≤ 40 mph	0.000	1.000	0.105	0.213
Speed ≥ 40mph	Proportion of speed greater than 40 mph	0.000	1.000	0.041	0.153

TABLE 3 Summary Characteristics for Independent Variables

Variable Names (N=11040)	Definition	Minimum	Maximum	Average	Standard Deviation
Roadway Characteristics					
Length	Length of the segment in mile	0.003	3.289	0.668	0.674
Average median width	Ln (Average median width in the segment, in feet)	0.000	4.282	3.047	0.451
Median with hard shoulder	Presence of median with hard shoulder (Dummy variable)	0.000	1.000	0.818	0.386
Median with soft shoulder	Presence of median with soft shoulder (Dummy variable)	0.000	1.000	0.394	0.489
Maximum number of lane	Maximum number of lane present in the segment	1.000	4.000	2.614	0.666
Minimum number of lane	Minimum number of lane present in the segment	1.000	4.000	2.351	0.675
Average number of lane	(Maximum number of lane + Minimum number of lane)/2	1.000	4.000	2.482	0.618
Lane drop	Lane drop in the segment, (Maximum number of lane - Minimum number of lane)	0.000	3.000	0.264	0.520
Average inside shoulder width	Ln of (average width of shoulder adjacent to the median in feet in segment)	0.000	2.325	0.587	0.580
Average outside shoulder width	Ln of (average width of shoulder adjacent to the outer lane in feet in segment)	0.693	2.565	1.385	0.323
Average sidewalk width	Ln of (average width of sidewalk in feet in segment)	0.000	2.428	1.740	0.506
Average posted speed limit	Ln (average posted speed limit in mile per hour in a segment)	3.434	3.964	3.793	0.103
Bike lane length	Ln of (total length of bike lane in a segment in feet)	0.000	10.323	1.482	3.134
Average bike lane length	Ln of (average length of bike lane in a segment in feet).	0.000	8.993	1.277	2.725
Intersection density	Ln of (total number of intersection in the segment / length of segment)	0.000	5.693	2.285	1.379
Land Use Attribute					
Proportion of residential area	Proportion of residential area within 1-mile buffer of the roadway segment	0.011	0.670	0.411	0.119

Proportion of industrial area	Proportion of industrial area within 1-mile buffer of the roadway segment	0.000	0.166	0.029	0.033
Proportion of institutional area	Proportion of institutional area within 1-mile buffer of the roadway segment	0.000	0.440	0.041	0.058
Proportion of office area	Proportion of office area within 1-mile buffer of the roadway segment	0.005	0.255	0.106	0.054
Proportion of recreational area	Proportion of recreational area within 1-mile buffer of the roadway segment	0.000	0.145	0.022	0.026
Land use mix	Land use mix $= \left[\frac{-\sum_k (p_k(lnp_k))}{lnN}\right]$, where k is the category of land-use, p is the proportion of the developed land area devoted to a specific land-use, N is the number of land-use categories within 1-mile buffer of the roadway segment	0.309	0.895	0.639	0.124
Proportion of urban area	Proportion of urban area within 1-mile buffer of the segment	0.553	1.000	0.991	0.053
Built Environment					
Business center	Number of business center within 1-mile buffer of the roadway segment	0.000	5.000	0.370	0.804
Commercial center	Number of commercial center within 1-mile buffer of the roadway segment	0.000	8.000	1.543	1.789
Educational center	Z score[3]: Number of educational center within 1-mile buffer of the roadway segment	-1.565	3.468	0.000	1.000
Recreational center	Z score: Number of recreational center within 1-mile buffer of the roadway segment	-0.844	5.959	0.000	1.000
Restaurant	Ln (Number of restaurant within 1-mile buffer of the roadway segment)	0.000	5.106	3.439	0.746
Shopping center	Ln (Number of shopping center within 1-mile buffer of the roadway segment)	0.000	5.252	3.634	0.819
Financial center	Ln (Number of financial center within 1-mile buffer of the roadway segment)	0.000	4.407	2.511	0.798
Parking facilities	Z score: Number of parking Facilities within 1-mile buffer of the roadway segment	-0.229	6.777	0.000	1.000

---

[3] Z-score represents the standardized form of the actual variable.

Traffic Characteristic					
AADT	Ln of (average annual daily traffic in the roadway segment/100)	0.595	9.200	5.221	2.113
Truck AADT	Ln of (average annual daily truck traffic in the roadway segment/100)	0.031	6.059	2.524	1.642
Proportion of heavy traffic	Total number of truck traffic/ Total number of vehicles in the roadway segment	0.019	0.094	0.048	0.013
Peak traffic portion	Total traffic in peak direction as a percentage	0.510	0.688	0.536	0.013
Hour AADT	Proportion of aadt occurring in an hour, in this case, it is 30th highest hour (represent as percentage)	9.000	9.500	9.003	0.034
Environmental Factor					
Maximum air temperature	Z score: Maximum air temperature at 2-meter depth in degree Celsius	-3.226	2.026	0.000	1.000
Minimum air temperature	Z score: Minimum air temperature at 2-meter depth in degree Celsius	-3.011	2.140	0.000	1.000
Average air temperature	Z score: Average air temperature at 2-meter depth in degree Celsius	-3.119	2.093	0.000	1.000
Average humidity	Z score: Average relative humidity at 2meter depth in percentage	-2.316	1.634	0.000	1.000
Maximum dew point temperature	Z score: Maximum dew point temperature at 2-meter depth in degree Celsius	-2.727	1.614	0.000	1.000
Minimum dew point temperature	Z score: Minimum dew point temperature at 2-meter depth in degree Celsius	-2.636	1.632	0.000	1.000
Average dew point temperature	Z score: Average dew point temperature at 2-meter depth in degree Celsius	-2.691	1.585	0.000	1.000
Average wind speed	Z score: Average wind speed at 10 m depth (mph)	-1.746	2.704	0.000	1.000
Maximum wind speed	Z score: Maximum wind speed at 10 m depth (mph)	-1.736	2.990	0.000	1.000
Temperature difference	Difference between average air and dew point temperature (Air-Dew point)	-1.434	3.051	0.000	1.000
Average Precipitation	Ln (Total amount of precipitation in the road segment in inches)	0.000	0.662	0.003	0.032
Rain	A dummy variable indicating the occurrence of rain	0.000	1.000	0.022	0.146

Average solar radiation rate	Ln (Average solar radiation rate)	0.001	6.815	5.153	1.498
Road Specific Attribute					
Lake	Dummy for Lake Underhill road	0.000	1.000	0.035	--
SR 15	Dummy for SR 15 road	0.000	1.000	0.016	--
SR 426	Dummy for SR 426 road	0.000	1.000	0.122	--
SR 434	Dummy for SR 434 road	0.000	1.000	0.242	--
SR 436	Dummy for SR 436 road	0.000	1.000	0.196	--
SR 50	Dummy for SR 50 road	0.000	1.000	0.207	--
SR 551	Dummy for SR 551 road	0.000	1.000	0.114	--
UBlvd	Dummy for University boulevard road	0.000	1.000	0.068	--

TABLE 4 Panel Mixed Generalized Ordered Probit Fractional Split (PMGOPFS) Model Results

Variable[4]	Propensity	Threshold between 20-25mph and 25-30	Threshold between 25-30mph and 30-35mph	Threshold between 30-35mph and 35-40mph	Threshold between 35-40mph and >40mph
	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)
Constant	-1.577 (-15.019)	-0.292 (-12.696)	-0.262 (-9.357)	-0.382 (-14.148)	-0.290 (-6.170)
Roadway Characteristics					
Length	0.558 (10.528)	---[5]	---	---	---
Lane drop	-0.307 (-9.303)	---	---	---	---
Standard deviation (at hourly resolution)	0.248 (12.400)				
Average inside shoulder width	0.171 (6.577)	---	---	---	---
Average sidewalk width	-0.109 (-1.946)	---	---	---	---
Intersection density	-0.097 (-4.042)	---	---	---	---
Average bike lane length	-0.232 (-2.729)	---	---	---	0.448 (3.420)
Traffic Characteristic					
AADT	-0.094 (-5.529)	---	---	---	---
Land Use Characteristic					
Industrial area	-2.216 (-2.353)	---	---	2.009 (2.438)	---
Shopping Center	---	---	0.148 (3.364)	---	---
Environmental Characteristic					
Average Precipitation	-0.522 (-2.212)	---	---	0.748 (2.588)	---
Road Specific Characteristics					
Lake	-0.608 (-13.511)	---	---	0.551 (5.989)	---
SR 15	-0.252 (-5.250)	---	0.485 (5.052)	---	---

---

4 Please see Table 3 for variable definitions and units

[5] ---= attribute insignificant at 90% significance level

30

SR 426	-0.048 (-1.778)	---	---	0.401 (10.025)	---
UBlvd	0.106 (2.865)	---	---	---	---
SR 551	---	-0.073 (-6.636)	---	---	---
Unobserved Effects					
Roadway specific unobserved effect	0.045 (1.957)				
Day specific unobserved effect	0.153 (3.477)				
Direction specific unobserved effect	0.020 (1.767)				
Log-Likelihood at convergence (N=11040): -17388.07, AIC:34836.14, BIC: 35066.48					

TABLE A.1 Ordered Probit Fractional Split (OPFS) Model Results

Variable[6] (N=11040)	Propensity	Threshold between 20-25mph and 25-30	Threshold between 25-30mph and 30-35mph	Threshold between 30-35mph and 35-40mph	Threshold between 35-40mph and >40mph
	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)
Constant	-1.638 (-38.571)	-0.902 (-21.500)	-0.149 (-3.554)	0.596 (14.178)	1.352 (31.343)
Roadway Characteristics					
Length	0.559 (28.265)	---	---	---	---
Lane drop	-0.211 (-10.704)	---	---	---	---
Average inside shoulder width	0.180 (21.715)	---	---	---	---
Average sidewalk width	-0.101 (-5.640)	---	---	---	---
Intersection density	-0.099 (-15.094)	---	---	---	---
Average bike lane length	-0.263 (-5.012)	---	---	---	--
Traffic Characteristic					
AADT	-0.101 (-15.685)	---	---	---	--
Land Use Characteristic					
Industrial area	-2.657 (-10.170)	---	---	---	---
Environmental Characteristic					
Average Precipitation	-0.610 (-2.292)	---	---	---	---
Road Specific Characteristics					
Lake	-0.835 (-9.354)	---	---	---	---
SR 15	-0.307 (-6.766)	---	---	---	---
SR 426	-0.234 (-9.294)	---	---	---	---
UBlvd	0.111 (2.927)	---	---	---	---
Log-Likelihood at convergence (N=11040): -17518.38, AIC:35072.76, BIC: 35210.97					

---

[6] Please see Table 3 for variable definitions and units

TABLE A.2 Generalized Ordered Probit Fractional Split (GOPFS) Model Results

Variable (N=11040)[7]	Propensity	Threshold between 20-25mph and 25-30	Threshold between 25-30mph and 30-35mph	Threshold between 30-35mph and 35-40mph	Threshold between 35-40mph and >40mph
	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)
Constant	-1.578 (-37.571)	-0.303 (-27.545)	-0.268 (-26.800)	-0.383 (-22.529)	-0.292 (-16.222)
Roadway Characteristics					
Length	0.555 (29.211)	---	---	---	---
Lane drop	-0.217 (-10.850)	---	---	---	---
Average inside shoulder width	0.173 (21.625)	---	---	---	---
Average sidewalk width	-0.103 (-5.722)	---	---	---	---
Intersection density	-0.101 (-16.833)	---	---	---	---
Average bike lane length	-0.175 (-3.241)	---	---	---	0.449 (4.582)
Traffic Characteristic					
AADT	-0.074 (-12.333)	---	---	---	---
Land Use Characteristic					
Industrial area	-1.409 (-5.337)	---	---	2.016 (4.978)	---
Shopping Center	---	---	0.147 (13.364)	---	---
Environmental Characteristic					
Average Precipitation	-1.163 (-3.969)	---	---	0.745 (1.725)	---
Road Specific Characteristics					
Lake	-0.778 (-17.289)	---	---	0.515 (5.598)	---
SR 15	-0.234 (-4.875)	---	0.487 (5.096)	---	---
SR 426	-0.179 (-6.630)	---	---	0.363 (9.075)	---
UBlvd	0.087 (-2.351)	---	---	---	---
SR 551	---	-0.303 (-27.545)	---	---	---
Log-Likelihood at convergence (N=11040): -17437.79, AIC:34927.58, BIC: 35127.21					

---

[7] Please see Table 3 for variable definitions and units

TABLE A.3 Elasticity Analysis at Road Level (All 3 Models)

Road	Variable	Model	Speed<=20mph	Speed >20mph and <=25mph	Speed >25mph and <=30mph	Speed >30mph and <=35mph	Speed >35mph and <=40mph	Speed >40mph
Lake	Length (+10%)	OPFS*	-2.422%	-0.677%	1.348%	3.834%	6.588%	9.810%
		GOPFS	-2.400%	-0.722%	1.263%	4.162%	8.146%	11.573%
		PMGOPFS	-2.449%	-0.849%	0.948%	3.600%	7.314%	10.651%
	Intersection density (+10%)	OPFS	2.344%	-0.384%	-1.448%	-2.306%	-3.077%	-3.909%
		GOPFS	2.414%	-0.378%	-1.404%	-2.497%	-3.630%	-4.576%
		PMGOPFS	2.423%	-0.204%	-1.173%	-2.188%	-3.235%	-4.135%
	Average length of bike lane (+10%)	OPFS	0.001%	0.000%	0.000%	-0.002%	-0.004%	-0.007%
		GOPFS	0.001%	0.000%	0.000%	-0.001%	-0.001%	-0.011%
		PMGOPFS	0.001%	0.001%	0.000%	-0.001%	-0.002%	-0.011%
	AADT (+10%)	OPFS	3.099%	0.375%	-1.865%	-4.238%	-6.586%	-9.069%
		GOPFS	2.823%	0.386%	-1.591%	-4.164%	-7.308%	-9.733%
		PMGOPFS	2.943%	0.597%	-1.285%	-3.759%	-6.817%	-9.223%
	Proportion of industrial Area (+10%)	OPFS	0.240%	-0.017%	-0.147%	-0.266%	-0.377%	-0.494%
		GOPFS	0.194%	-0.011%	-0.112%	-0.182%	-0.750%	-0.962%
		PMGOPFS	0.207%	0.005%	-0.097%	-0.165%	-0.733%	-0.952%
SR 15	Length (+10%)	OPFS	-1.532%	0.003%	1.194%	2.506%	3.912%	5.556%
		GOPFS	-1.548%	-0.077%	1.448%	3.038%	4.175%	5.476%
		PMGOPFS	-1.433%	0.097%	1.491%	2.875%	3.894%	5.139%
	Intersection density (+10%)	OPFS	3.659%	-0.396%	-2.885%	-5.279%	-7.564%	-9.975%
		GOPFS	3.848%	-0.282%	-3.292%	-6.825%	-9.122%	-11.644%
		PMGOPFS	3.326%	-0.483%	-3.236%	-6.509%	-8.628%	-10.949%
	Average length of bike lane (+10%)	OPFS	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
		GOPFS	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
		PMGOPFS	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
	AADT (+10%)	OPFS	4.464%	-0.399%	-3.522%	-6.591%	-9.563%	-12.712%
		GOPFS	4.155%	-0.186%	-3.668%	-7.403%	-9.844%	-12.486%
		PMGOPFS	3.848%	-0.530%	-3.818%	-7.324%	-9.648%	-12.250%
	Proportion of industrial Area (+10%)	OPFS	0.125%	-0.004%	-0.098%	-0.198%	-0.302%	-0.422%
		GOPFS	0.103%	0.001%	-0.093%	-0.153%	-0.454%	-0.598%
		PMGOPFS	0.098%	-0.007%	-0.100%	-0.158%	-0.454%	-0.592%

SR 426	Length (+10%)	OPFS	-3.604%	-1.384%	0.416%	2.744%	6.369%	14.662%
		GOPFS	-3.606%	-1.479%	0.461%	3.055%	8.728%	14.836%
		PMGOPFS	-3.555%	-1.572%	0.186%	2.687%	8.053%	13.571%
	Intersection density (+10%)	OPFS	3.421%	0.591%	-1.088%	-2.476%	-3.562%	-4.580%
		GOPFS	3.603%	0.711%	-1.116%	-2.617%	-3.874%	-5.228%
		PMGOPFS	3.394%	0.840%	-0.834%	-2.303%	-3.594%	-4.908%
	Average length of bike lane (+10%)	OPFS	0.232%	0.109%	0.003%	-0.165%	-0.502%	-1.390%
		GOPFS	0.151%	0.077%	-0.003%	-0.128%	-0.113%	-2.845%
		PMGOPFS	0.199%	0.111%	0.008%	-0.152%	-0.240%	-2.677%
	AADT (+10%)	OPFS	6.972%	1.489%	-1.953%	-5.161%	-8.275%	-12.270%
		GOPFS	6.477%	1.582%	-1.822%	-4.884%	-8.552%	-12.314%
		PMGOPFS	6.483%	1.877%	-1.379%	-4.521%	-8.316%	-12.039%
	Proportion of industrial Area (+10%)	OPFS	0.869%	0.149%	-0.260%	-0.616%	-0.945%	-1.342%
		GOPFS	0.717%	0.148%	-0.198%	-0.318%	-1.606%	-2.284%
		PMGOPFS	0.732%	0.200%	-0.160%	-0.294%	-1.602%	-2.264%
University Blvd	Length (+10%)	OPFS	-3.987%	-2.201%	-0.516%	1.401%	3.554%	6.490%
		GOPFS	-3.834%	-2.119%	-0.445%	1.404%	3.304%	6.085%
		PMGOPFS	-3.833%	-2.092%	-0.424%	1.391%	3.243%	5.958%
	Intersection density (+10%)	OPFS	4.526%	1.920%	0.119%	-1.536%	-3.061%	-4.741%
		GOPFS	4.588%	1.842%	0.028%	-1.557%	-2.988%	-4.666%
		PMGOPFS	4.310%	1.745%	0.036%	-1.462%	-2.819%	-4.429%
	Average length of bike lane (+10%)	OPFS	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
		GOPFS	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
		PMGOPFS	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
	AADT (+10%)	OPFS	7.253%	3.566%	0.538%	-2.554%	-5.673%	-9.425%
		GOPFS	6.415%	3.114%	0.383%	-2.283%	-4.905%	-8.309%
		PMGOPFS	6.488%	3.123%	0.370%	-2.293%	-4.901%	-8.302%
	Proportion of industrial Area (+10%)	OPFS	1.766%	0.529%	-0.123%	-0.563%	-0.828%	-0.973%
		GOPFS	1.340%	0.407%	-0.129%	-0.196%	-0.987%	-0.967%
		PMGOPFS	1.373%	0.413%	-0.135%	-0.207%	-0.993%	-0.973%

* Note: OPFS= Ordered probit fractional split model, GOPFS= Generalized ordered probit fractional split model,

PMGOPFS= Panel mixed generalized ordered probit fractional split model.