HOW BICYCLING SHARING SYSTEM USAGE IS AFFECTED BY LAND USE AND URBAN FORM: ANALYSIS FROM SYSTEM AND USER PERSPECTIVES

Ahmadreza Faghih-Imani

McGill University
Department of Civil Engineering and Applied Mechanics
817 Sherbrooke St. W., Room 492, Montreal, Quebec H3A 2K6, Canada
Ph: 514-398-6823, Fax: 514-398-7361
E-mail: seyed.faghihimani@mail.mcgill.ca

Naveen Eluru

University of Central Florida
Department of Civil, Environmental and Construction Engineering
12800 Pegasus Drive, Room 301D, Orlando, Florida 32816, USA
Ph: 407-823-4815; Fax: 407-823-3315
Email: naveen.eluru@ucf.edu

Rajesh Paleti

Old Dominion University
Department of Civil and Environmental Engineering
135 Kaufman Hall, Norfolk, Virginia 23529, USA
Ph: 757-683-5670
Email: rpaleti@odu.edu

Abstract

There is a rapid growth of bicycle-sharing systems (BSS) around the world. Cities are supporting these systems as a more sustainable transport mode for short trips. Given the relatively recent adoption of BSS, there is substantial interest in understanding how these systems impact urban transportation. In this paper, we examine the functioning of the hugely successful New York City CitiBike system. We focus on the interaction of BSS with land-use and built environment attributes and the influence of weather condition and temporal characteristics on BSS usage. Towards this end, CitiBike system is analyzed along two dimensions: (1) at the system level, we examine the hourly station level arrival and departure rates using a linear mixed model and (2) at the trip level, we investigate users' destination station choice preferences after they pick up a bicycle from a station employing a random utility maximization approach. The results highlight clear spatial and temporal differences in the usage of CitiBike by users with annual membership and users with temporary passes. Overall, our analysis provides a framework and useful insights for cities that are planning to install a new bicycle sharing system or to expand an existing system.

1. *Bicycle Sharing Systems*

1.1. Background

Bicycle-sharing systems offer a potential alternative and complementary mode of transportation. These systems are recognized to have traffic and health benefits such as flexible mobility, physical activity, and support for multimodal transport connections (Shaheen et al., 2010). A bicycle-sharing system (BSS) is intended to provide increased convenience because individuals can use the service without the costs and responsibilities associated with owning a bicycle for short trips within the service area of the system. Further, BSS frees the user from the need to secure their bicycles avoiding bicycle theft issues (Rietveld and Daniel, 2004; van Lierop et al., 2015). At the same time, the decision to make a trip can be made in a short time frame providing an instantaneously accessible alternative for a one-way or a round trip. These systems can enhance accessibility to public transportation systems by improving the last mile connectivity (Jäppinen et al., 2013). Moreover, installing BSS promotes active transportation that can enhance physical activity levels to obtain better health outcomes. Furthermore, recent observed trends in travel behavior among the millennial generation (or millennials) demonstrate that the younger generation is willing to drive less. They are environmentally conscious and inclined towards shared transportation systems (Davis et al., 2012; Dutzik and Baxandall, 2013).

There is significant evidence from the travel behavior data in the United States to support BSS installation in urban areas. According to data from the 2009 National Household Travel Survey (NHTS), about 37.6% of the trips by private vehicles in the United States are less than 2 miles long. The NHTS data also indicates that about 73.6% of bicycle trips are less than 2 miles long. Even if a small proportion of the shorter private vehicle trips (around dense urban cores) are substituted with BSS trips it offers substantial benefits to individuals, cities and the environment. A well designed and planned BSS can serve as an access/egress mode for public transportation. Cities, by installing BSS, are focusing on inducing a modal shift to cycling, and subsequently, decrease traffic congestion and air pollution.

BSS research can be broadly categorized along two major dimensions: (1) System level and (2) User level. At the <u>system level</u>, the focus is on understanding BSS usage from the perspective of the BSS operator. The analysis is intended to offer insights on spatio-temporal variation in BSS demand (characterized typically as arrivals and departures at the station). The system level models will allow system operators to predict the demand across the BSS and offer opportunities to reduce any capacity issues (either full or empty stations). Further, the demand representation will also offer solutions to minimize rebalancing operations (such as adding bikes to empty stations and/or removing bikes from full stations) in the system. The quantitative model developed would accommodate for the influence of BSS infrastructure installed, transportation network infrastructure, built environment and urban form, meteorological data and temporal characteristics on BSS usage.

On the other hand, the <u>user level</u> analysis examines individual level preferences for BSS usage. Specifically, the analysis would focus on behavioral considerations such as what factors motivate the adoption of BSS, are users considering shifting their travel mode from car alternatives or what factors influence the distance travelled by bicycle (or choice of destination). In this direction of research, the emphasis is on identifying various individual level market trade-offs (such as distance versus number of activity opportunities) that will promote BSS usage and improve the overall experience. To be sure, the reader would recognize that both these dimensions

have to converge together for the establishment of an efficient BSS. Given the complexity involved in considering the two levels in a unified framework, it is beneficial to consider these dimensions separately and subsequently coalesce the insights from the analyses efforts. The current paper follows this direction of research examining revealed usage data from New York City's BSS, CitiBike. The study builds on our earlier work for Montreal (system level - Faghih-Imani et al., 2014) and Chicago (user level - Faghih-Imani and Eluru, 2015) BSS.

1.2. Earlier Research

The earlier research on BSS can be broadly classified along two dimensions: (1) system and (2) user. Under the <u>system perspective</u>, earlier quantitative studies employed actual bicycle usage data to capture the determinants of BSS usage (Rixey, 2013; Gebhart and Noland 2014; Faghih-Imani et al., 2014; Han et al. 2014; Rudloff and Lackner 2014; Wang et al., 2015; Faghih-Imani and Eluru, 2016a, Faghih-Imani and Eluru, 2016b, Faghih-Imani et al., 2017). In these studies, usage is usually characterized as arrivals (depositing bicycles) and departures (removal of bicycles). These studies examine the influence of number of BSS stations and stations' capacity, length of bicycle facilities, streets and major roads, presence of metro and bus stations, restaurants, businesses and universities, temperature and humidity, and time of day, day of the week and month on BSS usage.

The studies focussed on the <u>user perspective</u> contribute to the literature by studying user behavior in response to BSS. Studies found that flexibility offered by BSS (such as no requirement of bicycle ownership and one-way trips) as well as having a BSS station closer to home location significantly encouraged individuals to use the system (Bachand-Marleau et al., 2012; Fuller et al., 2011). Another study showed that allowing daily users (versus annual pass users) to use the system resulted in increased BSS usage on weekends and overall usage increase at a number of stations (Lathia et al., 2012). Several studies highlighted the importance of encouraging people to shift from car to BSS to reduce total vehicle kilometers traveled (Fishman et al., 2014). Moreover, studies showed that BSS's implementation in the city could motivate new segments of the society to cycle and thus increase the overall bicycling mode share (Buck et al., 2013). Schoner and Levinson (2014) modeled the station choice for the trip origin for Nice Ride Minnesota BSS using survey data to study how people use BSS and underscored the difference between preferences of workers and non-workers.

Although past studies have provided several useful insights, the BSS literature is still in its infancy and there are several dimensions that are unexplored in earlier research. Specifically, it is important to quantify the influence of bicycle infrastructure, built environment and land-use attributes on BSS usage while controlling for the meteorological and temporal characteristics. Further, along the user dimension, it is also important to examine BSS users' behaviour at a trip level to understand individual preferences. This paper documents research that analyzes BSS from both user and system perspectives to provide useful insights on how these new growing public systems influence the urban transportation. At the system level, the paper develops quantitative models to study station level usage defined as hourly arrivals and departures. As opposed to the simple linear regression model to study the dependent variables, a linear mixed model structure that allows for accommodating common unobserved factors that influence station level usage (arrivals and departures) across multiple observations is employed.

From a user perspective, this study examines bicyclists' destination choice preferences by employing multinomial logit models. Specifically, we analyze factors influencing the users' decision process in selecting a BSS station as a destination. For this purpose, a random utility-based multinomial logit model with random sampling is employed. There have been several location choice studies in traditional travel demand literature that adopt a random utility maximization approach for understanding destination/location preferences (Sivakumar and Bhat, 2007; Waddell et al., 2007; Chakour and Eluru, 2014; Faghih-Imani and Eluru, 2015). The current work adapted this approach to the BSS data.

For both analyses, we employ data from New York's CitiBike system. CitiBike is the largest and the fastest growing system in North America with increasing ridership that offers an excellent case study to understand BSS usage and flows. For the system level analysis, the current paper employs a week of data for every station significantly increasing the data employed compared to our earlier work on the Montreal system. Furthermore, the data from Montreal system did not have any information on membership status. Hence, the BSS demand modeled was grouped at a station level as a single usage metric. The CitiBike system data provides a more accurate representation of demand. For the user level, our earlier analysis explored the Chicago Divvy system - a substantially smaller system relative to the CitiBike system. The difference in urban form and non-motorized transportation preferences across Chicago and New York warrant a separate investigation of the influence of exogenous factors on destination choice.

The rest of the paper is organized as follows: Section 2 provides a description of the data employed in our analysis and approach to generate dependent and independent variable. In Section 3 and 4, methodology, model fit measures and results for system level and user level are described respectively. Finally, Section 5, concludes the paper.

2. *Data*

2.1. Data Source

New York's CitiBike system is the latest major public BSS around 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 (Figure 1). The system covers the city's major commercial business districts and some residential areas with an average daily ridership of 34,000 trips. New York City is the most populous city in the US and a host to millions of visitors every year. In 2013, the mode share of cycling in New York City increased to 1% from about 0.5% in 2007 (Kaufman et al., 2015). The city's dense and walkable urban form provides a good opportunity for the success of a well-designed BSS.

The data used in our research was obtained primarily from the CitiBike website (https://www.citibikenyc.com/system-data). The CitiBike website provides trip dataset for every month of operation since July 2013. The trip dataset includes information about origin and destination stations, start time and end time of trips, user types i.e. whether the user was a subscriber of the system with annual membership or a customer with a temporary pass, and the age and gender for members' trips. Additionally, the stations' capacity and coordinates as well as trip duration are also provided 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).

2.2. Dependent Variable Representation

2.2.1. *System level*

The main objective of the system level analysis exercise was to quantify the influence of various factors on arrival and departure rates at the BSS station using a general statistical modeling framework that other regions can adopt. The estimated models enable the prediction of changes to demand profiles (arrival and departure rates) in response to capacity reallocation or new station installation. Earlier studies showed that there are significant differences between the BSS usage patterns of annual members and customers with temporary passes (Lathia et al., 2012; Buck et al., 2013; Faghih-Imani and Eluru, 2015). So, in this paper, we distinguish between the trips made by annual members and daily customers instead of grouping them together.

The first step in data assembly for system-level analysis is sample formation to calculate the dependent variables for the analysis (arrivals and departures) from disaggregate trip data. Number of trips originated from and destined to one station are equal to the number of departures and arrivals for that station, respectively. Thus, we aggregated the number of trips originated from/destined to one station by different type of users at an hourly level to obtain hourly arrivals and departures by members and daily customers at a station level. Further, we normalized stations' arrivals and departures with station capacity to account for the influence of station capacity on demand. In our modeling efforts, we employed logarithm of the hourly normalized arrivals and departures as the dependent variable to ensure that model predictions are non-negative. We randomly selected 7 days from the month of September, 2013; i.e. the peak month of the usage in 2013 for each station to obtain a reasonable sample size for our analysis[1]. Thus, the final sample size consisted of 55,440 records (330 stations × 24 hours × 7 days). Four separate models were developed to examine the arrival and departure rates at every station by annual members and daily customers.

2.2.2. *User level*

The user level analysis examined BSS behavior at a trip level to analyze bicyclists' destination preferences. Specifically, we studied the decision process involved in identifying destination locations after picking up the bicycle at a BSS station. The analysis process considered that an individual who picks a bicycle at one of the stations makes destination station choice based on a host of attributes including individual's age and gender, time period of the day, and destination attributes such as distance from the origin station, points of interest, bicycle infrastructure, land use and built environment variables. The decision process was studied using a random utility maximization approach where individuals choose the destination that offers them the highest utility from the universal choice set of stations in the study region. The information will allow

---

[1] The reader would note that the 7 days are randomly selected for each station in a way that the final sample properly covers the entire month of September.

urban planners and BSS operators to enhance their understanding of decision maker preferences and enable them to re-orient the urban form to facilitate BSS usage and non-motorized usage in general. Additionally, the framework developed will allow us to identify BSS stations that have very high arrivals – thus allowing the BSS operators to optimally rebalance their vehicle fleet in the urban region.

For our analysis, we focus on the trips in the month of September. Again, we separated trips made by members and daily customers; about 86% of all the trips were made by members. The sample formation exercise involved a series of steps. First, trips with missing or inconsistent information were removed. Second, trips longer than 90 minutes in duration (only 0.5% of all the trips) were deleted considering that the trips longer than 90 minutes are not typical bicycle-sharing rides and could also be a result of misplacing the bicycle when returning it to the station. At the same time, trips that had the same origin and destination were also eliminated. For trips with the same origin and destination, it is possible that the bicycle was not functioning well and the users returned them to the origin station. Further, to obtain a reasonable sample size for model estimation, 20,000 trips were randomly selected for each user type. This sample size was adopted to maintain a reasonable data processing and model estimation related computational effort.

CitiBike system has 330 stations across the city. From each origin station, individuals have 329 other stations to choose to return the bicycle to. However, considering all the stations in the universal choice set will result in substantial computational burden. Hence, for the purpose of the modeling effort, for every destination choice record, a sample of 30 alternatives from the universal choice set including the chosen alternative was randomly selected. The process of random sampling does not affect the parameter estimates in multinomial logit models (see McFadden, 1987). The random sampling approach is consistent with the earlier research in destination choice modelling (for example see Pozsgay and Bhat, 2002; Scott et al., 2005; Scott and He, 2012; Faghih-Imani and Eluru, 2015). With the sampled choice set, information for the 30 stations was augmented with the individual trip records.

2.3. Independent Variable Generation

The independent variables considered in our analysis can be categorized into four groups: (1) weather, (2) temporal, (3) spatial variables and (4) trip attributes. It must be noted that trip attributes were only included in the destination choice models. Weather variables include hourly temperature, relative humidity, and the hourly weather condition represented as a dummy variable indicating whether or not it was raining. The temporal variables considered in the analysis aim to capture average time-of-day and day-of-the-week effects over and beyond other variable effects in the models. Considering the start time of the trips for departures and end time of the trips for arrivals, five time periods were created: AM (7:00-10:00), Midday (10:00-16:00), PM (16:00-20:00), Evening (20:00-24:00), and Night (0:00-7:00) to capture the time-of-day effect on usage. For the destination choice models, the same time periods were used considering the start time of the trips. The influence of weekend vs. weekday was also taken into account.

Several variables were considered under the spatial variables group. Specifically, population density was calculated at census block level and employment density at zip code level whereas other attributes were considered at a station buffer level. A 250 meter buffer around each station was found to be an appropriate walking distance considering the distances between CitiBike stations and the dense urban form of New York City; a typical New York City block is about 60

meter (Kaufman et al., 2015). The length of bicycle routes and streets in the 250 meter buffer around the stations were calculated in order to examine the impact of street network and cycling facilities. In addition, the number and capacity of CitiBike stations in the 250 meter buffer were computed to capture the effect of neighbouring stations. Also, the presence of subway and Path train stations in the 250 meter buffer were generated to examine the influence of public transit on BSS usage. The number of restaurants (including coffee shops and bars), and area of recreational parks in the buffer region were also considered as point-of-interest attributes near CitiBike stations.

Trip attributes considered in destination choice model include the street network distance between the origin and destination of every trip. This distance was computed using the shortest path between origin and destination stations to investigate the travel distance influence along with other attributes. While the actual trip might involve a different route, the shortest distance would be a surrogate indicator of the actual distance traveled. Moreover, for the users with annual membership, the gender and age information were available and were considered in our analysis. It must be mentioned that several exogenous variables such as gender, age or weather variables cannot be directly included within the destination choice model structure since these variables do not change across alternatives. Therefore, the interaction effects of such variables with distance variable were considered in our modelling effort to uncover relative sensitivities of different population segments to distance. To provide an illustration of the data compiled, a descriptive summary of the sample is provided in Table 1.

3. *System Level Analysis: Determinants of Bicycle-Sharing System Usage*

3.1. Linear Mixed Models

The most common methodology employed to study continuous dependent variables such as arrival and departure flows is the linear regression model. However, the traditional linear regression model is not appropriate to study data with multiple repeated observations. The arrivals and departures were observed at the same station at an hourly level for each station. Hence to recognize this, a multilevel linear model that explicitly recognizes the dependencies associated with the bicycle flow variable originating from the same CitiBike station was employed. Specifically, a linear mixed modeling approach that builds on the linear regression model while incorporating the influence of repeated observations from the same station was adopted. The linear mixed model collapses to a simple linear regression model in the absence of any station specific effects. A brief description of the linear mixed model is provided below (see Faghih-Imani et al., 2014 for a similar approach).

Let q = 1, 2, …, Q be an index to represent each station, d = 1, 2, …, D be an index to represent the various days on which data was collected and t = 1, 2, …, 24 be an index for hourly data collection period. The dependent variable (arrival or departure rate over station capacity) is modeled using a linear regression equation which, in its most general form, has the following structure:

$$y_{qdt} = \beta X_{qdt} + \varepsilon_{qdt}$$

where y_{qdt} is the normalized arrival or departure rate as the dependent variable, X is a L×1 column vector of attributes and the model coefficients, β, is an L×1 column vector. The random error term, ε, is assumed to be normally distributed across the dataset.

6

The error term may consist of three components of unobserved factors: a station-specific component, a day-specific component, and an hour-of-the-day component. Due to the substantial size of the data and the number of independent variables considered in this study, it was prohibitively burdensome, in terms of run time, to estimate the combined influence of the three components simultaneously. Thus, the analysis considered the station and the time-of-day to be related common unobserved effects. In this structure, the data can be visualized as 24 records for each Station-Day combination for a total of 2310 observations (330 stations × 7 days). Estimating a full covariance matrix (24 x 24) was computationally intensive while providing very little intuition. Hence, the covariance matrix (Ω) was parameterized. To be specific, for estimating a parsimonious specification, the correlation structure was assumed to be a first-order autoregressive moving average with three parameters σ, ρ, and φ as follows:

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \varphi\rho & \varphi\rho^2 & \cdots & \varphi\rho^{23} \\ \varphi\rho & 1 & \cdots & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \varphi\rho^{23} & \cdots & \cdots & \cdots & 1 \end{pmatrix}$$

The parameter σ represents the error variance of ε, φ represents the common correlation factor across time periods, and ρ represents the dampening parameter that reduces the correlation with time. The correlation parameters φ and ρ, if significant, highlight the impact of station specific effects on the dependent variables. The models were estimated in SPSS using the Restricted Maximum Likelihood Approach (REML) that is slightly different from maximum likelihood (ML) approach. The REML approach estimates the parameters by computing the likelihood function on a transformed dataset. The approach is commonly used for linear mixed models (Harville, 1977).

3.2. Model Fit Measures

In this paper, two model frameworks were estimated for arrivals and departures: (1) a linear regression model and (2) a linear mixed model. The final model selection was based on the restricted log-likelihood metrics. The model estimation process was guided by considerations of parsimony and intuitiveness. The two model frameworks were compared using the *log-likelihood ratio* (LR) test. For the arrivals model, the LR test statistic was significant at any reasonable level of significance (the LR test-statistic value for members' and daily customers' models were 22207 and 7150, respectively, significantly higher than the corresponding chi-square value for two additional degrees of freedom (φ and ρ)). Similarly, for the departures model, the LR test statistic was significant at any reasonable level of significance (the LR test-statistic value were 20088 and 7790 for members' and daily customers' models respectively). These model fit comparisons clearly highlight the suitability of the mixed modeling approach employed in this analysis for examining the determinants of CitiBike usage in New York City.

3.3. Results

This section discusses the results of linear mixed model estimation to understand the different effects of meteorological, spatial and temporal elements on the bicycle usage in the CitiBike BSS.

It must be noted that several specifications were considered but only the statistically significant results for arrival and departure rates are presented in Table 2.

3.3.1. Weather variables

As expected, there is a positive correlation between temperature and the arrival and departure rates for members and non-members. On the other hand, humidity and rainy weather variables have negative impacts on the arrival and departure rates. People are less likely to ride a bicycle in rainy or very humid time periods. Overall, weather attributes have similar impact on usage of BSS in annual members' and daily customers' models.

3.3.2. Temporal variables

Annual members tend to bicycle more on weekdays than weekends, as highlighted by the negative coefficient of the weekend variable. On the contrary, daily customers are more likely to use CitiBike system on weekends. This might be indicative of the recreational nature of trips made by daily customers. The interpretation of the time-of-day variables needs to be judiciously undertaken due to the presence of interaction effects with population and job density variables. Nevertheless, it can clearly be observed that the CitiBike system is more predominantly used during the Midday and PM period relative to other times of the day. Also, the results clearly indicate different temporal patterns of the CitiBike usage for annual members and daily customers.

3.3.3. Land use and built environment variables

In this section, the results corresponding to land use and built environment variables are explained. The bicycle flows and usage of the BSS increase when there are more bicycle facilities (bicycle lanes, bicycle paths, etc.) nearby a CitiBike station (in agreement with the findings of Buck and Buehler, 2012). CitiBike users, especially annual members, often combine their trip mode with the subway or train which is reflected in the positive impact of the presence of subway and path train stations near CitiBike stations in the results (similar results can be seen in Nair et al., 2013). In general, the number of restaurants in the vicinity of a CitiBike station increases the usage of that station (similar to the findings of Wang et al., 2015). The park variable yielded interesting results. For the daily customers, there is an increase in usage when the CitiBike stations are near parks with a more pronounced effect on weekends. On the other hand, for annual members, the parks area variable has a negative impact on weekdays but on weekends there is a positive effect on members' arrivals and departures. As indicated earlier, the effect of population and job density were incorporated in the models at the census track and zip code level, respectively. As expected, the population and job density variables have a positive impact on arrivals and departures for both user types (see Rixey, 2013; Wang et al., 2015, for similar results). The interactions of these two variables with AM and PM provide evidence in support of the usage of CitiBike system for daily work commute trips. Moreover, it can be observed that the daily customers' usage is less sensitive to population and job density variables than the usage of annual members.

4. User Level Analysis: Examination of BSS Users' Destination Choice Preferences

4.1. Multinomial Logit Model

The use of Multinomial Logit Model (MNL) is common to study location choice in transportation and related literature (Rashidi et al., 2012; Zolfaghari et al., 2013; Chakour and Eluru, 2014; Faghih-Imani and Eluru, 2015). A brief description of the MNL model employed is provided below.

Let $s = 1, 2, ..., 30$ be an index to represent each station, $q = 1, 2, ..., Q$ be an index to represent the BSS users. Then, the random utility formulation takes the following form:

$$u_{qs} = \beta' X_{qs} + \varepsilon_{qs} \tag{1}$$

Where u_{qs} is the utility obtained by user q by selecting station s from the choice set of 30 stations. X_{qs} is the vector of attributes and β is the model coefficients to be estimated. The random error term, ε_{qs}, is a disturbance term for individual and station combination and assumed to be independent and identically Gumbel-distributed across the dataset. The BSS user q will choose the station as destination that offers the highest utility. With this notation, the probability expression takes the typical multinomial logit form given by:

$$P_{qs} = \frac{\exp(\beta' X_{qs})}{\sum_{s=1}^{30} \exp(\beta' X_{qs})} \tag{2}$$

The log-likelihood function can be defined as:

$$L = \sum_s \sum_q \ln(P_{qs})^{d_{qs}} \tag{3}$$

where $d_{qr} = 1$ if individual q chose station r and 0 otherwise.

By maximizing this log-likelihood function, the model parameters β are estimated. The maximum likelihood model estimation is programmed in GAUSS matrix programming language.

4.2. Results

This section discusses the results of multinomial logit model estimation to understand the different factors influencing users' choice of destination in the New York City's CitiBike bicycle-sharing system. The final Log-likelihood values of the station destination choice multinomial logit model for the annual member and daily customer samples were -59,461.6 and -59,767.3, respectively. The corresponding values for the equal probability model were -68023.9. The *log-likelihood ratio* (LR) test-statistic of comparison between the final models and the equal probability models for the two user types were 17,125.6 and 16,695.2, respectively. These LR test statistics are significantly higher than the critical chi-square values corresponding to 15 and 12 additional degrees of freedom for the annual member and daily customer, respectively. The improvement in the data fit clearly illustrates the superiority of the MNL based destination choice models. The model specification process was guided by intuition and parsimony considerations. The statistically significant results for members' and daily customers' destination choice models are presented in Table 3. As expected, there were distinct impacts of several contributing factors in the decision making of customers and members towards destination station choice.

4.2.1. Land-use and built environment characteristics

In this section, we discuss the results for land-use and built environment variables. The positive coefficients for station capacity variable in members' and daily customers' models demonstrate that stations with higher capacity are more likely to be chosen as they are likely to have more available docking stations. Moreover, it is possible that people tend to remember larger stations. Daily customers tend to choose stations with longer bicycle paths nearby as highlighted by the positive coefficient of the bicycle facility variable in daily customers' model. However, the variable has no significant effect for members. It is possible that daily customers are less familiar with the city street network and are generally more cautious; thus prefer using city's bicycle routes. On the contrary, the length of rails variable has negative impact on the propensity of choosing a station in members' model; it is expected as railway tracks typically act as barriers to pedestrian and bicyclist movements.

The CitiBike stations near subway system tend to be chosen as destination by members indicating the last mile connectivity offered by BSS for public transit users. However, the results were opposite for daily customers as shown by the negative coefficient of subway station and path train station variables. This shows that for regular members, BSS is likely to complement existing public transit services whereas for daily customers BSS serves as a substitute for existing public transit services. Since the purpose of daily customers' trips are more likely to be for recreational activities, the CitiBike stations in the vicinity of parks are also more likely to be chosen by daily customers as highlighted by area of parks in the buffer variable. The positive impact of park variable is higher in weekends. Interestingly, the parks variable for annual members has contrasting effect for trips during weekdays and weekends. During weekdays, bikers are less likely to choose a destination near parks while during weekends, CitiBike stations near parks are more likely to be chosen. As expected, annual members are more inclined towards CitiBike stations with higher number of restaurants in their vicinity while the corresponding effect is negative for daily customers.

For members, population density variable has positive impact on likelihood of choosing a CitiBike station (except during the AM time period) while it has negative impact for daily customers during all time periods. CitiBike stations with lower population density are more likely to be chosen during AM period for short-term users as well as annual members. This is understandable given that trips in AM period mostly originate from home zones and are destined to work areas. Job density variable has negative impact in models for daily customers and annual members during all time periods except in the mornings. The coefficients of population and job density variables in AM and PM periods clearly demonstrate the use of CitiBike system for daily commute to and from work in the mornings and evenings, particularly by annual members. Further, the results show that there is clear distinction in the use of CitiBike system by daily customers and annual members. This provides further support to the hypothesis that separate behavior models are appropriate for daily customers and annual members.

4.2.2. Trip level attributes

The most important variable in destination station decision making process in a BSS is expected to be the distance of the trip between origin and destination. For members, the analysis examined the distance of trips as well as age, gender and weather attribute effects while for daily customers we lack user-specific disaggregate information in dataset. As expected, the network distance

10

variables have negative impact on likelihood of choosing a station as destination for both annual members and short-term users. Moreover, gender, age and weather effects were also considered in members' model estimation. It is important to note that since the user and weather attribute remains the same for all the destination station alternatives, these effects were captured by interacting these variables with distance variables. Interestingly, age was not found to influence distance sensitivity. However, this is not completely unintuitive given that annual members self-select themselves from the pool of active bicyclists. The gender impact, on the other hand, offered interesting results. The results show that female members are more likely to have longer trips. This might be due the fact that in New York CitiBike system, only about 24.7% of members are female. It is possible that women who join are actually regular bicyclists and are more likely to be fit and pursue longer trips. However, this result might be different in other BSS especially in bicycle-friendlier cities. Annual members are less likely to make longer trips at the time of high humidity or rainy weather conditions. However, the weather attributes do not have significant impact on daily customers' destination choice as they probably do not buy the temporary pass at the time of adverse weather conditions.

4.3. *Policy Analysis*

In order to better understand the magnitude of the effects of variables on destination choice, Figure 2 illustrates the utility function trade-offs between origin and destination network distance and other attributes such as the length of bicycle facilities in the buffer, the number of restaurants and destination station capacity. As is indicated by the model estimates, the utility decreases when the trips' distance between the origin and destination increases; while it increases with increase in the length of bicycle facilities near a station, number of restaurants and station capacity. It can be observed from the three-dimensional relationship that the negative impact of distance is compensated to some degree by the positive impact of bicycle facilities, number of restaurants and station capacity. This is illustrated by how the utility for various distances remains the same with an appropriate increase in the other two attributes. For example, moving a destination station from 1 km to 2 km farther from origin without changing other variables would result in about 0.49 unit reduction in utility for members and non-members. Now, in order to maintain the attraction of that station constant (i.e. keep the utility constant), adding to existing bicycle routes or adding more capacity can compensate for that reduction in utility. To offset the utility reduction caused by increased distance of 1 km for daily customers, an increase in bicycle facility length by about 11.8 km and, for members an increase in number of restaurants by about 950 units is necessary. In terms of capacity, the 1 km increase in distance can be offset by increasing the capacity by 33 and 25 for members and non-members, respectively. Hence, one could argue that adding to bicycle capacity is an easier proposition. Of course, if the changes were made simultaneously only a 2 km increase in bicycle route length in conjunction with a capacity increase of 21 for daily customers can offset the 1 km increase in distance. Overall, the analysis showed the significant distance effects on users' decision-making process; further investigation of the distance impact can be done by segmenting the distance effect for different population segments. Nevertheless, the negative impact of trip distance variable can be marginally offset by the positive impact of other factors. The figures and the subsequent analysis illustrate the applicability of the proposed model for system operators for reallocating capacity or installing new capacity while regional planners can adopt the model to enhance land-use to encourage shared bicycling usage.

5. *Conclusions*

Recently, bicycle sharing systems (BSS) have become more prevalent. With the fast-growing installation of BSS infrastructure across the world there is substantial interest in understanding how these systems impact the urban transportation system. This paper evaluated the impact of BSS by examining BSS usage along both system and user dimensions by using revealed usage data from New York City's CitiBike system. The analyses were adapted from our earlier research for Montreal system level analysis (Faghih-Imani et al., 2014) and Chicago user level analysis (Faghih-Imani and Eluru, 2015), respectively. For the system level analysis, the current paper employs a larger and more nuanced data sample providing a more accurate representation of demand. For the user level, New York City with its unique urban form and non-motorized transportation preferences warrants a separate investigation of the influence of exogenous factors on CitiBike destination choice behavior.

Analyzing CitiBike system by a multilevel model estimation approach provided intuitive results for both arrival and departure rates. It was observed that people were more likely to use a BSS under good weather conditions. While during the weekends the bicycle usage reduces, the arrival and departure rates of stations near parks were increased. The bicycle flows were expected to increase when there are subway or path train stations near CitbiBike stations. The number of restaurants in the vicinity of a station significantly influenced the arrival and departure rates of the CitiBike station. Population and job density variables positively affected the bicycle flows while the interaction effects of these variables provided evidence that CitiBike system is being primarily used for daily work commute.

Examining CitiBike system users' destination preferences employing a Multinomial Logit Model provided several useful insights for both members and daily customers. It was observed that daily customers tend to choose stations with longer bicycle paths nearby. In terms of the destination station, the stations with higher capacity were more likely to be chosen. The network distance between origin and destination station had a negative impact on the likelihood of choosing a station as the destination for CitiBike users. During AM period, stations with higher job density and stations with lower population density were more likely to be chosen. CitiBike stations in the vicinity of parks were more likely to be chosen by daily customers while less likely to be selected by annual members during weekdays. Interestingly, the parks variable for annual members had a positive effect for trips during weekends. The utility function trade-off analysis showed that the negative impact of distance is compensated to some degree by the positive impact of bicycle facilities and station capacity.

The results of our analysis of CitiBike system from users were in agreement with the results of our analysis from the system perspective. Both results demonstrated that there are clear spatial and temporal differences between annual members and daily customers' usage patterns of the system. The user level analysis also indicates that the CitiBike system is mainly used for work commute purposes by annual members. The weather attributes had a negative impact on the usage of Citibike system. Overall, our analysis provides a framework and useful findings for Cities that are planning to install a new bicycle sharing-system or to expand an existing system.

To be sure, the study is not without limitations. In this paper, data from the month of September was considered for analysis. For user level analysis, a more representative sample covering the entire year would be appropriate to understand seasonal variations in BSS usage. For user level analysis, choice of destination might not be very strongly influenced by season. Finally, the linear mixed model correlation structure considered in our paper is one of several possibilities. In future research efforts, other more flexible structures can be explored to improve data fit and prediction capabilities.

References

Bachand-Marleau, J., B. Lee, and A. El-Geneidy, 2012. 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, pp. 66-71.

Buck, D., Buehler, R., 2012. Bike lanes and other determinants of capital bikeshare trips. Paper presented at the 91st Transportation Research Board Annual Meeting 2012, Washington, DC.

Buck, D., D. Buehler, P. Happ, B. Rawls, P. Chung, and N. Borecki, 2013. 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, pp. 112–119.

Chakour V. and N. Eluru, 2014. Analyzing Commuter Train User Behavior: A Decision Framework for Access Mode and Station Choice, *Transportation*, Vol. 41 (1), pp. 211-228.

City of Chicago Data portal, https://data.cityofchicago.org.

Davis, B., T. Dutzik and P. Baxandall, 2012. Transportation and the new generation: Why young people are driving less and what it means for transportation policy, Frontier Group.

Dutzik, T. and P. Baxandall, 2013. A new direction: Our changing relationship with driving and the implications for America's future. U.S. PIRG Education Fund and Frontier Group.

Faghih-Imani A., N. Eluru, A. El-Geneidy, M. Rabbat, and U. Haq, 2014. How does land-use and urban form impact bicycle flows: Evidence from the bicycle-sharing system (BIXI) in Montreal, *Journal of Transport Geography*, Vol. 41, pp. 306-314.

Faghih-Imani A., and N. Eluru, 2015. Analyzing Bicycle Sharing System User Destination Choice Preferences: An Investigation of Chicago's Divvy System, Journal of Transport Geography, Vol. 44, pp. 53-64.

Faghih-Imani, A., and N. Eluru, 2016a. 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, pp. 218-227.

Faghih-Imani, A., and N. Eluru, 2016b. Determining the Role of Bicycle Sharing System Infrastructure Installation Decision on Usage: Case Study of Montreal BIXI System, *Transportation Research Part A: Policy and Practice*, Vol. 94, pp. 685-698.

Faghih-Imani, A., R. Hampshire, M. Lavanya and N. Eluru, 2017. An Empirical Analysis of Bike Sharing Usage and Rebalancing: Evidence from Barcelona and Seville, *Transportation Research Part A: Policy and Practice*, Vol. 97, pp. 177-191.

Fishman, E., S. Washington and N. Haworth, 2014. 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, pp. 13-20.

Gebhart, K. and R. Noland, 2014. The Impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, Vol. 41, pp. 1205-1225.

Han, Y., E. Côme, and L. Oukhellou, 2014. Towards Bicycle Demand Prediction of Large-Scale Bicycle Sharing System. In Transportation Research Board 93rd Annual Meeting, No. 14-2637.

Harville, D.A., 1977. Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association*, Vol. 72, pp. 320-338.

Jäppinen, S., T. Toivonen, and M. Salonen, 2013. Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography*, Vol. 43, pp. 13-24.

Lathia, N., S., Ahmed, and L., Capra. Measuring the impact of opening the London shared bicycle scheme to casual users. *Transportation Research Part C: Emerging Technologies*, Vol. 22, 2012, pp. 88-102.

McFadden, D. Modeling the Choice of Residential Location. In Spatial Interaction Theory and Planning Models (A. Karlqvist et al., eds.), North Holland Publishers, Amsterdam, Netherlands,1978.

Nair, R., Miller-Hooks, E., Hampshire, R., and A., Busic, 2013. Large-Scale Vehicle Sharing Systems: Analysis of Velib. *International Journal of Sustainable Transportation*, Vol. 7, pp. 85-106.

NHTS, 2009. U.S. Department of Transportation, Federal Highway Administration, National Household Travel Survey 2009. URL: http://nhts.ornl.gov.

Pozsgay, M.A., and C.R. Bhat. Destination choice modeling for home-based recreational trips: analysis and implications for land use, transportation, and air quality planning. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1777, 2001, pp. 47-54.

Rashidi, T. H., J. Auld, and A. K. Mohammadian. A behavioral housing search model: Two-stage hazard-based and multinomial logit approach to choice-set formation and location selection. *Transportation Research Part A: Policy and Practice*, Vol. 46, 2012, pp. 1097-1107.

Rietveld, P. and V. Daniel, 2004. Determinants of bicycle use: do municipal policies matter? *Transportation Research Part A: Policy and Practice* Vol. 38, pp. 531-550.

Rixey, R., 2013. 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, pp. 46-55.

Rudloff, C. and Lackner, B., 2014. Modeling Demand for BikeSharing Systems: Neighboring Stations as Source for Demand and Reason for Structural Breaks. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2430, pp. 1-11.

Schoner, J. E. and D. M. Levinson, 2014. Which Station? Access Trips and Bike Share Route Choice. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, DC.

Scott, D.M., and S.Y. He. Modeling constrained destination choice for shopping: a GIS-based, time-geographic approach. *Journal of Transport Geography*, Vol. 23, 2012, pp. 60-71.

Scott, D.M., P.A. Coomes, and A.I. Izyumov. The Location Choice of Employment-based Immigrants among US Metro Areas. *Journal of Regional Science*, Vol. 45, 2005, pp. 113-145.

Shaheen, S., S. Guzman and H. Zhang, 2010. Bikesharing in Europe, the Americas, and Asia Past, Present, and Future. *Transportation Research Record: Journal of the Transportation Research Board,* No. 2143, pp. 159-167.

Sivakumar, A., and C.R. Bhat, 2007. A Comprehensive, Unified, Framework for Analyzing Spatial Location Choice, *Transportation Research Record: Journal of the Transportation Research Board,* No. 2003, pp. 103-111.

van Lierop, D., M. Grimsrud, and A. El-Geneidy, 2015. Breaking into bicycle theft: Insights from Montreal, Canada. *International Journal of Sustainable Transportation*, Vol. 9, pp. 490-501.

Waddell, P., C.R. Bhat, N. Eluru, L. Wang and R.M. Pendyala, 2007. Modeling the Interdependence in Household Residence and Workplace Choices, *Transportation Research Record: Journal of the Transportation Research Board,* No. 2003, pp. 84-92.

Wang, X., G. Lindsey, J. Schoner and A. Harrison, 2015. 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, pp 1-9.

Zolfaghari, A., A. Sivakumar, and J. Polak, 2013. Simplified probabilistic choice set formation models in a residential location choice context. *Journal of Choice Modelling*, Vol. 9, pp. 3-13.
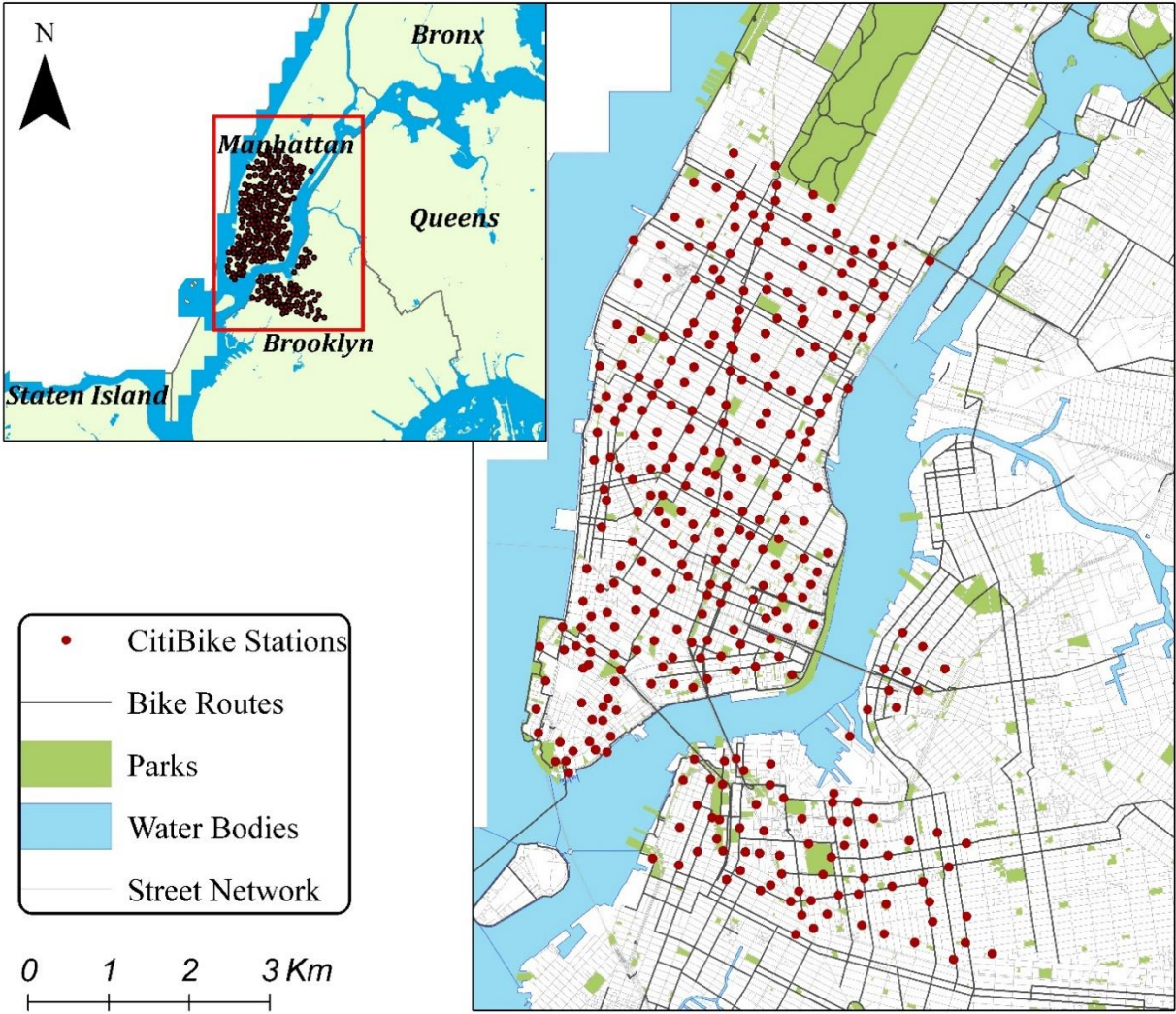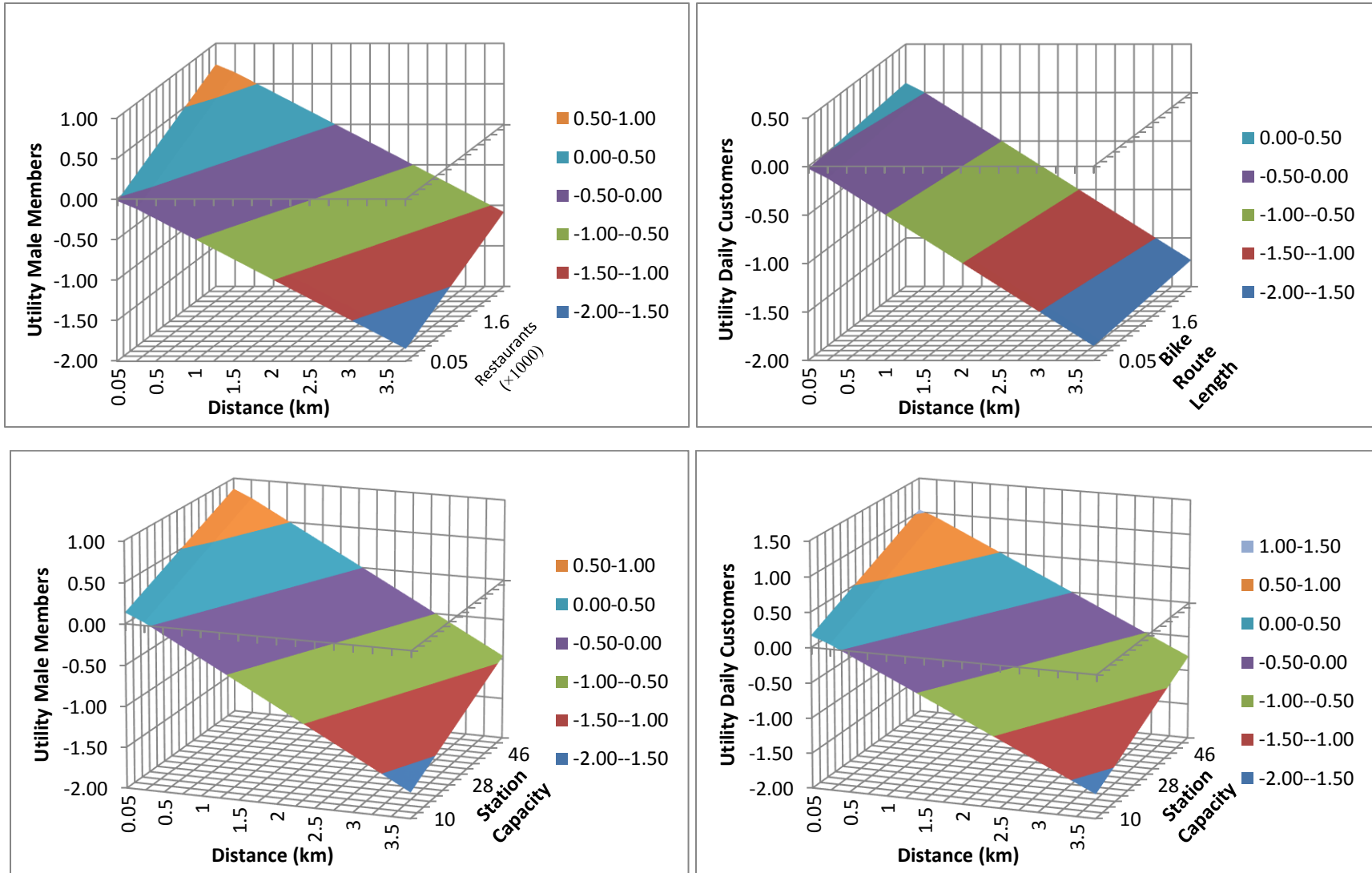
Figure 1 CitiBike stations in New York City

Figure 2: The Variation of Utility as a function of Distance, Bike Route Length, Number of Restaurants and Station Capacity

Table 1 Descriptive Summary of CitiBike Sample Characteristics

Continuous Variables	Min	Max	Mean	Std. Deviation
Hourly Arrivals (Annual Members)	0	83.00	3.74	5.26
Hourly Arrivals (Daily Customers)	0	39.00	0.61	1.50
Hourly Departures (Annual Members)	0	102.00	3.74	5.36
Hourly Departures (Daily Customers)	0	34.00	0.61	1.49
Temperature (°C)	8.3	34.4	19.64	4.82
Relative Humidity (%)	27.0	94.2	60.96	16.01
Length of Bicycle Facility in 250m Buffer (m)	0	1022.7	314.95	178.82
Area of Parks in 250m Buffer (m^2)	0	95209.9	10181.87	15169.65
Number of Restaurants in 250m Buffer	0	545	54.35	92.21
Number of CitiBike stations in 250m Buffer	0	4.00	1.24	1.01
Capacity of CitiBike stations in 250m Buffer	0	169.00	43.93	38.93
Station Capacity	3.00	67.00	34.35	10.76
Pop Density (people per m^2 ×1000)	0.01	67.20	24.87	14.68
Job Density (jobs per m^2 ×1000)	0	432.52	55.83	53.83
Trip Distance (km)	0.05	12.68	3.97	2.30
Trip Duration (min)	1.02	89.57	12.79	8.87
Members Age	16.00	96.00	37.33	10.95

Categorical Variables	Percentage
Rainy Weather	2.6
Weekends	30.0
Subway Station in 250m Buffer	49.7
Path Train Station in 250m Buffer	4.2
Female Members	24.7

Table 2 Usage Models Estimation Results

Parameter	Members				Daily Customers			
	Arrival Rate		Departure Rate		Arrival Rate		Departure Rate	
	β	t-stat	β	t-stat	β	t-stat	β	t-stat
Intercept	-4.1125	-64.417	-4.0918	-65.571	-4.4767	-138.092	-4.5210	-135.102
Weather Variables								
Temperature	0.0152	8.166	0.0093	5.126	0.0135	13.553	0.0117	11.596
Relative Humidity	-0.0062	-12.596	-0.0055	-11.226	-0.0026	-9.037	-0.0015	-5.316
Rainy Weather	-0.2289	-9.588	-0.2537	-10.392	-0.1006	-5.743	-0.1423	-8.092
Time Variables								
Weekend	-0.1363	-5.352	-0.1304	-5.286	0.1981	15.367	0.1969	15.133
AM	1.0530	39.490	1.1937	49.343	0.1491	15.867	0.1775	18.710
Midday	1.1489	71.426	1.1717	72.701	0.4697	50.894	0.5010	53.126
PM	1.3321	45.598	1.3098	58.049	0.4706	27.388	0.4897	25.547
Evening	0.7252	36.562	0.6950	35.133	0.1503	13.178	0.1425	12.287
Land use and Built Environment Variables								
Length of Bicycle Facility in 250m Buffer	0.0474	2.833	0.0701	4.318	0.3246	3.834	0.2343	2.738
Length of Rail in 250m Buffer	-	-	-0.0329	-2.160	-	-	-	-
Subway Station in 250m Buffer	0.1130	5.770	0.1574	8.195	-	-	0.0183	1.827
Path Train Station in 250m Buffer	0.3386	7.098	0.3394	7.341	0.0545	2.269	0.0659	2.705
Area of Parks in 250m Buffer	-3.4696	-4.517	-3.4465	-4.628	1.8468	4.750	1.3923	3.548
Area of Parks in 250m Buffer *Weekend	2.8711	2.080	2.5926	1.939	2.4714	3.538	3.0273	4.294
Number of Restaurants in 250m Buffer	0.7966	7.447	0.8604	8.302	0.1058	1.953	0.1380	2.524
Population Density	13.6325	19.914	12.2601	18.549	0.7519	2.186	1.1313	3.214
Population Density* AM	-3.6563	-4.748	4.5982	5.845	-	-	-	-
Population Density* PM	2.5911	3.485	-	-	1.7109	3.400	-0.8667	-1.703
Job Density	0.7037	3.683	0.9388	5.093	0.3656	3.848	0.3289	3.349
Job Density * AM	4.0054	19.076	-	-	-	-	-	-
Job Density * PM	0.5158	2.545	2.2434	10.858	-	-	0.5137	3.702
ARMA Correlation Parameters								
σ	0.9093	102.289	0.9121	106.195	0.3796	141.575	.3831	140.154
ρ	0.8272	214.291	0.8172	189.676	0.8370	162.812	.8251	156.126
φ	0.5541	116.164	0.5295	108.721	0.2733	56.614	.2911	59.916

Table 3: Destination Choice Model Estimation Results

Parameter	Members		Daily Customers	
	Coefficient	t-statistic	Coefficient	t-statistic
Built Environment Variables				
Destination Station Capacity	0.0162	21.103	0.0193	24.984
Length of Bicycle Facility in 250m Buffer	-	-	0.0416	3.235
Length of Rails in 250m Buffer	-0.0931	-5.317	-	-
Presence of Subway Station in 250m Buffer	0.0291	1.811	-0.1265	-7.857
Presence of Path Train Station in 250m Buffer	-	-	-0.0636	-1.877
Area of Parks in 250m Buffer	-1.1213	-1.732	11.8365	19.91
Area of Parks in 250m Buffer *Weekend	3.1184	2.496	1.3668	1.67
Number of Restaurants in 250m Buffer	0.5102	6.896	-0.2127	-2.467
Population Density	5.0537	8.666	-7.6587	-11.304
Population Density * AM	-13.9936	-10.854	-4.4922	-2.047
Population Density * PM	-	-	4.4869	4.131
Job Density	-0.9931	-4.096	-1.0731	-6.085
Job Density * AM	4.352	11.435	3.4029	5.882
Job Density * PM	-1.8706	-4.997	-	-
Trip Attributes				
Distance	-0.4931	-26.965	-0.4941	-104.279
Distance*Female	0.0447	3.885	-	-
Distance*Humidity	-0.0754	-2.482	-	-
Distance*Rainy	-0.1323	-1.999	-	-