A Finite Mixture Modeling Approach to Examine New York City Bicycle Sharing System (CitiBike) Users' Destination Preferences

Ahmadreza Faghih-Imani
Postdoctoral Fellow
Department of Civil Engineering
University of Toronto
E-mail: a.faghihimani@utoronto.ca

Naveen Eluru*
Associate Professor
Department of Civil, Environmental and Construction Engineering
University of Central Florida
Ph: 407-823-4815; Fax: 407-823-3315
E-mail: naveen.eluru@ucf.edu

*Corresponding author

An abridged version of the paper was presented at the 2017 Annual Transportation Research Board Conference

Abstract

Given the recent growth of bicycle-sharing systems (BSS) around the world, it is of interest to BSS operators/analysts to identify contributing factors that influence individuals' decision processes in adoption and usage of bicycle-sharing systems. The current study contributes to research on BSS by examining user behavior at a trip level. Specifically, we study the decision process involved in identifying destination locations after picking up the bicycle at a BSS station. In the traditional destination/location choice approaches, the model frameworks implicitly assume that the influence of exogenous factors on the destination preferences is constant across the entire population. We propose a Finite Mixture Multinomial Logit (FMMNL) model that accommodates such heterogeneity by probabilistically assigning trips to different segments and estimate segment-specific destination choice models for each segment. Unlike the traditional destination choice based Multinomial Logit (MNL) model or Mixed Multinomial Logit (MMNL), in an FMMNL model, we can consider the effect of fixed attributes across destinations such as users' or origins' attributes in the decision process. Using data from New York City bicycle-sharing system (CitiBike) for 2014, we develop separate models for members and non-members. We validate our models using hold-out samples and compare our proposed FMMNL model results with the traditional MNL and MMNL model results. The proposed FMMNL model provides better results in terms of goodness of fit measures, explanatory power and prediction performance.

Keywords: Bicycle Sharing Systems, CitiBike New York, Finite Mixture Model, Multinomial Logit Model, Destination Choice

1. INTRODUCTION

In recent years, the adoption of the bicycling mode of transportation for commuting and leisure has experienced significant growth. Nationally, the mode share for commuting by bicycle has increased 46% since 2005 (McLeod, 2014). Coupled with the increasing adoption of bicycling, analysis of millennials' travel behavior suggests increased willingness among millennials to use shared transportation systems rather than private vehicles (Davis et al., 2012). These recent trends have encouraged urban regions to invest substantially in bicycle-sharing systems (BSS). Over 1000 cities have a BSS in operation or under construction (Meddin, and DeMaio, 2016). BSS offer a practical, physically active and sustainable mode of transportation for short to medium distance trips in dense urban regions. If shorter private vehicle trips can be substituted by BSS trips, significant benefits in terms of reducing traffic congestion, GHG emissions and air pollution will be accrued. A well designed bicycle-sharing system can also potentially provide a solution for the first/last mile problem of other public transportation systems (Jäppinen et al., 2013). Earlier research efforts indicate that BSS have assisted in normalizing the image of the bicycle as a daily mode of travel in public perception and thus have contributed to expanding the demographics of cyclists (Goodman et al., 2014). Further, the presence of BSS in a city might enhance the driver awareness towards cyclists and thus increase the safety for cyclists (Murphy and Usher, 2015).

Considering all the benefits of bicycle-sharing systems, it is not surprising that there is a rapid growth of these systems around the world. Accordingly, the number of studies on bicycle-sharing systems have increased in recent years (for a review of recent literature on BSS, please see Fishman, 2016). The relevant literature in the burgeoning area of BSS research can be classified based on the data employed in the analysis as: (1) Survey compiled data and (2) BSS operation data. The *first stream of studies* employs survey compiled data to examine BSS users travel behavior and choices. Bachand-Marleau et al. (2012) conducted a survey in Montreal, Canada, and found that convenience and having a BSS station closer to home location are the main motivators for individuals to use BSS. Fishman et al. (2015a) examined the factors influencing the user's membership of BSS in two major Australia cities (Melbourne and Brisbane) and identified riding frequency, age, proximity to docking station as significant contributing factors for membership. Buck et al. (2013) highlighted the differences between BSS short-term users and annual members and regular cyclists in Washington, DC. Several studies investigated the impact of BSS on cyclists' safety and prevalence of using helmets by BSS users (Kraemer et al., 2012; Graves et al., 2014; Fishman and Schepers, 2016). Murphy and Usher (2015) employed survey data conducted in Dublin, Ireland, and underlined the gender gap and income equity issues with regard to accessing and using BSS. Further, the impact of BSS on mode choice and modal shift to BSS was analyzed in several studies (Buck et al., 2013; Martin and Shaheen, 2014; Murphy and Usher, 2015). It is found that BSS mostly substituted trips made by public transport and by walking. However, the overall impact of BSS on increasing active transportation is found to be positive (Fishman et al., 2015b). The overall benefits of BSS in terms of car use and air pollution reduction are strongly dependent on the design of the system, the need for the bicycle redistribution in the system and the how the rebalancing operations are performed by the system operators (Fishman et al., 2014).

The *second stream of studies*, and of particular relevance to our study, develops quantitative frameworks to understand BSS employing real operation data provided by the BSS operator or downloaded through automated scripts from BSS operator websites. In this context, most of the earlier studies focused on the relationship between BSS usage and demand with bicycling infrastructure, land use and built environment, public transportation infrastructure, temporal and meteorological attributes. For example, it is found that increasing BSS infrastructure (number of stations and capacity) or bicycle routes around BSS stations would increase the usage of BSS (Faghih-Imani et al., 2014; Wang et al., 2015). Studies demonstrate that population and job density, as well as point of interests (such as retail stores, and universities), have a positive impact on the ridership of BSS (Rixey, 2013; Faghih-Imani et al., 2014; Faghih-Imani et al., 2017). A subset of studies examined the link between BSS usage and other public transportation systems (Nair et al., 2013; Faghih-Imani et al., 2014; Faghih-Imani and Eluru, 2015; González et al., 2015). The use of BSS for the daily commute to work is identified in several studies (O'Brien et al., 2014; Faghih-Imani et al., 2014). Further, it is demonstrated by several studies that BSS users prefer routes with bicycle facilities such as bicycle lanes and shorter trips with all else same (Faghih-Imani and Eluru, 2015; González et al., 2015). The weather attributes influence on BSS usage is highlighted by several earlier research efforts (Gebhart and Noland, 2014, Faghih-Imani et al., 2014). Further, a small subset of studies employed advanced models to highlight the importance of recognizing the spatial correlation and the self-selection impact of BSS infrastructure installation decision process in modeling BSS usage (Faghih-Imani and Eluru, 2016a; Faghih-Imani and Eluru, 2016b).

Within this group of studies, several studies also focus on operational issues of BSS including identifying problematic stations (stations that are full or empty) and efficiency of operator rebalancing program. For example, Fricker and Gast (2014), studied the effect of the randomness of user decisions on the number of problematic stations. Vogel and Mattfeld, (2011) and Bouveyron et al. (2015) developed cluster analysis and found different categories of stations within the bicycle-share systems. Several studies analyzed various methods for optimizing bicycle rebalancing operations and repositioning trucks' routing schemes (Vogel and Mattfeld, 2011; Nair et al. 2013; Raviv et al., 2013; Pfrommer et al., 2014; Forma et al., 2015).

2. CURRENT STUDY IN CONTEXT

The current study belongs to the second stream of research that employs operational data from BSS. The study contributes to research on BSS by examining user behavior at a trip level to provide a better understanding of the system for the BSS planners/operators. Specifically, the destination station choice of individuals after picking up the bicycle at a BSS station is studied using a random utility maximization approach. Such approach is common in location choice studies in transportation literature (Waddell et al. 2007; Chakour and Eluru 2014; Faghih-Imani and Eluru, 2015). For example, Faghih-Imani and Eluru (2015) employed a random utility based multinomial logit model to study destination choice preferences for the Chicago's Divvy BSS.

In these traditional approaches, the model frameworks implicitly assume that the influence of exogenous factors on the destination preferences is constant across the entire population. To illustrate this, consider the destination station choice behavior for two users (U1 and U2) with the same attributes except for trip start time period. U1 starts her trip in AM time period while U2 starts her trip in the PM time period. Now let us consider the influence of "job density" and "restaurant density" in the vicinity of the destination alternatives on destination station preferences. U1 beginning her trip in the AM period is more likely to be positively affected by "job density" while being minimally affected by "restaurant density". U2, on the other hand, is likely to be positively influenced by "restaurant density" and either minimally (or even negatively) affected by "job density". This is an illustration of how based on the trip start time, impact of exogenous variables can be substantially different across users. The illustration provided is a case of one variable (trip start time) moderating the influence of other variables ("job density" and "restaurant density"). The reader would recognize that it is possible that multiple variables might serve as a moderating influence on a reasonably large set of exogenous variables. If such distinct profile of exogenous variables across users is not considered, a restrictive assumption that all exogenous variables have the same effect on user destination choice process is imposed.

A commonly proposed solution to address the restrictive homogeneity assumption is the clustering of the sample population based on exogenous variables and developing cluster specific models. However, based on the number of exogenous variables of interest, the number of mutually exclusive sample clusters could increase very rapidly thus increasing the number of models to be developed (see Eluru et al., 2012 for more discussion). Further, the large number of mutually exclusive clusters might result in samples with fewer records. Another alternative approach to address homogeneity is to employ mixed versions of the traditional models that accommodate for unobserved heterogeneity across the population. These approaches are focussed on the error component of the model and usually require extensive simulation for model estimation. However, one disadvantage is that they do not capture the heterogeneity corresponding to observed variables (systematic heterogeneity) in the modeling framework. Also, the large number of alternatives in destination choice models further reduce the appeal for such approaches.

A third approach to accommodate heterogeneity is to undertake an endogenous segmentation or a finite mixture model approach. In this approach, the sample population is allocated probabilistically to different segments, and segment-specific destination choice models are estimated. The segment membership is achieved based on a multivariate set of exogenous variables within a closed form approach. Thus the finite mixture approach addresses the limitations of the two previous methods by accommodating for heterogeneity within a closed form model. The proposed approach also addresses a specific limitation of the traditional Multinomial Logit (MNL) model, exogenous segmentation or mixed models. In a destination choice based MNL model and its variants such as Mixed Multinomial Logit model (MMNL), only attributes that vary across alternatives within the choice set can be considered for influencing utility i.e. the model is limited to destination attributes. The consideration of socio-demographics and other attributes fixed across all alternatives (such as origin variables or temporal and meteorological characteristics) is possible only through their interaction with destination attributes. However, within the Finite Mixture Multinomial Logit (FMMNL) framework, we can account for such user level attributes (that are fixed across alternatives) through the segmentation component of the model. The use of FMMNL framework has increased in the transportation literature over the past few years. Studies employed FMMNL to examine travel mode choice (Kemperman and Timmerman, 2009; Vij et al. 2013; Ma et al. 2015), vehicle ownership (Anowar et al. 2014), residential location (Walker and Li, 2007; Ettema, 2010), activity participation (Sobhani et al. 2013), and driver injury severity (Eluru et al. 2012; Yasmin et al; 2014).

The current study extends Faghih-Imani and Eluru's (2015) work by developing an FMMNL model on data from the New York City bicycle-sharing system (CitiBike) for 2014. We develop two separate models for members and non-members. In addition, we consider a pooled dataset of members and non-members and compare the pooled model estimation with separate models. Further, we validate our estimated models with hold-out samples and compare the prediction performance of the proposed FMMNL models with the traditional MNL and MMNL models. The broad set of exogenous variables considered in the FMMNL model include temporal and weather characteristics (such as time period of the day or temperature), trip attributes (such as trip distance), users' attributes (such as age and gender), and origin and destination characteristics (such as BSS capacity, and built environment attributes around the stations).

The remainder of the paper is organized as follows. The next section presents the data and the sample formation steps. Section 4 describes the methodology used and model structure. The model results and validation are presented in Section 5. Finally, Section 6 concludes the paper.

3. DATA

New York's CitiBike system is one of the major public BSS in the world and the largest in the United States. The service was launched in May 2013 with 330 stations and 6,000 bicycles in the lower half of Manhattan and some part of northwest of Brooklyn. The system covers the city's major commercial business districts and some residential areas with an annual ridership of more than 8 million trips in 2014. The city's dense and walkable urban form provides a good opportunity for the success of a well-designed BSS. The users can either buy the annual membership for or a temporary daily pass. The first 45 minutes (30 minutes) of each ride are included in the annual membership (daily pass) price with additional charges for each additional 15 minutes.

The trip itinerary dataset of New York's CitiBike system was obtained from its website (https://www.citibikenyc.com/system-data). The dataset has the information for every trip made by the CitiBike system since beginning operations in 2013. For every trip, origin and destination stations, start time and end time, and type of user information are provided in the dataset, as well as, the age and gender of annual members. Further, the CitiBike stations' coordinates and capacity are also available in the dataset. The built environment attributes such as bicycle routes and subway stations were derived from New York City open data (https://nycopendata.socrata.com) while the socio-demographic characteristics of resident population were gathered from US 2010 census and the weather information corresponding to the Central Park station was retrieved from the National Climatic Data Center (http://www.ncdc.noaa.gov/data-access).

3.1. Sample Formation

We employed data for trips for the year 2014. Several steps were followed to generate the final samples for model estimation. First, trips with missing or inconsistent information as well as very long trips (longer than 2 hours) were deleted (only 0.5% of all the trips). Second, modeling trips with the same origin and destination is beyond the scope of this study and thus were also excluded from the sample. Further, trips made by annual members and users with daily passes were separated; about 90% of all the trips were made by annual members. Further, to maintain a reasonable sample for data processing and model estimation related computational effort, 10,000 trips by members were randomly selected from the entire year of 2014. For non-members, we chose two different sample sizes for two different model estimations: 1) a sample size of 10,000 trips for non-member only model; 2) a sample size of 1087 trips to add to the 10,000 member trips to obtain an overall pooled sample representing the CitBike trips (90.2% members and 9.8% non-members). The sample sizes followed the recommendations of an earlier study that investigated the impact of different sample sizes on BSS analyses (Faghih-Imani and Eluru, 2017).

There are 332 CitiBike stations in New York City in 2014. Since we focus on trips that were destined outward, there are 331 stations in the universal choice set of destinations. Three samples of 30, 60 and 120 alternatives (destination stations) including the chosen alternative were randomly selected for analysis[1].

3.2. Independent Variables Generation

The independent variables considered in our analysis can be categorized into three groups: (1) weather and temporal characteristics, (2) spatial variables and (3) trip attributes. Weather variables include hourly temperature and relative humidity. In order to capture the time of the day effect on usage, based on the start time of the trips, five time periods were created - AM (6:00-10:00), Midday (10:00-16:00), PM (16:00-20:00), Evening (20:00-24:00), and Night (0:00-7:00). Day of the week impact also was considered in the model estimation by a categorical variable indicating weekends. Moreover, for the users with an annual membership, the gender, and age information were available.

Several variables were considered from the spatial variables group. The distance from each station to Washington Square Park was computed as a measure to identify the impact of Central Business District (CBD). We employed the population information at census block level and the employment data at zip code level to generate population density and job density variables. For the rest of the attributes a 250-meter buffer around each station is used. The 250-meter buffer seems an appropriate walking distance based on the distances between CitiBike stations and the dense urban form of New York City (Kaufman et al., 2015). The set of variables generated at buffer level include the length of bicycle routes, streets and railways in the buffer, the number and capacity of CitiBike stations in the buffer, the presence of subway and Path train stations in the buffer, the number of restaurants (including coffee shops and bars), and area of park in the buffer.

Trip attributes considered in destination choice model include the street network distance between the origin and destination of every trip. The shortest distance is computed based on the street network around the stations (excluding highways). Although the actual trip might take place on another route, the shortest distance can act as a reasonable surrogate for the actual distance travelled. Further, a categorical variable indicating whether the trip needed to cross a bridge (i.e. origin or destination in two boroughs of Manhattan or Brooklyn) or not also was generated. Table 1 presents a descriptive summary of sample characteristics.

4. METHODOLOGY

A brief description of the FMMNL model employed in our study is provided below[2].

Let us consider S homogenous segments of trips (the optimal number S is to be determined). The utility for assigning a trip q (1,2,..Q) to segment s is defined as:

---

[1] To be sure, random sampling of alternative might affect parameter estimates in the FMMNL model. Random sampling does not introduce a bias in the estimation process for simple multinomial logit model. However, based on recent research by Guevara and Ben-Akiva (2013) there is evidence to suggest that the naïve estimator (i.e. employing random sampling based estimation) offers reasonable accuracy in model estimation for MMNL model. We conducted a comparison exercise with different number of alternatives for the FMMNL model and observed relatively similar parameters with increasing choice set size (similar to MMNL).

[2] For sake of brevity, we did not include the mathematical formulations for MNL and MMNL models.

$$U_{qs}^{*} = \beta_s' z_q + \xi_{qs} \tag{1}$$

z_q is a (M x 1) column vector of attributes that influences the propensity of belonging to segment s, β_s' is a corresponding (M x 1) column vector of coefficients and ξ_{qs} is an idiosyncratic random error term assumed to be identically and independently Gumbel-distributed across trips q and segment s. Then the probability that trip q belongs to segment s is given as:

$$P_{qs} = \frac{\exp(\beta_s' z_q)}{\sum_s \exp(\beta_s' z_q)} \tag{2}$$

Now let us assume k (1,2, … K, in our study K=30) represents destination station choices, then the random utility formulation takes the following form when a trip probabilistically assigned to a segment *s* and station *k* is chosen as destination:

$$U_{qk} \,|\, s = \alpha_s' \, x_q + \varepsilon_{qk} \tag{3}$$

x_q is a (L x 1) column vector of attributes that influences the utility of destination choice model. α is a corresponding (L x 1)-column vector of coefficients and ε_{qk} is an idiosyncratic random error term assumed to be identically and independently Gumbel distributed across the dataset. Then the probability that trip q chooses station k as destination within the segment s is given as:

$$P_q(k) \,|\, s = \frac{\exp(\alpha_s' x_q)}{\sum_k \exp(\alpha_s' x_q)} \tag{4}$$

Within the finite mixture framework, the overall probability of trip q to be destined to station k is given as:

$$P_q(k) = \sum_{s=1}^{S} (P_q(k) \,|\, s)(P_{qs}) \tag{5}$$

Therefore, the log-likelihood function for the entire dataset is:

$$L = \sum_{q=1}^{Q} \log(P_q(k_q^*)) \tag{6}$$

where k_q^* represents the chosen destination station for trip q. By maximizing this log-likelihood function, the model parameters β and α are estimated. GAUSS matrix programming language is used to code the maximum likelihood model estimation.

5. ANALYSIS AND DISCUSSION

We estimated the traditional MNL and MMNL models and used their results as a starting point for our FMMNL estimation. The FMMNL estimation process starts with a model with only two segments. Then, we continue to add segments to model until it does not significantly improve model performance. The model performance is evaluated using Bayesian Information Criterion (BIC). For a given empirical model, BIC = Kln(Q) - 2ln(L) where K is the number of parameters, Q is the number of observations and ln(L) is the log–likelihood value at convergence. The model with the lowest value of BIC is preferred. Thus, the final and optimal number of segments corresponds to the model with lowest BIC value. We separately estimated the models for members and non-members on samples of 10,000 trips. In addition, we estimated one pooled model of 10,000 trips by members and 1087 trips by non-members. The separate models out-performed the pooled model in terms of model goodness of fit. The better fit produced by the separate models can be attributed to two factors: (1) gender and age variables could be utilized for members in the separate models. The information was ignored in the pooled model as it was available only partially (for members only) and (2) the pooled model by implicitly restricting the influence of exogenous factors to be the same for members and non-members ignores the behavioral differences across the two user types. While the pooled model performed reasonably, for the sake of brevity in this paper, we focus on two separate models by user membership type. The reader would note that the FMMNL model was tested with multiple choice set sizes (30, 60 and 120). The specification was quite stable and offered minor differences across the various sample sizes. Hence, we restricted ourselves to presenting the results of sample size 30 in our paper.

For members and non-members, the FMMNL with three segments provided the best fit. We also estimated the corresponding traditional MNL models and MMNL models and used them as benchmarks to evaluate our FMMNL model performances. Table 2 summarizes the models' performance. For members, the final log-likelihood (BIC) value of the FMMNL model with three segments is -28132.6 (56762.5) while the corresponding values for the traditional MNL model is -28451.0 (57012.5) and for MMNL model is -28421.2 (56980.6). Similarly, for non-members, the final log-likelihood (BIC) value of the FMMNL model with three segments is -28589.8 (57594.1) while the corresponding value for the traditional MNL model is -29270.4 (58679.0) and for MMNL model is -29231.4 (58637.7). The improvement in the data fit illustrates the improved data fit provided by the FMMNL based destination choice models; providing evidence in favor of the presence of segments within the spectrum of trips especially for the non-member users. The results confirm our hypothesis that BSS users' decision process heterogeneity can be investigated through segmentation of trips. Moreover, the FMMNL allows us to capture behaviorally richer contributing factors.

5.1. Model Estimation Results

The estimation results for traditional MNL and MMNL models are presented in Table 3. In traditional MNL and MMNL models, we can only estimate parameters for destination station attributes. The FMMNL model estimation results are presented in Table 5 and 6. Table 5 presents the segmentation component and Table 6 presents the destination choice component of the FMMNL model. For the sake of brevity, this section focuses on the discussion of the results of the FMMNL model with three segments for members and non-members to understand the different factors influencing users' choice of destination in the New York City's CitiBike bicycle-sharing system.

5.1.1. Behavioral interpretation

To provide insights on the segments identified by FMMNL, it is useful to examine the segment characteristics. Specifically, we can estimate the trips' share across the segments as well as the distribution of independent variables within each segment (see Anowar et al., 2014 or Bhat 1997 for more details on computation procedures). These estimates are presented in Table 4.

For members, the probabilities of trips belonging to the different segments are: segment 1 (39.4%), segment 2 (18.8%) and segment 3 (41.8%). A closer examination of the mean values of exogenous variables will allow us to understand the membership of these segments. Segment 1 is likely to have users' bicycling in AM, Midday and PM periods toward destinations with high job density indicating the higher likelihood of ***daily commuters*** being allocated to this segment. Segment 2 is primarily composed of ***recreational trips*** of younger individuals (average age under 30) as indicated by higher evening and weekend participation with trips destined to lower job density areas. Segment 3 can be considered an ***everyday bicyclist*** group who show high bicycling on weekdays (AM, PM, and midday) and weekends. To be sure, these labels are indicative of the potential identifiers for each segment based on the average values of exogenous variables. The difference of mean values for some exogenous variables might be very small to draw clear distinctions.

For non-members, the composition of the three segments are: segment 1 (17.7%), segment 2 (43.0%) and segment 3 (39.3%). The non-members trips are usually made in better weather conditions compared to members' trips and are of longer duration. Based on the exogenous variable comparison, segment 1 is mostly ***weekend long trips*** as indicated by higher values of travel duration and weekend variable values. The second segment corresponds to ***casual short trips*** based on higher rates for evening and duration. Finally, the third segment accounts for ***midday trips***.

5.1.2. The segmentation component

The segmentation component determines the probability that a trip is assigned to one of the three segments identified. In our modeling effort, we select the everyday bicyclist segment (for members) and the midday trips' segment (for non-members) as the base for the segmentation model.

5.1.2.1. Time and weather variables:

The negative coefficients for AM, Midday, PM variables indicate a high propensity for trips in Evening and Night period for members' recreational trips segment. The probability of a non-members' trip to be in the casual short segment is negatively affected by Midday and PM period variables. As expected, the weekend variable has a negative impact on the utility of members' daily commute segment while the same variable has a positive influence for non-members' weekend long segment. Temperature variable exhibits an interesting influence on the segmentation process. The reader would note that in New York City the temperature is rarely too high to discourage bicycling. For members, higher temperature days encourage trip allocation to everyday bicyclist segment i.e. individuals exhibit tendency to use BSS in spite of it being a weekend. For non-members, as expected, higher temperature accounts for increased allocation to the weekend long segment.

5.1.2.2. <u>Origin built environment attributes:</u>

Trips starting from origins near transit stations are more likely to be assigned to daily commute and recreational segments for members while such trips are less likely to be allocated to non-members weekend long trip segment. Trips originating from stations in areas with higher population density are more likely to be members' recreational trips or non-members' casual short trips. The population density variable has also a negative impact on the utility of members' daily commute and non-members' weekend long trips' segments. When the origin station distance to CBD increases, as expected we are less likely to see non-members' casual short trips and more likely to have members' recreational and daily commute trips. The origin station elevation and the area of parks near origin station variables also influence trip allocation to the segments.

5.1.2.3. <u>User attributes:</u>

The age and gender attributes are only available for members and thus are only included in members' model estimation. As expected, younger users are more likely to engage in recreational trips while female members are less likely to use BSS for the daily commute.

5.1.3. Destination choice component

5.1.3.1. <u>Destination bicycle infrastructure attributes:</u>

The destination choice component quantifies the influence of contributing factors on users' destination station preferences. The station capacity variable does not have a significant impact for recreational trips while it has a positive coefficient for the other segments for both members and non-members' models. The results are expected as the large stations are more likely to have available docks to return the bicycle, thus they are more likely to be chosen as a destination. The number of stations and the capacity of stations within the buffer variables are aimed to capture the impact of neighboring stations on destination choice. It must be noted that the overall impact of number and capacity of stations in the buffer should be examined since as the number of stations in the buffer increases we simultaneously increase the capacity in the buffer. Overall, an increase in the number of stations in the buffer results in an increase in destination likelihood for daily commute segment. The length of bicycle facilities in the buffer has positive coefficients in two segments of non-members model, highlighting the importance of bicycle infrastructure in encouraging people to cycle (similar to findings of Faghih-Imani and Eluru, 2016b). The findings overall illustrate how based on the segment considered the influence of the exogenous variables can vary substantially; thus supporting our hypothesis that allowing for such systematic heterogeneity provides increased flexibility.

5.1.3.2. <u>Destination built environment attributes:</u>

Members in daily commute segment and non-members in midday trips segment are more likely to choose destination near transit stations while the impact is negative for weekend long trips. The length of railways in buffer variable represents a negative influence for non-members' weekend long trips while for the non-members' other two segments, it represents an attraction. Stations near parks are more likely to be selected by non-members across all segments and less likely to be selected by members' daily commuters. Members' recreational and everyday bicyclist trips and non-members' weekend long trips are more likely to be made downhill as indicated by the negative coefficient of elevation variable in these segments. The insignificance of the parameter for daily commuters is notable. Individuals are unlikely to have a choice in elevation along their commute. The negative coefficients of distance to CBD variable for members in everyday bicyclist segment and non-members in the casual short segment indicate that those users are more likely to choose stations that bring them closer to CBD. On the other hand, non-members in weekend long trips segment choose stations that take them farther from CBD.

Population density variable has negative impacts on the utility of choosing a station in members' daily commute and non-members' weekend long segments while has a positive impact in recreational trips and everyday bicyclist segments. Recreational trips are less likely to be destined to stations in high job density area while daily commute and everyday bicyclist trips are prone to be destined to a station in high job density area in AM period. Further, in non-members casual short segment, the population density variable has a negative coefficient in AM period and job density variable has a negative coefficient in PM period clearly indicating an occasional commute trend in this segment of non-members trips. Overall, the coefficients for population and job density variables in different segments are consistent with findings of earlier studies (see Rixey 2013; Faghih-Imani et al., 2014; Wang et al., 2015) and support our interpretation of the segments estimated.

5.1.3.3.<u>Trip attributes:</u>

The trip distance is expected to be one of the most significant variables in destination station decision-making process of BSS users. In general, it is expected that individuals do not use the BSS for very short trips or for very long distance trips. Thus, to better model the distance impact on the utility of choosing a station, we distinguish the very short distance and very far distance by indicator variables and a continuous variable for distance in between. To do so, we tested various thresholds for the trip distance to characterize the short, medium, and long distance trips. At the end, we employed indicator variables identifying stations within 750m or farther than 4km of origin and a continuous form of distance for the stations within 750m to 4km from the origin. As expected, for all segments the coefficients estimated for the network distance variables are negative, indicating that the trip distance has a negative impact on the likelihood of choosing a station as a destination. We also investigate the interaction of distance variable with gender and temperature variables. The results show that when the temperature increases, it is expected that users pursue longer trips in daily commute and recreational trips segments. Further, female members are more likely to have longer trips in recreational trips and everyday bicyclist segments (in agreement with the findings of Faghih-Imani and Eluru 2015). The weekend long trips segment of non-members has the lowest magnitude for long distance trips (over 4km). The coefficients of distance variables and specifically long distance variable are another indicator in support of our behavioral interpretation of segments.

People are less likely to choose a station that requires them to cross a bridge between Manhattan and Brooklyn boroughs except in the non-members midday segment. The result is an indication of how a separation (by water) imposes a subtle boundary for activity travel participation envelope.

5.2. Model Validation

The performance of the models estimated is evaluated using hold-out samples: 3000 trips by members and 3000 trips by non-members. The same data processing exercises and choice set generation for estimation samples were undertaken for the validation samples. The utility and the probability of choosing a station are computed for 30 stations of choice set for each of the 3000 trips. We generate several measures to evaluate the prediction performance: a) the predictive log-likelihood (the log-likelihood for the predicted probabilities of the sample), b) the probability of correct prediction (correct prediction is defined as assigning the highest probability to the chosen alternative) and c) average probability of the chosen station. We compute these measures for the FMMNL model and for the equivalent traditional MNL and MMNL models. Moreover, we generate the performance measures for the entire validation sample as well as several sub-samples within the validation sample. The validation results are presented in Table 5.

The predictive log-likelihood for FMMNL models of members and non-members are -8493.0 and -8535.4 while the corresponding values for the traditional MNL are -8543.9 and -8705.1, and for MMNL are -8547.0 and -8712.3, respectively (It must be noted that the predictive log-likelihood for an equal probability model is -10203.6). Furthermore, the FMMNL models perform better than both traditional and MMNL models in all sub-samples except for members' sub-sample of trips that require passing a bridge between Manhattan and Brooklyn. Interestingly, the FMMNL models significantly outperform the traditional MNL models in prediction for long trips highlighting the importance of capturing the heterogeneity for long distance variable. Moreover, comparing to the observed destination choice, the chance of correct destination station prediction for members and non-members for FMMNL model is about 14.5 and 17.2, which is about 0.5 and 1.2 units higher than the corresponding value for traditional MNL and MMNL models, respectively (it must be noted that a prediction without any model to choose a destination out of 30 stations has a chance of being correct at about 3.33%). The average of predicted probability for the observed destination station is 0.0806 and 0.0846 for FMMNL members and non-members models while the corresponding value for MMNL is 0.0803 and 0.0763 and for traditional MNL models is 0.0787 and 0.0764, respectively. The validation exercise indicates that in addition to having a richer explanatory power, the proposed FMMNL model performs relatively well in terms of prediction.

6. CONCLUSION

There is a rapid growth of bicycle-sharing systems around the world in recent years. This paper contributes to research on BSS by examining user behavior at a trip level. Specifically, we study the decision process involved in identifying destination locations after picking up the bicycle at a BSS station. The assumption that the influence of exogenous variables remains the same for the entire spectrum of trips might lead to biased estimation. Thus, in this study, rather than homogenizing the entire spectrum of trips we examine the presence of potential segments within the trip population. Specifically, by developing a Finite Mixture Multinomial Logit Model (FMMNL), we allow for various segments and segment specific destinations choice models to enhance our understanding of the destination choice behavior. The proposed approach can account for fixed attributes (such as origin variables or temporal and meteorological characteristics) through the segmentation component of the model while the traditional MNL model is limited to attributes that vary across alternatives within the choice set (destination attributes). We estimate our FMMNL on data from the New York City bicycle-sharing system (CitiBike) for 2014. We develop two separate models for members and non-members.

The optimal number of segments for members' and non-members' FMMNL was three segments. The comparison of the proposed FMMNL models with corresponding traditional MNL models illustrates the superiority of the FMMNL based destination choice models. The results indicate the presence of segments within the spectrum of trips thus confirming that better understanding of BSS users' decision process is possible through segmentation of trips. Our estimation process identifies three segments of daily commute trips, recreational trips and everyday bicyclist trips for members and three segments of weekend long trips, casual short trips, and midday trips for non-members. The estimated coefficients in segmentation and destination choice components of the FMMNL support our interpretation of the segments computed. Further, we validate our proposed FMMNL model with hold-out samples. The validation results indicate the superiority of the prediction performance of the FMMNL model compared to the traditional MNL model. The use of the FMMNL framework to analyze users' destination choice process also provides richer explanatory power which would be very useful for BSS operators/analysts to better understand the individuals' decision processes in adoption and usage of BSS in order to enhance their service offerings.

The model estimates generated have several applications. For example, the developed model will allow BSS operators to examine the impact of travel distance, bicycle infrastructure, land use, and built environment on destination preferences of different types of users. The detailed specification will allow BSS operators to evaluate the impact of different changes in the system or built environment on users' travel behaviour. Specifically, the BSS operators can differentiate such impacts for different segments of trips. The model will also provide guidance on how the expansion of the existing bicycle-sharing system will affect the current station demand by providing potential destination locations to be used from the newly proposed bicycle stations.

The study is not without limitations. In traditional transportation demand models, traffic analysis zones (TAZs) are considered as destination choices. The aggregation of destination location to the TAZ level results in loss of exact destination information. However, within the confines of data availability and modelling complexity, this is the accepted norm for destination (or even residential and workplace location). In our particular context, given the mode of travel is restricted to bicycling, our destination choice models that consider station as the choice are less affected by aggregation bias. To elaborate, once the bicycle is returned, customers are likely to walk to their actual destination. Thus, for most customers, the destination is very close to the station. Hence, our approach of considering the station as the destination is a reasonable compromise to reduce model complexity. However, attempts to consider destination choice at finer spatial resolution is of value in future efforts. Further, in the dataset, it is not possible to identify trips made by the same individual, thus we were not able to account for the panel structure of the data. The availability of docks in destination stations and their neighbouring stations can influence users' decisions to choose a station. Future work can use a web crawler to capture snapshot data of stations' state in real-time from the operator website and use that information along with the trip dataset in examining destination choice preferences. The endogeneity impact of station capacity on destination choice should be examined in future efforts.

ACKNOWLEDGMENTS

The authors would like to acknowledge insightful feedback on an earlier version of the manuscript from two anonymous reviewers and the editor.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Anowar S., S. Yasmin, N. Eluru, and L. Miranda-Moreno. Analyzing Car Ownership in Quebec City: A Comparison of traditional and Latent Class Ordered and Unordered Models. *Transportation*, Vol. 41, 2014, pp. 1013-1039.
2. Bachand-Marleau, J., B. Lee, and A. El-Geneidy. Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2314, 2012, pp. 66-71.
3. Bhat, C. R. Endogenous Segmentation Mode Choice Model with an Application to Intercity Travel. *Transportation Science*, Vol. 31, 1997, pp. 34-48.
4. Bouveyron, C., E. Côme, and J. Jacques. The discriminative functional mixture model for a comparative analysis of bike sharing systems. *Annals of Applied Statistics*, Vol. 9, 2015, pp. 1726-1760.
5. Buck, D., D. Buehler, P. Happ, B. Rawls, P. Chung, and N. Borecki. Are bikeshare users different from regular cyclists? a first look at short-term users, annual members, and area cyclists in the Washington, D.C., region. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2387, 2013, pp. 112–119.
6. Chakour V. and N. Eluru. Analyzing Commuter Train User Behavior: A Decision Framework for Access Mode and Station Choice, *Transportation*, Vol. 41, 2014, pp. 211-228.
7. Davis, B., T. Dutzik and P. Baxandall. Transportation and the New Generation: Why Young People Are Driving Less and What It Means for Transportation Policy, Frontier Group, 2012.
8. Eluru, N., M. Bagheri, L. Miranda-Moreno, and L. Fu. A latent class modelling approach for identifying vehicle driver injury severity factors at highway-railway crossings. *Accident Analysis & Prevention*, Vol. 47, 2012, pp. 119-127.
9. Ettema, D.F. The impact of telecommuting on residential relocation and residential preferences, a latent class modeling approach. *Journal of Transport and Land Use*, Vol. 3, 2010, pp. 7-24.
10. Faghih-Imani, A., Hampshire, R., Marla, L., & Eluru, N. An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice*, Vol. *97*, 2017, 177-191.
11. Faghih-Imani, A., and N. Eluru, "Determining the Role of Bicycle Sharing System Infrastructure Installation Decision on Usage: Case Study of Montreal BIXI System," Transportation Research Part A 94C (2016a) pp. 685-698
12. Faghih-Imani A., and N. Eluru. Analyzing bicycle sharing system user destination choice preferences: An investigation of Chicago's Divvy system. *Journal of Transport Geography*, Vol. 44, 2015, pp. 53-64.
13. Faghih-Imani A., and N. Eluru. Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. *Journal of Transport Geography*, Vol. 54, 2016b, pp. 218-227.
14. Faghih-Imani, A., and N. Eluru. Examining the impact of sample size in the analysis of bicycle-sharing systems. *Transportmetrica A: transport science*, Vol. 13, (2017), pp 139-161.
15. Faghih-Imani A., N. Eluru, A. El-Geneidy, M. Rabbat, and U. Haq. How land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography*, Vol. 41, 2014, pp. 306-314.
16. Fishman, E. and Schepers, P. Global bike share: what the data tells us about road safety. *Journal of Safety Research*, Vol. 56, 2016, pp. 41-45.

17. Fishman, E. Bikeshare: A review of recent literature, *Transport Reviews*, Vol. 36, 2016, pp. 92-113.
18. Fishman, E., S. Washington and N. Haworth. Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia, *Transportation Research Part D: Transport and Environment*, Vol. 31, 2014, pp. 13-20.
19. Fishman, E., S. Washington and N. Haworth. Bikeshare's impact on active travel: Evidence from the United States, Great Britain, and Australia. *Journal of Transport & Health,* Vol. 2, 2015b, pp. 135-142.
20. Fishman, E., Washington, S., Haworth, N. and A. Watson. Factors influencing bike share membership: An analysis of Melbourne and Brisbane. *Transportation research part A: policy and practice*, Vol. 71, 2015a, pp.17-30.
21. Forma, I.A., Raviv, T. and M. Tzur. A 3-step math heuristic for the static repositioning problem in bike-sharing systems. *Transportation research part B: methodological*, Vol. 71, 2015, pp. 230-247.
22. Fricker, C., and N. Gast. Incentives and redistribution in homogeneous bike-sharingsystems with stations of finite capacity. *EURO Journal on Transportation and Logistics*, 2014, pp. 1–31.
23. Gebhart, K. and R. Noland. The Impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, Vol. 41, 2014, pp. 1205-1225.
24. González, F., C. Melo-Riquelme, and L. de Grange. A combined destination and route choice model for a bicycle sharing system. *Transportation*, Vol. 43, 2015, pp 407-423.
25. Goodman, A., J., Green, and J. Woodcock. The role of bicycle sharing systems in normalising the image of cycling: An observational study of London cyclists. *Journal of Transport & Health*, Vol. 1, 2014, pp. 5-8.
26. Graves, J.M., Pless, B., Moore, L., Nathens, A.B., Hunte, G. and F.P. Rivara. Public bicycle share programs and head injuries. *American journal of public health*, Vol. 104, 2014, pp. 106-111.
27. Guevara, C. A. and M. E. Ben-Akiva. Sampling of alternatives in Logit Mixture models. *Transportation Research Part B: Methodological*, Vol. 58, 2013, pp. 185-198.
28. Jäppinen, S., T. Toivonen, and M. Salonen. Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography*, Vol. 43, 2013, pp. 13-24.
29. Kaufman, S. M., L. Gordon-Koven, N. Levenson and M. L. Moss. Citi Bike: The First Two Years. The Rudin Center for Transportation Policy and Management, 2015. Accessed online on June 6th 2016 at: http://wagner.nyu.edu/rudincenter/wp-content/uploads/2015/06/Citi_Bike_First_Two_Years_RudinCenter.pdf.
30. Kemperman, A. and H. Timmerman. Influences of built environment on walking and cycling by latent segments of aging population. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2134, 2009, pp.1-9.
31. Kraemer, J.D., Roffenbender, J.S. and L. Anderko. Helmet wearing among users of a public bicycle-sharing program in the District of Columbia and comparable riders on personal bicycles. *American journal of public health*, Vol. 102, 2012, pp. 23-25.
32. Ma, T.Y., Gerber, P., Carpentier, S. and S. Klein. Mode choice with latent preference heterogeneity: a case study for employees of the EU institutions in Luxembourg. T*ransportmetrica A: Transport Science*, Vol. 11, 2015, pp. 441-463.

33. Martin, E.W. and S.A. Shaheen. Evaluating public transit modal shift dynamics in response to bikesharing: a tale of two US cities. *Journal of Transport Geography*, Vol. 41, 2014, pp. 315-324.
34. McLeod, K. Where We Ride: Analysis of Bicycle Commuting in American Cities, League of American Bicyclists, 2014. Accessed online on July 19th 2016 at: http://bikeleague.org/sites/default/files/Where_We_Ride_2014_data_web.pdf
35. Meddin, R. and P. DeMaio. The Bike-Sharing World Map, http://www.bikesharingworld.com Accessed 20th June, 2016.
36. Murphy, E., and J., Usher. The role of bicycle-sharing in the city: Analysis of the Irish experience. *International Journal of Sustainable Transportation*, Vol. 9, 2015, pp. 116-125.
37. Nair, R., Miller-Hooks, E., Hampshire, R. C., and A. Bušić. Large-Scale Vehicle Sharing Systems: Analysis of Vélib'. *International Journal of Sustainable Transportation*, Vol. 7, 2013, pp. 85–106.
38. O'Brien, O., Cheshire, J., and M. Batty. Mining bicycle sharing data for generating insights into sustainable transport systems. *Journal of Transport Geography*, Vol. 34, 2014, pp. 262-273.
39. Pfrommer, J., Warrington, J., Schildbach, G. and M. Morari, Dynamic vehicle redistribution and online price incentives in shared mobility systems. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, 2014, pp.1567-1578.
40. Raviv, T., Tzur, M. and I.A. Forma. Static repositioning in a bike-sharing system: models and solution approaches. EURO Journal on Transportation and Logistics, Vol. 2, 2013, pp.187-229.
41. Rixey, R. Station-level forecasting of bikesharing ridership: Station network effects in three U.S. systems. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2387, 2013, pp. 46-55.
42. Sobhani A., N. Eluru and A. Faghih-Imani. A latent segmentation based multiple discrete continuous extreme value model. *Transportation Research Part B: Methodological*, Vol. 58, 2013, pp. 154-169.
43. Vij, A., Carrel, A. and J.L. Walker. Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A: Policy and Practice*, Vol. 54, 2013, pp.164-178.
44. Vogel, P., and D.C. Mattfeld. Strategic and operational planning of bike-sharing systems by data mining–a case study. *In International Conference on Computational Logistics*, pp. 127-141. Springer Berlin Heidelberg, 2011.
45. Waddell, P., C.R. Bhat, N. Eluru, L. Wang and R.M. Pendyala. Modeling the Interdependence in Household Residence and Workplace Choices. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2003, 2007, pp. 84-92.
46. Walker, J.L. and J. Li. Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, Vol. 9, 2007, pp.77-101.
47. Wang, X., G. Lindsey, J. Schoner and A. Harrison. Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development*, Vol. 142, 2015, pp 1-9.
48. Yasmin. S., N. Eluru, C. R. Bhat and R. Tay. A Latent Segmentation Generalized Ordered Logit Model to Examine Factors Influencing Driver Injury Severity, *Analytic Methods in Accident Research*, Vol. 1, 2014, pp. 23-38.

Table 1 Descriptive Summary

	Members Dataset				Non-Members Dataset			
Continuous Variables	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
Trip Distance (km)	0.05	10.32	2.00	1.33	0.05	10.78	2.31	1.39
Trip Duration (min)	1.02	117.83	11.92	8.44	1.70	119.80	22.95	15.33
Temperature (°C)	-15	33.3	16.35	9.09	-11.1	33.3	19.72	6.94
Relative Humidity (%)	13	100	56.34	18.06	13	100	54.34	16.72
Members Age	17.00	90.00	37.99	11.25	-	-	-	-
Origin Attributes								
Length of Bicycle Facility in 250m Buffer (m)	0	3355.3	1049.92	582.56	0	3355.3	1060.15	588.23
Length of Railways in 250m Buffer (m)	0	9705.14	77.77	451.30	0	9705.14	124.39	711.48
Area of Parks in 250m Buffer (m^2)	0	95209.9	8634.99	12793.99	0	95209.9	13399.34	18359.73
Number of Restaurants in 250m Buffer	0	545	63.70	99.87	0	545	53.43	85.17
Number of CitiBike stations in 250m Buffer	0	4.00	1.29	0.95	0	4.00	1.21	1.02
Capacity of CitiBike stations in 250m Buffer	0	169.00	48.44	37.85	0	169.00	45.05	39.60
Station Capacity	3.00	67.00	38.15	10.81	3.00	67.00	37.76	10.81
Population Density (people per $m^2 \times 1000$)	0.01	67.20	26.86	14.78	0.01	67.20	24.00	14.07
Job Density (jobs per $m^2 \times 1000$)	0	432.52	65.67	47.01	0	432.52	62.46	50.49
Distance to CBD (m)	0	7939.95	2423.91	1369.82	0.00	7939.95	2745.74	1444.26
Destination Attributes								
Length of Bicycle Facility in 250m Buffer (m)	0	3355.3	1027.08	591.25	0	3355.3	1059.78	589.47
Length of Railways in 250m Buffer (m)	0	9705.14	86.11	503.56	0	9705.14	121.89	655.38
Area of Parks in 250m Buffer (m^2)	0	95209.9	10186.22	15159.82	0	95209.9	13803.77	18685.06
Number of Restaurants in 250m Buffer	0	545	54.20	92.11	0	545	52.14	86.54
Number of CitiBike stations in 250m Buffer	0	4.00	1.24	1.01	0.00	4.00	1.17	1.00
Capacity of CitiBike stations in 250m Buffer	0	169.00	44.15	38.85	0.00	169.00	43.30	38.60
Station Capacity	3.00	67.00	34.40	10.79	3.00	67.00	37.52	10.98
Population Density (people per $m^2 \times 1000$)	0.01	67.20	24.90	14.69	0.01	67.20	23.77	13.96
Job Density (jobs per $m^2 \times 1000$)	0	432.52	55.98	53.79	0	432.52	60.23	48.78
Distance to CBD (m)	0	7728.00	2434.85	1376.65	0.00	7939.95	2763.52	1460.47
Categorical Variables	**Percentage**				**Percentage**			
Weekends	21.5				46.93			
Transit Station in 250m Buffer of Origin	59.56				55.96			
Transit Station in 250m Buffer of Destination	58.06				55.89			
Female Members	22.7				-			

Table 2 Summary of Models' Performance

	Model	Log-likelihood (LL)	Number of Parameters	Number of Observations	BIC
Members	MNL	-28451.0	12	10000	57012.5
	MMNL	-28421.2	15	10000	56980.6
	FMMNL (2 segments)	-28266.2	38	10000	56882.4
	FMMNL (3 segments)	-28132.6	54	10000	56762.5
Non-members	MNL	-29270.4	15	10000	58679.0
	MMNL	-29231.4	19	10000	58637.7
	FMMNL (2 segments)	-28770.2	31	10000	57826.0
	FMMNL (3 segments)	-28589.8	45	10000	57594.1

Table 3 MNL and MMNL Estimation Results

	MNL				MMNL			
	Members		Non-members		Members		Non-members	
Variables	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Destination Bicycle Infrastructure Attributes								
Station Capacity	0.188	16.62	0.224	18.63	0.182	15.10	0.214	17.40
Standard Deviation	-	-	-	-	0.050	1.82	-	-
Number of Other Citibike Station in Buffer	-	-	-0.232	-6.97	-	-	-0.234	-6.91
Capacity of Other Citibike Station in Buffer	-	-	0.002	2.78	-	-	0.003	2.90
Length of Bicycle Facility in Buffer	-	-	0.074	6.93	-	-	0.067	6.09
Standard Deviation	-	-			-	-	0.152	4.77
Destination Built Environment Attributes								
Transit Station in Buffer	-	-	-0.088	-3.68	-	-	-0.061	-2.48
Area of Parks in Buffer	-0.023	-1.81	0.233	26.22	-0.024	-1.88	0.225	23.08
Standard Deviation	-	-	-	-	-	-	0.052	2.46
Number of Restaurants in Buffer	-	-	-0.038	-3.34	-	-	-0.073	-3.77
Standard Deviation	-	-	-	-	-	-	0.167	5.01
Elevation	-	-	-0.054	-4.77	-	-	-0.121	-7.89
Standard Deviation	-	-	-	-	-	-	0.323	12.81
Distance to CBD	-0.088	-5.59	-	-	-0.109	-6.40	-	-
Standard Deviation	-	-	-	-	0.139	3.55	-	-
Population Density	0.107	6.57	-0.114	-8.29	0.108	6.55	-0.114	-8.23
Population Density * AM	-0.241	-8.65	-	-	-0.244	-8.67	-	-
Population Density * PM	-0.047	-1.92	0.065	2.83	-0.047	-1.92	0.065	2.78
Job Density	-0.093	-5.94	-	-	-0.097	-6.13	-	-
Job Density * AM	0.331	12.57	0.082	1.94	0.336	12.55	0.095	2.23
Trip Attributes								
Distance < 750m	-1.048	-23.81	-0.408	-8.83	-1.169	-23.79	-0.426	-9.04
Distance (750m< &<4km)	-0.870	-36.52	-0.450	-32.68	-0.946	-33.78	-0.459	-32.76
Standard Deviation	-	-	-	-	0.254	11.08	-	-
Distance (750m< &<4km) *Temperature (*10^{-2} °C)	0.317	2.82	-	-	0.317	2.73	-	-
Distance > 4km	-3.875	-80.54	-2.991	-58.55	-3.982	-76.55	-3.040	-58.23
Trip between Manhattan & Brooklyn	-	-	-0.285	-6.45	-	-	-0.268	-6.01

Table 4 Segments Characteristics

	Members			Non-Members		
Mean Values of Segmentation Variables	**Segment 1**	**Segment 2**	**Segment 3**	**Segment 1**	**Segment 2**	**Segment 3**
Trips Share	39.42%	18.81%	41.78%	17.55%	42.80%	39.65%
Female	0.12	0.23	0.33	-	-	-
Age	40.58	29.16	39.53	-	-	-
Temperature (°C)	14.66	17.22	17.55	20.48	19.39	19.72
Relative Humidity (%)	57.47	51.88	57.28	54.00	54.31	54.51
AM	0.30	0.00	0.24	0.04	0.09	0.04
Midday	0.34	0.07	0.38	0.54	0.39	0.56
PM	0.33	0.31	0.34	0.33	0.32	0.31
Evening	0.03	0.50	0.03	0.07	0.16	0.07
Night	0.01	0.12	0.01	0.02	0.04	0.02
Weekend	0.13	0.25	0.28	0.55	0.44	0.47
Distance (km)	2.00	1.91	2.03	2.45	2.27	2.28
Duration (minute)	11.62	11.34	12.46	23.99	21.73	23.79
Origin Attributes						
Population Density (×100)	2.34	2.97	2.88	2.12	2.75	2.14
Job Density (×100)	7.23	6.37	6.03	5.62	6.46	6.29
Area of Parks in Buffer (×100)	0.76	0.77	1	1.59	0.86	1.75
Transit Station in Buffer	0.62	0.72	0.51	0.34	0.59	0.62
Distance to CBD (km)	2.67	2.26	2.27	3.13	2.33	3.02
Destination Attributes						
Population Density (×100)	2.5	3.04	2.71	2.31	2.49	2.29
Job Density (×100)	7.09	5.67	6.4	5.58	6.23	6.00
Area of Parks in Buffer (×100)	0.82	0.9	0.87	1.45	1.26	1.48
Transit Station in Buffer	0.60	0.51	0.59	0.52	0.57	0.57
Distance to CBD (km)	2.59	2.35	2.33	2.94	2.57	2.89

Table 5 Segmentation Component of FMMNL Models

	Members Model (Everyday bicyclist as base)				**Non-Members Model** (Midday trips as base)			
	Daily Commute		**Recreational Trips**		**Weekend long Trips**		**Casual short**	
	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat	Coefficient	t-stat
Constant	0.9898	3.329	4.3046	2.244	-0.9516	-4.569	1.1901	4.598
Temporal and Weather Variables								
AM	-	-	-25.7896	-2.710	-	-	-	-
Midday	-	-	-21.0315	-2.821	-	-	-1.2306	-7.439
PM	-	-	-14.9275	-2.746	-	-	-0.7451	-4.252
Weekend	-0.9566	-4.830	-	-	0.2772	2.806	-	-
Temperature	-3.9635	-4.124	-13.0792	-2.317	1.8192	2.252	-1.8169	-1.960
Origin Built Environment Attributes								
Transit Station in Buffer	0.2986	1.926	7.6789	2.715	-0.9368	-8.187	-	-
Population Density	-0.2349	-3.184	2.0758	2.310	-0.1696	-2.790	0.3073	4.305
Station Distance to CBD	0.3318	3.162	1.6478	2.034	-	-	-0.4512	-5.207
Elevation	0.1400	2.142	-	-	-0.3891	-5.818	-	-
Area of Parks in Buffer	-0.2012	-2.804	-1.6557	-2.078	-	-	-0.4179	-6.279
User Attributes								
Age	-	-	-9.0351	-2.792	-	-	-	-
Female	-1.2716	-4.627	-	-	-	-	-	-

Table 6 Destination Choice Component of FMMNL Models

Destination Choice Component	Members Model						Non-Members Model					
	Daily commute		Recreational trips		Everyday cyclists		Weekend long trips		Casual short trips		Midday trips	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Destination Bicycle Infrastructure Attributes												
Station Capacity	0.2271	7.991	-	-	0.2124	7.102	0.5406	11.190	0.1506	5.647	0.1821	6.028
Number of Other Citibike Station in Buffer	0.2055	2.626	-	-	-0.3636	-4.127	-	-	-	-	-0.5433	-6.793
Capacity of Other Citibike Station in Buffer	-0.0059	-3.092	-	-	0.0043	1.919	-	-	-	-	0.0082	4.156
Length of Bicycle Facility in Buffer	-	-	-	-	-	-	0.1980	3.860	-	-	0.0886	3.366
Destination Built Environment Attributes												
Transit Station in Buffer	0.1909	3.623	-	-	-	-	-1.2697	-9.451	-	-	0.2900	4.640
Area of Parks in Buffer	-0.1288	-3.637	-	-	0.0672	2.334	0.1166	2.686	0.0859	3.095	0.4170	17.534
Number of Restaurants in Buffer	-	-	-	-	-	-	-0.1013	-2.842	-	-	-0.1389	-3.447
Elevation	-	-	-0.0524	-1.724	-0.0833	-3.277	-1.2394	-13.465	0.0616	2.238	0.1220	4.108
Distance to CBD	-	-	-	-	-0.1988	-5.565	0.3247	3.119	-0.2126	-4.905	-	-
Population Density	-0.1445	-4.562	0.2475	8.719	0.2144	5.981	-0.5294	-10.565	-	-	-	-
Population Density * AM	-	-	-	-	-0.2772	-4.967	-	-	-0.2065	-3.393	-	-
Population Density * PM	-	-	-	-	-	-	-	-	-	-	0.0949	2.035
Job Density	-	-	-0.1739	-4.988	-	-	-	-	-	-	-	-
Job Density * AM	0.2588	5.487	-	-	0.2619	4.878	-	-	-	-	-	-
Job Density * PM	-	-	-	-	-0.1913	-3.567	-	-	-	-	-	-
Trip Attributes												
Distance < 750m	-1.2162	-9.484	-1.1482	-10.408	-0.9408	-8.418	-	-	-0.4727	-3.668	-0.9241	-7.483
Distance (750m< &<4km)	-0.8861	-17.987	-1.0841	-15.805	-1.0224	-16.140	-	-	-0.3770	-10.688	-0.8954	-17.004
Distance (750m< &<4km) *Temperature (*10^{-2} °C)	1.4139	5.133	1.1675	3.857	-	-	-	-	-	-	-	-
Distance (750m< &<4km) *Female	-	-	0.2139	3.404	0.2240	3.808	-	-	-	-	-	-
Distance > 4km	-5.2758	-14.649	-4.1769	-28.512	-2.7995	-21.564	-0.7352	-6.794	-1.8740	-13.334	-7.6876	-18.828
Trip between Manhattan & Brooklyn	-	-	-0.3619	-2.348	-1.2159	-9.260	-3.1396	-8.295	-0.9090	-6.743	1.9612	13.201

Table 7 Models Validation Results

	FMMNL (3 segments)		MNL		MMNL	
	Members' Model	Non-members' Model	Members' Model	Non-members' Model	Members' Model	Non-members' Model
Predictive Log-Likelihood						
Overall	-8494.0	-8535.4	-8543.9	-8705.1	-8547.0	-8712.3
AM	-1945.8	-594.2	-1953.1	-606.9	-1956.3	-607.9
Midday	-2522.8	-3916.6	-2539.7	-4004.0	-2538.7	-4009.3
PM	-2825.1	-2833.1	-2844.2	-2891.8	-2844.0	-2891.7
Evening	-999.9	-940.9	-1004.4	-951.2	-1005.5	-951.7
Female	-1951.6	--	-1959.3	--	-1960.2	--
Trip between Manhattan & Brooklyn	-511.3	-929.8	-466.9	-989.1	-470.4	-986.8
Distance < 750m	-775.7	-702.5	-779.4	-702.5	-775.0	-705.6
Distance (750m< &<4km)	-6503.9	-6412.7	-6506.5	-6512.0	-6520.6	-6512.5
Distance > 4km	-1214.4	-1420.2	-1258.0	-1490.6	-1251.4	-1494.2
Percentage of Correct Prediction	14.47	17.40	13.93	16.23	13.97	16.07
Average Probability of Predicted Chosen Station	0.0806	0.0846	0.0787	0.0764	0.0803	0.0763