**Latent Segmentation Based Count Models: Analysis of Bicycle Safety in Montreal and Toronto**

**Shamsunnahar Yasmin**

Department of Civil, Environmental & Construction Engineering

University of Central Florida

Tel: 407-823-4815, Fax: 407-823-3315

Email: [shamsunnahar.yasmin@mail.mcgill.ca](mailto:shamsunnahar.yasmin@mail.mcgill.ca)

**Naveen Eluru\***

Associate Professor

Department of Civil, Environmental & Construction Engineering

University of Central Florida

Tel: 407-823-4815, Fax: 407-823-3315

Email: [naveen.eluru@ucf.edu](mailto:naveen.eluru@ucf.edu)

**Abstract**

The study contributes to literature on bicycle safety by building on the traditional count regression models to investigate factors affecting bicycle crashes at the Traffic Analysis Zone (TAZ) level. TAZ is a traffic related geographic entity which is most frequently used as spatial unit for macroscopic crash risk analysis. In conventional count models, the impact of exogenous factors is restricted to be the same across the entire region. However, it is possible that the influence of exogenous factors might vary across different TAZs. To accommodate for the potential variation in the impact of exogenous factors we formulate latent segmentation based count models. Specifically, we formulate and estimate latent segmentation based Poisson (LP) and latent segmentation based Negative Binomial (LNB) models to study bicycle crash counts. In our latent segmentation approach, we allow for more than two segments and also consider a large set of variables in segmentation and segment specific models. The formulated models are estimated using bicycle-motor vehicle crash data from the Island of Montreal and City of Toronto for the years 2006 through 2010. The TAZ level variables considered in our analysis include accessibility measures, exposure measures, sociodemographic characteristics, socioeconomic characteristics, road network characteristics and built environment. A policy analysis is also conducted to illustrate the applicability of the proposed model for planning purposes. This macro-level research would assist decision makers, transportation officials and community planners to make informed decisions to proactively improve bicycle safety – a prerequisite to promoting a culture of active transportation.

Key words: bicycle crashes, population heterogeneity, latent segmentation Poisson model, latent segmentation Negative Binomial model

**1. BACKGROUND**

Active forms of transportation such as walking and bicycling have the lowest carbon footprint on the environment and improve the physical health of pedestrians and bicyclists. With growing concern of worsening global climate change and increasing obesity among adults in developed countries, it is hardly surprising that transportation decision makers are proactively encouraging the adoption of active forms of transportation for short distance trips. For instance, bicycling, as a transport mode, is experiencing increased patronage and support in most Canadian cities where personal automobiles are the most common mode of transportation. In fact, between 1996 and 2006, a 42% increase in the number of daily bike commuters was observed in Canada (Pucher et al., 2011).

However, transportation safety concerns related to active transportation users form one of the biggest impediments to their adoption as a preferred alternative to private vehicle use for shorter trips. Earlier research reveals that the likelihood of being involved in a collision increases as the number of cyclists on the road increases (Wei and Lovegrove, 2013). Also, the risk of being injured in a collision while cycling could be about seven times higher than a motorist (Reynolds et al., 2009). Thus, traffic crashes and the consequent injury and fatality remain a detriment for cycling, leading to low bicycle mode share, specifically in North American communities (Wei and Lovegrove, 2013). Any effort to reduce the social burden of these crashes and encourage people to use bicycle for their daily short trips would necessitate the implementation of policies that enhance safety for bicyclists. An important tool to identify the critical factors affecting occurrence of bicycle crashes is the application of planning level crash prediction models.

* 1. **Earlier Research**

Traffic crashes aggregated at a certain planning scale, for any given time interval, are non-negative integer valued events. Naturally, these integer counts are examined employing count regression approaches. The traditional Poisson regression and Negative Binomial (NB) models have been the workhorses in examining the crash count events in safety literature. A number of research efforts have examined transportation (vehicle, pedestrian and bicyclist) related crash frequency (see Lord and Mannering (2010) for a detailed review). These studies have been conducted for different modes − vehicle (automobiles and motorbikes), pedestrian and bicycle and for different scales - micro (such as intersection and segment) and macro-level (such as traffic analysis zone, county, census tract). It is beyond the scope of the paper to review all the research on transportation crash frequency (for example see Brüde and Larsson, 1993; Turner et al., 2006; Loo and Tsui, 2010; Carter and Council, 2007; Jung et al., 2014; Dong et al., 2014 for micro-level studies). In our paper, we focus on studies examining crashes at the planning/macro-level.

A summary of earlier studies investigating crash frequency at a macro-level is presented in Table 1. The information provided in the table includes the methodological approach employed, the spatial aggregation level considered and the variable categories considered in the analysis from six variable categories − accessibility measures, exposure measures, sociodemographic characteristics, socioeconomic characteristics, road network characteristics and built environment. The following observations can be made from the table. First, the most prevalent spatial unit considered at the macro-level analysis is Traffic Analysis Zone (TAZ[[1]](#footnote-1)). Second, NB model is the most frequently used statistical technique for examining crashes at the aggregate level. Third, very few studies (5 out of 33) explored bicycle crash frequency at the planning level. Fourth, none of the studies have employed latent segmentation based approach in examining crash frequency at macro-level.

With respect to macro-level bicycle crash frequency, the overall findings from earlier research efforts are usually consistent. The most commonly identified variables that contribute to the increase in bicycle crash risk include: (1) accessibility measures such as transit accessibility and number of bus stops, (2) exposure measures such as households with no cars, population density and total bicycle commuters, (3) sociodemographic characteristics such as proportion of young population and African population, (4) socioeconomic characteristics such as low-income population and per capita expenditure on alcohol, (5) road network characteristics such as street connectivity, total number of intersections and total on-street bike lanes and (6) built environment characteristics such as neighborhood compactness, higher mix of land use and proximity to academic buildings.

**1.2 Current Study in Context**

The overview of earlier literature indicates that, in recent years, examining crash frequency at the macro-level has seen a revival of interest among safety researchers. However, there is paucity of research focusing on macro-level bicycle crashes. Therefore, it is important to investigate zonal level bicycle crashes to identify critical factors and propose implications to facilitate proactive safety-conscious planning. A critical component in the process of identifying the contributing risk factors is the application of appropriate econometric models. As indicated in Table 1, the most prevalent formulation to study macroscopic crash frequency is the NB model. NB model allows for overdispersion and thus provides a natural enhancement over the traditional Poisson model and is easy to estimate with a closed form structure to accommodate for unobserved heterogeneity. However, NB model (and Poisson model) typically restricts the impact of exogenous variables to be same across the entire population of crash events – population homogeneity assumption. But, the impact of control variables on bicycle crash frequency might vary across TAZs based on different attributes. Ignoring such heterogeneous impact of variables might result in incorrect coefficient estimates.

To account for systematic heterogeneity, researchers have employed a clustering technique (Karlaftis et al., 1998). In this approach, target groups are divided in to different clusters based on a multivariate set of factors and separate models are estimated for each cluster. However, the approach requires allocating data records exclusively to a particular cluster, and does not consider the possible effects of unobserved factors that may moderate the impact of observed exogenous variables. Additionally, this approach might result in very few records in some clusters resulting in loss of estimation efficiency. An alternative approach to accommodate for population heterogeneity is to employ random parameter count models (Ukkusuri et al., 2011). However, in this approach the focus is on incorporating unobserved heterogeneity through the error term which necessitates extensive amount of simulation for model estimation while also not considering for observed heterogeneity.

A possible work around to accommodate for population heterogeneity is the application of latent segmentation based approach (or sometimes also referred to as finite mixture model). In this approach TAZs are allocated probabilistically to different segments and a segment specific model is estimated for each segment. Such an endogenous segmentation scheme is appealing for many reasons: First, each segment is allowed to be identified with a multivariate set of exogenous variables, while also limiting the total number of segments to a number that is much lower than what would be implied by a full combinatorial scheme of the multivariate set of exogenous variables. Second, the probabilistic assignment to segments explicitly acknowledges the role played by unobserved factors in moderating the impact of observed exogenous variables. Finally, this approach is semiparametric and hence, there is no need to specify a distributional assumption for the coefficients as is required in random parameter models (Greene and Hensher, 2003; Yasmin et al., 2014).

To be sure, the latent segmentation approach has been employed recently in safety literature for examining traffic crash count events at micro-level (Park et al., 2010; Park and Lord, 2009; Zou et al., 2014). However, the role of such population heterogeneity, in the context of macro-level crash count models has not been investigated in the existing literature. The microscopic models were developed with either a fixed weight parameter or segmentation model with a very small number of parameters (in the segmentation and segment specific models) were estimated citing model estimation complexity challenges. Further, earlier micro-level studies restricted the number of segments to two without any model selection exercise. The current study enhances the methodology from earlier finite mixture based count models in two ways: (1) we consider a large set of exogenous variables in the segmentation and segment specific models, and (2) we estimate more than two segments of latent segmentation and provide a clear framework for model selection[[2]](#footnote-2).

In summary, the current study makes a threefold contribution to literature on crash frequency in general and bicycle crash safety in particular. First, the study formulates and estimates latent segmentation based count models that accommodates for population heterogeneity. The current paper is the first effort in safety literature for examining crash count events where a latent segmentation model that is completely generic is estimated. We allow for a flexible segment membership function and test for the presence of multiple segments in the model estimation. In the current study context, we demonstrated the application by employing data for bicycle crash count events of two urban regions. It is worthwhile to mention here that, such a generalized approach can also be implemented for examining crash count events for other road users, such as motor vehicles and pedestrians as well. Second, the count models are estimated at the TAZ level employing a comprehensive set of exogenous variables by using data from two different cities of Canada: Montreal and Toronto. Examining bicycle crash count data and evaluation of validation across two datasets would allow us to illustrate the importance of incorporating population heterogeneity in identifying critical factors contributing to macro-level bicycle crash count events for different urban regions. Finally, based on the model results we identify important exogenous variables that influence bicycle crash counts.

The rest of the paper is organized as follows. Section 2 provides details of the econometric model frameworks used in the analysis. In Section 3, the study areas and data are described, respectively. The model estimation results are presented in Section 4. Elasticity effects, spatial representation and potential policy implications are discussed in Section 5. Section 6 concludes the paper.

**2. ECONOMETRIC FRAMEWORK**

In the latent segmentation based approach, bicycle crash count records for TAZs are probabilistically assigned to relatively homogenous (but latent to the analyst) segments based on various explanatory variables. Within each segment, the effects of exogenous variables on the number of crashes occurring across the TAZ over a given period of time are fixed in the segment. Hence, the latent segmentation based model consists of two components: (1) assignment component and (2) segment specific count model component. The general structure for all latent segmentation based count models involves specifying these two components. For the ease of presentation, we describe the general mathematical structure first and then identify the different modeling structures for various models in the subsequent discussion.

Let us assume that be the index for segments , be the index for TAZ and be the index for crashes occurring over a period of time in a TAZ . The assignments of TAZ to different segments are modeled as a function of a column vector of exogenous variable by using the random utility based multinomial logit model (see Wedel et al., 1993; Bago d'Uva, 2006; Eluru et al., 2012; Yasmin et al., 2014 for similar formulation) as:

|  |  |
| --- | --- |
|  |  |

where, is the probability of TAZ to be assigned to segment , is a vector of attributes and is a conformable parameter vector to be estimated. The assignment process is the same for all latent class models.

Within any latent segmentation approach, the unconditional probability of can be given as:

|  |  |
| --- | --- |
|  |  |

where corresponds to the probability of count in segment *s*. The exact probability function for depends on the count model chosen for the segment specific model. In our research effort, we have considered Poisson and NB approach in specifying .

The probability distribution for Poisson is given by:

|  |  |
| --- | --- |
|  |  |

where is the expected number of crashes occurring in TAZ over a given period of time in segment . We can express as a function of explanatory variable by using a log-link function as: , where is a vector of parameters to be estimated specific to segment . However, one of the most restrictive assumptions of Poisson regression, often being violated, is that the conditional mean is equal to the conditional variance of the dependent variable.

The variance assumption of Poisson regression is relaxed in NB by adding a Gamma distributed disturbance term to Poisson distributed count data (Jang, 2005). Given the above setup, the NB probability expression for conditional on belonging to segment can be written as:

|  |  |
| --- | --- |
|  |  |

where, is the Gamma function and is the NB dispersion parameter specific to segment .

Finally, the log-likelihood function for the latent segmentation based count model can be written as:

|  |  |
| --- | --- |
|  |  |

Substitution of by Equations 3 and 4 into Equation 5 results in latent segmentation based Poisson (LP) and latent segmentation based NB (LNB) models, respectively. The parameters to be estimated in the model of Equation 5 are: and for each latent segmentation based model along with for LP model and and for LNB models. The parameters are estimated using maximum likelihood approaches. The model estimation is achieved through the log-likelihood functions programmed in Gauss. In the application of these models, determining the appropriate number of segments is a critical issue with respect to interpretation and inferences. Therefore, we estimate these models with increasing numbers of segments until an addition of a segment does not add value to the model in terms of data fit and model interpretation.

It is important to mention here that the estimation of latent segmentation based models using quasi-Newton routines can be computationally unstable (see Bhat, 1997 for more discussion). There may exist multiple local optimal solutions in such models with non-linear in parameters’ specification. Therefore, the estimation of such models requires employing good starting values for obtaining the global optimum in maximum likelihood estimation. To deal with the issue of local optima, we have estimated latent segmentation based models with multiple starting parameters and chose the model with best likelihood. The estimation results were stable with different starting values (Sobhani et al., 2013 for related discussion on latent segmentation model estimation).

**3. DATA**

**3.1 Study Areas**

Our study areas include: (1) the Island of Montreal associated with 837 TAZs with a population of about 1.8 million and covers an area of approximately 499 km2 and (2) the City of Toronto associated with 672 TAZs with a population of about 2.6 million and covers an area of approximately 630 km2 (Statistics Canada, 2011). Montreal is an old city and is characterized by a heterogeneous built environment with a dense old city near the original port. The transportation system of the city is characterized by a highly developed highway system, as well as transit system that includes a heavy-rail metro, commuter trains and an extensive bus network. Montreal has more than 430 km of bicycle facilities where the mode share for bicycle mode is 2.4% (Pucher et al., 2011). On the other hand, Toronto has one of the most extensive public transit systems in North America and the transit network includes bus, subway and streetcar system. The city has 1.7% of mode share for cycling with approximately 11 km of bike lanes/paths per 100,000 population (Pucher et al., 2011). Both these cities are home to a thriving bike culture and are investing substantially in enhancing existing bicycle infrastructure and building additional bicycling infrastructure to encourage more bicycle usage (Pucher, 2005). Therefore, given this change in encouraging more people to travel by bicycle, it is important to identify the factors contributing to bicycle-motor vehicle crashes at the planning level to make cycling safer and a more attractive mode of transportation.

**3.2 Data Description**

This study is focused on bicycle-motor vehicle crash data at the zonal level. Data for our empirical analysis are sourced from the two most populous cities in Canada, Montreal and Toronto, for the year 2006 through 2010. The datasets for Montreal and Toronto are downloaded from the newspaper data archives of Montreal Gazette (<http://www.montrealgazette.com/>) and The Globe and Mail (<http://www.theglobeandmail.com/>), respectively. The datasets were obtained by the newspapers from the official crash databases maintained by the Societé d'assurance automobile du Québec (SAAQ) for Montreal and the City of Toronto’s Traffic Safety Unit for Toronto. The geocoded crash data are aggregated at the level of TAZ for each year for both cities. For the five years, Montreal has a record of 4,185 bicycle crashes with an average of 0.73 crashes (ranging from 0 to 28 crashes) per year per TAZ. On the other hand, the city of Toronto has an average of 1.63 crashes (ranging from 0 to 23) per year per TAZ with a total record of 5,475 bicycle crashes for the five years period.

In addition to the crash databases, the explanatory attributes considered in the empirical study are also aggregated at the TAZ level. For the empirical analysis, we selected variables that can be grouped into six broad categories[[3]](#footnote-3): accessibility measures, exposure measures, sociodemographic characteristics, socioeconomic characteristics, road network characteristics and built environment. For both cities, these data are extracted from the Geographic information system (GIS) data archive of Transportation Research at McGill (TRAM) of McGill University, Canada (<http://tram.mcgill.ca/>)[[4]](#footnote-4).

For Montreal, accessibility measures considered include number of bus stops and bus route length; exposure measures considered include transit commuters, walk commuters, other mode commuters, male bike commuters, female bike commuters, number of vehicles and length of bike lanes; sociodemographic characteristics considered include dependence (defined as proportion of youth and elderly relative to working adults), non-permanent residents, African population, Asian population, European population and population with diploma degree; socioeconomic characteristics considered include population without income, commuting time, median commuting time and median TAZ income; road network characteristics considered include number of dead-ends, number of one-way link, number of intersections and length of highway; and finally built environment considered include number of bars, lot coverage, number of schools, distance from central business district (CBD), land use mix[[5]](#footnote-5) and presence of university. For Toronto. accessibility measures considered include number of bus stops and number of metro stops; exposure measures considered include number of private cars, household density, length of bike route and total TAZ area; sociodemographic characteristics considered include population less than 25 years old; socioeconomic characteristics considered include number of employed person and average TAZ income; road network characteristics considered include number of intersections and length of local road; and finally built environment considered include distance from CBD and land use mix.

Table 2 offers a summary of the sample characteristics of the exogenous factors in the estimation dataset. To conserve on space, we present the sample characteristics only for the Montreal Island. Table 2 represents the definition of variables considered for final model estimation along with the zonal minimum, maximum and average values of continuous variables and percentages of indicator variables for Montreal. The final specification of the model development was based on removing the statistically insignificant variables in a systematic process based on statistical significance (90% significance level). The specification process was also guided by prior research and parsimony considerations. In estimating the models, several functional forms and variable specifications are explored. The functional form that provided the best result is used for the final model specifications and, in Table 2, the variable definitions are presented based on these final functional form of variables.

**4. EMPIRICAL ANALYSIS**

**4.1 Model Selection**

The empirical analysis of bicycle crash frequency involves the estimation of four models: (1) Poisson, (2) Negative Binomial (NB), (3) Latent Segmentation based Poisson (LP) and, (4) Latent Segmentation based Negative Binomial (LNB) model. Prior to discussing the estimation results, we compare the performance of these models in this section. The model comparisons are undertaken in two stages. First, we determine the appropriate number of latent segments for the estimated latent segmentation based count models. Second, we compare the unsegmented models with the more general latent segmentation based count models in order to assess the importance of accounting for heterogeneity in estimating zonal level crash frequency models.

*Determining the Appropriate Number of Latent Segments*

The estimation of the latent segmentation based model involves probabilistic assignment of TAZs into a given number of segments based on the available exogenous variables. Determining the appropriate number of segments in estimating these models is a critical issue with respect to interpretation and inferences. Among different traditionally used information criterion (IC) (Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), adjusted BIC), BIC imposes substantially higher penalty on over-fitting and is the most commonly used IC for identifying the appropriate number of classes for latent segmentation based analysis (Nylund et al., 2007). Therefore, we estimated LP and LNB models with increasing numbers of segments and computed the BIC values for each of these models. The BIC for a given empirical model is equal to:

|  |  |
| --- | --- |
|  |  |

where is the log likelihood value at convergence, is the number of parameters, and is the number of observations. The model with the lower BIC is the preferred model.

The calculated BIC values for both the island of Montreal and the city of Toronto are presented in the first row panel of Table 3 along with the log-likelihood at convergence for all the estimated latent segmentation based models. For Montreal, from Table 3 we can see that LP III has a lower BIC than LP II model. However, the expected mean crash (sample share) of one of the segments of LP III model is 121.01 (3.82%) whereas the observed mean crash of the sample is 0.73. Thus, we can recognize that LP III model does not satisfy the behavioral interpretation in terms of resulting expected number of bicycle-motor vehicle crashes for Montreal. Hence, we only consider the LP II model as plausible. Overall, LNB II has a lower BIC, hence we have selected LNB with two segments as the preferred model among all estimated latent segmentation based models for Montreal. For Toronto, after extensively testing for four segments in LP and LNB approaches we found that the models collapse to the three segment LP and LNB models, respectively. Among all estimated latent segmentation based models LNB III model provides superior fit based on BIC measures. Thus, we selected LNB with three segments as the preferred model for Toronto.

*Comparing the Unsegmented and Segmented Models*

To compare the performance of estimated Poisson, NB and the best fitted LP and LNB models, BIC measure is used. The BIC values are computed as shown in Equation 6. The computed BIC values are presented in the second row panel of Table 3. The comparison exercise clearly highlights the superiority of the LNB model in terms of data fit compared to all the other models for Montreal (LNB II) as well as for Toronto (LNB III). This comparison exercise suggests that bicycle crash count events for Montreal and Toronto are better characterized by the NB model with heterogeneity across different TAZs.

**4.2 Estimation Results**

In explaining the effect of exogenous variables, we will restrict ourselves to the discussion of the LNB II model for Montreal[[6]](#footnote-6). For simplicity, we will refer LNB II model as LNB model in the following sections. Table 5 presents the estimation results of the LNB model. Following Bhat (1997), we first present an intuitive discussion of the segmentation component followed by the discussion of segmentation component parameters and crash frequency component parameters specific to segment 1 and 2 of LNB model for Montreal.

*Characterizing the Segments in the LNB Model*

To delve into the segmentation characteristics, the model estimates are used to generate information on: 1) sample share across the two segments, and 2) expected mean of crash count events within each segment. These estimates are presented in the first row panel of Table 4. From the estimates, it is evident that the probability of TAZs being assigned to segment 2 is substantially higher than the probability of being assigned to segment 1 (0.61 versus 0.39). Further, the expected number of bicycle-motor vehicle crash events conditional on their belonging to a particular segment offer contrasting results indicating that the two segments exhibit distinct crash risk profiles in the current research context. From Table 4, it is clear that expected mean of crash count events for TAZs assigned to segment 1 is much higher than the observed sample mean (1.24 versus 0.73) while mean of crash count events for TAZs assigned to segment 2 is lower than the observed sample mean (0.41 versus 0.73). Thus, we may label segment 1 as the “high risk segment” and segment 2 as the “low risk segment”.

The latent segmentation component determines the overall prevalence of each segment, as well as the probability of a TAZ being assigned to one of the two latent segments based on explanatory variables. In our empirical analysis, the explanatory variables that affect the allocation of TAZs to segments include length of bike lane, length of highway, distance from CBD, land use mix and number of intersections. Further, to illustrate the characteristics of each segment, we compute the mean values of variables in the segmentation components (see Bhat (1997) for details on computing these means). The results are presented in the second row panel of Table 4 along with the overall sample shares of these variables. The differences in the mean values of the segmentation variables indicate that the variables representing number of intersections and distance from CBD and kilometers of designated bike lane offer the most substantial differences across the two segments. Based on these estimates (means and differences in the mean values), we can argue that zones within segment 1 are more likely to be characterized by higher number of intersections near urban core with lower bike lane facility length, while segment 2 is more likely to consist of zones away from urban core with fewer intersections and more designated bike lanes. The characteristics of these segment specific variables are presented in the following section.

*Segment Membership Component*

The results in Table 5 provide the effects of these control variables, using the high risk segment (segment 1) as the base segment. Thus, a positive (negative) sign for a variable in the segmentation component indicates that TAZs with the variable characteristics are more (less) likely to be assigned to the low risk segment relative to the high risk segment. The positive sign on the constant term does not have any substantive interpretation, and simply reflects the larger size of the low risk segment compared to the high risk segment.

From the estimation results of segment membership components, we can observe that TAZs with more kilometers of designated bike lane length are more likely to be assigned to low risk segment. The result associated with highway length reflects that an increase in total kilometer length in highway increases the likelihood of assigning TAZs to lower risk segment.

The possibility of being allocated to low risk segment increases with increasing distance from CBD to the TAZ. The TAZs with higher land use mix are less likely to be assigned to the low risk segment. The result while seeming counter-intuitive is also a reflection of increased bicycling exposure. An increase in total number of intersections in a TAZ decreases the likelihood of assigning the TAZ to the lower risk segment.

*Crash Risk Component: High Risk Segment (Segment 1)*

The bicycle crash risk component within the high risk segment (segment 1) is discussed in this section by variable groups. A positive (negative) sign for a variable in the crash count component of Table 5 indicates that an increase in the variable is likely to result in more (less) bicycle-motor vehicle crashes.

Accessibility measures: With respect to the accessibility measures, none of the variables are found to affect bicycle crash risk in the high risk segment.

Exposure measures: In the high risk segment, the results for the number of commuters based on different commute modes reveal that TAZs with higher number of transit and walk commuters, proxies for bicycle activities, increase the likelihood of bicycle-motor vehicle collisions. The result associated with other mode (taxi, motorcycle, paratransit) commuters reflects lower probability of bicycle crash risk with higher number of other mode commuters. As found in previous studies (Kim et al., 2010), our study also found that more vehicles within a TAZ leads to higher probability of bicycle crashes. The reader would note that if more detailed information on destination of the commuter is available it will allow us to enhance the model results.

Sociodemographic characteristics: In terms of sociodemographic characteristics, dependence variable (defined as proportion of youth and elderly relative to working adults) reveals a lower probability of bicycle-motor vehicle crash risk for higher values of the variable. Increased number of African population in zones are positively associated with increased number of bicycle crashes. The estimation results also indicate that the TAZ with greater number of European people are likely to experience less bicycle crashes.

Socioeconomic characteristics: The only socioeconomic characteristic influencing bicycle crash risk for the high risk segment is the zonal level median income. Crashes are positively associated with medium zonal median income relative to zone of low and high median income.

Road network characteristics: For segment 1, none of the variables within road network characteristics are found to significantly influence bicycle crash risk in the current study context.

Built environment: The result associated with lot coverage, a proxy for neighborhood compactness, reflects that an increase in lot coverage increases the likelihood of bicycle crash risk. We also found that presence of university has a positive correlation with bicycle crash risk. Surprisingly, crashes are negatively associated with higher number of schools in a zone.

*Crash Risk Component: Low Risk Segment (Segment 2)*

The crash risk component within the low risk segment (segment two) is discussed in this section. The LNB model corresponding to low risk segment provides variable impacts that are significantly different, in magnitude as well as in sign (for a few variables), from the impacts observed for the control variables in high risk segment.

Accessibility measures: Société de transport de Montréal (STM) bus stops and STM bus route length are two of the accessibility measures that significantly influence bicycle crash risk for the low risk segment. An increase in the number of STM bus stops increases the likelihood of bicycle-motor vehicle crashes at the TAZ level. Our analysis also shows that TAZs with more STM bus routes are likely to be positively correlated with higher bicycle crash risk.

Exposure measures: In segment two, increased presence of transit and walk commuters are associated with higher bicycle crash risk as in segment one but the magnitude of the impact is more pronounced in the second segment. As expected, bicycle crash risk is also found to be higher for the TAZs with more male and female cyclist commuters. However, within the two cyclist commuter categories, the male category has a larger impact relative to female cyclist category.

Sociodemographic characteristics: In terms of sociodemographic characteristics, the estimation results indicate that the TAZs with higher number of Asians are likely to experience lower number of bicycle crashes while zones with higher number of non-permanent residents are likely to experience more bicycle crashes. Our study also shows that more people with diploma degree within a TAZ leads to lower probability of bicycle crashes.

Socioeconomic characteristics: Several socioeconomic characteristics considered are found to be significant determinants of bicycle-motor vehicle crash risk. For the low risk segment, increased bicycle-motor vehicle collision is found to be associated with greater number of population without income at the zonal level. As in segment one, medium level median zonal income in segment two (with lower impact) is also positively correlated with bicycle crash risk relative to low and high level median income.

The impact of median travel time to work is also investigated. Median commuting time less than 11 minutes is likely to result in a lower probability of bicycle-motor vehicle crash risk compared to zones with median commuting time ≥ 11 minutes. Higher number of commuters commuting between 7:00 am - 9:00 am and after 9:00 am are associated with lower bicycle crash risk.

Road network characteristics: With respect to road network characteristics, the model estimation result indicates an expected negative correlation of more dead-ends with bicycle-motor vehicle crashes (see Kim et al., 2010 for similar result). An increase in total number of one-way links in a TAZ increases the likelihood of bicycle crash risk.

Built environment: With regards to built environment, the results reveal that bicycle crashes are positively associated with higher number of bars in the neighborhood. In segment 2, presence of university indicator variable has negative impact on bicycle crash risk. Further, the results of the LNB reveal that the presence of more schools in a TAZ is positively associated with bicycle crash risk. The effects of both the presence of university and number of schools variables are strikingly different in the low risk segment compared to the impacts of these variables in the high risk segment. The different impacts in the two segments for these variables highlight how the same variable can have distinct influence on crash risk based on the segment to which the zone is allocated.

**4.3 Predictive Performance Evaluation**

In an effort to assess the predictive performance of the estimated models, computation of several in-sample goodness-of-fit measures are also carried out. In doing so, performance of LNB model is compared with the predictive performance of NB model for verifying the improvement of incorporating population heterogeneity in estimating macro-level bicycle-motor vehicle crash count models. To evaluate the in-sample predictive performance of NB and LNB models, we employ three different fit measures: mean prediction bias (MPB), mean absolute deviation (MAD) and mean squared prediction error (MSPE). MPB represents the magnitude and direction of average bias in model prediction. The model with the lower MPB provides better prediction of the observed data and is computed as:

|  |  |
| --- | --- |
|  |  |

where, and are the predicted and observed, number of crashes occurring over a period of time in a TAZ ( be the index for TAZ ) and is the number of TAZs, respectively. On the other hand, MAD describes average misprediction of the estimated models. The model with lower MAD value closer to zero provides better average predictions of observed data. MAD is defined as:

|  |  |
| --- | --- |
|  |  |

Finally, MSPE quantifies the error associated with model predictions and is defined as:

|  |  |
| --- | --- |
|  |  |

The smaller the MSPE, the better the model predicts the observed data.

Table 6 presents the values for the three measures for NB and LNB models for Montreal and Toronto. The resulting fit measures for comparing the predictive performance clearly indicate that LNB model offers superior fit compared to NB model for Montreal as well as for Toronto. These prediction results further confirm the benefit of accommodating population heterogeneity in modeling bicycle crash counts at a TAZ level.

**5. POLICY ANALYSIS**

**5.1 Elasticity Effects**

The parameter effects of exogenous variables in Table 5 do not directly provide the magnitude of the effects on zonal level bicycle crash counts. For this purpose, we compute aggregate level “elasticity effects” of exogenous variables having positive impacts in either of the segments of the estimated LNB model for the Island of Montreal. We investigate the effect as the percentage change in the expected total zonal bicycle crash counts per year due to the change in exogenous variable for the overall sample as well as for each segment separately to emphasize policy repercussions based on most critical contributory factors. The computed elasticities are presented in Table 7 (see Eluru and Bhat (2007) for a discussion on the methodology for computing elasticities). The corresponding segment level elasticities (column 2 and 3 of Table 7) represent the aggregate percentage change of each segment contributing to overall percentage change in bicycle-motor vehicle crash counts at zonal level. Further, the overall sample elasticity (column 4 of Table 7) are calculated as the summation of segment level elasticities. In calculating the expected percentage change of crash counts, we increase the value of continuous variable by 10% for each TAZ. The numbers in the Table 7 may be interpreted as the percentage change in the expected total zonal bicycle crash counts per year due to the change in exogenous variable. For instance, the elasticity effects for Transit commuters for In-Sample data indicates that, the expected mean crashes will increase by 34.819% with an increase in 10% of transit commuters.

The following observations can be made based on the elasticity effects presented in Table 7. First, the most significant variable in terms of increase in the expected number of bicycle-motor vehicle crash counts include: presence of university, transit commuters, medium median TAZ income, number of vehicles and walk commuters. Second, the results in Table 7 indicate that there are considerable differences in the elasticity effects across two segments, which illustrate the importance of allowing for population heterogeneity in examining aggregate level bicycle-motor vehicle crash counts. Third, it is interesting to note that TAZs belonging to segment 1 have higher elasticities for variables related to presence of university, number of vehicles and medium median TAZ income. On the other hand, TAZs belonging to segment 2 have higher elasticities for variables such as transit commuters, population without income and STM bus stops. Finally, the elasticity analysis assists in providing a clear picture of attribute impacts on zonal level bicycle crash counts. The elasticity analysis conducted provides an illustration on how the proposed model can be applied to determine the critical factors contributing to increase in bicycle crash counts.

**5.2 Spatial Distribution**

The model findings have important implications in terms of countermeasures for zonal level bicycle safety planning. Specifically, the findings can be used to identify targeted countermeasures for the TAZs with greater number of bicycle crashes. To illustrate the spatial distribution of bicycle-motor vehicle crash frequency, we present the predicted per year crash counts across different TAZs from LNB model for Montreal in Figure 1. From the spatial crash distribution we can see that bicycle-motor vehicle crashes are dispersed throughout the island with evidence of clustering for high crash zones. Also we can see that risk of getting involved in bicycle-motor vehicle crashes is higher in most of the zones near CBD. This spatial illustration can easily be used to prioritize TAZs based on which the safety treatment can be implemented for the most significant contributory factors (as presented in Table 7) in enhancing bicycle safety features of these high crash risk zones.

The calculated predictions have also important implications in developing proactive safety-conscious planning tools for improving overall bicycle safety at zonal level. For instance, proactive safety planning decision tools can be identified to address most important contributory factors of bicycle-motor vehicle crash counts based on allocation of TAZs to different segments. The spatial distribution of TAZs to segment 1 (high risk) for LNB model is presented in Figure 2. This map can easily be used to identify TAZ specific proactive safety management strategies reflecting the allocation of zones to different segments. The development of such spatial profiles will allow planners to identify high risk zones for screening and treatment purposes.

**6. CONCLUSIONS**

This paper formulates and estimates econometric models by building on the traditional count regression models to investigate factors affecting bicycling crashes at the Traffic Analysis Zone (TAZ) level for two Canadian cities – Montreal and Toronto. To accommodate for the potential variation in the impact of exogenous factors we formulated latent segmentation based count models. The entire set of alternative modeling approaches considered for this investigation include: Poisson, Negative Binomial (NB), Latent Segmentation based Poisson (LP) and Latent Segmentation based Negative Binomial (LNB) model. For the empirical analysis we selected bicycle-motor vehicle crash datasets from the Island of Montreal and from the City of Toronto for the years 2006 through 2010. The models were estimated using a comprehensive set of exogenous variables - accessibility measures, exposure measures, sociodemographic characteristics, socioeconomic characteristics, road network characteristics and built environment. The comparison of the estimated latent segmentation based models, based on information criterion metrics, highlighted the superiority of the LNB model with two segments for Montreal and LNB model with three segments for Toronto in terms of data fit compared to the other estimated models. According to our results, the impacts of exogenous variables among different segments were different (for some variables) in magnitude as well as in sign for both cities.

In an effort to further assess the predictive performance of the estimated models, computation of several in-sample goodness-of-fit measures were also carried out. We employed three different fit measures: mean prediction bias (MPB), mean absolute deviation (MAD) and mean squared prediction error (MSPE). The resulting fit measures for comparing the predictive performance of estimated models clearly indicated that LNB offer superior fit compared to NB model for Montreal and Toronto. These prediction results further confirmed the benefit of accommodating population heterogeneity in modeling bicycle crash counts at a TAZ level. The model estimates were also augmented by conducting policy analysis including elasticity analysis along with a spatial representation of model outcomes for the overall sample as well as for each segment separately to emphasize policy repercussions based on most critical contributory factors. The policy analysis conducted provided an illustration on how the proposed model can be applied to determine the critical factors contributing to increase in bicycle crash counts.

The variable effects obtained from our models have important implications in terms of enforcement, engineering and educational strategies. In terms of engineering measures, bicycle facilities that separate cyclists from motor vehicle and increase visibility of cyclists to motorists (for instance: bike box at intersections, special bike turn lanes, removable barrier between bicycle and motor vehicle flows or reallocating space from motor vehicle to bicycles) should be provided in the vicinity of universities. Moreover, traffic calming measures have potential to reduce bicycle-motor vehicle collisions both in transit- and auto-oriented neighborhoods. With respect to education, our results endorse public awareness efforts and education campaigns specifically for economically deprived areas. Rigorous traffic education for safe cycling are also needed for both non-motorists and motorists of zones with medium median TAZ income and TAZs with more transit and walk commuters. Besides, stricter enforcement of traffic regulations for motorists and non-motorists near universities are needed for cyclists’ safety in these areas.

From our analysis, it is evident that analysts should explore the estimation of latent segmentation models for crash frequency analysis. To be sure, the study is not without limitations. In our research effort, we explored factors contributing to bicycle-motor vehicle crash frequency at zonal level. However, in our analysis, we have not considered spatial correlation among adjacent zones in examining the critical factors. It will be an interesting exercise to model the impact of spatial effects on segment specific crash count models. It would be a methodologically challenging exercise to consider spatial correlation within a latent segmentation model structure for crash frequency analysis. Further, we develop a macro-level analysis framework that explore bicycle crashes from a macro-perspective. A finer resolution micro-level analysis would complement our proposed approach in accommodating the effects for bicyclist exposure, locational attributes of crashes (intersection, mid-block etc.) and bicycle infrastructure in examining bicycle safety. Finally, it would be interesting to examine the influence of bicycle sharing systems that have been in operation in Montreal and Toronto in future research efforts.

**Acknowledgement**

The authors would like to gratefully acknowledge Prof. Ahmed El-Geneidy’s TRAM website for providing access to the GIS data and also would like to acknowledge help provided by Adham Badran, and Michelle Pinto in preparing the datasets. The authors would also like to acknowledge critical inputs from anonymous reviewers on a previous version of the paper.

**References**

Abdel-Aty, M., Lee, J., Siddiqui, C., & Choi, K. (2013). Geographical unit based analysis in the context of transportation safety planning. Transportation Research Part A: Policy and Practice, 49(0), 62-75.

Abdel-Aty, M., Siddiqui, C., Huang, H., & Wang, X. (2011). Integrating trip and roadway characteristics to manage safety in traffic analysis zones. Transportation Research Record: Journal of the Transportation Research Board, 2213, 20-28.

Aguero-Valverde, J., & Jovanis, P. P. (2006). Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis & Prevention, 38(3), 618-625.

Amoros, E., Martin, J. L., & Laumon, B. (2003). Comparison of road crashes incidence and severity between some French counties. Accident Analysis & Prevention, 35(4), 537-547.

Andersson, A.-L., & Bunketorp, O. (2002). Cycling and alcohol. Injury, 33(6), 467-471.

Bago d'Uva, T. (2006). Latent class models for utilisation of health care. Health economics, 15(4), 329-343.

Bhat, C. R. (1997). Endogenous segmentation mode choice model with an application to intercity travel. Transportation Science, 31(1), 34-48.

Bhat, C. R., Guo, J. Y., & Sardesai, R. (2005). Non-Motorized Travel in the San Francisco Bay Area. University of Texas, Austin, 17-27.

Brüde, U., & Larsson, J. (1993). Models for predicting accidents at junctions where pedestrians and cyclists are involved. How well do they fit? Accident Analysis & Prevention, 25(5), 499-509.

Carter, D., & Council, F. (2007). Factors contributing to pedestrian and bicycle crashes on rural highways. Paper presented at the Transportation Research Board 86th Annual Meeting Paper.

Chaney, R. A., & Kim, C. (2014). Characterizing Bicycle Collisions by Neighborhood in a Large Midwestern City. Health Promotion Practice, 15(2), 232-242.

Chen, C., Lin, H., & Loo, B. P. (2012a). Exploring the impacts of safety culture on immigrants’ vulnerability in non-motorized crashes: a cross-sectional study. Journal of Urban Health, 89(1), 138-152.

Chen, L., Chen, C., Srinivasan, R., McKnight, C. E., Ewing, R., & Roe, M. (2012b). Evaluating the safety effects of bicycle lanes in New York City. American Journal of Public Health, 102(6), 1120-1127.

Cho, G., Rodríguez, D. A., & Khattak, A. J. (2009). The role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. Accident Analysis & Prevention, 41(4), 692-702.

Cottrill, C. D., & Thakuriah, P. V. (2010). Evaluating pedestrian crashes in areas with high low-income or minority populations. Accident Analysis & Prevention, 42(6), 1718-1728.

De Guevara, F. L., Washington, S. P., & Oh, J. (2004) Forecasting crashes at the planning level: Simultaneous negative binomial crash model applied in Tucson, Arizona. Transportation Research Record: Journal of the Transportation Research Board, 1987, 191-199.

Dong, C., Clarke, D. B., Richards, S. H., & Huang, B. (2014). Differences in passenger car and large truck involved crash frequencies at urban signalized intersections: An exploratory analysis. Accident Analysis & Prevention, 62(0), 87-94.

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.

Eluru, N., Bagheri, M., Miranda-Moreno, L. F., & Fu, L. (2012). A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accident Analysis & Prevention, 47(0), 119-127.

Givoni, M., & Rietveld, P. (2007). The access journey to the railway station and its role in passengers’ satisfaction with rail travel. Transport Policy, 14(5), 357-365.

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: Contrasts with mixed logit. Transportation Research Part B: Methodological, 37(8), 681-698.

Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2010b). Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. Accident Analysis & Prevention, 42(2), 676-688.

Hadayeghi, A., Shalaby, A., & Persaud, B. (2003). Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. Transportation Research Record: Journal of the Transportation Research Board, 1840, 87-95.

Hadayeghi, A., Shalaby, A., & Persaud, B. (2007). Safety Prediction Models: Proactive Tool for Safety Evaluation in Urban Transportation Planning Applications. Transportation Research Record: Journal of the Transportation Research Board, 2019, 225-236.

Hadayeghi, A., Shalaby, A., & Persaud, B. (2010a). Development of planning-level transportation safety models using full Bayesian semiparametric additive techniques. Journal of Transportation Safety and Security, 2(1), 45-68.

Hamann, C., & Peek-Asa, C. (2013). On-road bicycle facilities and bicycle crashes in Iowa, 2007–2010. Accident Analysis & Prevention, 56(0), 103-109.

Herslund, M.-B., & Jørgensen, N. O. (2003). Looked-but-failed-to-see-errors in traffic. Accident Analysis & Prevention, 35(6), 885-891.

Huang, H., Abdel-Aty, M., & Darwiche, A. (2010). County-Level Crash Risk Analysis in Florida. Transportation Research Record: Journal of the Transportation Research Board, 2148, 27-37.

Hunter, W., Stewart, J., & Stutts, J. (1999). Study of Bicycle Lanes versus Wide Curb Lanes. Transportation Research Record: Journal of the Transportation Research Board, 1674, 70-77.

Jang, T. Y. (2005). Count data models for trip generation. Journal of Transportation Engineering, 131(6), 444-450.

Jung, S., Jang, K., Yoon, Y., & Kang, S. (2014). Contributing factors to vehicle to vehicle crash frequency and severity under rainfall. Journal of safety research, 50(0), 1-10.

Karlaftis, M. G., & Tarko, A. P. (1998). Heterogeneity considerations in accident modeling. Accident Analysis & Prevention, 30(4), 425-433.

Kim, K., Pant, P., & Yamashita, E. (2010). Accidents and Accessibility. Transportation Research Record: Journal of the Transportation Research Board, 2147, 9-17.

LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. Accident Analysis & Prevention, 32(5), 651-658.

Lee, J., Abdel-Aty, M., & Jiang, X. (2014). Development of zone system for macro-level traffic safety analysis. Journal of Transport Geography, 38(0), 13-21.

Lee, J., Abdel-Aty, M., & Jiang, X. (2015). Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. Accident Analysis & Prevention, 78(0), 146-154.

Levine, N., Kim, K. E., & Nitz, L. H. (1995). Spatial analysis of Honolulu motor vehicle crashes: II. Zonal generators. Accident Analysis & Prevention, 27(5), 675-685.

Li, Z., Wang, W., Liu, P., Bigham, J. M., & Ragland, D. R. (2013). Using Geographically Weighted Poisson Regression for county-level crash modeling in California. Safety Science, 58, 89-97.

Loo, B. P., & Tsui, K. (2010). Bicycle crash casualties in a highly motorized city. Accident Analysis & Prevention, 42(6), 1902-1907.

Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice, 44(5), 291-305.

Love, D. C., Breaud, A., Burns, S., Margulies, J., Romano, M., & Lawrence, R. (2012). Is the three-foot bicycle passing law working in Baltimore, Maryland? Accident Analysis & Prevention, 48(0), 451-456.

MacNab, Y. C. (2004). Bayesian spatial and ecological models for small-area accident and injury analysis. Accident Analysis & Prevention, 36(6), 1019-1028.

Martinez, R., & Veloz, R. A. (1996). A challenge in injury prevention--the Hispanic population. Academic emergency medicine: official journal of the Society for Academic Emergency Medicine, 3(3), 194-197.

Moeinaddini, M., Asadi-Shekari, Z., & Zaly Shah, M. (2014). The relationship between urban street networks and the number of transport fatalities at the city level. Safety Science, 62, 114-120.

Mura, T. (1992). Visual search in intersections: An underlying mechanism. IATSS Research, 16, 4249.

Naderan, A., & Shahi, J. (2010). Aggregate crash prediction models: Introducing crash generation concept. Accident Analysis & Prevention, 42(1), 339-346.

Ng, K.-s., Hung, W.-t., & Wong, W.-g. (2002). An algorithm for assessing the risk of traffic accident. Journal of safety research, 33(3), 387-410.

Noland, R. B. (2003). Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. Accident Analysis & Prevention, 35(4), 599-611.

Noland, R. B., & Quddus, M. A. (2004a) Analysis of pedestrian and bicycle casualties with regional panel data. Transportation Research Record: Journal of the Transportation Research Board, 1897, 28-33.

Noland, R. B., & Quddus, M. A. (2004b). A spatially disaggregate analysis of road casualties in England. Accident Analysis & Prevention, 36(6), 973-984.

Noland, R. B., & Quddus, M. A. (2004c). A spatially disaggregate analysis of road casualties in England. Accident Analysis & Prevention, 36(6), 973-984.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling, 14(4), 535-569.

Pai, C.-W. (2011). Overtaking, rear-end, and door crashes involving bicycles: An empirical investigation. Accident Analysis & Prevention, 43(3), 1228-1235.

Park, B.-J., & Lord, D. (2009). Application of finite mixture models for vehicle crash data analysis. Accident Analysis & Prevention, 41(4), 683-691.

Park, B.-J., Lord, D., & Hart, J. D. (2010). Bias properties of Bayesian statistics in finite mixture of negative binomial regression models in crash data analysis. Accident Analysis & Prevention, 42(2), 741-749.

Park, J., Abdel-Aty, M., Lee, J., & Lee, C. (2015). Developing crash modification functions to assess safety effects of adding bike lanes for urban arterials with different roadway and socio-economic characteristics. Accident Analysis & Prevention, 74, 179-191.

Parkin, J., & Meyers, C. (2010). The effect of cycle lanes on the proximity between motor traffic and cycle traffic. Accident Analysis & Prevention, 42(1), 159-165.

Pucher, J. (2005). Cycling trends & policies in Canadian cities John Pucher & Ralph Buehler. World Transport Policy & Practice, 11(1), 43-61.

Pucher, J., Buehler, R., & Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. Transportation Research Part A: Policy and Practice, 45(6), 451-475.

Pulugurtha, S. S., & Thakur, V. (2015). Evaluating the effectiveness of on-street bicycle lane and assessing risk to bicyclists in Charlotte, North Carolina. Accident Analysis & Prevention, 76(0), 34-41.

Quddus, M. A. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data. Accident Analysis & Prevention, 40(4), 1486-1497.

Rasanen, M., & Summala, H. (2000). Car Drivers' Adjustments to Cyclists at Roundabouts. Transportation Human Factors, 2(1), 1-17.

Reynolds, C., Harris, M. A., Teschke, K., Cripton, P. A., & Winters, M. (2009). The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environmental Health, 8(1), 47.

Schneider, R. J., Ryznar, R. M., & Khattak, A. J. (2004). An accident waiting to happen: A spatial approach to proactive pedestrian planning. Accident Analysis & Prevention, 36(2), 193-211.

Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. Accident Analysis & Prevention, 45, 382-391.

Sobhani A., N. Eluru and A. Faghih-Imani (2013), "A latent segmentation based multiple discrete continuous extreme value model," Transportation Research Part B 58, 154-169

Stamatiadis, N., & Puccini, G. (2000). Socioeconomic descriptors of fatal crash rates in the Southeast USA. Injury Control and Safety Promotion, 7(3), 165-173.

Statistics Canada. (2011). 2011 Census of Population. Ottawa: Statistics Canada.

Strauss, J., & Miranda-Moreno, L. F. (2013). Spatial modeling of bicycle activity at signalized intersections. Journal of Transport and Land Use, 6(2), 47-58.

Strauss, J., Miranda-Moreno, L. F., & Morency, P. (2013). Cyclist activity and injury risk analysis at signalized intersections: A Bayesian modelling approach. Accident Analysis & Prevention, 59(0), 9-17.

Turner, S., Francis, T., Roozenburg, A., & Land Transport, N. (2006). Predicting accident rates for cyclists and pedestrians: Land Transport New Zealand.

Ukkusuri, S., Hasan, S., & Aziz, H. (2011). Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency. Transportation Research Record: Journal of the Transportation Research Board, 2237, 98-106.

Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. Safety Science, 50(4), 1141-1151.

Walker, I. (2007). Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. Accident Analysis & Prevention, 39(2), 417-425.

Wedel, M., DeSarbo, W. S., Bult, J. R., & Ramaswamy, V. (1993). A latent class Poisson regression model for heterogeneous count data. Journal of Applied Econometrics, 8(4), 397-411.

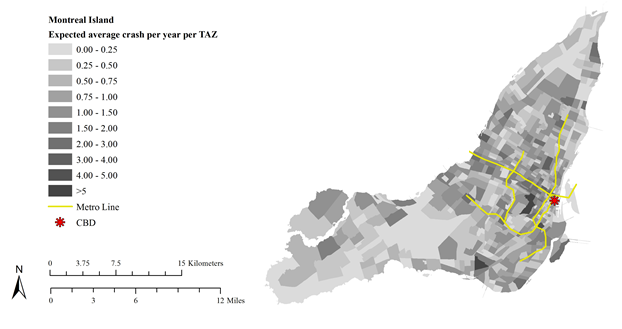
Wei, F., & Lovegrove, G. (2013). An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. Accident Analysis & Prevention, 61, 129-137.

Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis & Prevention, 41(1), 137-145.

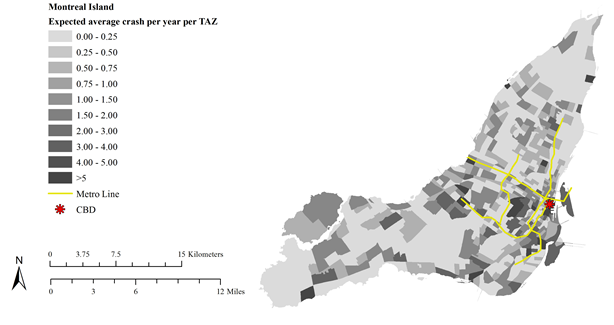
Yasmin, S., Eluru, N., Bhat, C. R., & Tay, R. (2014). A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic Methods in Accident Research, 1, 23-38.

Zou, Y., Zhang, Y., & Lord, D. (2014). Analyzing different functional forms of the varying weight parameter for finite mixture of negative binomial regression models. Analytic Methods in Accident Research, 1, 39-52.

**FIGURE 1 Spatial Distribution of Expected Bicycle-Motor Vehicle Crash Frequency for LNB model for Montreal Island**



**FIGURE 2 Spatial Distribution of TAZs to be assigned to Segment 1 (High Risk) for LNB model for Montreal Island**



**TABLE 1 Summary of Existing Macro-Level Crash Frequency Studies**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Paper** | **Methodological Approach** | **Unit of Analysis** | **Spatial Unit** | **Crash level** | **Independent Variable Considered** | | | | | |
| **Accessibility measures** | **Exposure Measures** | **Sociodemographic characteristics** | **Socioeconomic characteristics** | **Road network characteristics** | **Built environment** |
| Noland and Quddus (2004a) | Negative binomial | Pedestrian  Bicyclist | Standard statistical regions | Fatal/Serious injury crashes, Slight injury crashes | --- | Yes | Yes | Yes | Yes | --- |
| Wier et al.  (2009) | Ordinary least square regression | Pedestrian crash | Census tract | Injury crashes | --- | Yes | Yes | Yes | Yes | Yes |
| Ukkusuri et al. (2012) | Negative binomial, Negative binomial with heterogeneity in  dispersion parameter, Zero-inflated negative binomial | Pedestrian crash | Census tract | Total crashes | Yes | Yes | Yes | Yes | Yes | Yes |
| Lee et al. (2015) | Multivariate Poisson lognormal conditional  autoregressive model | Motor vehicle, bicycle and pedestrian crashes | Traffic analysis zone | Total crashes |  | Yes | Yes | Yes | Yes |  |
| Wei and Lovegrove (2013) | Negative binomial | Bicycle crash | Traffic analysis zone | Total crashes | Yes | Yes | Yes | Yes | Yes | Yes |
| Siddiqui et al. (2012) | Negative binomial,  Bayesian log-normal model | Pedestrian and bicycle crash | Traffic analysis zone | Total pedestrian crashes, Total bicycle crashes | --- | Yes | Yes | Yes | Yes | Yes |
| Cho et al. (2009) | Path analysis | Pedestrian and bicycle crash | Community analysis zones | Total crashes | Yes | Yes | --- | --- | Yes | Yes |
| MacNab (2004) | Bayesian spatial and ecological regression model | Young (Age 0-25) pedestrian and Bicyclist crash | Local health areas | Injury crashes | --- | Yes | --- | Yes | Yes | Yes |
| Noland (2003) | Random Effect negative binomial | Crash | State | Fatal crashes, Injury crashes | --- | Yes | Yes | Yes | Yes | --- |
| Noland and Quddus (2004b) | Negative binomial | Crash | County | Total crashes, Fatal crashes | --- | Yes | Yes | Yes | Yes | --- |
| Karlaftis et al. (1998) | Cluster analysis, Negative binomial | Aged driver crash |  | Total crashes, Urban crashes, Rural crashes | --- | Yes | --- | Yes | Yes | --- |
| Huang et al. (2010) | Bayesian spatial model | Crash | County | Total crashes, Severe crashes | --- | Yes | Yes | Yes | Yes | --- |
| Amoros et al. (2003) | Negative binomial | Crash | County | Total crashes, Fatal crashes | --- | --- | --- | --- | Yes | --- |
| Ukkusuri et al. (2011) | Random parameter negative binomial | Pedestrian crash | Census tract | Total crashes | Yes | Yes | Yes | --- | Yes | Yes |
| Cottrill et al. (2010) | Poisson Regression with heterogeneity | Pedestrian crash | Census Tract | Total crashes | Yes | Yes | Yes | Yes | Yes | Yes |
| Quddus (2008) | Negative Binomial, Spatial autoregressive model, Bayesian hierarchical model | Motorized vehicle crash, Non-motorized vehicle crash, Pedestrian crash | Census ward | Fatal crashes, Serious injury crashes, Slight  Injury crashes | --- | Yes | Yes | --- | Yes | --- |
| Naderan and Shahi (2010) | Negative binomial | Crash | Traffic analysis zone | Total crashes, PDO crashes, Injury crashes, Fatal crashes | --- | Yes | --- | --- | --- | --- |
| Ng et al. (2002) | Cluster analysis, Negative Binomial | Crash | Traffic analysis zone | Total crashes, Fatal crashes, Pedestrian crashes | --- | --- | --- | --- | --- | Yes |
| Abdel-Aty et al. (2011) | Negative binomial | Crash | Traffic analysis zone | Total crashes, Severe crashes, Peak hour crashes, Pedestrian and Bicycle crashes | --- | Yes | --- | --- | Yes | --- |
| Aguero-Valverde and Jovanis (2006) | Full Bayes hierarchical model | Crash | County | Fatal crashes,  Injury crashes | --- | Yes | Yes | Yes | --- | --- |
| Stamatiadis and Puccini 2000 | Quasi induced exposure method | Single vehicle crash Multi vehicle Crash | State | Fatal crashes | --- | Yes | Yes | Yes | --- | --- |
| Noland (2003) | Negative binomial | Crash | State | Fatal injury crashes | --- | Yes | Yes | Yes | Yes | --- |
| LaScala et al. (2000) | Spatial autocorrelation corrected regression | Pedestrian crash | Census tract | Injury crashes | --- | --- | Yes | Yes | Yes | Yes |
| Abdel-Aty et al. (2013) | Poisson-lognormal | Crash  Pedestrian crash | Traffic analysis zones  Block groups Census  tracts | Total crashes, Severe crashes, Pedestrian crashes | --- | Yes | Yes | Yes | Yes | --- |
| Levine et al. (1995) | Spatial lag | Crash | Census block | Total crashes | --- | Yes | --- | Yes | Yes | --- |
| Lee et al. (2014) | Bayesian Poisson Lognormal | Crash | Traffic analysis zone  Traffic safety analysis zone | Total crashes, Severe crashes | --- | Yes | Yes |  | Yes | --- |
| Li et al. (2013) | Geographically  Weighted Poisson Regression | Crash | County | Fatal crashes | --- | Yes | Yes | Yes | Yes | --- |
| Noland and Quddus (2004c) | Negative binomial | Crash | Census wards | Fatal crashes, Serious injury crashes, Slight injury crashes | --- | Yes | Yes | Yes | Yes | Yes |
| De Guevara et al. (2004) | Simultaneous negative binomial | Crash | Traffic analysis zone | Fatal crashes, Injury crashes, PDO crashes | Yes | Yes | Yes | Yes | Yes | Yes |
| Hadayeghi et al. (2003) | Negative binomial  Geographically weighted regression | Crash | Traffic zone | Total crashes, Severe crashes | --- | Yes | Yes | Yes | Yes | --- |
| Moeinaddini et al. (2014) | Negative binomial | Crash | City | Fatalities per million inhabitants | --- | --- | --- | --- | Yes | --- |
| Hadayeghi et al. (2007) | Negative binomial | Crash | Traffic analysis zone | Total crashes, Severe crashes | Yes | Yes | Yes | Yes | Yes | Yes |
| Hadayeghi et al. (2010a) | Geographically Weighted Poisson Regression, Full Bayesian Semiparametric Additive | Crash | Traffic analysis zone | Total crashes, Severe crashes | Yes | Yes | Yes | Yes | Yes | Yes |
| Hadayeghi et al. (2010b) | Geographically Weighted Poisson Regression | Crash | Traffic analysis zone | Total crashes, Severe crashes | Yes | Yes | Yes | Yes | Yes | Yes |

**TABLE 2 Sample Statistics for Montreal Island**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variables Name** | **Definition** | **Zonal** | | |
| **Minimum** | **Maximum** | **Average** |
| **Dependent variable** |  |  |  |  |
| Crashes per year per TAZ | Total number of crashes per year per TAZ | 0.000 | 28.000 | 0.819 |
| **Accessibility measures** |  |  |  |  |
| STM bus stops | Ln(Total Société de transport de Montréal (STM) bus stops in TAZ) | 0.000 | 10.697 | 7.816 |
| STM bus route length | Ln(Total Société de transport de Montréal (STM) bus line kilometer in TAZ) | 0.000 | 4.127 | 1.980 |
| **Exposure measures** |  |  |  |  |
| Transit commuters | Ln(Total transit commuters in TAZ) | -2.567 | 7.444 | 5.913 |
| Walk commuters | Ln(Total walk commuters in TAZ) | -2.005 | 7.069 | 4.416 |
| Other mode commuters | Ln(Total other mode (taxi, motorcycle, paratransit) commuters in TAZ) | -5.412 | 4.605 | 1.356 |
| Male bike commuters | Ln(Total male bike commuters in TAZ) | -4.616 | 5.273 | 1.877 |
| Female bike commuters | Ln(Total female bike commuters in TAZ) | -4.079 | 5.298 | 0.943 |
| Number of vehicles | Ln(Total number of vehicles in TAZ) | 0.432 | 8.991 | 6.390 |
| Length of bike lane | Ln(Total length of designated bike lane kilometer on road in TAZ) | -8.580 | 1.491 | -0.136 |
| **Sociodemographic characteristics** | | | | |
| Dependence | Ratio of youth (19 years or younger) and elderly (65 years or more) to working age persons | 0.000 | 1.557 | 0.547 |
| Non-permanent resident | Ln(Total non-permanent resident in TAZ) | -4.375 | 7.682 | 3.592 |
| African population | Ln(Total African resident in TAZ) | -0.991 | 8.145 | 5.267 |
| Asian population | Ln(Total Asian resident in TAZ) | -2.186 | 8.693 | 5.577 |
| European population | Ln(Total European resident in TAZ) | -3.053 | 8.255 | 5.843 |
| Population with diploma degree | Ln(Total population with diploma degree in TAZ) | -3.935 | 7.204 | 5.163 |
| **Socioeconomic characteristics** | | | | |
| Population without income | Ln(Total population without income in TAZ) | -3.523 | 6.142 | 4.629 |
| Commuters commuting between 7:00 am and 9:00 am | Ln(Total commuters of TAZ commuting between 7:00 am and 9:00 am) | -0.913 | 7.844 | 6.565 |
| Commuters commuting after 9:00 am | Ln(Total commuters of TAZ commuting between 7:00 after 9:00 am) | -1.954 | 6.856 | 5.544 |
| **Road network characteristics** | | | | |
| Number of dead-ends | Ln(total number of dead-ends in TAZ) | 0.000 | 3.367 | 0.477 |
| Number of one-way link | Ln(total number of one-way link in TAZ) | 0.000 | 4.920 | 2.801 |
| Number of intersections | Ln(total number of intersection in TAZ) | 0.000 | 5.746 | 3.061 |
| Length of highway | Ln(total length of highway kilometer in TAZ) | -6.465 | 2.757 | -0.109 |
| **Built environment** |  |  |  |  |
| Number of bars | Ln(total number of bars in TAZ) | 0.000 | 3.045 | 0.130 |
| Lot Coverage | Building foot print area of TAZ/Total area of TAZ | 0.000 | 0.583 | 0.169 |
| Number of school | Number of schools in TAZ | 0.000 | 5.000 | 0.661 |
| Distance from CBD | Ln(distance from CBD to the TAZ (kilometer)) | -2.083 | 3.516 | 1.892 |
| Land use mix | Land use mix = , where is the category of land-use, is the proportion of the developed land area devoted to a specific land-use, is the number of land-use categories in a TAZ | 0.000 | 0.999 | 0.494 |
| **Variables** | | **Sample Share** | | |
| **Frequency (Percentage)** | | |
| **Median TAZ income** | | | | |
| Low (<$40,000) |  | 1565 (37.395) | | |
| Medium ($40,000-$80,000) |  | 460 (10.992) | | |
| High (≥$80,000) |  | 2160(51.613) | | |
| **Median commuting time** |  |  |  |  |
| Less than 11 minutes |  | 135(3.20) | | |
| Equal and more than 11 minutes | | 4050(96.8) | | |
| **Presence of university** |  |  |  |  |
| Presence of university building | | 295 (7.00) | | |
| No university buildings are present | | 3890 (93.0) | | |

**TABLE 3 Measures for Model Selection**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Determining the Appropriate Number of Latent Segments** | | | | | | |
| **Cities** | **Montreal** | | | **Toronto** | | | |
| Number of cases | 4185 | | | 3360 | | | |
| **Models** | **Number of parameters** | **Log-likelihood**  **at Convergence** | **BIC** | **Number of parameters** | **Log-likelihood**  **at Convergence** | **BIC** | |
| LP II | 39 | -4382.91 | 9091.05 | 23 | -4813.44 | 9813.62 | |
| LP III | 46 | -4281.80 | 8947.20 | 32 | -4764.61 | 9789.05 | |
| LNB II | 40 | -4253.84 | 8841.26 | 24 | -4797.67 | 9790.23 | |
| LNB III | 50 | -4226.98 | 8870.91 | 30 | -4759.51 | 9762.61 | |
| **Comparing the Unsegmented and Segmented Models** | | | | | | |
| **Models** | **Number of parameters** | **Log-likelihood**  **at Convergence** | **BIC** | **Number of parameters** | **Log-likelihood**  **at Convergence** | **BIC** |
| **Poisson** | 31 | -4993.30 | 10245.10 | 14 | -4941.82 | 9997.32 |
| **NB** | 24 | -4345.20 | 8890.55 | 13 | -4832.49 | 9770.53 |
| **LNB** | 40 | -4253.84 | 8841.26 | 30 | -4759.51 | 9762.61 |

*NB = Negative Binomial, LP II = Latent Segmentation based Poisson with two segments, LP III = Latent Segmentation based Poisson with three segments, LNB II = Latent Segmentation based Negative Binomial with two segments, LNB III = Latent Segmentation based Negative Binomial with three segments, BIC = Bayesian Information Criterion.*

**TABLE 4 Segment Characteristics and Mean Values of Segmentation Variables for LNB model for Montreal Island**

|  |  |  |  |
| --- | --- | --- | --- |
| **Components** | | **Segments** | |
| **Segment 1** | **Segment 2** |
| Sample shares | | 0.390 | 0.610 |
| Observed mean of crash events | | 0.730 | |
| Expected mean of segment level crash events | | 1.240 | 0.407 |
| **Mean Values of Segmentation Component Variables in Each Segment** | | | |
| **Variable Names\*** | **Overall Sample** | **Segment 1** | **Segment 2** |
| Length of bike lane | -0.136 | -0.315 | -0.021 |
| Length of highway | -0.109 | -0.189 | -0.058 |
| Distance from CBD | 1.893 | 1.708 | 2.011 |
| Land use mix | 0.494 | 0.564 | 0.450 |
| Number of intersections | 3.061 | 3.523 | 2.767 |

***\**** *Variable definitions are presented in Table 2*

**TABLE 5 LNB Estimates for Montreal Island**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Segment Components** | | | | |
| **Variable Names\*** | **Segment 1** | | **Segment 2** | |
| **Estimate** | **t-stat** | **Estimate** | **t-stat** |
| Constant | -- | -- | 6.361 | 5.690 |
| Length of bike lane | -- | -- | 0.767 | 2.680 |
| Length of highway | -- | -- | 0.493 | 2.833 |
| Distance from CBD | -- | -- | 1.490 | 5.521 |
| Land use mix | -- | -- | -2.439 | -3.659 |
| Number of intersections | -- | -- | -2.240 | -6.300 |
| **Crash Count Component** | | | | |
| Constants | -5.693 | -9.143 | -1.546 | -1.638 |
| **Accessibility measures** |  |  |  |  |
| STM bus stops | -- | -- | 0.292 | 2.800 |
| STM bus route length | -- | -- | 0.233 | 2.680 |
| **Exposure measures** |  |  |  |  |
| Transit commuters | 0.305 | 3.459 | 1.198 | 5.093 |
| Walk commuters | 0.136 | 2.209 | 0.355 | 3.318 |
| Other mode commuters | -0.157 | -5.588 | -- | -- |
| Male bike commuters | -- | -- | 0.153 | 3.298 |
| Female bike commuters | -- | -- | 0.094 | 2.343 |
| Number of vehicles | 0.648 | 8.584 | -- | -- |
| **Sociodemographic characteristics** |  |  |  |  |
| Dependence | -0.797 | -2.916 | -- | -- |
| Non-permanent resident | -- | -- | 0.190 | 3.475 |
| African population | 0.091 | 1.992 | -- | -- |
| Asian population | -- | -- | -0.085 | -1.778 |
| European population | -0.214 | -2.983 | -- | -- |
| Population with diploma degree | -- | -- | -0.193 | -2.556 |
| **Socioeconomic characteristics** |  |  |  |  |
| Population without income | -- | -- | 0.414 | 4.032 |
| Median TAZ income (Base: Low and High income) | | | | |
| Medium | 0.509 | 4.998 | 0.301 | 2.323 |
| Median commuting time (Base: Equal and more than 11 minutes) | | | | |
| Less than 11 minutes | -- | -- | -2.802 | -2.902 |
| Commuters commuting between 7:00 am and 9:00 am | -- | -- | -1.523 | -6.979 |
| Commuters commuting after 9:00 am | -- | -- | -0.577 | -2.263 |
| **Road network characteristics** | |  |  |  |
| Number of dead-ends | -- | -- | -0.429 | -3.259 |
| Number of one-way link | -- | -- | 0.206 | 2.781 |
| **Built environment** |  |  |  |  |
| Number of bars | -- | -- | 0.452 | 3.952 |
| Lot-coverage | 1.878 | 2.709 | -- | -- |
| Presence of university building in TAZ (Base: No university buildings are present) | | | | |
| Presence of university building | 0.835 | 5.033 | -0.440 | -1.839 |
| Number of school | -0.166 | -2.785 | 0.166 | 2.616 |
| Dispersion parameter | 0.836 | | 1.645 | |

***\**** *Variable definitions are presented in Table 2*

**TABLE 6 Predictive Performance Evaluation**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Measures of fit** | **Montreal** | | **Toronto** | |
| **NB** | **LNB** | **NB** | **LNB** |
| **MPB** | 0.010 | 0.002 | 0.012 | 0.006 |
| **MAD** | 1.159 | 0.827 | 1.107 | 1.073 |
| **MSPE** | 2.486 | 2.382 | 2.859 | 2.608 |

*MPB = Mean prediction bias, MAD = Mean absolute deviation, MSPE = Mean squared prediction error, NB = Negative Binomial, LNBI = Latent Segmentation based Negative Binomial model*

**TABLE 7 Elasticity Effects for LNB model for Montreal Island**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable Names\*** | **Segment 1** | **Segment 2** | **In-Sample** |
| **STM bus stops** | 0.000 | 4.4924 | 4.492 |
| **STM bus route length** | 0.000 | 1.383 | 1.383 |
| **Transit commuters** | 10.764 | 24.055 | 34.819 |
| **Walk commuters** | 3.485 | 4.033 | 7.518 |
| **Male bike commuters** | 0.000 | 0.922 | 0.922 |
| **Female bike commuters** | 0.000 | 0.364 | 0.364 |
| **Number of vehicles** | 28.625 | 0.000 | 28.625 |
| **Non-permanent resident** | 0.000 | 1.870 | 1.870 |
| **African population** | 2.700 | 0.000 | 2.700 |
| **Population without income** | 0.000 | 4.669 | 4.669 |
| **Medium Median TAZ income** | 25.820 | 6.652 | 32.472 |
| **Number of one-way link** | 0.000 | 1.379 | 1.379 |
| **Number of bars** | 0.000 | 0.325 | 0.325 |
| **Lot-coverage** | 1.823 | 0.000 | 1.823 |
| **Presence of university** | 62.091 | -8.011 | 54.081 |
| **Number of schools** | -7.902 | 3.928 | -3.974 |

***\**** *Variable definitions are presented in Table 2*

**APPENDIX A Segment Characteristics and Mean Values of Segmentation Variables for LNB model for the City of Toronto**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Components** | | **Segments** | |  |
| **Segment 1** | **Segment 2** |  |
| Sample shares | | 0.127 | 0.229 | 0.644 |
| Observed mean of crash events | | 1.629 | | |
| Expected mean of segment level crash events | | 0.363 | 2.500 | 1.231 |
| **Mean Values of Segmentation Component Variables in Each Segment** | | | |  |
| **Variable** | **Overall Sample** | **Segment 1** | **Segment 2** | **Segment 3** |
| Ln(total number of intersection in TAZ) | 3.430 | 2.585 | 3.896 | 3.431 |
| Ln(total area of TAZ in Hectare) | 4.401 | 4.905 | 4.106 | 4.406 |
| Ln(total length of local road kilometer in TAZ) | 1.364 | 1.313 | 1.257 | 1.412 |
| Land use mix | 0.448 | 0.261 | 0.558 | 0.446 |
| Ln(Total length of bike route kilometer in TAZ) | 2.131 | 0.990 | 2.127 | 2.357 |
| Ln(Number of households in TAZ per hectare) | 2.442 | 0.431 | 2.674 | 2.755 |

**APPENDIX B LNB Estimates for the City of Toronto**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Segment Components** | | | | | | |
| **Variable** | **Segment 1** | | **Segment 2** | | **Segment 3** | |
| **Estimate** | **t-stat** | **Estimate** | **t-stat** | **Estimate** | **t-stat** |
| Constant | -- | -- | 3.745 | 0.888 | 5.678 | 1.240 |
| Ln(total number of intersection in TAZ) | -- | -- | 9.907 | 5.095 | 2.783 | 2.447 |
| Ln(total area of TAZ in Hectare) | -- | -- | -7.417 | -4.252 | -4.025 | -2.730 |
| Ln(total length of local road kilometer in TAZ) | -- | -- | -3.013 | -4.119 | -2.391 | -3.497 |
| Land use mix | -- | -- | 4.669 | 3.425 | -- | -- |
| Ln(Total length of bike route kilometer in TAZ) | -- | -- | -- | -- | 2.905 | 2.141 |
| Ln(Number of households in TAZ per hectare) | -- | -- | -- | -- | 2.110 | 4.496 |
| **Crash Count Component** | | | | | | |
| Constants | -3.700 | -7.190 | 0.033 | 0.098 | -3.450 | -11.321 |
| **Accessibility measures** |  |  |  |  |  |  |
| Ln(Total number of bus stops in TAZ) | 1.141 | 7.851 | 0.249 | 4.574 | 0.468 | 10.455 |
| Total number of metro stops in TAZ | -- | -- | 0.148 | 3.383 | -- | -- |
| **Exposure measures** |  |  |  |  |  |  |
| Ln(Total number of private cars in TAZ) | -1.709 | -4.720 | -1.116 | -13.573 | -1.310 | -18.156 |
| **Sociodemographic characteristics** |  |  |  |  |  |  |
| Ln(Total population less than 25 year old in TAZ) | -- | -- | -0.700 | -9.089 | -- | -- |
| **Socioeconomic characteristics** |  |  |  |  |  |  |
| Ln(Total number of employed person in TAZ) | 1.429 | 4.028 | 2.247 | 21.011 | 1.481 | 18.874 |
| Ln(Average TAZ income) | -- | -- | -0.369 | -9.705 | -- | -- |
| **Built environment** |  |  |  |  |  |  |
| Ln(Distance of TAZ from CBD in kilometer) | -- | -- | -0.170 | -2.966 | -- | -- |
| Dispersion parameter | 0.354 | 1.643 | 0.066 | 3.082 | 0.187 | 4.986 |

1. TAZ is a traffic related geographic entity delineated by state and/or local transportation officials. The geographic unit is used for recording traffic-related data (for instance journey-to-work, place of work records) and are most widely used for conventional transportation planning. The size of a zone varies and the layout of the zonal system is usually defined based on the similarities in socioeconomic and/or land use conditions. TAZ is the most frequently used as spatial unit for macroscopic crash risk analysis (Abdel-Aty et al., 2013), and hence TAZ is considered as the aggregate unit of analysis in the current study context. [↑](#footnote-ref-1)
2. The framework proposed in our analysis has been employed for discrete outcome - ordered or unordered - analysis earlier (see Bhat, 1997; Eluru et al., 2012, Yasmin et al., 2014). [↑](#footnote-ref-2)
3. The explanatory variables are considered by reviewing previous macro-level bicycle crash count studies. A list of the variables identified from those studies are as follows: total population, number of households, length of different (major, arterial, local) roadways, average vehicle age, population of different age groups, vehicle miles travelled, number of commuters by different modes, number of population from different races, households without vehicles, truck percentages, number of signals, number of lanes, urban /rural zone indicator, employed/unemployed residents, median households income, number of households with retired persons, park density, bus route density, sidewalk density, sidewalk coverage, number of recent immigrants, hospital beds and female lone parent families. [↑](#footnote-ref-3)
4. Exposure measures, sociodemographic characteristics, socioeconomic characteristics data are mainly compiled by TRAM from 2011 National Household Survey of Canada. Accessibility measures and land use including road network data are generally compiled from 2010 Société de transport de Montréal (STM) for Montreal, 2010 Toronto Transit Commission (TTC) for Toronto, and 2011 general land use file for both cities available from TRAM. [↑](#footnote-ref-4)
5. Land use mix is defined as: , where k is the category of land-use, is the proportion of the developed land area devoted to a specific land-use *k*, *N* is the number of land-use categories in a TAZ. In our study, five land use types were considered including residential, commercial, industrial, government and institutional, and park facilities. Institutional land use refers to land uses that cater to community’s social and educational needs (schools, town hall, police station) while park facilities refer to land used for recreational or entertainment purposes. The value of this index ranges from zero to one - zero (no mix) corresponds to a homogenous area characterized by single land use type and one to a perfectly heterogeneous mix). [↑](#footnote-ref-5)
6. The results for Toronto are presented in the Appendix (Table A and B). Specifically, Appendix A has details of the segment membership for the Toronto region while Appendix B provides model estimation results for the Toronto region. [↑](#footnote-ref-6)