European Transport \ Trasporti Europei (2017) Issue 66, Paper n° 8, ISSN 1825-3997

Count and Ordinal Data Models of Pedestrian Trip Frequency

Vaishali Patankar[1], K.V.Krishna Rao[2*]

[1]*Ph.D. Student, Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai, India*
[2]*Professor, Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai, India*

Abstract

The purpose of pedestrian trip frequency prediction is to reasonably forecast the number of trips to and from each traffic analysis zone in future, and to provide the correlates of walking behavior. This paper evaluates the performance of count and ordinal data models of pedestrian trip frequency prediction by employing a comparative empirical analysis. The study applies to examine the weekday home based work and non-work walk trips using the activity travel diary data collected in Mumbai Metropolitan Region, India. The results are informative in determining whether both model structures produce similar results and in exploring whether reliance on techniques in past pedestrian trip generation models has produced biased significance levels and parameter estimates, and to determine increased predictive accuracy for robust estimates. Predictive analytics indicate that the ordered logit models have better ability to replicate the observed pedestrian trip generation patterns than negative binomial regression models.

Keywords:walking; count data; negative binomial; ordered logit; pedestrian..

1. Introduction

The purpose of pedestrian trip frequency prediction is to reasonably forecast the number of pedestrian trips to and from each traffic analysis zone in future for the pedestrian-scale trip generation model. The pedestrian trip frequency models provides a behavioral multivariate analysis framework capturing the trade-offs among factors influencing individual walking behavior for understanding the correlates of walking behavior to control, manage and shape individual travelerbehavior and aggregate pedestrian travel demand. Indeed, encouraging walking for commuting (work) and non-commuting (non-work) purposes is part of a broader policy plan to alleviate traffic congestion and air pollution in metropolitan areas as well as to stem health problems such as obesity caused by physically inactive lifestyles and to enhance quality of life.

The literature on econometrics provides an extensive suite of model structures for modeling the discrete choice on pedestrian trip frequency generation. The state-of-the-

* Corresponding author: K. V. Krishna Rao (kvkrao@iitb.ac.in)

practice approaches utilize these model structures to model pedestrian trip generation. However, these approaches have not considered a comparative empirical analysis of count and ordinal data models to evaluate their relative performance and the predictive accuracy for unbiased robust estimates for pedestrian trip generation.

2. Previous research

Past empirical studies of pedestrian trip generation provided various approaches considering several determinants of pedestrian trip making activity to characterize walking behavior. Linear regression models were developed to relate land use data and socio-demographics to pedestrian volumes in areas of high density (Pushkarev and Jeffrey, 1971; Behnam and Patel, 1977; Matlick, 1996; Ercolano et al., 1997). However, linear regression structures does not recognize the nonnegativity and integer nature of the dependent variable – walk trips (Jang, 2005).To overcome the shortcomings of linear regression, researchers have used various count and ordinal data models to model pedestrian trip generation. Targa and Clifton (2005) provides a Poisson distribution model to explore the relationship between land use and travel behavior by testing the effects of several urban form/design characteristics and travel attitudes on the frequency of walking. The major shortcoming of the Poisson regression model is the assumed equality of the conditional mean and variance functions for the walk trips which tend to be over-dispersed (Cao et al., 2006). The Negative Binomial (NB) regression model which considers overdisperson is used to study the influence of built environment, individual attitudes and perceptions, and residential self-selection on pedestrian behavior (Handy, 1996; Shay et al., 2006; Baran et al., 2008; Cao et al., 2006). Kim and Susilo (2013) compared NB regression with poisson and logistic regression for modeling pedestrian trip generation. It was found that the NB performed well over other models. Binary logistic regression was used to understand the correlates of walking behavior, although the binary choice formulation of logistic regression loses information on the multi-level discrete choice countable data (Agrawal and Schimek, 2007; Krizek and Johnson, 2006).Handy et al. (2006) provides a quasi-longitudinal design on explaining the role of attitudes and built environment on walking. A NB model and an Ordered Probit model were used to determine the strolling frequency and the change in walking frequency respectively.

The utilization of ordinal data models is well established in the transportation planning literature for bicycle trip frequency, car ownership and trip generation models (Sener et al., 2009; Bhat and Pulugurta, 1998; Lim and Srinivasan, 2011; Handy et al., 2006). This study applies the ordinal data models to estimate pedestrian trip frequency and compares its performance with the count data models which have been widely used for pedestrian trip frequency prediction.

This paper contributes to the existing research by adopting a comparative empirical analysis of count and ordinal data models to evaluate the performance of pedestrian trip generation models. In this study, we use the standard NB model to represent the count data model, and Ordered Logit (OL) model to represent ordered choice ordinal model. This comparative study of NB and OL models is carried out with an aim to determine whether both models produce similar results and in exploring whether reliance on techniques in past pedestrian trip frequency generation models has produced biased significance levels and parameter estimates for understanding the determinants of pedestrian trip frequencies for weekday home based walking trips; and to determine increased predictive accuracy to calculate robust estimates from pedestrian trip generation models.

3. Data and sample formation

3.1 Data

The primary source of data used in this analysis is the 2010 Greater Mumbai region Activity Travel Survey. This fifteen-day survey for the Greater Mumbai region was designed and administered by Subbarao (2013), as a part of his Ph.D. research work at Indian Institute of Technology Bombay. The survey collected detailed activity and travel information for all household members from about 126 households for a fifteen-day period. The survey data is available as three main files: (1) the activity data file, (2) the person data file, and the (3) the household data file. The activity data file provides detailed information collected on activity episodes included the type of activity (based on a 9-category classification system), the location type, start and end times of activity participation, and the geographic location of activity participation. Travel episodes were characterized by the mode used, and the start and end times of travel. The person data file has information on the demographic characteristics (for example, gender, age, education level, driving license status, employment status, etc.) of the survey respondents. The household file has household-level characteristics such as location of home, tenure, vehicle ownership, etc. for the households responding to the survey (Subbarao, 2013).

3.2 Sample Formation and Analysis Context

Since the sample didn't show any variability across the weekdays from the fifteen day survey data as intuited and the analysis is focused on a weekday, the appropriate subset of data for a single weekday was extracted from the overall cleaned data from the fifteen day survey data.

The final dataset has information for 126 households. It has 347 individual household members, thus, constituting 347 individual observations. The walk travel activity file for these individuals provide information on 898 walk trip activity episodes undertaken by the household members for a representative weekday. These out-of-home walk travel activity episodes were first classified into home-based and non-home based episodes based on the location of activity participation. Each of these home-based and non-home based episodes was classified into one of work (commuting) and non-work/other (non-commuting) activities. The work walk trips also include school trips for children. The work and non-work walk trips also include access/egress trips for public transportation. Out of 898 weekday walk trips, 350 (38.98%) are Home based work (HBW), 342 (38.08%) are Home based non-work (HBO), 182 (20.27%) are Non-home based work (NHBW) and 24 (2.67%) are Non-home based non-work (NHBO) walk trips. Work walk trips constitute the major share for weekday walk trips.90% of the 347 individual household members' sample, accounting to data on 312 individuals is used for estimating the two different model components of pedestrian trip frequency (Dataset 1). The remaining 10% is hold out for validation (Dataset 2).

The following two components of pedestrian trip generationare identified for this analysis: (1) Weekday HBW pedestrian trip generation (2) Weekday HBO pedestrian trip generation.Data was aggregated appropriately to the person level to identify the work and non-work walk trip activity participation choices undertaken by the individual household member.Finally, separate data sets were created with the data appropriately structured for the estimation of Weekday HBW pedestrian trip generation and HBO pedestrian trip generation as the dependent variables. For each observation, a total of 22 possible explanatory variables were collected as described in Table 1. Table 1 gives the statistics on the dependent and the explanatory variables considered for this analysis.

The scale of distribution of the dependent variable on pedestrian trip frequency generation was collapsed into fewer categories. This was done because the share of individuals with more than four walk trips was very small and hence we assigned individuals with more than four walk trips to the four walk trips category. Thus, we specify five pedestrian trip frequency alternatives for the two MMR data sets on weekday walk trips: zero, one, two, three, and four walk trips. The recoded distribution of the dependent variable for HBW and HBO is shown in Figure 1. It seems that the distribution is not smooth, i.e. in thiscase 1 trip was less likely than 2 trips.

4. Model structures

The model structures describes the two different behavioral mechanisms for modeling the discrete variable on pedestrian trip frequency. Negative Binomial Regression Model (NBRM) represents the event based count data modelling, whereas, the Ordered Logit Model (OLM) structure considers the utility maximization principle to model the ordinal nature of pedestrian trip frequency. These two different models are described as follows:

In the following two alternate models, the dependent variable y_i is walk trip frequency and the x_i represents all the causal independent variables like individual socio-demographics, household demographics etc.

The form of NB probability distribution is given as:

$$P(y_i) = \frac{e^{-\mu_i e^{\varepsilon_i}} (\mu_i e^{\varepsilon_i})^{y_i}}{y_i!} \tag{1}$$

Where ε_i is an error term, and e^{ε_i} is gamma-distributed error term with mean 1 and variance α^2. The addition of ε_i makes the variance to be different from the mean as follows:

$$\mathrm{Var}(y_i) = E(y_i)[1 + \alpha E(y_i)] = E(y_i) + \alpha E(y_i)^2 \tag{2}$$

where α is also called the dispersion parameter. If α approaches zero, then the conditional mean and conditional variance become equal (the unobserved heterogeneity vanishes) and the NB model collapses into the Poisson model.

Table 1: Description of the dependent and explanatory variables.

Variable	Definition	Mean	Std.
Dependent variable			
Weekday HBW	Frequency of HBW walk trips for weekday	1.01	1.27
Weekday HBO	Frequency of HBO walk trips for weekday	0.93	1.32
Personal Characteristics			
Age	Age of the individual	34.02	15.76
Male	1 if male, otherwise 0	0.54	0.50
Education	Highest Grade of School Completed (=1 Graduate, =	0.46	0.50
Driving license status	1 if having driving license, otherwise 0	0.35	0.48
Occupation level	1 if full time, otherwise 0	0.70	0.46
Travel Pass	1 if having a travel pass, otherwise 0	0.24	0.43
Public mode	1 if using public mode, otherwise 0	0.55	0.50
Work place timing	1 if fixed work hours, otherwise 0	0.35	0.48
Household Characteristics			
Household Size	Number of members in the household	3.94	1.15
Residence type			
Apartment	1 if residence type is Apartment, otherwise 0	0.67	0.47
Independent house	1 if residence type is Independent house, otherwise 0	0.03	0.18
*Chawl**	1 if residence type is Chawl, otherwise 0	0.24	0.43
Slum(Informal settlements)	1 if residence type is slum, otherwise 0	0.07	0.25
Ownership type			
Owned	1 if ownership type is owned, otherwise 0	0.76	0.43
Rented	1 if ownership type is rented, otherwise 0	0.19	0.39
Government Quarter	1 if ownership type is government quarter, otherwise	0.05	0.23
Accommodation level	Number of rooms in the household	3.48	1.43
Children	1 if household has children, otherwise 0	0.29	0.45
Students	Number of students in the household	1.06	0.88
Workers	Number of workers in the household	1.60	0.78
Vehicles	Number of vehicles in the household	0.94	0.89
Car ownership	1 if household has car, otherwise 0	0.31	0.46
Two-wheeler ownership	1 if household has two-wheeler, otherwise 0	0.47	0.50
Bike ownership	1 if household has bike, otherwise 0	0.03	0.18
Household Income level	Annual income of the household	4.27	1.84
Neighborhood Characteristics			
Population Density per square km			
PD4K (Reference)	1 if Pop. Density is 4000-8999 in the region,	0.32	0.47
PD9K	1 if Pop. Density is 9000-11999 in the region,	0.05	0.21
PD12K	1 if Pop. Density is 12000-14999 in the region,	0.05	0.21
PD15K	1 if Pop. Density is 15000-19999 in the region,	0.10	0.30
PD20K	1 if Pop. Density is 20000-44999 in the region,	0.45	0.50

PD45K	1 if Pop. Density is 45000 or more in the region,	0.04	0.19
Walkscore	Walkability and Accessibility Index	70.61	12.25

* Chawl – A large building divided into many separate tenements, offering cheap basic accommodation to laborers

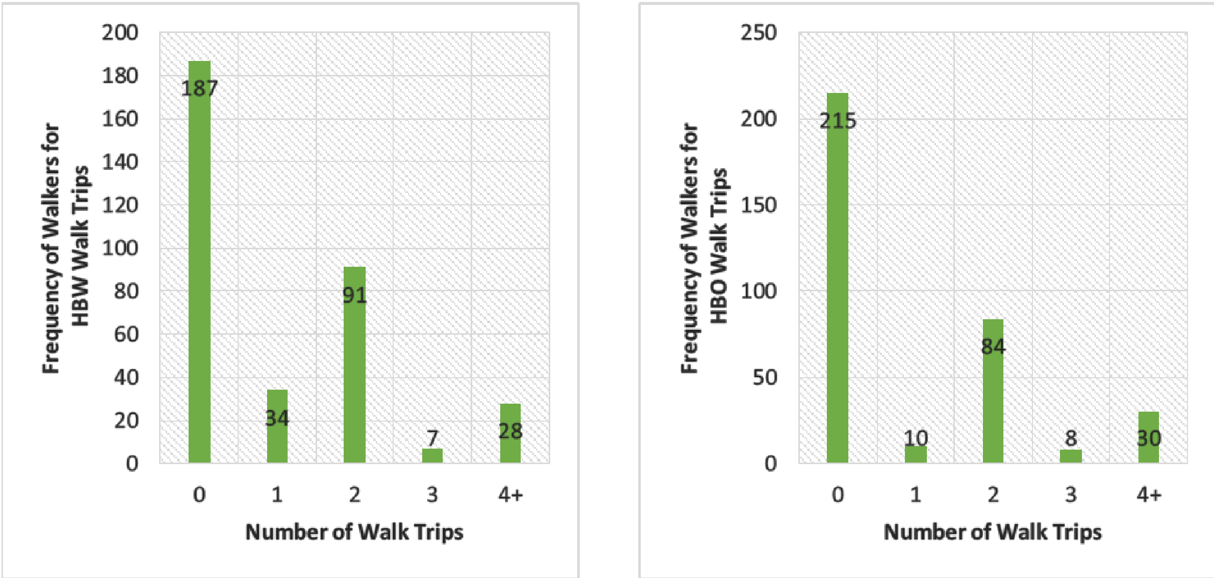


Figure 1:Weekday HBW and HBO walk trip distribution.

The formulation integrating ε_i is as follows:

$$P(y_i) = \frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{\mu_i\alpha + 1}\right)^{\alpha^{-1}} \left(\frac{\mu_i\alpha}{\mu_i\alpha + 1}\right)^{y_i} \tag{3}$$

where$\Gamma(.)$ is a gamma function.

The NB model is estimated by the standard maximum likelihood method. The corresponding likelihood function is:

$$L(\beta, \alpha) = \prod_{i=1}^{n} \frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)} \left(\frac{1}{\mu_i\alpha + 1}\right)^{\alpha^{-1}} \left(\frac{\mu_i\alpha}{\mu_i\alpha + 1}\right)^{y_i} \tag{4}$$

The likelihood function is the product of density over all n sample cases. The function is maximized to obtain coefficient estimates for β and α [see Cameron and Trivedi,1998; Lawless, 1987 for additional details].

The alternative approach used to model pedestrian trip frequency generation is the OLM. The dependent variable on walk trips is assumed to have an ordered scale. It is based on a latent regression with utility maximization framework (Greene, 2002; Bhat and Pulugurta,1998; McKelvey and Zavoina,1975; Washington et al.,2003). In this approach, the latent (unobserved) propensity of an individual to make walk trips is related to a vector of explanatory factors as follows:

$$y_i^* = \beta' x_i + \epsilon_i \tag{5}$$

Where y_i^* is an unobserved variable, β is a vector of model parameters, and ϵ_i is a random disturbance. The discrete ordered observations y_i = 0,1,2,…J are generated according to the following mechanism:

$$\begin{aligned} y_i &= 0 \text{ if } y_i^* < \mu_0, \\ y_i &= 1 \text{ if } \mu_0 < y_i^* \leq \mu_1, \\ y_i &= 2 \text{ if } \mu_1 < y_i^* \leq \mu_2, \\ y_i &= \cdots, \\ y_i &= J \text{ if } y_i^* > \mu_{J-1} \end{aligned} \tag{6}$$

Where the μ's are a set of threshold parameters that are estimated together with the vector of parameters $\beta$with the restriction that $\mu_0 < \mu_1 < \mu_2 < \cdots < \mu_{J-1}$. For identification purposes, it is also necessary to impose one additional restriction that $\mu_0 = 0$. With the above assumptions, the probabilitities observing particular values of y_i are given by:

$$\begin{gathered} P[y_i = J] = P_{iJ} = \Lambda(\mu_J - \beta' x_i) - \Lambda(\mu_{J-1} - \beta' x_i), \\ \text{where } \Lambda(z) = \exp[-\exp(-z)] \end{gathered} \tag{7}$$

$\Lambda(.)$ represents the standard cumulative gumbel distribution[for details see Greene, 2010].

5. Estimation results

This section of the paper presents the models estimated for work and non-work walk trips from the 2010 MMR activity travel diary data. All estimations were performed by using the maximum-likelihood methods in LIMDEP version 9 (Greene, 2007). Several different variable specifications and functional forms of variables were explored through exploratory analysis. Cross-Tabulation, Factor Analysis and Pearson's Correlation Coefficient matrix are examined to see the correlation effects among variables.The final variable specification for each model was obtained on the basis of a systematic process of eliminating variables found to be statistically insignificant, parsimony in representing variable effects, as well as intuitive considerations and results from earlier studies.All variables of interest were retained in the prediction of the final model irrespective of their less statistical significance to understand the directionality of their impact as policy variables. Because the intent of this analysis is to compare the relative performance of alternate structures, all the variables which are relevant to NBRM arespecified in the OLM and vice-versa. In all models, a positive coefficient indicates that the corresponding factor is associated with a greater number of trips. Tables 2 and 3 present results of estimation for each of the trip purposes (HBW and HBO respectively) for weekday.

Table 2:Model Estimation for Weekday HBW Walk Trips.

Variable	NBRM		OLM	
	Coeff	t-Stat	Coeff	t-Stat
Constant	0.74747	1.48	1.37694	1.35
Age	-.03659***	-6.41	-.06283***	-6.28
Male	0.18488	1.50	0.3937	1.62
Education	0.0796	0.43	0.34836	1.05
Public mode	.28236**	2.06	.66881**	2.49
Work Place Timings	.47868***	2.73	.72716**	2.46
Chawl	.34023**	2.03	.81761**	2.45
Rented	-.47654**	-2.48	-.86997**	-2.39
Government Quarter	-0.04529	-0.18	-0.28148	-0.50
Students	.15589**	2.05	.30174**	2.05
Vehicles	-.39574***	-3.88	-.79039***	-3.99
HH Income level	-0.02236	-0.47	-0.01221	-0.13
PD9K	-0.53044	-1.36	-0.95187	-1.30
PD12K	0.46767	1.45	0.8015	1.19
PD15K	0.00301	0.01	-0.1983	-0.44
PD20K	0.04816	0.24	-0.00707	-0.02
PD45K	-0.20237	-0.57	-0.28467	-0.35
Walk Score	0.00023	0.04	-0.0005	-0.04
LL(β)	-374.768		-310.999	
LL(c)	-374.798		-371.964	
ρ2	0.0001		0.164	
Alpha (α)	0.02584	0.24		
AIC	785.6		664.0	

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

The two alternate methods produce almost identical results in terms of the relative importance and statistical significance of the independent variables for both trip purposes. However, for work trips, it is observed that the directionality of built environment variables is different from the two models although at very low significance level. The OL model is consistent with the theory of random utility maximization which accounts travel as a derived demand (Boarnet and Crane, 2001).

Therefore, for the purpose of hypothesis testing, it is more appropriate to consider the parameter estimate and statistical significance of OL models.

For the purpose of prediction, model selection test is carried out to see the increase in predictive accuracy amongst the two model structures which can be utilized to determine the robust estimates for pedestrian trip generation models.

Table 3:Model Estimation for Weekday HBO Walk Trips.

Variable	NBRM		OLM	
	Coeff	t-Stat	Coeff	t-Stat
Constant	-0.25236	-0.26	-0.71689	-0.65
Age	.02569***	3.47	.04032***	4.91
Male	-0.20164	-1.00	-0.19481	-0.78
Education	-0.29327	-1.05	-0.39813	-1.22
Public mode	-.42639**	-2.14	-.64113**	-2.44
Work Place Timings	-.59761**	-2.41	-.97295***	-3.18
Chawl	0.31618	1.15	.57074*	1.67
Rented	0.23326	0.81	0.30688	0.86
Government Quarter	0.61013	1.36	0.89665	1.53
Students	-0.03607	-0.31	-0.08855	-0.56
Vehicles	-0.09721	-0.71	-0.13408	-0.71
HH Income level	-0.06406	-0.89	-0.07368	-0.77
PD9K	-0.41278	-0.65	-0.60129	-0.79
PD12K	-0.46295	-0.78	-0.6003	-0.70
PD15K	-0.12018	-0.31	-0.04598	-0.10
PD20K	0.02556	0.08	0.0603	0.15
PD45K	-0.6188	-0.96	-0.99638	-1.17
Walk Score	0.00101	0.08	0.0016	0.12
LL(β)	-384.044		-289.208	
LL(c)	-401.386		-324.598	
ρ2	0.043		0.109	
Alpha (α)	.89249***	3.03		
AIC	838.8		620.4	

Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

6. Modelselection

6.1 Likelihood Ratio Index

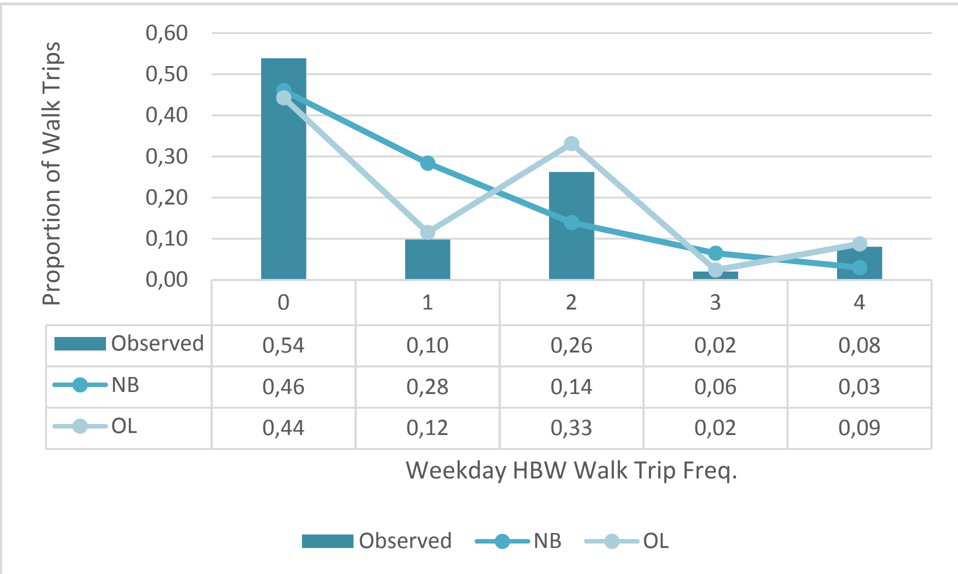
The likelihood ratio index ρ^2 (similar to R^2 in linear regression analysis) is computed as a measure of overall statistical fit. A higher ρ^2 value indicates a better model (Ben-Akiva and Lerman, 1985). The two model structures are compared by using the likelihood ratio index values. For both trip purposes, it is followed that the OL models perform the best compared to the NB models.

6.2 Akaike Information Criterion

Akaike information criterion (AIC), is also used as agoodness of fit measure. The model with the smallest AIC is considered as the best fit (Akaike, 1998; Flury, 1988). The information on AIC indicates the superiority of OL models over NB.

6.3 Evaluation Criteria for Data Fit

The correspondence between conditional means and actual data is used as a measure of fit to inform how well the model predicts the data.The calibrated models for the two different model components of pedestrian trip frequency are applied to Dataset 1 of the respective model components and the conditional means of the alternatives is calculated. The observed vs. calculated proportion for HBW and HBO pedestrian trip frequency models for weekday is presented in Figure 2 and 3 respectively.



	0	1	2	3	4
Observed	0,54	0,10	0,26	0,02	0,08
NB	0,46	0,28	0,14	0,06	0,03
OL	0,44	0,12	0,33	0,02	0,09

Figure 2: Observed and calculated HBW pedestrian trip frequency distribution.

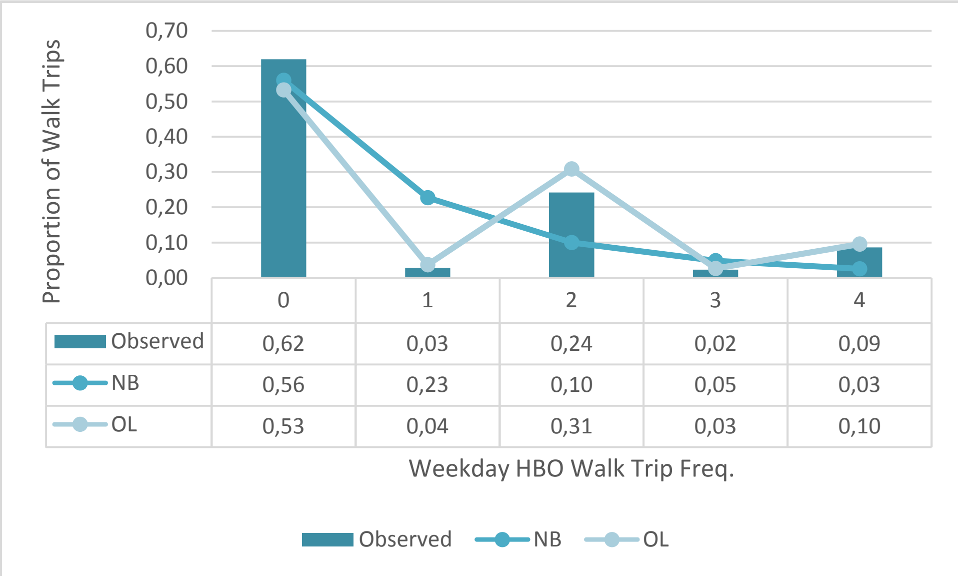


Figure 3: Observed and calculated HBO pedestrian trip frequency distribution.

From the above figures, it is observed that the OL models fit the observed trip generation patterns rather accurately for both HBW and HBO trip purposes

6.4 Vuong Test

Tables 2 and 3 shows the results of estimation. Based on estimation, the Vuong test can be used as a model selection test to find the better of two rival nonnested models. That is, the model reflecting the data best can be chosen through the Vuong test. If $f_1(y_i|x_i)$ *and* $f_2(y_i|x_i)$are predicted probability distributions for each observed trip $y_i$of Models 1 and 2 respectively, then the Vuong statisticcan be calculated as follows(Vuong, 1989):

$$V = \frac{\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} m_i\right]}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}}, \text{Where } m_i = ln\left(\frac{f_1(y_i|x_i)}{f_2(y_i|x_i)}\right)$$

In above equation, V, $\bar{m}$, and n are the Vuong Statistic, mean, and number of observation, respectively.This is the standard statistic for testing the hypothesis that E $[m_i]$ equals zero. Vuong shows that V has a limiting standard normal distribution. The statistic is bidirectional. If $|V|$ is less than 1.96 at 95% confidence level, then the test does not favor one model or the other. Otherwise, large values favor Model 1 whereas small (negative) values favor Model 2. Carrying out the test requires estimation of both models and computation of both sets of predicted probabilities (Greene, 2002).

Table 4 presents the results of model selection tests.The results clearly indicate the predictive superiority of the OL models over the other method (see the large positive values of the Vuong statistics compared to NB).

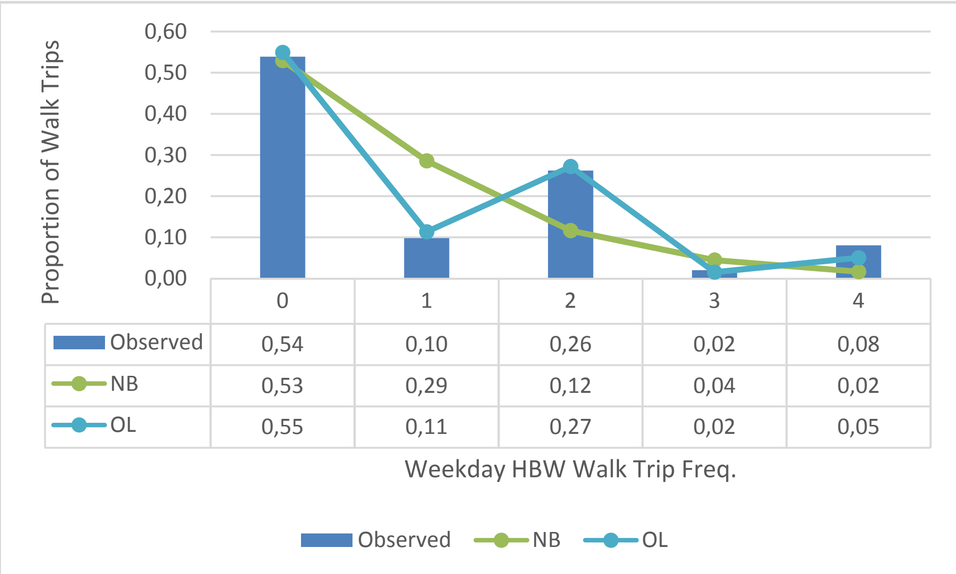
Table 4:Vuong Test Results.

Trip Purpose	Rival Models	Vuong Statistic	Selection
Weekday HBW	OL versus NB	6.67	OL
Weekday HBO	OL versus NB	8.75	OL

6.5 Model Forecasting Performance

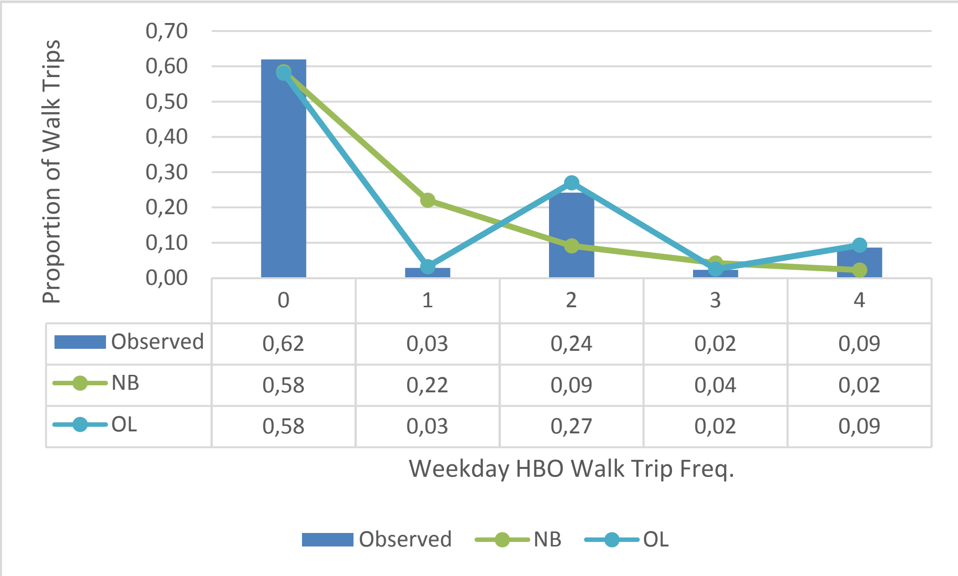
The calibrated models estimated using Dataset 1 for the two different model components of pedestrian trip frequency are applied to Dataset 2 of the respective model components and the predicted share of the alternatives is determined. The observed vs. predicted proportion for HBW and HBO pedestrian trip frequency models for weekday is presented in Figure 4 and 5 respectively.

From Figure 4 and 5, it is observed that the predictions from the OL models replicate the observed trip generation patterns rather accurately for all trip purposes.



	0	1	2	3	4
Observed	0,54	0,10	0,26	0,02	0,08
NB	0,53	0,29	0,12	0,04	0,02
OL	0,55	0,11	0,27	0,02	0,05

Figure 4: Observed and predicted HBW pedestrian trip frequency distribution.



	0	1	2	3	4
Observed	0,62	0,03	0,24	0,02	0,09
NB	0,58	0,22	0,09	0,04	0,02
OL	0,58	0,03	0,27	0,02	0,09

Figure 5: Observed and predicted HBO pedestrian trip frequency distribution.

7. Optimal model

Based on the results of above likelihood ratio index test, Akaike information criterion, evaluation criteria for data fit, Vuong test and model forecasting performance test; OLM is selected as the optimal model amongst the alternate model structures for both trip purposes.

7.1 Analysis of Optimal model

The identification of slope coefficients, that is, a change for any independent variable should have the same effect on the odds of answering each category of the dependent variable, is an important assumption of the OL models. It is called as the parallel regression assumption, also known as the proportional odds assumption (Long, 1996). A Wald test proposed by Brant in 1990 allows both an overall test that all vector of coefficients are same in every equation and tests of the equality of individual coefficients by comparing slope coefficients of the J - 1 binary choice models implied by the ordered regression model. If the probability of the test (p-value) is less than a critical value (usually .05), the hypothesis that slope coefficients remain identical across the response values is rejected; in other words, there is strong evidence that the slope coefficients violate the assumption (Brant, 1990).

We conducted the Brant test to check whether the two ordered logit models violated the parallel regression assumption. The results of Brant tests are given in Table 5. From the table, there is strong evidence for rejecting the parallel regression assumption for the independent variables which has p-values < .05. For HBW model, these variables are Age, Education, Public Mode, Work Place Timings, Rented, Students, Vehicles, HH Income Level, PD9K, PD12K, PD15K, PD20K, and PD45K. Other variables, including Male, Chawl, Government Quarter, and Walk Score do not violate the assumption and should remain the same across pedestrian trip frequency levels. Thus, for HBW model, the parallel regression assumption was fulfilled for 4 of all 17 variables. For HBO model, the variables which has p-values < .05, are students, PD9K, PD12K. Thus, these independent variables violates the parallel regression assumption and should differ across pedestrian trip levels. Other variables, including Age, Male, Education, Public Mode, Work Place Timings, Chawl, Rented, Government Quarter, Vehicles, HH Income Level, PD15K, PD20K, PD45K, and Walk Score do not violate the assumption and should remain the same across pedestrian trip frequency levels. Thus, for HBO model, the parallel regression assumption was fulfilled for 14 of all 17 variables.

It is common in OL models for one or more coefficients to differ over the outcome levels and, thus, OL models is restricted when this parallel regression assumption is violated (Long, 1996; Williams, 2006;Peterson and Harrell, 1990). To relax the restriction, the Partial Proportional Odds Model (PPOM) was proposed by Peterson and Harrell (1990). Also, Long & Freese (2001) introduced a Generalized Ordered Logit Regression (GOLR) model that extends the regular OLRegression model. The suitability of PPOM and GOLR models for modeling the frequency of pedestrian trips especially to work would be an appropriate topic for future research.

8. Marginal effects

Table 6and 7show the Marginal Effects for Weekday HBW and HBO pedestrian trip frequency models respectively. The sign of the coefficients are as expected. The marginal effect for alternative 0 is interpreted as follows: An increase in age characteristic for the 0 alternative by 1 unit will increase the choice probability for the 0 alternative by .01552, ceteris paribus. This is in the direction expected. The remaining marginal effects represent the changes in the probabilities for competing alternatives. A 1-unit increase in the age characteristic will decrease the probability of selecting the alternative 1 by .00152, ceteris paribus. Similarly, a 1-unit change in the age characteristic will, according to the estimated model, decrease the probabilities by 0.00152, 0.01024, 0.00088, and 0.00258 for the 1, 2, 3, and 4 alternatives respectively, ceteris paribus. As the choice probabilities must sum to one, the marginal effects which represent the change in the choice probabilities are mathematically constrained to sum to zero, thus representing a net zero change over all alternatives. The signs of the coefficients are consistent across 1 to 4 levels but reversed compared to 0 pedestrian trip frequency. This is because 0 is effectively the reverse of the other pedestrian trip frequencies and this model assigns ratings to pedestrian trip levels. Similarly, the marginal effects for other variables of interest can be interpreted.

Table 5:Brant Test for Parameter Homogeneity.

Variable	DF	HBW		HBO	
		Chi-Square	p-Value	Chi-Square	p-Value
All	51	7134.41	0	602.35	0
Age	3	89.3800	0.0000	0.8900	0.8289

Male	3	1.7800	0.6191	3.1200	0.3741
Education	3	9.2800	0.0258	1.5800	0.6640
Public mode	3	31.7200	0.0000	5.5100	0.1382
Work Place Timings	3	12.4300	0.0061	0.4600	0.9268
Chawl	3	1.5000	0.6819	5.2700	0.1529
Rented	3	1399.6100	0.0000	1.8400	0.6056
Government Quarter	3	5.4400	0.1426	2.1400	0.5442
Students	3	68.7200	0.0000	9.8800	0.0196
Vehicles	3	13.9400	0.0030	1.7900	0.6162
HH Income level	3	43.6300	0.0000	1.8400	0.6069
PD9K	3	594.6000	0.0000	282.7400	0.0000
PD12K	3	749.7800	0.0000	132.6200	0.0000
PD15K	3	1303.2100	0.0000	3.5400	0.3155
PD20K	3	2114.1500	0.0000	0.4900	0.9210
PD45K	3	379.3500	0.0000	1.2700	0.7352
Walk Score	3	2.2600	0.5199	1.4100	0.7038

Table 6: Marginal Effects for weekday HBW pedestrian trip frequency.

Variables	0	1	2	3	4
Age	.01522***	-.00152***	-.01024***	-0.00088	-.00258***
Male	-0.09064	0.00926	0.0609	0.00523	0.01525
Education	-0.07447	0.00722	0.04998	0.00438	0.01289
Public mode	-.14969**	.01592*	.10045**	0.00852	.02481*
Work Place Timings	-.17396**	.01366**	.11534**	0.01095	.03400*
Chawl	-.17413**	.01152**	.11468**	0.01144	0.0365
Rented	.20759***	-.03118*	-.13953***	-0.01001	-.02687***
Government Quarter	0.05701	-0.00678	-0.03848	-0.00308	-0.00866
Students	-.07436**	.00743*	.05000**	0.00432	.01261*
Vehicles	.17598***	-.01758***	-.11835***	-0.01022	-.02984**
HH Income level	0.21473	-0.03735	-0.14391	-0.0094	-.02407**
PD9K	-0.18346	0.00599	0.11768	0.01345	0.04633
PD12K	0.03585	-0.00397	-0.02417	-0.002	-0.00571
PD15K	-0.01367	0.00136	0.00919	0.00079	0.00232
PD20K	0.0351	-0.00393	-0.02367	-0.00195	-0.00555
PD45K	0.0005	-.49786D-04	-0.00034	-.28940D-04	-.84501D-04
Walk Score	0.00673	-0.00067	-0.00453	-0.00039	-0.00114

Note: ***, **, * ==> Significance at 1%, 5%, 10% level. The marginal effects of less statistically significant variables of interest are retained to know the directionality of their impact on pedestrian trip frequency.

9. Conclusions

This study seeks to contribute to the area of pedestrian travel behavior modeling by providing a comparative framework to analyze the two different model structures.It adds substantive empirical evidence on the value of using ordered logit models for pedestrian trip frequency modeling when the distribution is not smooth especially in small samples. Ordered Logit models are consistent with the utility maximization theory which considers travel as a derived demand.Therefore, to resolve the ambiguity raised due to difference in parameter estimates, it is more appropriate to use OL models for hypothesis testing. Predictive analytics indicate that the ordered logit models have better ability to replicate the observed trip generation patterns than negative binomial regression models for both trip purposes. It therefore appears appropriate to employ ordered logit models amongst the two statistical approaches for pedestrian trip generation forecasting.

Empirical results provide useful insight into effect of socio-demographic characteristics and built environment attributes to pedestrian walk trip making activity behavior. In this regard, this paper contributed to the existing research by adopting a multilevel analysis structure to examine a comprehensive set of sociodemographic characteristics and built environment attributes associated with pedestrian walk trip frequency. The empirical analysis was based on a sample of individuals who reside in the Mumbai Metropolitan Region.

Table 7: Marginal Effects for weekday HBO pedestrian trip frequency.

Variables	0	1	2	3	4
Age	-.00890***	.00031***	.00537***	0.00069	.00253***
Male	0.05593	-0.00194	-0.03359	-0.00434	-0.01607
Education	0.09633	-0.00349	-0.05832	-0.00739	-0.02712
Public mode	.14865**	-.00486**	-.08786**	-0.01165	-.04428**
Work Place Timings	.20138***	-.00838**	-.12484***	-0.01492	-.05325***
Chawl	-0.11603	.00331*	0.06682	0.00933	0.03658
Rented	-0.07205	0.00213	0.04189	0.00576	0.02228
Government Quarter	-0.21218	0.0028	.10680**	0.0184	0.08417
Students	0.01605	-0.00057	-0.00968	-0.00124	-0.00456
Vehicles	0.02837	-0.001	-0.01711	-0.00219	-0.00807
HH Income level	0.12912	-0.00644	-0.08369	-0.00905	-0.02995
PD9K	0.11266	-0.00539	-0.07241	-0.00801	-0.02686
PD12K	0.02413	-0.00091	-0.01475	-0.00184	-0.00664
PD15K	-0.01557	0.00055	0.00939	0.0012	0.00443
PD20K	0.1796	-0.00999	-0.11889	-0.01209	-0.03863
PD45K	-0.00036	.12624D-04	0.00022	.27662D-04	0.0001
Walk Score	0.01953	-0.00069	-0.01178	-0.00151	-0.00555

Note: ***, **, * ==> Significance at 1%, 5%, 10% level. The marginal effects of less statistically significant variables of interest are retained to know the directionality of their impact on pedestrian trip frequency.

The study elicited information from a sample of current walkers and non-walkers from the overall population of interest. Specifically, we focus on modeling weekday HBW and HBO walk trip participation choices of individuals.There are several important findings from the empirical analysis. The two alternate methods produce almost identical results in terms of the relative importance and statistical significance of the independent variables for both trip purposes. However, for work trips, it is observed that the directionality of built environment variables is different from the two models although at very low significance level. The use of OL is justified to resolve such ambiguity and to provide unbiased estimates.

In summary, theordered logit models can be applied to estimate the walk trip frequency models for work and non-work which follows an ordinal, discrete and non-negative integer as a dependent variable and to examine the impact of socio-demographics and environmental factors in shaping the daily walk trip activity participation choices of individuals.

References

Agrawal, A.W., Schimek,P. (2007) “Extent and Correlates of Walking in the USA”,*Transportation Research D* 12(8), pp. 548-563.

Akaike, H. (1998) “Information theory and an extension of the maximum likelihood principle”, In S. Kotz, N.L.Johnson (eds) *Breakthroughs in Statistics*, Springer, New York.

Baran, P.K., Rodríguez,D.A., Khattak,A.J.(2008) “Space Syntax and Walking in a New Urbanist and Suburban Neighbourhoods”,*Journal of Urban Design* 13(1), pp. 5-28.

Behnam, J., Patel,B.G.(1977) “A Method for Estimating Pedestrian Volume in a Central Business District”,*Transportation Research Record: Journal of the Transportation Research Board,Transportation Research Board of the National Academies, Washington, D.C.* 629, pp. 22-26.

Ben-Akiva, M., Lerman,S.R. (1985)*Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, Mass.

Bhat, C.R., Pulugurta,V. (1998) “A Comparison of Two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions”,*Transportation Research B* 32(1), pp. 61-75.

Boarnet, M., Crane, R. (2001) “The influence of land use on travel behavior: specification and estimation strategies”,*Transportation Research Part A: Policy and Practice* 35, pp. 823-845

Brant, R. (1990) “Assessing Proportionality in the Proportional Odds Model for Ordered Logistic Regression” *Biometrics*46(4), pp. 1171-1178.

Cameron, A.C., Trivedi,P.K.(1998)*Regression Analysis of Count Data*, Cambridge University Press, Cambridge, United Kingdom.

Cao, X.,Handy,S.L, Mokhtarian,P.L. (2006) "The Influences of the Built Environment and Residential Self-Selection on Pedestrian Behavior: Evidence from Austin, TX",*Transportation* 33(1), pp.1-20.

Ercolano, J.M., Olson,J.S., Spring,D.M.(1997) "Sketch-Plan Method for Estimating Pedestrian Traffic for Central Business Districts and Suburban Growth Corridors",*Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington D.C.* 1578, pp. 38-47.

Flury, B. (1988)*Common principal components and related multivariate methods,* John Wiley and Sons, New York.

Greene, W.H. (2002)*Econometric Analysis*, Prentice-Hall Inc., Upper Saddle River, NJ.

Greene, W.H.(2007)*Limdep Version 9.0: Econometric Modeling Guide*, Econometric Software Inc., Plainview, NY.

Greene, W.H., Hensher,D.A. (2010)*Modeling Ordered Choices: A Primer,*Cambridge University Press, New York.

Handy, S. (1996) "Urban Form and Pedestrian Choices: Study of Austin Neighborhoods",*Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington D.C.* 1552, pp. 135-144.

Handy, S., Cao,X., Mokhtarian,P. L. (2006) "Self-selection in the Relationship between the Built Environment and Walking: Evidence from Northern California",*Journal of the American Planning Association* 72(1), pp. 55-74.

Hensher, D.A., Rose,J.M., Greene,W.H.(2005)*Applied Choice Analysis*, Cambridge University Press, New York.

Jang, T.Y. (2005) "Count Data Models for Trip Generation" *ASCE: Journal of Transportation Engineering* 131(6), pp. 444–450.

Kim, N.S., Susilo,Y.O. (2013) "Comparison of Pedestrian Trip Generation Models" *Journal of Advanced Transportation* 47(4):399-412.

Krizek, K.J., Johnson,P.J. (2006) "Proximity to Trails and Retail: Effects on Urban Cycling and Walking" *Journal of the American Planning Association* 71(1), pp. 33-42.

Lambert, D. (1992) "Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing" *Technometrics* 34(1), pp. 1-14.

Lawless, J.F. (1987) "Negative Binomial and Mixed Poisson Regression" *The Canadian Journal of Statistics* 15(3), pp. 209-225.

Lim, K., Srinivasan,S. (2011) "Comparative Analysis of Alternate Econometric Structures for Trip Generation Models"*Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington D.C.*2254, pp. 68-78.

Long, J. S. (1996)*Regression Models for Categorical and Limited Dependent Variables,*Sage PublicationsInc., Thousand Oaks, California.

Long, J. S., Freese,J.(2001)*Regression models for categorical dependent variables using STATA*, STATA Press,College Station, TX.

Matlick, J.M. (1996) "Forecasting Pedestrian Use and Flows: If We Build it, Will They Come?" *Pro Bike/Pro Walk '96 : Forecasting the Future* 69, pp. 315-319.

McKelvey, R.D., Zavoina,W.(1975) "A Statistical Model for the Analysis of Ordinal Level Dependent Variables" *Journal of Mathematical Sociology* 4(1), pp. 103-120.

Peterson, B., Harrell,F. E. (1990) "Partial Proportional Odds Models forOrdinal Response Variables"*Applied Statistics* 39(2), pp. 205–217.

Pushkarev, B., Jeffrey,M.Z. (1971) "Pedestrian Travel Demand" *Highway Research Record, Highway Research Board, National Research Council, Washington D.C.*HS-011 999(355), pp. 37-53.

Sener I.N., Eluru N., Bhat C.R. (2009) "Who are bicyclists? Why and how much are they bicycling?",*Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington D.C.*2134, pp. 63-72

Shay, E., Fan, Y., Rodríguez, D., Khattak,A.(2006) "Drive or Walk?: Utilitarian Trips within a Neotraditional Neighborhood",*Transportation Research Record: Journal of the Transportation Research Board, Transportation Research Board of the National Academies, Washington D.C.* 1985, pp. 154-161.

Subbarao, S.S.V. (2013)Activity Based Travel Demand Analysis for a Mega City in a Developing Country,*Ph.D. Thesis*, Civil Engineering Department, Indian Institute of Technology Bombay, Mumbai, MH

Targa, F., Clifton,K. (2005) "Built Environment and Nonmotorized Travel: Evidence from Baltimore City using the NHTS",*Journal of Transportation and Statistics*8(3), pp. 55-70.

Vuong, Q. (1989) "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses",*Econometrica* 57(2), pp. 307-333.

Washington, S.P., Karlaftis,M.G.,Mannering,F.L. (2003)*Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC, Boca Raton, FL.

Williams, R. (2006) "Generalized Ordered Logit/Partial Proportional OddsModels for Ordinal Dependent Variables",*Stata Journal* 6(1), pp. 58–82.