

# **MODELING VEHICLE OPERATING SPEED ON URBAN ROADS IN MONTREAL: A PANEL MIXED ORDERED PROBIT FRACTIONAL SPLIT MODEL**

**Naveen Eluru\***

Assistant Professor

Department of Civil Engineering and Applied Mechanics

McGill University

Ph: 514 398 6823, Fax: 514 398 7361

Email: [naveen.eluru@mcgill.ca](mailto:naveen.eluru@mcgill.ca)

**Vincent Chakour**

Master's Student

Department of Civil Engineering and Applied Mechanics

McGill University

Ph: 514 398 6823, Fax: 514 398 7361

Email: [vincent.chakour@mail.mcgill.ca](mailto:vincent.chakour@mail.mcgill.ca)

**Morgan Chamberlain**

Master's Student

Department of Civil Engineering and Applied Mechanics

McGill University

Ph: 514 398 6823, Fax: 514 398 7361

Email: [morgan.chamberlain@mail.mcgill.ca](mailto:morgan.chamberlain@mail.mcgill.ca)

**Luis F. Miranda-Moreno**

Assistant Professor

Department of Civil Engineering & Applied Mechanics

McGill University

Ph: 514 398 6589, Fax: 514 398 7361

Email: [luis.miranda-moreno@mcgill.ca](mailto:luis.miranda-moreno@mcgill.ca)

Submission Date: May 12 2013

\* Corresponding author

## ABSTRACT

Vehicle operating speed measured on roadways is a critical component for a host of analysis in the transportation field including transportation safety, traffic flow modeling, roadway geometric design, vehicle emissions modeling, and road user route decisions. The current research effort contributes to literature on examining vehicle speed on urban roads methodologically and substantively. In terms of methodology, we formulate a new econometric model framework for examining speed profiles. The proposed model is an ordered response formulation of a fractional split model. The ordered nature of the speed variable allows us to propose an ordered variant of the fractional split model in literature. The proposed formulation allows us to model the proportion of vehicles traveling in each speed interval for the entire segment of roadway. We extend the model to allow the influence of exogenous variables to vary across the population. Further, we develop a panel mixed version of the fractional split model to account for the influence of site-specific unobserved effects. The paper contributes substantively by estimating the proposed model using a unique dataset from Montreal consisting of weekly speed data (collected in hourly intervals) for about 50 local roads and 70 arterial roads. We estimate separate models for local roads and arterial roads. The model estimation exercise considers a whole host of variables including geometric design attributes, roadway attributes, traffic characteristics and environmental factors. The model results highlight the role of various street characteristics including number of lanes, presence of parking, presence of sidewalks, vertical grade, and bicycle route on vehicle speed proportions. The results also highlight the presence of site-specific unobserved effects influencing the speed distribution. The parameters from the modeling exercise are validated using a hold-out sample not considered for model estimation. The results indicate that the proposed panel mixed ordered probit fractional split model offers promise for modeling such proportional ordinal variables.

**Key words:** Fractional split models, Mixed ordered response fractional split models, vehicle operating speed, Montreal

# 1 INTRODUCTION

Vehicle operating speed measured on roadways is a critical component for a host of analysis in the transportation field including transportation safety, traffic flow modeling, roadway geometric design, vehicle emissions modeling, and road user route decisions (TRC 2011). Vehicle operating speed is an important precursor for identifying potentially dangerous roadway facilities (often referred to as a surrogate safety measure). Roadway facilities with very high observed speeds are linked with crashes resulting in severe injuries and fatalities (Traffic Safety Facts 2009; Eluru and Bhat 2007; Yasmin and Eluru, 2013). In terms of traffic flow modeling, vehicle speed is one of the most crucial traffic flow parameters; when high vehicle speeds are observed the flow patterns resemble free-flow conditions while very low vehicle speeds imply high congestion on the facility. In fact, vehicle speed profiles across the day provide information on whether a roadway facility is under-utilized or if the roadway facility has insufficient capacity. Moreover, many traffic flow modules (analytical and simulation based) require vehicle speed as input (Park et al., 2010).

Vehicle operating speed also has an important role in geometric design of roadways. Roadway geometric design is undertaken based on an assumed design speed; if the vehicle operating speeds are significantly different from the design speeds considered, the geometric design might be inefficient and in some cases unsafe. Another area of transportation research that has started using vehicle speed is the emission modeling community. With the recent focus on global climate change, measuring transport emissions has received increased attention; an integral part of modeling vehicle emissions is generating accurate estimates of vehicle speed. Furthermore, road users use vehicle speed (closely linked to travel time) as a measure to compare various route options; in fact, roadways where vehicle speeds are perceived to be lower are less preferred by road users (TRC 2011).

The current research effort contributes to literature on modeling vehicle speed on urban roads. Consider, for instance, that speed measure in an urban region is compiled as number of vehicles within three speed categories: (1) less than 30kmph, (2) 30-60 kmph and (3) greater than 60 kmph. The information compiled provides two measure: (1) Total vehicle volume and (2) vehicular volume in each speed category. The data thus compiled offers substantial challenges to model the speed variable. Traditional approaches to modeling speed are not readily useful to examine such speed category data. In our research, we reformulate the above information as: (1) total vehicles on the roadway segment and (2) proportion of vehicles in each speed category. The modeling of the first component of the analysis can be readily undertaken through linear regression/count models. The focus of the current research effort is on the second measure - proportion of vehicles in each speed category.

In our study we propose to examine the influence of exogenous variables (including geometric design attributes, roadway attributes, traffic characteristics and environmental factors) on the proportion of vehicles in each speed category by developing a new econometric framework. Specifically, we propose an ordered response equivalent of the fractional split model (for an unordered version of the fractional split models see Sivakumar and Bhat, 2002, Papke and Woolridge, 1996). The ordered nature of the speed variable allows us to propose an ordered variant of the fractional split model. The ordered fractional split model considers the proportion

of vehicles traveling in various speed ranges for the entire segment of roadway as the dependent variable. The model implicitly recognizes that the proportions across all speed ranges sum to unity.

We further extend the ordered fractional split model to incorporate the influence of unobserved variables on speed proportion and also the possibility of exogenous variable effects varying across the dataset. Towards this end, we formulate a mixed version of the fractional split model. Further, within the mixed version of the model we also recognize that multiple observations from the same facility are bound to be correlated. We incorporate site-specific unobserved effects in a panel version of the model. In fact, the formulated model is the first instance of a panel mixed ordered fractional split model not only in transportation literature but also in econometrics literature.

The paper contributes substantively to vehicle speed literature by estimating the unique model using a dataset from Montreal consisting of weekly speed data (collected in hourly intervals) for about 50 local roads and 70 arterial roads. We estimate separate models for local roads and arterial roads. The model estimation exercise considers a whole host of variables including geometric design attributes, roadway attributes, traffic characteristics and environmental factors. To validate the model developed, we examine the model prediction on a hold-out sample not considered for model estimation.

The remainder of the paper is organized as follows. In Section 2, we provide a review of existing literature on vehicle speed modeling and discuss the context for our study. Section 3 presents the methodology of the panel mixed ordered response fraction split model. The data collection and assembly procedures are outlined in section 4. The empirical results from the model estimation are presented in Section 5. Section 6 concludes the paper and identifies limitations of the current study and proposes direction for future research.

## **2 EARLIER RESEARCH AND CURRENT STUDY IN CONTEXT**

Given the wide ranging applications involving vehicle speed it is not surprising that a number of studies have investigated vehicle speed on roadways. The earlier research efforts can be broadly classified based on the application focus of these studies into two categories: (1) Macro (low resolution) studies and (2) Micro (High resolution) studies. Macro studies refer to earlier research efforts that consider a single measure of the entire speed profile (such as the 85<sup>th</sup> percentile speed or the mean) neglecting the overall distribution of the speed at a site. Micro studies refer to approaches that consider the entire speed profile in their analysis.

### **2.1 Macro studies**

A number of earlier studies have focussed on driver behaviour along horizontal curves on rural highways (see TRC 2011 for a review of such models from various parts of the world; Donnell 2001). These studies employ the 85<sup>th</sup> percentile speed or the mean speed as the dependent variable of analysis. Most of these studies identify the radius of curvature to be the key variable in predicting operating speeds. Kanellaidis et al. (1990) developed a model based on degree of curvature on the basis of 85<sup>th</sup> percentile speed. Similarly, using 8 sites in Utah, Islam and

Seneviratne (1994) developed three speed models for speeds at the beginning, middle, and end of horizontal curves. Krammes et al. (1995) developed three models of operating speed for 138 horizontal curves in the United States. These models were developed to evaluate the consistency of design in horizontal curves and highlighted the degree of curvature, length of the curve, and the deflection angle as the relevant variables. Later, McFadden et al. (2001) used Krammes et al. (1995) data set to develop two backward propagating artificial neural network (ANN) models. They compared these models to the OLS regression models previously developed by Krammes et al. (1995) and determined the results to be comparable.

Models that consider the vehicle speeds on tangents of horizontal curves have also been developed. The interaction between the tangents and the curves has been found to have a distinct impact on operating speed. Fambro et al. (2000) studied vertical tangent curve sites and found that site distance played a large role in influencing driver speeds, and that volume seemed to have little impact on speed reductions. Furthermore, this study found the mean operating speeds at the study sites to be notably higher than the design speeds. Andueza (2000) found the relevant variables in a rural speed model to be the radii of consecutive curves, the tangent length before the curve, and the minimum sight distance for the horizontal curve. Polus et al. (2000) studied 162 rural sites and found that speeds on short tangents were controlled by the curves immediately before and after the tangent length. Speed along longer tangents was impacted by geometry, speed limit and speed enforcement.

Research in urban areas is lacking in comparison to speed research on rural highways. In Fayetteville, Arkansas, Gattis et al. (1999) studied the link between average operating speed and street width on urban streets. They found that street width was less relevant to determining speed than the street's function and suggested that the opportunity for uninterrupted travel distance was a more likely cause of high speeds than the actual geometry and classification of the road itself. Fitzpatrick et al. (2003) in a study of 78 urban and suburban sites around the United States found that of a variety of potential variables including roadway geometry, traffic control, and roadside features, the only statistically significant variable was the posted speed limit.

Ali et al. (2007) studied the free-flow speeds of vehicles on four-lane urban streets in Virginia. They found that the geometric design features that influenced free-flow speed were median width and segment length – signal spacing ratio (defined as the ratio of the study segment length to the maximum signal spacing). The posted speed limit was also significant. Interestingly, sites with a median were also found to typically have lower speeds than comparative sites without medians. Poe and Mason (2000) examined speed on urban collectors in Pennsylvania by developing a mixed model approach using both fixed and random effects. The study considered the influence of geometry and land use, with a mixed model developed for the midpoints of the studied curves recognizing the presence of common unobserved factors for repeated measures at the same site. They found several influential factors on speed at these sites: lane width, degree of curvature, grade, and roadside hazard rating.

## **2.2 Micro studies**

Tarris et al., (1996) conducted a one of its kind study where a comparison of vehicle speed measures at the aggregate level (such as 85<sup>th</sup> percentile) and measures at individual vehicle level. The study concluded that the aggregation models though provide higher data fit values are likely

to produce incorrect or biased results. The model considering operating speed was deemed to perform better than the aggregate model. Of course, the cost of conducting such data collection very often is prohibitively expensive. Figueroa and Tarko (2005) proposed and estimated a model to consider the entire speed profile by customizing the regression equation for required percentile. This research was conducted along rural highways and used panel data to account for multiple observations at each of 158 sites. The models developed allowed for the definition of any desired percentile speed and addressed the impacts on mean speed and dispersion simultaneously. Hewson, (2008) generalized this approach by employing quantile regression to speed modeling. The approach is flexible allowing the modeling of the 85<sup>th</sup> percentile or any other quantile. The approach was employed to test for change in the 85<sup>th</sup> percentile speed using data collected as part of a community watch program from Northamptonshire. Qin et al. (2010) employed the same approach to identify crash prone locations from Wisconsin. .

Another stream of research efforts have examined disaggregate speed data and focussed on matching the speed profile using parametric distributions - such as normal, log-normal or other variants (see Park et al., 2010 for a review of such studies). Recent studies have focussed on mixture based models that allow for fitting multi-modal parametric distributions to the speed data (Ko and Guensler, 2005; Park et al., 2010). These studies are a substantial improvement compared to previous research in modeling speed on roadway facilities. In particular, these studies explicitly control for vehicle flow conditions (proportion of heavy vehicles) and environmental conditions. However, these models examine the speed profile at a single site in their analysis. Naturally, these studies cannot examine the influence of geometric design, and roadway attributes on vehicle speed profiles.

## **2.3 Current study in context**

The majority of earlier research studies modeling vehicle speed have focussed on examining the 85<sup>th</sup> and 50<sup>th</sup> percentile speed as a function of various exogenous variables. These approaches though useful do not examine the speed profile in detail i.e. the speed data available is used to generate the 85<sup>th</sup> or 50<sup>th</sup> percentile value and remainder of the data is discarded. It is important to propose methods that consider entire speed distribution on urban regions. Of course, the consideration of speed-profiles for continuous speed distributions is far from straight forward. More so, if the data is available in speed categories i.e. instead of collecting vehicle speeds for each vehicle, data collection methods only count the number of vehicles in each speed range for a specified time interval. Effectively, the approach discretizes the speed distribution into speed categories (for example less than 30kmph, 30-60 kmph and greater than 60 kmph). Collecting data in speed categories happens often in practice when using certain types of popular technologies used in short-term (temporal) data collection (including pneumatic tubes and traffic analysers such as NC-100/200 sensors). The earlier approaches for speed modeling are not directly applicable to the speed category database. Of course, we can compile the 85<sup>th</sup> percentile and 50<sup>th</sup> percentile information through some computations from the data; however, we are likely to induce bias in the estimates by aggregating such speed range data.

The discretization of the variable does not lend itself to any discrete modeling approaches because unlike the discrete modeling approaches where only one of the possible alternatives are chosen, in our discretization case we have possible non-zero values (ranging between 0 and 1 for each category) for multiple categories. In some ways, we are modeling the proportion of vehicles

in each category (this is straightforward once we assume the total vehicle flow is available and we compute the proportion for all speed categories). In our current study, we propose an ordered response fractional split model that provides probability for proportion of vehicles in each speed category. Similar to the traditional ordered response model, a latent propensity is computed for each road segment with higher propensity indicating higher likelihood for higher speeds. In the traditional ordered response structure the thresholds are employed to generate one chosen alternative. However, in our case we have non-zero probabilities for all speed ranges i.e. we need to apportion the probability to different speed categories. The conventional maximum likelihood approaches are not suited for such models. Hence, we resort to a quasi-likelihood approach (proposed by Gouriéroux, et al., 1984; Papke and Woolridge, 1996; Sivakumar and Bhat, 2002). The proposed approach is the ordered response extension of the binary probit model proposed by Papke and Woolridge, (1996).

Of course, the ordered response fractional split model (or the unordered fractions split model employed in literature) capture the influence of exogenous variables on the dependent variable. Consider the influence of the number of lanes on vehicle speed. We expect vehicular speed to be positively influenced by number of lanes. This is an example of an exogenous variable that is observed that influences speed proportion. The effect of this variable is immediately considered in the proposed fractional split model. Now, consider another variable - the presence of parking on the roadway. Typically, the presence of parking is likely to reduce the vehicle speed; however, the influence is mainly a function of parking turnover rate (how often vehicles are entering and leaving the parking spots) and occupancy (percentage of parking spots occupied). The availability of such detailed information to the analyst is unlikely. Hence, considering that the presence of parking has the same influence on vehicle speed on all roadway facilities is erroneous because that leads to ignoring the impact of parking turnover rate and occupancy (missing information). Therefore, it is useful to allow for the effect of parking on vehicle speed to vary across the population by considering the parameter of the parking variable to follow a distribution across the population (see Eluru et al., 2008 for a similar discussion on influence of unobserved effects on modeling injury severity). Traditionally, such effects are captured in discrete choice literatures through mixed models. In our work, we propose a mixed extension of the ordered response fractional split model.

In a similar vein, the repeated measures at a particular roadway facility introduce common unobserved factors that influence vehicle speed. For instance, consider two identical roadway facilities. However, at one of the facilities, there is a large obstruction reducing the sight distance available to the drivers. This is the case of an unobserved variable that reduces vehicle speed across all observations at that facility. Hence we need a mechanism to capture such unobserved road facility specific impacts. In fact, many studies on vehicle speed have discussed the importance of site-specific effects on speed distribution modeling but only Poe and Mason (2000) and Tarris et al. (1996) facilitated the inclusion of random effects in these studies.

In summary, the current study proposes the ordered response fractional split model. The proposed formulation is further extended to (1) capture the impact of exogenous variables to vary across the population (similar to random coefficients ordered response model) and (2) incorporate the influence of site specific unobserved effects on the proportion variable (similar to

a panel random coefficients ordered response model). The detailed econometric methodology for the proposed approach is discussed subsequently.

### 3 METHODOLOGY

The formulation for the Panel Mixed Ordered Probit Fractional Split model (PMOPFS) is presented in this section.

#### 3.1 Model Structure

Let  $q$  ( $q = 1, 2, \dots, Q$ ) be an index to represent site,  $p$  ( $p = 1, 2, \dots, P$ ) be an index for different data collection period at site  $q$  and let  $k$  ( $k = 1, 2, 3, \dots, K$ ) be an index to represent speed category. The latent propensity equation for vehicle speed at the  $q$ th site and  $p$ th interval:

$$y_{qp}^* = (\alpha' + \delta_q') z_{qp} + \xi_q, \quad (1)$$

This latent propensity  $y_{qp}^*$  is mapped to the actual speed category proportion  $y_{qp k}$  by the  $\psi$  thresholds ( $\psi_0 = -\infty$  and  $\psi_K = \infty$ ).  $z_{qp}$  is an ( $L \times 1$ ) column vector of attributes (not including a constant) that influences the propensity associated with vehicle speed.  $\alpha$  is a corresponding ( $L \times 1$ )-column vector of mean effects, and  $\delta_q$  is another ( $L \times 1$ )-column vector of unobserved factors moderating the influence of attributes in  $z_q$  on the vehicle speed propensity for site  $q$ .  $\xi_q$  is an idiosyncratic random error term assumed to be identically and independently standard normal distributed across individuals  $q$ .

To complete the model structure in Equation (1), we need to specify the structure for the unobserved vector  $\delta_q$ . In the current paper, we assume that the  $\delta_q$  elements are independent realizations from normal population distributions.

#### 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. The parameters to be estimated in the Equation (1) are the  $\alpha$ ,  $\psi$  thresholds, and the variance terms  $\delta_q$ . To estimate the parameter vector, we assume that

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

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

$$P_{qpk} = \{ G[\psi_k - \{(\alpha' + \delta_q') z_q\}] - G[\psi_{k-1} - \{(\alpha' + \delta_q') z_q\}] \}$$

The proposed model ensures that the proportion for each speed category is between 0 and 1 (including the limits). Then, the quasi likelihood function (see Papke and Woolridge for a discussion on asymptotic properties of quasi likelihood proposed), for a given value of  $\delta_q$  vector may be written for site  $q$  for various repetitive measures as:

$$L_q(\alpha, \psi \mid \delta_q) = \prod_{p=1}^P \prod_{k=1}^K \left\{ G[\psi_k - \{(\alpha' + \delta'_q)z_q\}] - G[\psi_{k-1} - \{(\alpha' + \delta'_q)z_q\}] \right\}^{d_{qk}} \quad (3)$$

where  $G(\cdot)$  is the cumulative distribution of the standard normal distribution and  $d_{qk}$  is the proportion of vehicles in speed category  $k$ . Finally, the unconditional likelihood function can be computed for site  $q$  as:

$$L_q(\alpha, \psi, \delta_q) = \int_{\delta_q} L_q(\alpha, \psi \mid \delta_q) dF(\delta_q) \quad (4)$$

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

$$L(\Omega) = \sum_q L_q(\alpha, \psi, \delta_q). \quad (5)$$

The likelihood function in Equation (4) involves the evaluation of a multi-dimensional integral of size equal to the number of rows in  $\delta_q$ . In the current paper, we use a quasi-Monte Carlo (QMC) method proposed by Bhat (2001) for discrete choice models to draw realizations for  $\delta_q$ . Within the broad framework of QMC sequences, we specifically use the Halton sequence in the current analysis<sup>1</sup>. In our analysis, a rigorous examination of the influence of the number of halton draws was conducted. The final model specifications were estimated for 2000 draws per roadway facility.

## 4 DATA

The database for speed data was compiled in Montreal urban region and consisted of two roadway facility types: (1) local roads and (2) arterials. The data collection process for the two roadway facility types are described below.

---

<sup>1</sup> The reader is referred to Eluru and Bhat 2007 and Eluru et al., 2008 for earlier instances of Halton draw based maximum likelihood estimation in the literature.

## 4.1 Data Collection

The hourly speed data for local streets comes from an important data collection campaign that was planned and executed in 2009 to evaluate the impact of speed limit and geometric design factors on operating speeds. This was linked to the reduction of posted travel speed along local streets from 50 km/h to 40 km/h – as part of the 2008 Montreal Transportation plan. As shown in Figure 1, boroughs on the west side of the Montreal Island have speed limits of 40km/hr while boroughs in the east side have speed limits of 50km/hr.

The local streets were divided into segments based on intersections. The length of each street segment was then determined. Street segments shorter than 200 meters were excluded from the selection process, since it is difficult to reach and travel at a speed of 50 km/hr a street segment that is shorter than this length. Also two-way streets with two lanes in both directions<sup>2</sup> were dropped from the study. A random number was assigned to every street in each group (40km/h vs 50km/h), followed by the selection of a random sample of sites. Before site visits, the selected sites were examined using satellite images and Google on-street views. Site visits were then conducted at all the selected sites to validate the appropriateness of these sites for data collection. Finally, a sample of 58 sites was retained. Figure 1 shows the locations of the 58 sites. Note that at least one site from each borough with 50 km/h are present in the sample and that all the boroughs from the 40 km/h group are present as well.

The data was collected by a local consulting firm using NC-100 traffic sensors. Observed vehicle speeds and volumes over the course of five days, including three weekdays, plus both a Saturday and a Sunday in October 2009 were collected. The traffic sensor provides speed data in increment of 10kmph categories. In addition, roadway and street design characteristics were also collected as part of the data collection campaign including number of lanes, whether the street is one-way, the number of sides of the street that have a sidewalk (0, 1, or 2), whether a bike route is present, the width of the street (measured in meters), and parking characteristics (represented as presence of parking on one or two sides).

Using a similar protocol, speed and volume hourly data was collected for major arterials as part of the data collection program for the modernization of the signalized intersections in Montreal. Data collection sites were strategically selected with a focus on major arterials. Figure 2 shows the geographical locations of the sites. A similar traffic sensor was used for this purpose. NC-200 traffic analyzers were used in most of the cases – in few cases, the NC-97 model was used. The data was collected during 7 days during the period May-October 2009.

## 4.2 Data Compilation

The data collected was thoroughly examined for consistency and facilities with missing information were eliminated. The final samples for local roads and arterial roads consisted of 49 and 71 facilities respectively. The sample characteristics for the local roads and arterial sites are provided in Tables 1 and 2 respectively. For local roads, we can see that the speed limits 40 and 50 are well represented. Most sites have at least one side walk. Parking is permitted on both sides for most streets. Most streets have one lane in each direction. Data collection was undertaken on

---

<sup>2</sup> Please note that local roads with two lanes in one direction and one lane in the other direction are considered.

weekdays and weekends. The sample statistics clearly indicate the limited information available for the arterial road dataset. Further, the characteristics compiled are also different from the local roads database. The most common speed limit is the 50 kmph. The sites with parking permitted constitute about 30%. The number of lanes are much higher (4 and higher constitute about 85%). The estimation dataset was prepared while generating a hold-out sample of sites for both local roads and arterials to evaluate the model performance after estimation.

The dependent variable was computed as the proportion of vehicle in each category for every hour. The speed categories considered in our analysis are  $\leq 20$ , 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-120,  $>120$ . The same classification was employed for local roads as well as arterials. For the night hours we aggregated the records from 12am-6am and created a single record as there were very few vehicles in these intervals. For all other records we employed hourly records. Thus each day resulted in 19 records per site. The distribution of the dependent variable for local roads and arterial roads is presented in Tables 3 and 4 respectively. Specifically, the tables present the overall proportion distribution and a time of day based proportion distribution. From the tables, it is evident that majority of the flow on local roads is in the speed range 30-70 kmph (80%). The temporal distribution of the variable indicates a rightward shift in vehicle speed proportion for the morning rush hour period (6-10am). On the other hand, it is quite interesting to note that for the evening rush hour the vehicle operating speed shifts leftward – highlighting increased congestion during the evening.

For arterial roads, the vehicle speed proportions offer distinct speed profiles. In the arterial road system, the lowest speed category ( $\leq 20$  kmph) has a reasonably large share (~23%). It is also interesting to see that within arterial roads, the temporal variation in the speed variable is much smaller compared to local roads. Another striking statistic from the table is the presence of about ~5% of the vehicles operating at a speed greater than 80 kmph on arterial roads – a potential danger for non-motorists and other vehicles.

## 5 EMPIRICAL ANALYSIS

### 5.1 Variable considered

We examined a host of variables in the model specification. The street characteristics variables considered for local roads included speed limit, distance to next exit, distance from last exit, number of lanes, width of lanes, total width of street, width per lane, no. of sidewalks (1=one side, 2=two sides), no. of sides parking is allowed (1=one side, 2=two sides), as well as a number of dummy variables (one way street, presence of bicycle route, presence of good pavement, vertical alignment grade higher than 10%, vertical alignment grade lower than 4%, presence of horizontal curve, and Adequate sight distance available). Various temporal variables were also tested: night time (12am to 6am), rush hour (6am to 10am and 3pm to 7pm), evening time (7pm to 12am), and day time (10am to 3pm), and weekday. Finally, the daily total number of cars passing on the street in one day was tested.

Fewer street characteristic variables were available for the arterial roads. The variables tested include speed limit, number of lanes, average width of lanes, total width of street, presence of a median (dummy), width of the median, and presence of parking (dummy). The temporal variables

are the same as the local speed model, with the addition of the month component. The total number of cars passing on the street in one day was also tested for this model.

The final specification was based on a systematic process of removing statistically insignificant variables and combining variables when their effects were not significantly different. The specification process was also guided by prior research, intuitiveness and parsimony considerations. We should also note here that we tested alternative functional forms including linear and spline (or piece-wise linear), and dummy variables for different ranges.

## 5.2 Model results

In the paper, for both local roads and arterial roads we estimated two models: (1) Panel Ordered Probit Fractional Split Model (POPFS) and (2) Panel Mixed Probit Fractional Split Model (PMOPFS). The PMOPFS model is a generalized version of the POPFS. Hence, a direct log-likelihood comparison is adequate to determine the model that offers enhanced data fit. The Log-likelihood values at convergence for the local roads database for two models with log-likelihood and parameters are: POPFS – (-7430.2, 17) and PMOPFS – (-7292.2, 19). The corresponding numbers for the arterial roads database are: POPFS – (-27965.5) and PMOPFS – (-25458.2, 18). The two additional parameters in the PMOPFS clearly enhance the log-likelihood function substantially indicating that the model that incorporates unobserved and site specific effects outperforms the simple fractional split model at any level of statistical significance for both local roads and arterial roads. For the sake of brevity we only focus on the PMOPFS model results in our discussion.

The results for PMOPFS model for local roads and arterial roads are presented in Tables 5 and 6 respectively. The results for the local roads and arterial roads are discussed simultaneously in the subsequent sections.

### 5.2.1 Street characteristics

For local roads, a host of street characteristics influence the proportion of vehicles in each speed category. As expected an increase in the number of lanes increases the proportion of vehicles assigned to higher speed categories (see Figuero and Tarko, 2005 for a similar result). Presence of parking has a negative impact of vehicle speed. This is expected because car drivers are routinely slowed down by vehicles entering or leaving the parking area. The one way dummy variable has a counter intuitive result. It is intuitive to expect that the lack of vehicles in the opposing lanes might increase vehicle speeds. However, in our model we find that on one-way streets vehicle speeds are lower than two way streets. The result needs to be explored further in future research. The number of sidewalks on average tends to increase vehicular speed. It is important to note that the number of sidewalks serve as a surrogate for pedestrian volumes (in the absence of pedestrian volume data). The variables impact varies substantially across the sites as evidenced by the significant standard deviation parameter. The large standard deviation indicates that the influence of sidewalks exhibit different influence on vehicle speed. The presence of bicycle route increases the vehicle speeds on the road. The variable is clearly a manifestation of the possibility that bicycle routes are typically installed on larger roads with

high vehicular flow. As is expected, presence of good pavement encourages faster driving while higher vertical grade reduces the vehicle speed at a site.

For arterial roads, as data on street characteristics was sparse, we find that only the number of lanes and speed limit variable offer significant variables. The number of lanes and speed limit variables have an expected impact on vehicle speed proportion. As the variable increases, we observe an increase in the latent propensity of vehicle speed indicating a rightward shift in the speed proportion. Further, the speed limit variable has a significant standard deviation parameter highlighting that across roadway facilities speed limit has a varying effect on speed proportion.

### 5.2.2 Temporal characteristics

For local roads, the only temporal variable that was significant in our analysis for local roads was the late night hours (12am-6am) indicator variable. The variable indicates that vehicle speed during these time periods is usually lower than other time periods. Probably, drivers are careful during these times on local roads.

In terms of temporal variable, arterial roads have a higher number of significant variables. The results indicate that during the rush hour, evening and late night times the vehicle operating speeds are higher than during the mid-day time period. The increase is highest for the late night time period (as can be seen from last row of Table 4).

### 5.2.3 Thresholds

The threshold parameters do not have any substantive interpretation and only serve as cut-off points in the two models. We estimated 9 thresholds for the 10 speed categories.

### 5.2.4 Site specific unobserved effects

In our models (for local roads and arterial roads) we estimated site specific unobserved effects. The significant standard deviation parameter on the propensity constant provides evidence toward supporting our hypothesis that it is necessary to incorporate these effects. The site specific standard deviation variable on latent propensity constant indicates that the thresholds parameters might shift for various sites based on the unobserved effects. If the presence of these unobserved effects were neglected the resultant estimates will be biased (Eluru and Bhat. 2007).

## 5.3 **Model Validation**

The applicability of the proposed framework for modeling vehicle speed proportions for local roads and arterial roads was examined by using the estimated parameters to predict the proportions on the hold-out sample. The prediction exercise was undertaken for the two hold-out samples from both facility types. The validation exercise involves employing the estimated models to predict the vehicle speed proportions for facilities in the hold-out sample. The predicted proportions are compared with the actual observed proportions in the data. The validation exercise was also undertaken to examine model performance in various sub-samples

of the data. The sub-samples considered for local roads include: (1) one way streets, (2) sidewalk presence, (3) good pavement condition, (4) speed limit is 40 kmph and (5) rush hour time period. For arterial roads, the sub-samples considered include: (1) lanes  $\geq 4$ , (2) lanes  $\geq 3$ , (3) presence of parking, (4) rush hour time period and (5) weekday. The validation results for local roads and arterials are presented in Tables 7 and 8 respectively.

In the table, for each sample, the predicted proportions are placed in the first row and the observed proportions are provided below. The number of records in the sample is also provided. For each sample, we compute the error percentage measure to evaluate the performance of our model. The error percentage measure is quite misleading because you observe a large error usually for cells with small observed proportions. Hence, to generate a more useful measure we compute the  $\chi^2$  test metric. The  $\chi^2$  test metric is computed as follows.

$$\chi^2 = \sum_{i=1}^{10} \frac{(\text{Observed proportion} - \text{Predicted Proportion})^2}{\text{Observed proportion}}$$

If the computed  $\chi^2$  test metric is lower than the  $\chi^2$  value at 95% significance for 9 degrees of freedom, we can conclude that the predicted distribution is not statistically different from the observed distribution. The  $\chi^2$  value at 95% for 9 degrees of freedom is 16.92. The computed  $\chi^2$  test metric is presented in the last column for both Tables.

The following observations can be made from the numbers presented in Tables 7 and 8. First, the error percentage for cells with small proportion (such as very high speed cells) is very high. However, for the cells that have largest shares (usually speed categories from 20 through 60) the error band is quite low. Second, the error band for the local roads for the samples range from 22.5 through 45.0. The corresponding band for arterial roads ranges from 21.4 through 30.0. Third, the chi-square test indicators show that there is no evidence to accept the hypothesis that the predicted distribution is significantly different from the observed distribution for local and arterial population samples and all sub-samples (except for one way subsample in the local road context). Overall, these numbers are quite satisfactory. Fourth, usually, larger sample size in the validation sample results in a smaller error range. In summary, the validation results clearly indicate that the model for vehicle speed proportions performs adequately for the two datasets under consideration.

## 6 CONCLUSIONS

Vehicle operating speed measured on roadways is a critical ingredient for a host of analysis in the transportation field including transportation safety, traffic flow, geometric design, vehicle emissions, and road user route decisions. Given the wide ranging applications involving vehicle speed it is not surprising that a number of studies have investigated vehicle speed on roadways. The current research effort contributes to literature on examining vehicle speeds on urban roads methodologically and substantively. In terms of methodology, we formulate a new econometric model framework for examining speed profiles. The proposed model is an ordered response formulation of a fractional split model. The ordered fractional split model considers the proportion of vehicles traveling in various speed ranges for the entire segment of roadway as the dependent variable. The model implicitly recognizes that the proportions across all speed ranges

sum to unity. We also formulate a mixed version of the fractional split model to account for the impact of unobserved variables on speed proportion. Further, within the mixed version of the model we also recognize that multiple observations from the same facility are bound to be correlated. Towards this end, we incorporate site-specific unobserved effects in a panel version of the model. In fact, the formulated model is the first instance of a panel mixed ordered fractional split model not only in transportation literature but also in econometrics literature.

The analysis conducted in our research is the first of its kind on vehicle operating speed. Typically, vehicle speed studies focus on highways and major arterials. In our study, we examine local roads in addition to arterial roads in Montreal region. The dataset from Montreal consisting of weekly speed data (collected in hourly intervals) for about 50 local roads and 70 arterial roads is employed to estimate the proposed Panel Mixed Probit Fractional Split Model. We estimate separate models for local roads and arterial roads. The model estimation exercise considers a whole host of variables including geometric design attributes, roadway attributes, traffic characteristics and environmental factors. The model results highlight the role of various street characteristics including number of lanes, presence of parking, presence of sidewalks, vertical grade, and bicycle route on vehicle speed proportions. The model results are slightly different across the two datasets employed in our analysis. The results also highlight the presence of significant standard deviation on coefficients (presence of side walk and speed limit) and site-specific unobserved effects. Neglecting the presence of these unobserved effects in modeling the influence of exogenous variable on vehicle speed would result parameter estimates that are biased. Further, the applicability of the modeling framework proposed was examined using a hold-out sample composing of records not considered for model estimation. The results from the validation exercise provide satisfactory results for both local roads and arterial roads. Overall, the results indicate that the proposed panel mixed ordered probit fractional split model offer promise for modeling such proportional ordinal variables.

To be sure, the study is not without limitations. In our analysis, the dependent variable is measured as proportion variable. To accurately determine the exact number of vehicles in the speed category, an overall volume on the local road facility is required. The analyst can generate this information through an add-on regression model for vehicle volumes on roadway facilities. Further, it might be useful to jointly model the roadway volumes and the speed proportion variable – a significant challenge in terms of modeling. Further, the arterial database included in our analysis has a smaller set of street characteristics available to us. The model frameworks developed would need to be enhanced substantially with additional information on arterials.

## **ACKNOWLEDGEMENTS**

The authors would like to thank the Montreal Department of Transportation for providing the data. We would also like to thank Adham Badram and Tyler Krieder for helping with data processing. The corresponding author would like to acknowledge financial support from Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors also would like to acknowledge useful feedback from two anonymous reviewers.

## **REFERENCES**

Ali, A., A. Flannery, and M. Venigalla (2007). Prediction Models for Free Flow Speed on Urban Streets. Presented at the 86th Annual Meeting of the Transportation Research Board,

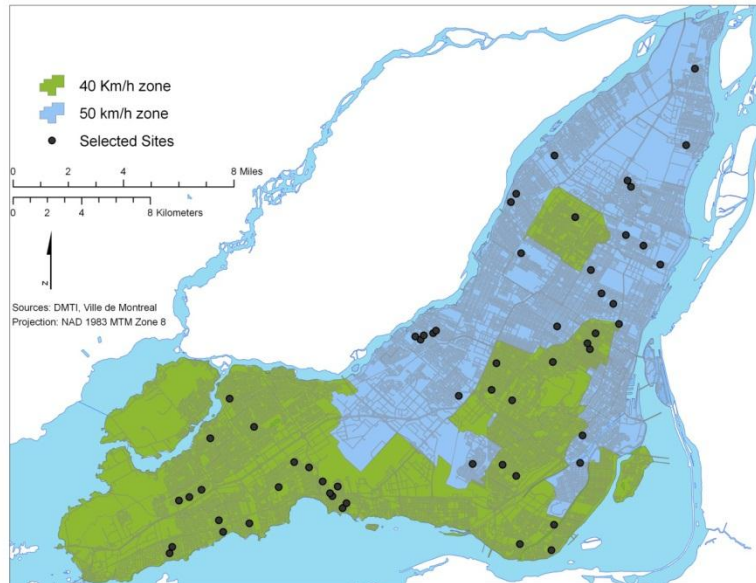
- Washington, D.C..
- Andueza, P.J. (2000) Mathematical Models of Vehicular Speed on Mountain Roads. *Transportation Research Record 1701*, TRB, National Research Council, Washington, D. C.
- Donnell, E.T., Ni, Y., Adolini, M., Elefteriadou, L., (2001). Speed prediction models for trucks on two-lane rural highways. *Transportation Research Record: Journal of the Transportation Research Board 1751* (-1), 44-55.
- Eluru, N., Bhat, C.R., (2007). A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis & Prevention* 39 (5), 1037-1049.
- Ewing, R., and Dumbaugh, E. (2009). The Built Environment and Traffic Safety: A Review of Empirical Evidence. *Journal of Planning Literature* 23, 347 – 367.
- Fambro, D. B., K. Fitzpatrick, and C. W. Russell. (2000). Operating Speed on Crest Vertical Curves with Limited Stopping Sight Distance. *Transportation Research Record: Journal of the Transportation Research Board, No. 1701*, TRB, National Research Council, Washington, D.C., pp. 25–31.
- Figuerola, A. M., and A. P. Tarko. (2005). Speed Factors on Two-Lane Rural Highways in Free-Flow Conditions. *Transportation Research Record: Journal of the Transportation Research Board, No. 1912*, Transportation Research Board of the National Academies, Washington, D.C., pp. 49–46..
- Fitzpatrick, K., P. Carlson, M. Brewer, M. Wooldridge, and S. Miaou. (2003). Design Speed, Operating Speed, and Posted Speed Practices. *NCHRP Report 504*, Transportation Research Board, Washington, D.C.
- Garber, N. and R. Gadiraju. (1989). Factors affecting speed variance and its influence on accidents. In *Transportation Research Record 1213*, TRB, National Research Council, Washington, D.C., pp. 64-71.
- Gattis, J.L., and Watts, A. (1999). Urban Street Speed Related to Width and Functional Class. *Journal of Transportation Engineering*, Vol. 125, No. 3.
- Gourieroux, C., Monfort, A., and A. Trognon (1984). Pseudo Maximum Likelihood Methods: Theory. *Econometrica*, Vol. 52, pp. 681-700.
- Hewson, P. (2008). Quantile regression provides a fuller analysis of speed data. *Accident Analysis & Prevention* 40 (2), 502-510.
- Islam, M. N., and P. N. Seneviratne. (1994). Evaluation of Design Consistency of Two-Lane Rural Highways. *ITE Journal*, Vol. 64, No. 2, pp. 28–31.
- Kanellaidis, G., Golias, J., and Efstathiadis, S. (1990) Drivers Speed Behaviour on Rural Road Curves. *Traffic Engineering & Control*.
- Ko, J., Guensler, R.L., (2005). Characterization of congestion based on speed distribution: a statistical approach using Gaussian mixture model. In: CD-ROM Proceedings of the 84th Annual Meeting of the Transportation Research Board, Washington, D.C.
- Krammes, R. A., Q. Brackett, M. A. Shafer, J. L. Ottesen, I. B. Anderson, K. L. Fink, K. M. Collins, O. J. Pendleton, and C. J. Messer. (1995). *Horizontal Alignment Design Consistency for Rural Two-Lane Highways*. Report No. FHWA-RD-94-034. Federal Highway Administration, U.S. Department of Transportation.
- McFadden, J., W. T. Yang, and S. R. Durrans. (2001). Application of Artificial Networks to Predict Speeds on Two-Lane Rural Highways. *Transportation Research Record: Journal of the Transportation Research Board, No. 1751*, TRB, National Research Council, Washington, D.C., pp. 9–17.

- Park, B.-J., Zhang, Y., Lord, D., (2010). Bayesian mixture modeling approach to account for heterogeneity in speed data. *Transportation Research Part B: Methodological* 44 (5), 662-673.
- Papke, L.E., and J.M. Wooldridge (1996). Econometric Methods for Fractional Response Variables with an Application to 401(k) Plan Participation Rates. *Journal of Applied Econometrics*, Vol. 11, pp. 619-632.
- Poe, C.M., and Mason, J.M. (2000). Analyzing Influence of Geometric Design on Operating Speeds Along Low-Speed Urban Streets. *Transportation Research Record* 1737, 18 – 25.
- Polus, A., K. Fitzpatrick, and D. B. Fambro. (2000). Predicting Operating Speeds on Tangent Sections of Two-Lane Rural Highways. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1737, TRB, National Research Council, Washington, D.C., pp. 50–57.
- Qin, X., Ng, M., Reyes, P.E. (2010). Identifying crash-prone locations with quantile regression. *Accident Analysis & Prevention* 42 (6), 1531-1537.
- Sivakumar, A. and C.R. Bhat (2002), A Fractional Split Distribution Model for Statewide Commodity Flow Analysis, *Transportation Research Record*, Vol. 1790, pp. 80-88.
- Tarris, J., C. Poe, J. M. Mason, and K. Goulias. (1996). Predicting Operating Speeds on Low-Speed Urban Streets: Regression and Panel Analysis Approaches. *Transportation Research Record* 1523, Transportation Research Board, National Research Council, Washington, D.C., pp. 46–54.
- TRC (2011), Modeling Operating Speed, Transportation Research Circular, Transportation Research Board, Washington.

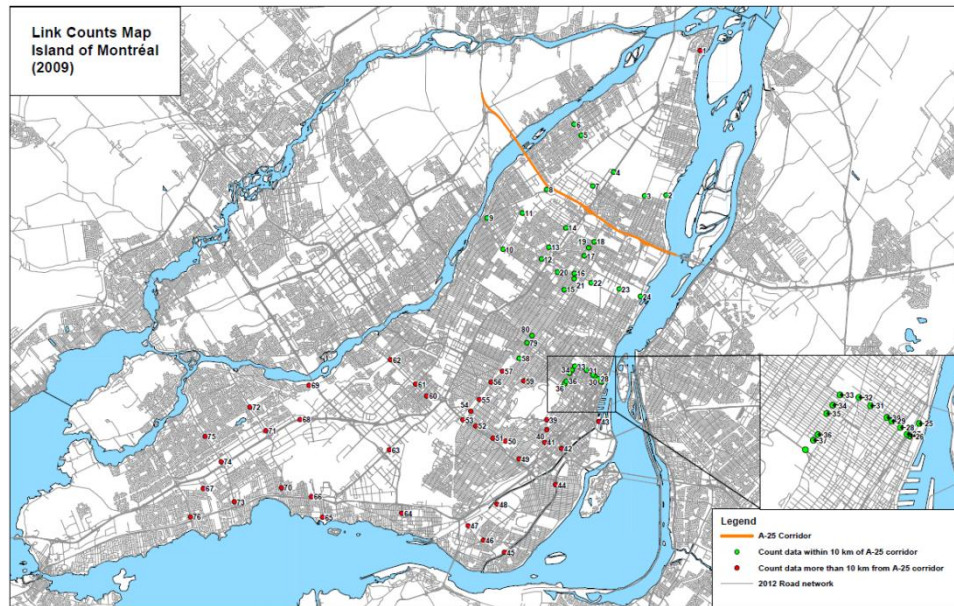
## **LIST OF FIGURES**

Figure 1: Spatial distribution of local road sites

Figure 2: Spatial distribution of arterial road sites



**Figure 1: Spatial distribution of local road sites**



**Figure 2: Spatial distribution of arterial road sites**

## **LIST OF TABLES**

- Table 1: Descriptive Statistics for Local Road Sample
- Table 2: Descriptive Statistics for Arterial Sample
- Table 3: Distribution of the speed proportion variable on local roads
- Table 4: Distribution of the speed proportion variable on arterial roads
- Table 5: PMOPFS Model estimates for Local roads
- Table 6: PMOPFS Model estimates for Arterials
- Table 7: Validation for Local Roads
- Table 8: Validation for Arterials

**Table 1: Descriptive Statistics for Local Road Sample**

|                                          |     |
|------------------------------------------|-----|
| <b><u>Site Summary (Local, N=49)</u></b> |     |
| <u>Street Characteristics</u>            |     |
| <i>Speed Limit</i>                       |     |
| 40km/h                                   | 23  |
| 50km/h                                   | 26  |
| <i>Sidewalk</i>                          |     |
| On one side                              | 6   |
| On both sides                            | 32  |
| Oneway streets                           | 16  |
| Parking permitted on street              |     |
| On one side                              | 2   |
| On both sides                            | 42  |
| Number of Lanes                          |     |
| 1                                        | 42  |
| 2                                        | 7   |
| <i>Days data was collected</i>           | (%) |
| Weekday                                  | 60  |
| Weekend                                  | 40  |

**Table 2: Descriptive Statistics for Arterial Sample**

| <b><u>Site Summary (Arterial, N=71)</u></b> |    |
|---------------------------------------------|----|
| <u>Street Characteristics</u>               |    |
| <i>Speed Limit</i>                          |    |
| 30km/h                                      | 1  |
| 50km/h                                      | 67 |
| 60km/h                                      | 1  |
| 70km/h                                      | 2  |
| Parking permitted on street                 | 28 |
| Number of Lanes                             |    |
| 2 or 3                                      | 11 |
| 4                                           | 35 |
| More than 4                                 | 25 |

**Table 3: Distribution of the speed proportion variable on local roads**

|                 | <b>≤ 20<br/>kmph</b> | <b>20 – 30<br/>kmph</b> | <b>30 – 40<br/>kmph</b> | <b>40 – 50<br/>kmph</b> | <b>50 – 60<br/>kmph</b> | <b>60 – 70<br/>kmph</b> | <b>70 – 80<br/>kmph</b> | <b>80 – 90<br/>kmph</b> | <b>90 – 120<br/>kmph</b> | <b>&gt; 120<br/>kmph</b> |
|-----------------|----------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|--------------------------|
| <b>Overall</b>  | 1.9                  | 5.2                     | 10.6                    | 21.4                    | 27.9                    | 19.1                    | 9.2                     | 2.8                     | 1.9                      | 0.1                      |
| <b>6am-10am</b> | 1.5                  | 4.1                     | 8.8                     | 18.6                    | 27.2                    | 22.0                    | 11.6                    | 3.5                     | 2.5                      | 0.1                      |
| <b>10am-3pm</b> | 2.1                  | 5.8                     | 11.8                    | 23.4                    | 28.7                    | 17.4                    | 7.4                     | 2.1                     | 1.2                      | 0.0                      |
| <b>3pm-7pm</b>  | 2.7                  | 7.0                     | 12.7                    | 23.7                    | 27.2                    | 16.2                    | 7.1                     | 2.1                     | 1.2                      | 0.0                      |
| <b>7pm-12am</b> | 2.1                  | 5.6                     | 11.3                    | 22.8                    | 28.5                    | 17.8                    | 7.9                     | 2.4                     | 1.6                      | 0.1                      |
| <b>12am-6am</b> | 1.2                  | 3.8                     | 8.7                     | 19.1                    | 27.5                    | 21.6                    | 11.5                    | 3.7                     | 2.8                      | 0.1                      |

**Table 4: Distribution of the speed proportion variable on arterial roads**

|                 | <b>≤ 20<br/>kmph</b> | <b>20 – 30<br/>kmph</b> | <b>30 – 40<br/>kmph</b> | <b>40 – 50<br/>kmph</b> | <b>50 – 60<br/>kmph</b> | <b>60 – 70<br/>kmph</b> | <b>70 – 80<br/>kmph</b> | <b>80 – 90<br/>kmph</b> | <b>90 – 120<br/>kmph</b> | <b>&gt; 120<br/>kmph</b> |
|-----------------|----------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|--------------------------|
| <b>Overall</b>  | 22.9                 | 16.6                    | 17.9                    | 15.8                    | 11.2                    | 6.8                     | 3.7                     | 2.3                     | 1.7                      | 1.2                      |
| <b>6am-10am</b> | 23.3                 | 16.6                    | 17.2                    | 16.5                    | 11.7                    | 6.6                     | 3.2                     | 2.0                     | 1.4                      | 1.5                      |
| <b>10am-3pm</b> | 22.1                 | 16.9                    | 19.0                    | 15.7                    | 11.4                    | 6.6                     | 3.5                     | 2.0                     | 1.7                      | 1.1                      |
| <b>3pm-7pm</b>  | 22.3                 | 17.3                    | 19.1                    | 15.8                    | 10.8                    | 6.2                     | 3.4                     | 2.1                     | 1.6                      | 1.4                      |
| <b>7pm-12am</b> | 23.3                 | 16.9                    | 18.1                    | 15.2                    | 10.8                    | 6.6                     | 3.5                     | 2.4                     | 2.0                      | 1.2                      |
| <b>12am-6am</b> | 23.4                 | 15.1                    | 16.1                    | 15.8                    | 11.4                    | 8.0                     | 4.6                     | 2.7                     | 1.9                      | 1.0                      |

**Table 5: PMOPFS Model estimates for Local roads**

| <b>Explanatory Variables</b>                             | <b>Coefficient</b> | <b>t-stat</b> |
|----------------------------------------------------------|--------------------|---------------|
| <u>Street Characteristics</u>                            |                    |               |
| <i>Number of lanes</i>                                   | 0.791              | 5.37          |
| <i>Number of sides parking is present</i>                | -0.203             | -2.68         |
| <i>One way street</i>                                    | -0.269             | -1.87         |
| <i>Number of sidewalks</i>                               | 0.142              | 2.10          |
| <i>Standard Deviation</i>                                | 0.121              | 2.33          |
| <i>Presence of bicycle route</i>                         | 0.406              | 3.51          |
| <i>Good pavement condition</i>                           | 0.310              | 2.72          |
| <i>High vertical grade</i>                               | -0.280             | -1.62         |
| <u>Time variables</u>                                    |                    |               |
| <i>Night hours (12am to 6am)</i>                         | -0.070             | -1.83         |
| <u>Standard Deviations</u>                               |                    |               |
| <i>Propensity constant Standard Deviation</i>            | 0.237              | 3.80          |
| <u><b>Threshold parameters</b></u>                       |                    |               |
| <i>Threshold 1 (<math>\leq 20</math> and 20-30 kmph)</i> | -0.025             | -0.10         |
| <i>Threshold 2 ( 20 – 30 and 30-40 kmph)</i>             | 0.511              | 1.93          |
| <i>Threshold 3 ( 30 – 40 and 40-50 kmph)</i>             | 1.032              | 3.88          |
| <i>Threshold 4 ( 40 – 50 and 50-60 kmph)</i>             | 1.516              | 5.64          |
| <i>Threshold 5 ( 50 – 60 and 60-40 kmph)</i>             | 1.964              | 7.14          |
| <i>Threshold 6 ( 60 – 70 and 70-40 kmph)</i>             | 2.332              | 8.19          |
| <i>Threshold 7 ( 70 – 80 and 80-40 kmph)</i>             | 2.623              | 8.96          |
| <i>Threshold 8 ( 80 – 90 and 90-120 kmph)</i>            | 2.895              | 9.59          |
| <i>Threshod 9 ( &gt; 120 kmph)</i>                       | 3.242              | 10.45         |
| <b>Total number of Cases</b>                             | 3903               |               |
| <b>Log Likelihood at Convergence</b>                     | -7292.2            |               |

**Table 6: PMOPFS Model estimates for Arterials**

| <b>Explanatory Variables</b>                             | <b>Coefficient</b> | <b>t-stat</b> |
|----------------------------------------------------------|--------------------|---------------|
| <u>Street Characteristics</u>                            |                    |               |
| <i>Number of lanes</i>                                   | 0.116              | 4.99          |
| <i>Speed limit</i>                                       | 0.074              | 12.28         |
| <i>Standard Deviation</i>                                | 0.007              | 20.527        |
| <u>Time variables</u>                                    |                    |               |
| <i>Night hours (12-6am) (Midday 10am-3pm is base)</i>    | 0.360              | 15.85         |
| <i>Rush hours (6-10am, 3-7pm)</i>                        | 0.069              | 4.75          |
| <i>Evening hours (7pm to 12am)</i>                       | 0.084              | 5.14          |
| <i>Weekday</i>                                           | 0.051              | 4.48          |
| <i>Summer Months (June-August)</i>                       | -0.125             | -2.39         |
| <u>Standard Deviations</u>                               |                    |               |
| <i>S.D. constant</i>                                     | 0.6199             | 28.172        |
| <u>Threshold parameters</u>                              |                    |               |
| <i>Threshold 1 (<math>\leq 20</math> and 20-30 kmph)</i> | 1.736              | 5.62          |
| <i>Threshold 2 ( 20 – 30 and 30-40 kmph)</i>             | 2.446              | 7.99          |
| <i>Threshold 3 ( 30 – 40 and 40-50 kmph)</i>             | 3.116              | 10.23         |
| <i>Threshold 4 ( 40 – 50 and 50-60 kmph)</i>             | 3.957              | 12.84         |
| <i>Threshold 5 ( 50 – 60 and 60-40 kmph)</i>             | 4.877              | 15.67         |
| <i>Threshold 6 ( 60 – 70 and 70-40 kmph)</i>             | 5.701              | 18.23         |
| <i>Threshold 7 ( 70 – 80 and 80-40 kmph)</i>             | 6.452              | 19.65         |
| <i>Threshold 8 ( 80 – 90 and 90-120 kmph)</i>            | 6.956              | 19.40         |
| <i>Threshod 9 ( &gt; 120 kmph)</i>                       | 8.403              | 20.43         |
| <b>Total number of Cases</b>                             | 15456              |               |
| <b>Log Likelihood at Convergence</b>                     | -25458.2           |               |

**Table 7: Validation for Local Roads**

| Sample                        | N   |           | p20    | p2030 | p3040 | p4050 | p5060 | p6070 | p7080 | p8090 | p90120 | pgt120 | $\chi^2$ -test<br>metric |
|-------------------------------|-----|-----------|--------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------------------------|
| Full<br>Sample                | 480 | Predicted | 27.6   | 19.6  | 19.8  | 14.9  | 9.2   | 4.5   | 2.1   | 1.1   | 0.7    | 0.5    | <b>9.57</b>              |
|                               |     | Observed  | 17.8   | 16.6  | 20.4  | 18.5  | 11.9  | 6.6   | 2.9   | 1.9   | 1.8    | 1.2    |                          |
|                               |     | % Error   | -55.2  | -17.7 | 3.1   | 19.3  | 22.3  | 32.3  | 29    | 39.4  | 58.2   | 61.5   |                          |
| Oneway                        | 160 | Predicted | 33.7   | 20.9  | 19.2  | 13.1  | 7.3   | 3.2   | 1.3   | 0.7   | 0.4    | 0.2    | <b>33.09</b>             |
|                               |     | Observed  | 14.9   | 17.3  | 23.2  | 20.3  | 13    | 6.9   | 2.7   | 1.1   | 0.5    | 0.2    |                          |
|                               |     | % Error   | -126.8 | -20.8 | 17.1  | 35.3  | 44    | 53.5  | 50.2  | 38.8  | 25.5   | -14.4  |                          |
| Sidewalk<br>Presence          | 320 | Predicted | 26.9   | 19.3  | 19.8  | 15.1  | 9.5   | 4.7   | 2.2   | 1.2   | 0.8    | 0.5    | <b>9.37</b>              |
|                               |     | Observed  | 17.6   | 16    | 20.1  | 18.1  | 12.5  | 7.1   | 3.2   | 2.2   | 1.6    | 1.4    |                          |
|                               |     | % Error   | -53    | -21.2 | 1.8   | 16.6  | 24.1  | 33.9  | 31.1  | 45.4  | 50.5   | 61.7   |                          |
| Good<br>Pavement<br>Condition | 160 | Predicted | 19.8   | 17.7  | 20.3  | 17.3  | 11.8  | 6.3   | 3.1   | 1.8   | 1.2    | 0.9    | <b>7.70</b>              |
|                               |     | Observed  | 16.5   | 13.7  | 15.7  | 17.6  | 12.5  | 9.1   | 4     | 3.2   | 3.6    | 2.7    |                          |
|                               |     | % Error   | -19.7  | -29.1 | -29.5 | 2.1   | 5.9   | 31.4  | 24.1  | 45.1  | 65.9   | 68.8   |                          |
| Speed<br>Limit is<br>40 kmph  | 240 | Predicted | 30.5   | 20.3  | 19.6  | 14.1  | 8.2   | 3.8   | 1.7   | 0.9   | 0.5    | 0.3    | <b>14.85</b>             |
|                               |     | Observed  | 17.1   | 18.4  | 22.9  | 19.8  | 10.8  | 5.5   | 2.1   | 0.9   | 1.4    | 0.7    |                          |
|                               |     | % Error   | -78.3  | -10.5 | 14.5  | 29    | 23.9  | 30.1  | 20.6  | 0.6   | 60.4   | 51.4   |                          |
| Rush<br>Hour                  | 240 | Predicted | 27.5   | 19.5  | 19.8  | 15    | 9.3   | 4.5   | 2.1   | 1.1   | 0.7    | 0.5    | <b>10.81</b>             |
|                               |     | Observed  | 16.5   | 17.7  | 20.9  | 19    | 12.1  | 6.2   | 2.9   | 1.7   | 1.4    | 1.3    |                          |
|                               |     | % Error   | -66.8  | -10.7 | 5.2   | 21.2  | 23.5  | 27    | 28.2  | 33.4  | 45.6   | 63     |                          |

**Table 8: Validation for Arterials**

| Sample              | N    |           | p20  | p2030 | p3040 | p4050 | p5060 | p6070 | p7080 | p8090 | p90120 | pgt120 | $\chi^2$ -test metric |
|---------------------|------|-----------|------|-------|-------|-------|-------|-------|-------|-------|--------|--------|-----------------------|
| Full Sample         | 1456 | Predicted | 0.9  | 3.7   | 10.4  | 26.4  | 33.7  | 18    | 5.6   | 1     | 0.3    | 0      | <b>7.10</b>           |
|                     |      | Observed  | 1.4  | 6.5   | 13.7  | 24.5  | 25.8  | 16.6  | 7.8   | 2.2   | 1.4    | 0.1    |                       |
|                     |      | % Error   | 37.5 | 42.9  | 23.9  | -7.5  | -30.4 | -8.3  | 27.6  | 53.6  | 75.2   | 96.4   |                       |
| Lanes $\geq 4$      | 896  | Predicted | 0.5  | 2.7   | 8.5   | 24.2  | 34.8  | 20.6  | 7     | 1.3   | 0.5    | 0      | <b>12.13</b>          |
|                     |      | Observed  | 1.6  | 6.4   | 10.8  | 18.3  | 25.7  | 21.2  | 10.8  | 3.1   | 2      | 0.1    |                       |
|                     |      | % Error   | 65.4 | 57.9  | 21.3  | -32.3 | -35.6 | 2.8   | 35.9  | 57.8  | 76.7   | 96.7   |                       |
| Lanes $\geq 3$      | 1120 | Predicted | 0.7  | 3.1   | 9.4   | 25.2  | 34.4  | 19.4  | 6.3   | 1.2   | 0.4    | 0      | <b>10.47</b>          |
|                     |      | Observed  | 1.4  | 6.8   | 12.6  | 20.3  | 25.5  | 19.3  | 9.5   | 2.7   | 1.7    | 0.1    |                       |
|                     |      | % Error   | 53.4 | 53.8  | 25.6  | -24.3 | -34.5 | -0.4  | 33.4  | 56.6  | 76.3   | 96.6   |                       |
| Presence of Parking | 784  | Predicted | 1.1  | 4.7   | 12.4  | 28.7  | 32.8  | 15.3  | 4.2   | 0.7   | 0.2    | 0      | <b>5.77</b>           |
|                     |      | Observed  | 1.1  | 6.3   | 17.1  | 32.9  | 27.4  | 10.7  | 3.1   | 0.7   | 0.4    | 0      |                       |
|                     |      | % Error   | -1.2 | 26.7  | 27.7  | 12.9  | -19.4 | -43.9 | -32.9 | 3.3   | 52.9   | 0      |                       |
| Rush Hours          | 728  | Predicted | 0.8  | 3.6   | 10.4  | 26.4  | 33.8  | 18    | 5.6   | 1     | 0.3    | 0      | <b>8.33</b>           |
|                     |      | Observed  | 1.6  | 7     | 13.7  | 23.5  | 25.4  | 17    | 7.9   | 2.2   | 1.4    | 0.1    |                       |
|                     |      | % Error   | 48.4 | 48.3  | 24.2  | -12.2 | -33.1 | -5.8  | 29.4  | 54.3  | 75.8   | 97     |                       |
| Weekday             | 1072 | Predicted | 0.8  | 3.6   | 10.3  | 26.2  | 33.8  | 18.2  | 5.7   | 1     | 0.4    | 0      | <b>6.53</b>           |
|                     |      | Observed  | 1.2  | 6.2   | 13.5  | 24.7  | 26.1  | 16.6  | 7.8   | 2.2   | 1.4    | 0.1    |                       |
|                     |      | % Error   | 33.6 | 41.8  | 23.9  | -6.2  | -29.5 | -9.2  | 26.6  | 53.7  | 75.3   | 96.8   |                       |