

**A FRAMEWORK FOR ESTIMATING BIKESHARE ORIGIN DESTINATION FLOWS  
USING A MULTIPLE DISCRETE CONTINUOUS SYSTEM**

**Bibhas Kumar Dey**

Doctoral Student

Department of Civil, Environmental and Construction Engineering  
University of Central Florida

Tel: 1-407-823-4815; Fax: 1-407-823-3315

Email: [bibhas.ce@knights.ucf.edu](mailto:bibhas.ce@knights.ucf.edu)

**Sabreena Anowar\***

Research Fellow

Sustainable Urban Mobility Research Laboratory

Engineering Systems and Design

Singapore University of Technology and Design (SUTD)

Tel: 1-407-823-4815; Fax: 1-407-823-3315

Email: [sabreena\\_anowar@sutd.edu.sg](mailto:sabreena_anowar@sutd.edu.sg)/[sabreena.anowar@ucf.edu](mailto:sabreena.anowar@ucf.edu)

**Naveen Eluru**

Professor

Department of Civil, Environmental and Construction Engineering  
University of Central Florida

Tel: 1-407-823-4815; Fax: 1-407-823-3315

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

\*Corresponding author

## **ABSTRACT**

Given the burgeoning growth in bikeshare system installations and their growing adoption for trip making, it is important to develop modeling frameworks to understand bikeshare demand flows in the system. The current study examines two choice dimensions for capturing the system level bikeshare system demand: (1) total station level demand and (2) distribution of bike flows from an origin station across the network. A linear mixed model is used to estimate the first choice and Multiple Discrete Continuous Extreme Value (MDCEV) model is used to analyze the latter. The data is drawn from the New York City bikeshare system (CitiBike) for six months (January through June, 2017). For our analysis, we examine demand and distribution patterns on a weekly basis controlling for a host of independent variables (trip, socio-demographics, bicycle infrastructure, land use and built environment, temporal and weather). Model validation exercise results revealed that the proposed model performs well for low demand destinations. A policy exercise evaluating destination choice behavior demonstrated how the impact of distance is compensated by additional bicycling infrastructure in the farther locations. The results from the study help bikesharing system planners and operators to better evaluate and improve bikeshare systems.

*Keywords:* Bikesharing system, Station level demand, Flow distributions, Multiple alternatives, Linear mixed model, Multiple Discrete Continuous Extreme Value (MDCEV) model

## 1. INTRODUCTION

Transportation field is undergoing a transformative change in response to several technological innovations resulting in the emergence and popularity of shared mobility systems such as bikesharing (such as CitiBike in New York City), carsharing (such as Zipcar or Car2Go), ride sourcing (such as Uber and Lyft), and ride-splitting (such as dynamic carpooling in urban regions). In addition to these sharing modes, there are other newly emerging transportation modes that are expected to penetrate the existing transportation fleet in the near future such as flying car (Eker et al., 2019, 2020), non-motorized modes e.g. electric bicycles and electric scooters (Wolf and Seebauer, 2014; Seebauer, 2015), dockless bikeshare (Peters and MacKenzie, 2019). Recent Transit Cooperative Research Program (TCRP) report (Feigon and Murphy, 2016), highlighted that adoption and usage of these tech-based alternative forms of transportation present an unprecedented opportunity to address the existing mobility shortcomings in urban regions. In fact, public transit and transportation planning agencies can enhance mobility and accessibility in a region by incorporating these shared transportation alternatives within their planning frameworks. Among the various shared forms of transport, bikesharing is a sustainable and affordable option (particularly in urban core regions) that could be an effective and promising solution to the first/last mile problem (Jäppinen et al., 2013). In our research, we focus our attention on developing a research framework to contribute to our understanding of bikeshare origin-destination (OD) flows.

About 1000 cities around the world have a bikeshare system in operation or in consideration for development (Meddin and DeMaio, 2016). As reported by Richter (2018), the number of public use bicycles in the world have nearly quadrupled between 2013 and 2016. Further, a recent National Association of City Transportation Officials (NACTO) report highlighted that of the 88 million trips made by bikeshare users in the US between 2010-2016, 28 million were trips in 2016 only (Dey et al., 2018). Given the burgeoning growth in bikeshare system installations and their growing adoption for trip making, it is important to develop modeling frameworks to understand bikeshare demand flows in the system. An important mechanism for enhancing system adoption and usage is the development of current performance metrics (see Fishman et al., 2013). As bikesharing is an emerging transportation mode, the current approaches being employed for analyzing system usage and performance measure are still in their infancy. In this study, we propose an enhanced framework to estimate usage dimensions of bikesharing at a system level.

To be sure, several earlier research efforts have explored approaches to model system level usage (Faghih-Imani and Eluru, 2015; Faghih-Imani et al., 2014; Rixey, 2013; Zhao et al., 2014). These research studies examine the impact of bicycling infrastructure, land use and built environment, public transportation infrastructure, temporal and meteorological attributes on bikeshare system usage (defined as station level arrivals and departures). These models can be viewed as analogous to the trip generation (production and attraction) models in the traditional trip-based modeling approach. While these models provide important insights on variables affecting bikeshare usage, they do not provide any information on the system level flows between the stations. To elaborate, the approaches provide trip end information without the trip distribution relationship. To address this shortcoming, recent research has developed destination choice models at an individual trip level (El-Assi et al., 2017; Faghih-Imani and Eluru, 2015, 2020). In these studies, for every individual trip the choice of destination given the origin station is analyzed using a random utility based approach. The models developed at an individual trip level can be employed to obtain aggregate estimates of trip distribution (analogous to the gravity model). However, such an aggregation approach is purely a statistical construct and lacks behavioral support.

In the current study, we remedy this drawback, by developing a model framework for bikeshare system usage as well as origin destination flows. Towards this end, we characterize

system demand as origin level demand (number of trips) and allocate these trips to various destination stations (number of trips from an origin to destination) in the system. For the first variable, a linear mixed model is developed while the second variable is analyzed using a multiple discrete continuous model system that implicitly recognizes that the total arrivals across (destination) stations should add up to the total number of trips leaving the origin stations. The proposed framework is implemented for the New York City bikeshare system (CitiBike). The data drawn for the exercise includes bikeshare trips from January 2017 through June 2017 for the CitiBike system.

The remainder of the paper is organized as follows. Section 2 provides a summary of the earlier studies on bikeshare and positions the current study. Data source and descriptive analysis together with econometric framework are presented in Section 3. Section 4 presents the model estimation results followed by model validation and policy analysis results. Finally, Section 7 concludes the paper.

## 2. EARLIER STUDIES

The recent growth of bikeshare systems around the world has resulted in a number of research efforts examining different aspects of bikeshare systems. These research efforts can be broadly categorized into two groups. The first group of studies is focused on understanding user behavior (such as reasons for adopting bikeshare) and satisfaction using online surveys or questionnaires (see for example Bachand-Marleau et al., 2012; Buck et al., 2013; Fishman et al., 2014; Fuller et al., 2011; Schoner and Levinson, 2013; Barbour et al., 2019; Fishman, 2016; Lu et al., 2018; Nikitas, 2018; Pal and Zhang, 2017; Caggiani et al., 2019; de Chardon, 2019; Nath and Rambha, 2019; Choi and Choi, 2020). These studies aid in formulating policies for promoting the bicycle mode as well as for attracting higher usage of the bikeshare systems. The second group is comprised of studies conducting quantitative analysis using bikeshare usage data. These studies attempt to understand user trip patterns and disentangle different factors that affect bikeshare demand. Given the focus of our current study, we restrict ourselves to the discussion of the second group of studies only; concentrating on the major research dimensions explored, methodological approaches employed, and major research findings from these studies.

The most common research dimensions explored in the previous studies include (a) system demand characterized as arrivals and departures from bikeshare stations (Caulfield et al., 2017; Faghih-Imani and Eluru, 2016a, b; Faghih-Imani et al., 2014, 2017a; Gebhart and Noland, 2014; Hyland et al., 2018; Noland et al., 2016; Rixey, 2013; Rudloff and Lackner, 2014; Wang et al., 2015; Yufei et al., 2014; Zhang et al., 2017; Kabra et al., 2019; Wang et al., 2020; Caggiani et al., 2020), (b) rebalancing demand (relocating bikes from overcrowded stations to those with shortage of bikes) (Bouveyron et al., 2015; Faghih-Imani et al., 2017b; Forma et al., 2015; Fricker and Gast, 2016; Nair et al., 2013; Pfrommer et al., 2014; Raviv et al., 2013; Vogel and Mattfeld, 2011; Pal and Zhang, 2017; Dell et al., 2016), and (c) destination station choice preferences of bikeshare users (El-Assi et al., 2017; Faghih-Imani and Eluru, 2015, 2020). The bikeshare systems analyzed are spread across a multitude of urban regions in different continents including New York (CitiBike), Montreal (BIXI), Paris (Velib), London (Santander Cycle), Chicago (Divvy), Hangzhou (Hangzhou Public Bicycle), Beijing (Beijing Public Bicycle), Zhongshan (Zhongshan Public Bike System), Melbourne (Melbourne Bikeshare), and Brisbane (CityCycle).

The data used in the analyses were either directly available in the bikeshare system provider website or were downloaded using automated scripts from the website. On the methodological front, the most commonly employed analytical approaches include linear regression (LR), mixed linear regression, panel ordered logit model, negative binomial count model, multinomial logit (MNL) model, mixed multinomial logit model, finite mixture MNL

model, and time series models and their variants (Buck et al., 2013; El-Assi et al., 2017; Faghih-Imani and Eluru, 2015; Faghih-Imani et al., 2014; Gebhart and Noland, 2014; Rixey, 2013; Rudloff and Lackner, 2014; Wang et al., 2015; Zhao et al., 2014). In addition to the conventional statistical methods, some studies have used machine learning methods and visual analytics to examine bikeshare demand, station usage, and other aspects (Hyland et al., 2018; Oliveira et al., 2016; Giot & Cherrier, 2014; Ashqar et al., 2017)<sup>1</sup>. The findings from the station demand studies suggest that bikeshare system usage, at a station level, is primarily influenced by bikeshare infrastructure (such as number of stations and station capacity), bicycling infrastructure (such as presence of bike lanes), land use and built environment (such as population density, job density, and points of interest), public transportation infrastructure (presence of bus/metro stops), and temporal and meteorological attributes (such as precipitation and temperature) (El-Assi et al., 2017; Faghih-Imani and Eluru, 2015, 2016a, b; Faghih-Imani et al., 2014; Gebhart and Noland, 2014; Rixey, 2013; Wang et al., 2015). Destination choice studies highlight that bikeshare users prefer shorter trips (El-Assi et al., 2017; Faghih-Imani and Eluru, 2015) and they make trade-offs on station distance with other conveniences such as access to points of interest and stations with larger capacity.

### *2.1. Current study in context*

From the literature review, it is evident that research on bikeshare systems is growing rapidly. However, several research questions remain to be answered. We build on the prior research and contribute to the burgeoning literature on bikeshare systems by examining system level demand along with its distribution. To elaborate, our emphasis is on understanding bikeshare demand at the stations and the flow of these bikes to their corresponding destinations. The framework would provide system operators not only an estimate of the system demand at a station level but also how these bike trips are distributed across the entire system. We identify two choice dimensions: (1) station level demand and (2) how bike flows from an origin station are distributed across the network. Station level demand is a continuous variable and can be easily analyzed using linear regression models and their advanced variants. On the other hand, the second choice variable is quite different. Specifically, for an origin station with a predefined demand, the choice process involves identifying the flows to all destination stations in the system. There are two major challenges associated with it. First, the destinations for bike flows from an origin are likely to involve multiple alternatives (as opposed to a single chosen alternative). Second, the potential universal alternative set includes all stations in the bikeshare system. The multiple discrete continuous (MDC) frameworks that follow Kuhn-Tucker (KT) approaches developed in the literature can be adapted to address this choice dimension. KT demand systems have been used in various empirical contexts including outdoor recreational demand (Phaneuf et al., 2000; von Haefen, 2004; von Haefen and Phaneuf, 2005), individual activity participation and time-use (Bhat, 2005; Nurul Habib and Miller, 2009; Pinjari and Bhat, 2010; Pinjari et al., 2009; Rajagopalan et al., 2009), household vehicle ownership and usage forecasting (Ahn et al., 2008; Bhat et al., 2009; Fang, 2008), and household budgetary allocation (Anowar et al., 2018; Ferdous et al., 2010; Rajagopalan and Srinivasan, 2008). For our current analysis, we adopt the methodology proposed by Bhat (2008)<sup>2</sup>.

---

<sup>1</sup> A number of studies in transportation literature adopted various modeling frameworks to study dependent variables with multiple dimensions such as fractional split model in vehicular speed (Bhowmik et al., 2019), bivariate or ordered probit model in injury severities and driving behavior (Fountas and Anastasopoulos, 2017; Fountas et al., 2018, 2019; Sarwar et al., 2017)

<sup>2</sup> The reader would note that fractional split models developed in recent years for several research studies (for example see Rahman et al., 2020; Bhowmik et al., 2018; Yasmin and Eluru, 2018 and Yasmin et al., 2016) offer an alternative approach to model destination flows. However, given the functional form of these models, in the presence of a large number of alternatives – as is the case in our context – the proportion allocated to these

The data for our analysis is drawn from New York City bikeshare system (CitiBike). Six months of bikeshare usage data from January 2017 through June 2017 was downloaded from the CitiBike website and rigorously processed to obtain weekly bikeshare usage patterns - station level weekly origin demand and the corresponding flow patterns to all destinations across the entire bikeshare system. In our case, the second choice dimension has 573 destination alternatives. To the best of the authors' knowledge this is the largest number of alternatives considered in a KT system in the literature.

### 3. MATERIAL AND METHODS

#### 3.1. Data

##### 3.1.1. Data source

New York's CitiBike system is one of the major public bikeshare systems around the world and the largest in the United States. The CitiBike system was launched in May 2013 with 330 stations and 6,000 bicycles in the lower half of Manhattan and some part of northwest Brooklyn. In 2017, the system expanded to 750 stations with 12,000 bicycles. According to CitiBike report, the number of annual subscribers were nearly 130,000 on July 2017. The trip itinerary dataset (from January 2017 to June 2017) of the CitiBike system is the primary data source employed (<https://www.citibikenyc.com/system-data>) in our study. The dataset provides information on start and end time of trips, their origin and destination, geographic coordinates of stations (latitude and longitude), travel time or trip duration, user types, and age and gender for members' trips. The trip data was augmented with other sources including: (1) built environment attributes derived from New York City open data (<https://nycopendata.socrata.com>); (2) socio-demographic characteristics at the census tract/zip code level gathered from US 2010 census data; (3) the weather information corresponding to the Central Park station retrieved from the National Climatic Data Center (<http://www.ncdc.noaa.gov/data-access>).

##### 3.1.2. Sample formation

A series of data cleaning and compilation exercises were undertaken for generating the sample data for estimation purposes. First, trips with missing or inconsistent information were removed. Second, trips longer than 2 hours in duration (around 0.5% of all trips) were deleted considering that the trips longer than 2 hours are not typical bicycle-sharing rides (see [Faghih-Imani and Eluru, 2015](#) for a similar approach). These trips could also be a result of misplacing the bicycle when returning it to the station. Third, 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 hence, the users returned them to the origin station. Therefore, we focus on trips that were destined outward only.

For the given study period (January 2017 to June 2017), the total number of available stations in CitiBike system was 644. Initially, we aggregated weekly trip data for each week (total 26 weeks) from each origin station to every possible destination station (643). The processing of large sample of trip data with other station level variables is substantially time-consuming and significantly increases the model run times. To obtain a reasonable sample size for model estimation, 5 weeks trip data for each origin were randomly selected. In the process, we ended up having 70 stations with no trips. So, we eliminated those 70 stations (about 10%

---

potentially unchosen alternatives could amount to be a significant value. Thus, it might be necessary to adopt an additional level of analysis with a binary choice model that determines whether a station is chosen or not and then for these chosen alternatives, a proportion is assigned.

trips) from both origin and destination choice set. Finally, we had 574 stations for analysis. The location of these stations (574 stations) is presented in Figure 1. We organized the dataset into two dimensions for our analysis; 1) For station level demand (aggregating total weekly trip at the origin level) and 2) Trip distribution from origin to destination (aggregating weekly trip at the O-D pair level). Figure 2 represents the independent and dependent variable data compilation procedure.

### 3.1.3. Independent variable generation

Several independent variables were generated in our study (see Figure 2). These can be grouped into four categories: 1) Trip attribute, 2) Bicycle and transportation infrastructure variables, 3) Weather attributes, 4) Temporal attributes and 5) Land use and built environment variables. Trip attribute includes the network distance between each origin-destination station pair estimated using the shortest path algorithm tool of ArcGIS software. While the actual trip might involve a different route, the shortest network distance would be an appropriate indicator of the distance traveled. Bicycle and transportation infrastructure attributes include CitiBike station attributes, bike route length, and public transit stations. For these attributes a 250-meter buffer around each station was created. The 250-meter buffer seems a reasonable walking distance based on the distances between CitiBike stations and the dense urban form of New York City (Kaufman et al., 2015). The variables created at the buffer level include length of bike routes (capturing the effect of availability of bicycle facilities on system usage), length of roads (minor and major roads). The number of CitiBike stations and total dock's capacity within 250 meter buffer (excluding the station considered and its capacity) were estimated to capture the impact of neighboring stations on cycling trips. Number of subway stations and bus stops in the 250 meter buffer were generated to examine the influence of public transit on cyclist's preference of destination station. Weather variables include average temperature, relative humidity and precipitation over the week. Several interaction variables were also created. Seasonality is the only temporal variable considered. We consider winter (January-March) and Spring (April-June) as dummy variables. Finally, several land use and built environment variables were considered including population density, job density and establishment density, the number of facilities (schools, colleges, hospitals), the number of point of interests (museums, shopping malls), and the number of restaurants (including coffee shops and bars), total area of parks and commercial space (office, industry, retail) within 250 meter buffer, station elevation, and distance of destination from Times Square. Population information was collected from US census 2010 and projected for 2017 at the census tract level. Job density data was estimated at the census tract level while establishment density was calculated at the zip code level for 2016. Non-motorized vehicle score (average of walk score and bike score) and transit score associated with each CitiBike station was considered at the census tract level.

### 3.1.4. Descriptive Analysis

A descriptive summary of the analysis sample is presented in Table 1. Some salient characteristics of the data are as follows. The average dock capacity in the CitiBike system is around 33 bicycles and the average network distance between origins and destinations is about 11 km. On average, 402 trips per week depart from each origin station. In order to better understand the trip generation and distribution in the CitiBike system, we generated the total number of weekly trips destined to each station from origin. The number of weekly trips generated and attracted at each station is presented in Figure 3. In Figure 3, the number of weekly trips generated (Figure 3a) and attracted (Figure 3b) to each station is categorized in five classes: Very Low (number of weekly trips less than 500), Low (500-1000), Medium

(1000-2000), High (2000-5000) and Very High (more than 5000). Overall, the visualization provides a brief overview of bicycle flows in NYC using the CitiBike system.

### 3.2. Econometric framework

#### 3.2.1. Linear mixed model for station level weekly origin demand

The station level weekly origin demand variable is a continuous value and can be analyzed using linear regression models. However, the traditional linear regression model is not appropriate for data with multiple repeated observations. In our empirical analysis, we observe the weekly demand at the same station for five weeks. Hence, we employ a linear mixed modeling approach that builds on the linear regression model while incorporating the influence of repeated observations for the same station. The linear mixed model collapses to a simple linear regression model in the absence of any station specific effects.

Let  $q = 1, 2, \dots, Q$  be an index to represent each station ( $Q = 574$ ),  $W = 1, 2, \dots, 5$  be an index to represent the various weeks of data compiled for each station. The dependent variable (weekly demand) is modeled using a linear regression equation which, in its most general form, has the following structure:

$$y_{wq} = \beta X_{wq} + \varepsilon_{wq} \quad (1)$$

where  $y_{wq}$  is the natural logarithm of weekly demand<sup>3</sup>,  $X$  is an  $L \times 1$  column vector of attributes and the model coefficients,  $\beta$ , is an  $L \times 1$  column vector.  $\beta$  includes fixed and random parameters considered in the model. The random error term,  $\varepsilon_{wq}$ , is assumed to be normally distributed across the dataset. In our analysis, the repetitions over weeks can result in common unobserved factors affecting the dependent variable. While a full covariance matrix can be estimated for the unobserved correlations, as we are selecting 5 random weeks from a sample of 26 weeks for each station, we decided to employ a simpler covariance structure. The exact functional form of the covariance structure assumed is shown below:

$$\Omega = \begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1 & \dots & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1^2 & \dots & \sigma_1 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1 & \sigma_1 & \dots & \sigma^2 + \sigma_1^2 \end{pmatrix} \quad (2)$$

The covariance structure restricts the covariance across all five records to be the same. The parameters estimated in this correlation structure are  $\sigma$  and  $\sigma_1$ . The parameter  $\sigma$  represents the error variance of  $\varepsilon$ ,  $\sigma_1$  represents the common correlation factor across weekly records. The models are estimated in SPSS using the Restricted Maximum Likelihood Approach (REML). 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).

#### 3.2.2. MDCEV model for destination choice

We consider the following functional form (Bhat and Eluru, 2010) for modeling destination preferences in this paper, based on a generalized variant of the translated Constant Elasticity of Substitution (CES) function:

---

<sup>3</sup> The reader would note that the log transformed variable distribution closely matches a normal distribution. The transformation is commonly applied for dependent variables with a large range in previous literature (Faghih-Imani et al., 2014, 2017b, Faghih-Imani and Eluru, 2016b; Rixey et al., 2013).



$$V(x) = \sum_{i=1}^I \frac{\gamma}{\alpha} \psi_i \left\{ \left( \frac{x_i}{\gamma} + 1 \right)^\alpha - 1 \right\} \quad (3)$$

where  $V(x)$  is a quasi-concave, increasing, and continuously differentiable function with respect to the bicycle flows ( $I \times 1$ )-vector ( $x_i \geq 0$  for all  $i$ ), and  $\psi_i$  associated with destination station  $i$ .  $\psi_i$  represents the baseline preference level ( $\psi_i > 0$  for all  $i$ ),  $\gamma$  is a translation parameter ( $\gamma$  should be greater than zero) which enables corner solutions while simultaneously influencing satiation and  $\alpha$  influences satiation ( $\alpha \leq 1$ ).

The KT approach employs a direct stochastic specification by assuming the function  $V(x)$  to be random over the population. A multiplicative random element is introduced to the baseline preference level for each good (in our case destination) as follows:

$$\psi(z_{iw}, \epsilon_{iw}) = \exp(\delta z_{iw} + \epsilon_{iw}) \quad (4)$$

where  $z_{iw}$  is a set of attributes characterizing destination station  $i$  during week  $w$ ,  $\delta$  corresponds to a column vector of coefficients, and  $\epsilon_{iw}$  captures idiosyncratic (unobserved) characteristics that impact the baseline preference for destination stations. The overall function from Equation (3) then takes the following form:

$$V(x) = \sum_{i=1}^I \frac{\gamma}{\alpha} \exp(\delta z_{iw} + \epsilon_{iw}) \left\{ \left( \frac{x_i}{\gamma} + 1 \right)^\alpha - 1 \right\} \quad (5)$$

Following (Bhat, 2005, 2008), consider a generalized extreme value distribution for  $\epsilon_{iw}$  and assume that  $\epsilon_{iw}$  is independent of  $z_{iw}$  ( $i = 1, 2, \dots, I$ ). The  $\epsilon_{iw}$ 's are also assumed to be independently distributed across alternatives with a scale parameter normalized to 1. Due to the common role of  $\gamma$  and  $\alpha$ , it is very challenging to identify both  $\gamma$  and  $\alpha$  in empirical application (see (Bhat, 2008) for detailed discussion). Hence, either  $\gamma$  or  $\alpha$  parameter is estimated. When the  $\alpha$  - profile is used, the function simplifies to:

$$V(x) = \sum_{i=1}^I \frac{1}{\alpha} \exp(\delta z_i + \epsilon_i) \{ (x_i + 1)^\alpha - 1 \} \quad (6)$$

When the  $\gamma$  - profile is used, the function simplifies to:

$$V(x) = \sum_{i=1}^I \gamma \exp(\delta z_i + \epsilon_i) \ln \left( \frac{x_i}{\gamma} + 1 \right) \quad (7)$$

In this study,  $\gamma$  - profile is used. Finally, the probability that an origin station has flows to the first  $M$  destination stations  $M \geq 1$  is:

$$P(e_1^*, e_2^*, e_3^*, \dots, e_M^*, 0, 0, \dots, 0) = \left[ \sum_{n=1}^M d_n \right] \left[ \sum_{n=1}^M \frac{1}{d_n} \right] \left[ \frac{\prod_{n=1}^M e^{V_n}}{(\prod_{m=1}^K e^{V_m})^M} \right] (M - 1)! \quad (8)$$

where  $(\sum_{n=1}^M d_n) (\sum_{n=1}^M 1/d_n)$  is defined as Jacobian form for the case of equal unit prices across goods (Bhat, 2008) where,  $d_n = \left( \frac{1-\alpha}{e_n^* + \gamma} \right)$ .

Unlike the traditional MDCEV model, in our context, the number of alternatives is substantially larger. Hence, we resort to estimating a generic parameter for each exogenous variable across alternatives (analogous to how multinomial logit based location choice models are estimated with a single utility equation).

#### 4. ESTIMATION RESULTS

In this section, estimation results from the two models are discussed – bikeshare demand model if followed by the trip distribution model results at destination level. The reader must note that we used same scaled parameter as presented in Table 1.

##### 4.1. Trip demand model

###### 4.1.1. Model fit measures

The empirical analysis began with estimating a simple linear regression model. This served as the benchmark for evaluating the model of the linear mixed model. The Log-likelihood ratio (LR) test statistic comparing these models was found to be 2015.0 which was higher than any corresponding chi-square value for 2 degrees of freedom ( $\sigma$  and  $\sigma_1$ ). Based on the LR test statistic, we can conclude that the linear mixed model outperforms the simple linear regression model and offers satisfactory fit for the station level demand<sup>4</sup>. Therefore, in the following section, we discuss the results from this model.

###### 4.1.2. Results

The linear mixed model estimation results are presented in Table 2.

#### **Bicycle infrastructure and transportation attributes**

Higher number of trips are likely to be generated from stations with higher capacity than lower capacity stations. Riders are willing to make more trips from stations well served by bicycle facilities such as bicycle lanes presumably because presence of bike lanes increases the accessibility of the station (see Buck and Buehler, 2012 for similar results). Overall, the results highlight the importance of station capacity and existing bicycle infrastructure on bikeshare demand. As expected, proximity of stations to subway stations positively impacts origin bike demand. This is plausible since bikeshare potentially serves as a last mile connection for some public transit users (similar results in Nair et al., 2013).

#### **Temporal attributes**

There is a negative relationship between winter season and total weekly bicycle departures from a station compared to spring season. The finding is in line with the findings reported in the literature – cold weather and snow are major deterrent to cycling trips, particularly in the North Eastern part of the US (Pucher et al., 2011).

#### **Land use and built environment attributes**

---

<sup>4</sup> The reader would note that due to inherent structure of the linear mixed models, traditional goodness of fit measures such as  $R^2$  are not readily applicable and require more involved approaches to computing the measure (see Nakagawa and Schielzeth, 2013 for more details).

Increased job density within the station buffer encourages increased bikeshare trips (see Rixey, 2013; Wang et al., 2015 for similar results). The result highlights the likely use of bicycle sharing systems for daily commute trips. Location of station in walk and bike friendly neighborhoods also drives bikeshare demand. Proximity of stations to different facilities (schools, colleges, hospitals, office) and recreational locations (point of interests such as Times Square, museums, amusement parks, shopping malls) increases station demand. Distance from Time Square is negatively associated with bikeshare flows.

### **Random parameters**

We tested for the presence of random effects for several variables. In our estimation, only one variable offered a significant estimate. Specifically, unobserved heterogeneity of the impact of length of bicycle lanes is significant highlighting that the value associated varies substantially across destinations.

### **Correlation parameters**

The correlation parameters are statistically significant highlighting the role of common unobserved factors influencing the origin stations.

## *4.2. Destination choice model*

### 4.2.1. Model fit measures

The final log-likelihood values for destination choice MDCEV model and equal probability MDCEV model are -1376961.379 and -1540196.38 respectively. The *log-likelihood ratio* (LR) test-statistic of comparison between the final model and the equal probability model is 326470.002. The LR test-statistic value is significantly higher than the corresponding chi-square value for 20 additional degrees of freedom. Based on these values, we can see that the MDCEV destination choice model offers a reasonable fit.

### 4.2.2. Results

The best fit model results of destination choice are presented in Table 3.

### **Trip attributes**

In the current research context, a negative coefficient was obtained for network distance of O-D pair. Intuitively, destinations further away are less appealing for cyclists. We also tried interaction of winter season with distance in the model. As expected, during cold weather the traveling further distance is more burdensome for bikeshare users.

### **Socio-demographic attributes**

Among socio-demographics, destination population, job and employment density variables significantly affect preferences for the destination. Stations located in census tracts with higher population density are more likely to be chosen as destination stations (see Faghih-Imani and Eluru, 2015, 2020; Rixey, 2013; Wang et al., 2015 for similar results). Similarly, job and establishment density also impact station choice positively. The result probably highlights that bicycle-sharing systems are likely to be used for daily commute trips (see Faghih-Imani et al., 2017a for similar result).

### **Bicycle infrastructure and Transportation attributes**

Stations with larger dock capacity are more likely to be chosen (similar results in El-Assi et al., 2017; Faghih-Imani and Eluru, 2015, 2020). An increase in the length of bicycle route within the 250-meter buffer of a destination station results in an increased likelihood of the station

being chosen as destination (similar to findings of El-Assi et al., 2017; Faghih-Imani and Eluru, 2015, 2016b, 2020) while a contrasting result (albeit with lower magnitude) is obtained for street length variable.

Literature suggests that in addition to their own attributes, neighboring station attributes also affect destination choice behavior. In our study, the number of stations and total dock capacity in the station buffer offer interesting results. The result is quite similar to what has been reported in earlier single discrete model (see Faghih-Imani and Eluru, 2015, 2020 for similar results). The positive impact associated with the number of neighboring stations on likelihood of choosing a station as destination is about 12 times larger than the negative impact of capacity of neighboring stations in the buffer. Hence, as long as the average capacity addition per station is under 12, neighboring stations increase demand. On the other hand, when larger stations exist in the 250 m buffer, they increase competition and reduce demand for the destination station. As the number of subway and bus stations in the buffer increases, we observe increased preference for that destination.

### **Land use and built environment attributes**

Intuitively, increased transit accessibility within the station buffer increases the station's likelihood of being chosen as destination. As expected, stations located in neighborhoods with high walk and bike accessibility – represented by higher non-motorized vehicle score - are preferred by cyclists. Cyclists prefer amenities around stations as indicated by the positive impact of number of restaurants and cafes in the vicinity of destination station. The CitiBike stations in the vicinity of parks are also more likely to be chosen. Individuals are likely to choose destination stations in a location with more facilities (such as museums, schools, colleges, university, hospitals). Visitors choose stations that bring them closer to Times Square as highlighted by negative coefficient of destination station distance to Times Square. Another important land use attributes that plays a significant role in choosing destination station is elevation of that station. People are less inclined to choose stations with steep slope for their trip. The presence of commercial area in the vicinity of destination station also increases the proclivity for the destination.

### **Satiation parameter**

As discussed earlier in the methodology section, the translation parameter  $\gamma$  captures the extent of decrease in marginal preference across different destination stations. The translation parameter  $\gamma$  is statistically significant at 95% level of significance, thereby implying that there are clear satiation effects in destination choice as distance of destination from Times Square increases. To elaborate, as the destination moves further away from Times Square, the satiation impacts are higher indicating fewer trips will be made to the destination.

## **5. VALIDATION ANALYSIS RESULTS**

For validation purpose, a hold-out sample was prepared following the same procedure used to extract the estimation sample. We randomly chose 5 weeks of data from the rest 21 weeks (a total of 26 weeks of data was available). The same approach of choice set generation for estimation sample is exercised for validation sample (574 origins x 5 weeks x 573 destinations). The difference in the log-likelihood for the predicted and equal probability model is 48118 units clearly highlighting the enhanced fit of the proposed model.

To further highlight the applicability of estimated model for predicting destination choice conditional on the origin, we estimated destined trips from each origin for each week at disaggregate level. Note that, zero trips to any destination for a week was also considered. For the performance evaluation, we compute the correctly classified predicted trips for each O-D pair for each week. The reader would note that for about 73% trips the prediction was correctly

classified (see Figure 4). Specifically, 78% of zero trips from an origin to all possible destinations in each week was classified correctly while the corresponding number of non-zero trips is 33%. The result indicates that predicted model performs better in case of destination stations with zero trips. Also, correct prediction was observed to be higher for the origin stations which have higher number of chosen alternative destinations (more than 70) (see Figure 5). The result makes intuitive sense. In cases where the number of destinations is fewer (say  $\leq 30$ ), the MDCEV allocation has to find a few alternatives from the universal set thus increasing potential scope for error.

## 6. POLICY ILLUSTRATION

To highlight the applicability of the proposed model system, we conducted two policy analysis exercises: (1) an innovative policy illustration and (2) estimation of elasticity effects. For the first exercise, we predicted changes in destination preferences with changes in bicycle infrastructure. Specifically, we increased the bike street length by 50% within the 250m buffer of the destination stations, compute the corresponding utility associated with choosing destinations, and demonstrate how the top 10 percentile of preferred destination stations alter in response to the change. For illustration purposes, we present the results for a randomly selected origin station (Station 3016) for a random week (see Figure 6). Figure 6 presents the preferred destinations in the top ten percentile, before (6a) and after (6b) increase. The results indicate that with increase in bike infrastructure, current preferred stations at the periphery are replaced with newer stations that are outside the base periphery (near lower Manhattan). The result is a manifestation of how the impact of distance on destination choice is compensated by additional bicycling infrastructure in the farther locations. While increasing bicycle infrastructure by 50% is far from straight forward, the analysis is an illustration of how the proposed model can be employed for policy analysis.

For the second exercise, elasticity effects computation considering changes in baseline preference function was used to evaluate the impact of exogenous variables on destination station choice. The elasticity effects are computed by evaluating the percentage change in baseline preference of an alternative in response to increasing the value of exogenous variables from best fit model by 10%, 25% and 50% respectively. The computed elasticities are presented in Figure 7. Based on elasticity effects results in Figure 7, following observations can be made. *First*, the elasticity estimate for station's capacity variable indicates that destination preference improves by 6.4, 16.80 and 36.62% in response to 10, 25 and 50% increase in station capacity respectively. *Second*, rank order of the top three significant variable in terms of changes for the preference without considering the sign of the impact include station's capacity, network distance and job density. *Third*, network distance between O-D can be considered as a proxy for travel time. Improving station connectivity by providing bicycle facilities can offer positive impetus to bike demand and flows. Overall, the elasticity analysis results provide an illustration on how the proposed model can be applied to determine the critical factors affecting bikeshare destination preferences.

## 7. CONCLUSION AND FUTURE RESEARCH

Given the burgeoning growth in bikeshare system installations and their growing adoption for trip making, it is important to develop modeling frameworks to understand bikeshare demand flows at the system level. The emergence of shared mobility options has changed the overall landscape of travel behavior in many metropolitan areas. However, current state-of-practice and travel demand models are not equipped to accurately examine the effects of these services. Developing more accurate and policy sensitive models, requires understanding the fundamentals of decision-making processes toward these new modes of travel. The current study proposes a model framework for investigating bikeshare system usage as along with the

origin-destination flows. We identify two choice dimensions: (1) station level demand and (2) how bike flows from an origin station are distributed across the network. A linear mixed model is considered for modeling weekly origin station demand while a multiple discrete continuous extreme value model (MDCEV) is employed to analyze flows from origin to multiple destinations.

The data for our analysis is drawn from New York City bikeshare system (CitiBike) for six months from January through June, 2017. For our analysis, we examine demand and distribution patterns on a weekly basis. A host of exogenous variables including trip attributes, socio-demographic attributes, bicycle infrastructure attributes, land use and built environment, temporal and weather attributes are considered. The model estimation results provide intuitive findings for both station level demand and destination choice behavior. Several attributes like job density, number of facilities and recreational points, transit and bike accessibility, dock capacity, bike length in vicinity, and census tract level variables (such as population density, job density, and establishment density) increase the preferences for a destination while distance to Times Square, and winter season decrease the likelihood of choosing a destination. In addition to model estimation, a model validation effort was conducted using a hold out sample. The data fit relative to the equal probability MDCEV model highlighted the significant improvement in data fit for the estimated model. Finally, we employed our MDCEV model for prediction to compute the demand for destination stations across the system. We estimated the number of trips at the disaggregate level for each O-D pair by week and computed the number of correctly classified trips based on our predictions. The prediction exercise illustrated the reasonable performance of the proposed model. To further augment the policy analysis, elasticity effects were computed by evaluating the percentage change in destination preferences in response to increasing the value of exogenous variables by 10%, 25% and 50% respectively. Based on the exercise, the top three significant variables in magnitude include station's capacity, network distance and job density.

To be sure, this paper is not without limitations. Given the large number of alternatives, the model run times were substantially long affecting number of specifications we can test. In our analysis, unobserved effects arising from repetitions in the MDCEV model were not captured. Another potential avenue for future research is the consideration of sampling for MDCEV models (similar to sampling in MNL models).

## ACKNOWLEDGEMENT

The authors would like to thank Ahmadreza Faghih-Imani for initial discussions on the idea of the paper and help with data assembly for CitiBike data.

## REFERENCES

1. Ahn, J., Jeong, G., & Kim, Y. (2008). A forecast of household ownership and use of alternative fuel vehicles: A multiple discrete-continuous choice approach. *Energy Economics*, 30(5), 2091-2104.
2. Anowar, S., Eluru, N., & Miranda-Moreno, L. F. (2018). How household transportation expenditures have evolved in Canada: a long term perspective. *Transportation*, 45(5), 1297-1317.
3. Ashqar, H. I., Elhenawy, M., Almannaa, M. H., Ghanem, A., Rakha, H. A., & House, L. (2017, June). Modeling bike availability in a bike-sharing system using machine learning. In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*(pp. 374-378). IEEE.
4. Bachand-Marleau, J., Lee, B. H., & El-Geneidy, A. M. (2012). Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use. *Transportation Research Record*, 2314, 66-71.

5. Barbour, N., Zhang, Y., Mannering, F. (2019). A statistical analysis of bike sharing usage and its potential as an auto-trip substitute. *Journal of Transport & Health* 12, 253-262.
6. Bhat, C. R. (2005). A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological*, 39(8), 679-707.
7. Bhat, C. R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological*, 42(3), 274-303.
8. Bhat, C. R., Sen, S., & Eluru, N. (2009). The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. *Transportation Research Part B: Methodological*, 43(1), 1-18.
9. Bhowmik, T., Yasmin, S., Eluru, N. (2018) A joint econometric approach for modeling crash counts by collision type. *Analytic methods in accident research* 19, 16-32.
10. Bhowmik, T., Yasmin, S., Eluru, N. (2019). A multilevel generalized ordered probit fractional split model for analyzing vehicle speed. *Analytic methods in accident research* 21, 13-31.
11. Bouveyron, C., Côme, E., & Jacques, J. (2015). The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics*, 9(4), 1726-1760.
12. Buck, D., & Buehler, R. (2012, January). Bike lanes and other determinants of capital bikeshare trips. Presented in *91st Transportation Research Board Annual Meeting*.
13. Buck, D., Buehler, R., Happ, P., Rawls, B., Chung, P., & Borecki, N. (2013). Are bikeshare users different from regular cyclists? A first look at short-term users, annual members, and area cyclists in the Washington, DC, region. *Transportation Research Record: Journal of the Transportation Research Board*, 2387, 112-119.
14. Caggiani, L., Camporeale, R., Marinelli, M., & Ottomanelli, M. (2019). User satisfaction based model for resource allocation in bike-sharing systems. *Transport Policy*, 80, 117-126.
15. Caggiani, L., Colovic, A., & Ottomanelli, M. (2020). An equality-based model for bike-sharing stations location in bicycle-public transport multimodal mobility. *Transportation Research Part A: Policy and Practice*, 140, 251-265.
16. Caulfield, B., O'Mahony, M., Brazil, W., & Weldon, P. (2017). Examining usage patterns of a bike-sharing scheme in a medium sized city. *Transportation Research Part A: Policy and Practice*, 100, 152-161.
17. Choi, Y., & Choi, E. J. (2020). Sustainable Governance of the Sharing Economy: The Chinese Bike-Sharing Industry. *Sustainability*, 12(3), 1195.
18. de Chardon, C. M. (2019). The contradictions of bike-share benefits, purposes and outcomes. *Transportation research part A: policy and practice*, 121, 401-419.
19. Dell, M., Iori, M., Novellani, S., & Stütze, T. (2016). A destroy and repair algorithm for the bike sharing rebalancing problem. *Computers & Operations Research*, 71, 149-162.
20. Dey, B. K., Anowar, S., Eluru, N., & Hatzopoulou, M. (2018). Accommodating exogenous variable and decision rule heterogeneity in discrete choice models: Application to bicyclist route choice. *PloS One*, 13(11), e0208309.
21. Eker, U., Ahmed, S. S., Fountas, G., & Anastasopoulos, P. C. (2019). An exploratory investigation of public perceptions towards safety and security from the future use of flying cars in the United States. *Analytic methods in accident research*, 23, 100103.
22. Eker, U., Ahmed, S. S., Fountas, G., & Anastasopoulos, P. C. (2019). An exploratory investigation of public perceptions towards safety and security from the future use of flying cars in the United States. *Analytic methods in accident research*, 23, 100103.

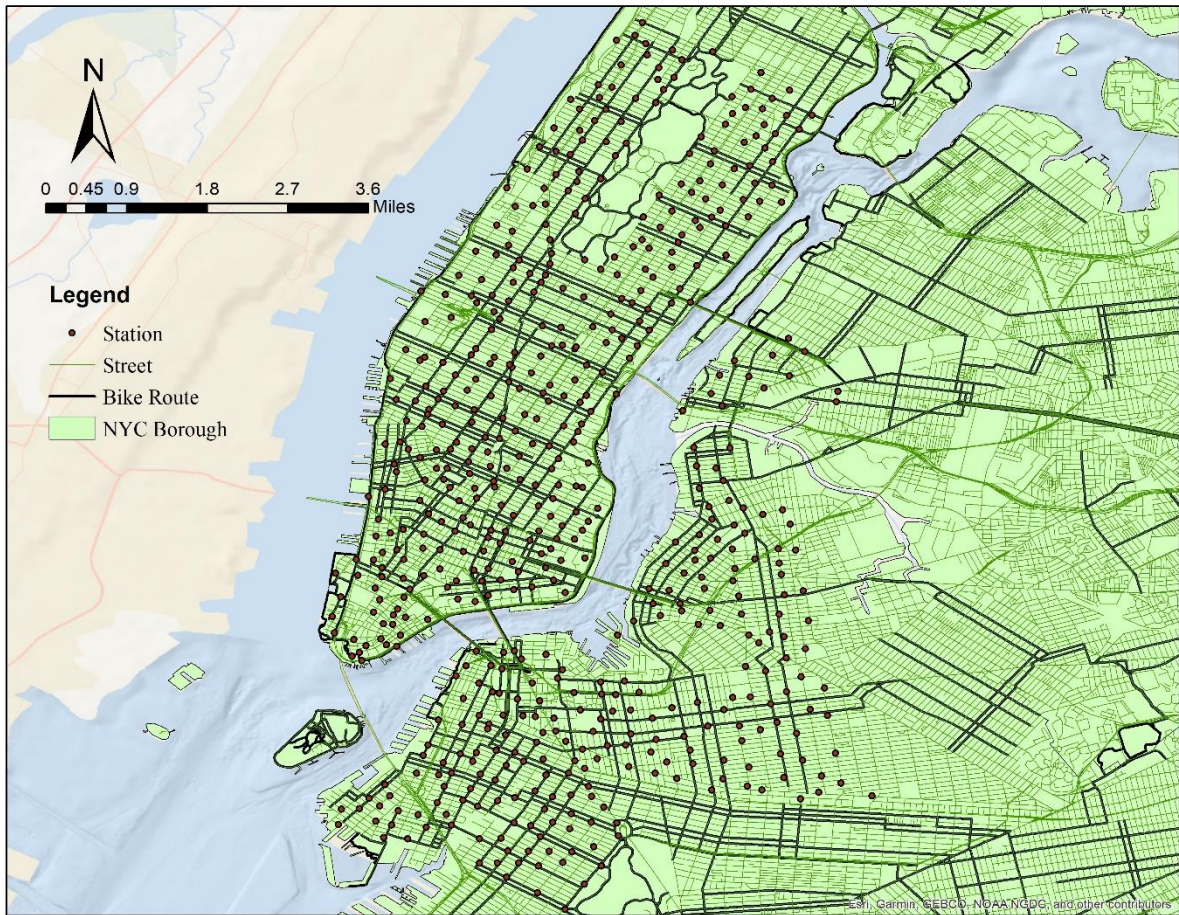
23. El-Assi, W., Mahmoud, M. S., & Habib, K. N. (2017). Effects of built environment and weather on bike sharing demand: a station level analysis of commercial bike sharing in Toronto. *Transportation*, 44(3), 589-613.
24. Faghieh-Imani, A., Eluru, N., El-Geneidy, A. M., Rabbat, M., & Haq, U. (2014). How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography*, 41, 306-314.
25. Faghieh-Imani, A., & Eluru, N. (2015). Analysing bicycle-sharing system user destination choice preferences: Chicago's Divvy system. *Journal of Transport Geography*, 44, 53-64.
26. Faghieh-Imani, A., & Eluru, N. (2016a). 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*, 94, 685-698.
27. Faghieh-Imani, A., & Eluru, N. (2016b). Incorporating the impact of spatio-temporal interactions on bicycle sharing system demand: A case study of New York CitiBike system. *Journal of Transport Geography*, 54, 218-227.
28. Faghieh-Imani, A., Eluru, N., & Paleti, R. (2017a). How bicycling sharing system usage is affected by land use and urban form: analysis from system and user perspectives. *European Journal of Transport and Infrastructure Research*, 17(3).
29. Faghieh-Imani, A., Hampshire, R., Marla, L., & Eluru, N. (2017b). An empirical analysis of bike sharing usage and rebalancing: Evidence from Barcelona and Seville. *Transportation Research Part A: Policy and Practice*, 97, 177-191.
30. Faghieh-Imani, A., & Eluru, N. (2020). A finite mixture modeling approach to examine New York City bicycle sharing system (CitiBike) users' destination preferences. *Transportation*, 47(2), 529-553.
31. Fang, H. A. (2008). A discrete-continuous model of households' vehicle choice and usage, with an application to the effects of residential density. *Transportation Research Part B: Methodological*, 42(9), 736-758.
32. Fountas, G., Anastasopoulos, P.C. (2017). A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. *Analytic methods in accident research* 15, 1-16.
33. Fountas, G., Anastasopoulos, P.C., Mannering, F.L. (2018). Analysis of vehicle accident-injury severities: A comparison of segment-versus accident-based latent class ordered probit models with class-probability functions. *Analytic methods in accident research* 18, 15-32.
34. Fountas, G., Pantangi, S.S., Hulme, K.F., Anastasopoulos, P.C. (2019). The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: A correlated grouped random parameters bivariate probit approach. *Analytic methods in accident research* 22, 100091.
35. Feigon, S., & Murphy, C. (2016). *Shared mobility and the transformation of public transit* (No. Project J-11, Task 21).
36. Ferdous, N., Pinjari, A. R., Bhat, C. R., & Pendyala, R. M. (2010). A comprehensive analysis of household transportation expenditures relative to other goods and services: an application to United States consumer expenditure data. *Transportation*, 37(3), 363-390.
37. Fishman, E., Washington, S., & Haworth, N. (2013). Bike share: a synthesis of the literature. *Transport Reviews*, 33(2), 148-165.
38. Fishman, E., Washington, S., & Haworth, N. (2014). Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. *Transportation Research Part D: Transport and Environment*, 31, 13-20.
39. Fishman, E. (2016). Bikeshare: A review of recent literature. *Transport Reviews* 36, 92-113.



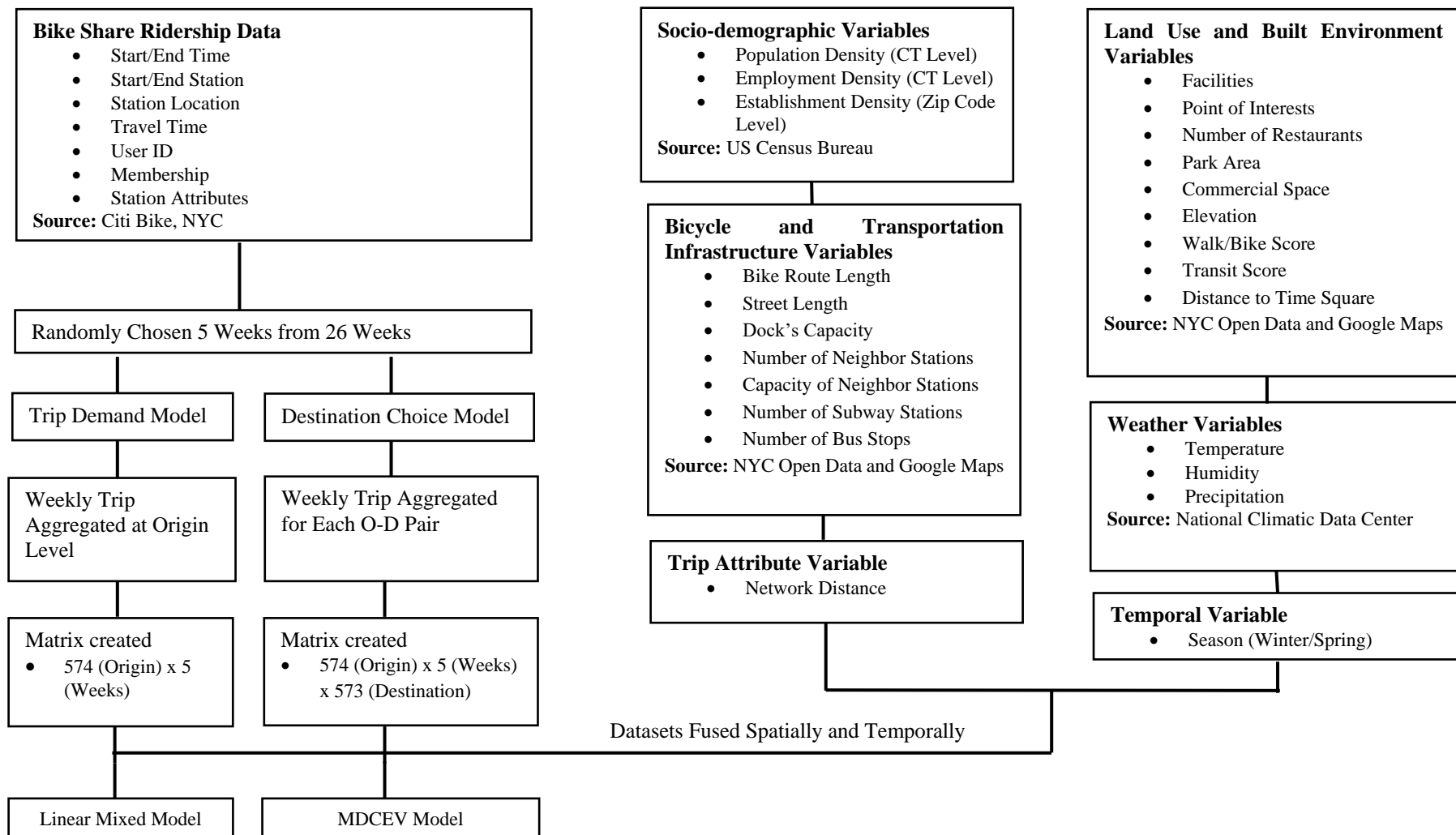
40. Forma, I. A., Raviv, T., & Tzur, M. (2015). A 3-step math heuristic for the static repositioning problem in bike-sharing systems. *Transportation Research Part B: Methodological*, 71, 230-247.
41. Fricker, C., & Gast, N. (2016). Incentives and redistribution in homogeneous bike-sharing systems with stations of finite capacity. *Euro Journal on Transportation and Logistics*, 5(3), 261-291.
42. Fuller, D., Gauvin, L., Kestens, Y., Daniel, M., Fournier, M., Morency, P., & Drouin, L. (2011). Use of a new public bicycle share program in Montreal, Canada. *American Journal of Preventive Medicine*, 41(1), 80-83.
43. Gebhart, K., & Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, 41(6), 1205-1225.
44. Giot, R., & Cherrier, R. (2014, December). Predicting bikeshare system usage up to one day ahead. In *2014 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS)* (pp. 22-29). IEEE.
45. Habib, K. M., & Miller, E. J. (2009). Modelling activity generation: a utility-based model for activity-agenda formation. *Transportmetrica*, 5(1), 3-23.
46. Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358), 320-338.
47. Hyland, M., Hong, Z., de Farias Pinto, H. K. R., & Chen, Y. (2018). Hybrid cluster-regression approach to model bikeshare station usage. *Transportation Research Part A: Policy and Practice*, 115, 71-89.
48. Jäppinen, S., Toivonen, T., & Salonen, M. (2013). Modelling the potential effect of shared bicycles on public transport travel times in Greater Helsinki: An open data approach. *Applied Geography*, 43, 13-24.
49. Kabra, A., Belavina, E., & Girotra, K. (2019). Bike-share systems: Accessibility and availability. *Management Science*.
50. Kaufman, S. M., Gordon-Koven, L., Levenson, N., & Moss, M. L. (2015). Citi Bike: The First Two Years.
51. Lu, M., Hsu, S.-C., Chen, P.-C., Lee, W.-Y. (2018). Improving the sustainability of integrated transportation system with bike-sharing: A spatial agent-based approach. *Sustainable cities and society* 41, 44-51.
52. Macioszek, E., Świerk, P., & Kurek, A. (2020). The Bike-Sharing System as an Element of Enhancing Sustainable Mobility—A Case Study based on a City in Poland. *Sustainability*, 12(8), 3285.
53. Meddin, R., & DeMaio, P. (2016). The bike-sharing world map. Retrieved from <http://www.bikesharingworld.com>.
54. Nair, R., Miller-Hooks, E., Hampshire, R. C., & Bušić, A. (2013). Large-scale vehicle sharing systems: analysis of Vélolib'. *International Journal of Sustainable Transportation*, 7(1), 85-106.
55. Nakagawa, S., Schielzeth, H. (2013) A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods in ecology and evolution* 4, 133-142.
56. Nath, R. B., & Rambha, T. (2019). Modelling Methods for Planning and Operation of Bike-Sharing Systems. *Journal of the Indian Institute of Science*, 1-26.
57. Nikitas, A. (2018). Understanding bike-sharing acceptability and expected usage patterns in the context of a small city novel to the concept: A story of 'Greek Drama'. *Transportation research part F: traffic psychology and behaviour* 56, 306-321.
58. Noland, R. B., Smart, M. J., & Guo, Z. (2016). Bikeshare trip generation in New York city. *Transportation Research Part A: Policy and Practice*, 94, 164-181.

59. Oliveira, G. N., Sotomayor, J. L., Torchelsen, R. P., Silva, C. T., & Comba, J. L. (2016). Visual analysis of bike-sharing systems. *Computers & Graphics*, 60, 119-129.
60. Pal, A., Zhang, Y. (2017). Free-floating bike sharing: Solving real-life large-scale static rebalancing problems. *Transportation Research Part C: Emerging Technologies* 80, 92-116.
61. Peters, L., MacKenzie, D. (2019) The death and rebirth of bikesharing in Seattle: Implications for policy and system design. *Transportation Research Part A: Policy and Practice* 130, 208-226.
62. Pfrommer, J., Warrington, J., Schildbach, G., & Morari, M. (2014). Dynamic vehicle redistribution and online price incentives in shared mobility systems. *IEEE Transactions on Intelligent Transportation Systems*, 15(4), 1567-1578.
63. Phaneuf, D. J., Kling, C. L., & Herriges, J. A. (2000). Estimation and welfare calculations in a generalized corner solution model with an application to recreation demand. *Review of Economics and Statistics*, 82(1), 83-92.
64. Pinjari, A. R., & Bhat, C. (2010). A Multiple Discrete–Continuous Nested Extreme Value (MDCNEV) model: formulation and application to non-worker activity time-use and timing behavior on weekdays. *Transportation Research Part B: Methodological*, 44(4), 562-583.
65. Pinjari, A. R., Bhat, C. R., & Hensher, D. A. (2009). Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B: Methodological*, 43(7), 729-748.
66. Pucher, J., Buehler, R., & Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A: Policy and Practice*, 45(6), 451-475.
67. Rajagopalan, B., Pinjari, A., & Bhat, C. (2009). Comprehensive model of worker nonwork-activity time use and timing behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2134, 51-62.
68. Rajagopalan, B., & Srinivasan, K. (2008). Integrating household-level mode choice and modal expenditure decisions in a developing country: Multiple discrete-continuous extreme value model. *Transportation Research Record: Journal of the Transportation Research Board*, 2076, 41-51.
69. Rahman, M., Yasmin, S., Eluru, N. (2020) A joint panel binary logit and fractional split model for converting route-level transit ridership data to stop-level boarding and alighting data. *Transportation Research Part A: Policy and Practice* 139, 1-16.
70. Raviv, T., Tzur, M., & Forma, I. A. (2013). Static repositioning in a bike-sharing system: models and solution approaches. *Euro Journal on Transportation and Logistics*, 2(3), 187-229.
71. Richter, F (2018, April). The Global Rise of Bike-Sharing. Retrieved from <https://www.statista.com/chart/13483/bike-sharing-programs/>.
72. Rixey, R. (2013). Station-level forecasting of bikesharing ridership: Station Network Effects in Three US Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 2387, 46-55.
73. Rudloff, C., & Lackner, B. (2014). Modeling demand for bikesharing systems: neighboring stations as source for demand and reason for structural breaks. *Transportation Research Record*, 2430, 1-11.
74. Sarwar, M.T., Fountas, G., Anastasopoulos, P.C. (2017). Simultaneous estimation of discrete outcome and continuous dependent variable equations: A bivariate random effects modeling approach with unrestricted instruments. *Analytic Methods in Accident Research* 16, 23-34.

75. Schoner, J., & Levinson, D. M. (2013). Which Station? Access Trips and Bike Share Route Choice.
76. Seebauer, S. (2015) Why early adopters engage in interpersonal diffusion of technological innovations: An empirical study on electric bicycles and electric scooters. *Transportation Research Part A: Policy and Practice* 78, 146-160.
77. Vogel, P., & Mattfeld, D. C. (2011). *Strategic and operational planning of bike-sharing systems by data mining—a case study*. Paper presented at the *International Conference on Computational Logistics*.
78. von Haefen, R. H. (2004). Empirical strategies for incorporating weak complementarity into continuous demand system models. *Unpublished, Department of Agricultural and Resource Economics, University of Arizona*.
79. von Haefen, R. H., & Phaneuf, D. J. (2005). Kuhn-Tucker demand system approaches to non-market valuation *Applications of Simulation Methods in Environmental and Resource Economics* (pp. 135-157): Springer.
80. Wang, X., Lindsey, G., Schoner, J. E., & Harrison, A. (2015). Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development*, 142(1), 04015001.
81. Wang, Z., Cheng, L., Li, Y., & Li, Z. (2020). Spatiotemporal Characteristics of Bike-Sharing Usage around Rail Transit Stations: Evidence from Beijing, China. *Sustainability*, 12(4), 1299.
82. Wolf, A., Seebauer, S. (2014). Technology adoption of electric bicycles: A survey among early adopters. *Transportation Research Part A: Policy and Practice* 69, 196-211.
83. Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M. (2016) Ordered fractional split approach for aggregate injury severity modeling. *Transportation Research Record* 2583, 119-126.
84. Yasmin, S., Eluru, N. (2018) A joint econometric framework for modeling crash counts by severity. *Transportmetrica A: transport science* 14, 230-255.
85. Yufei, H., Oukhellou, L., & Come, E. (2014). Towards bicycle demand prediction of large-scale bicycle sharing system.
86. Zhang, Y., Thomas, T., Brussel, M., & Van Maarseveen, M. (2017). Exploring the impact of built environment factors on the use of public bikes at bike stations: Case study in Zhongshan, China. *Journal of Transport Geography*, 58, 59-70.
87. Zhao, J., Deng, W., & Song, Y. (2014). Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. *Transport Policy*, 35, 253-264.

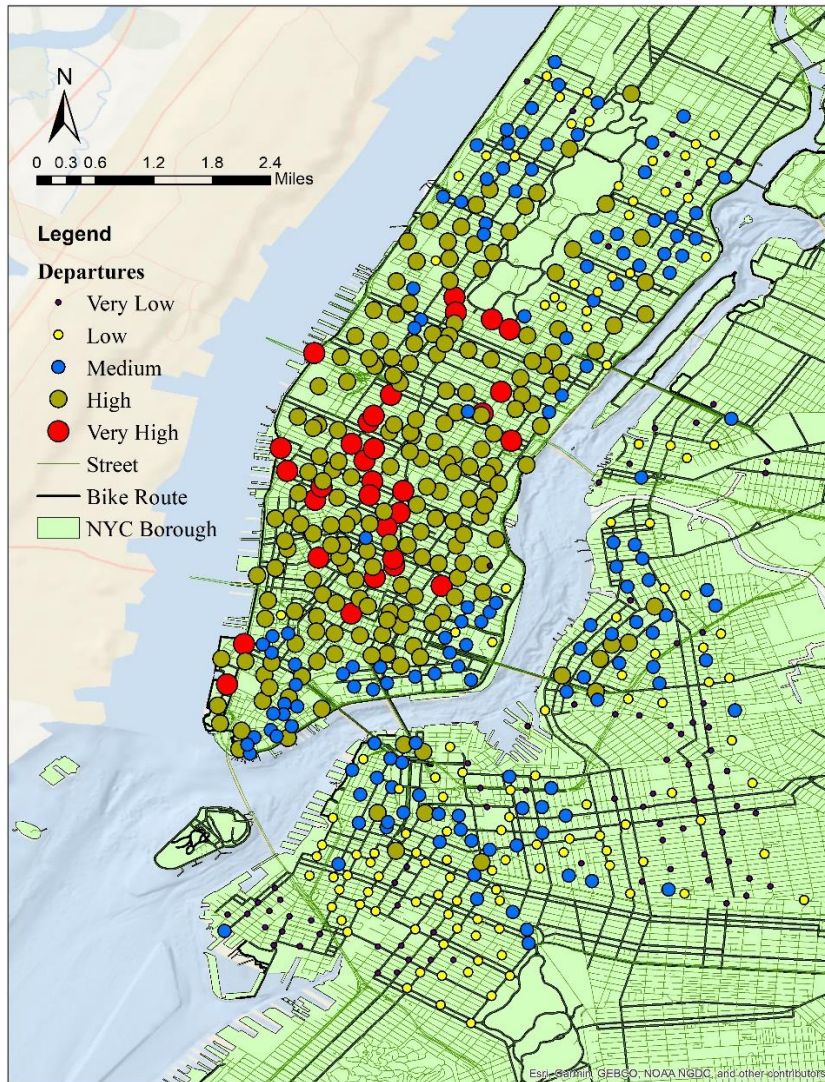


**Figure 1** NYC's bicycle-sharing system (CitiBike)

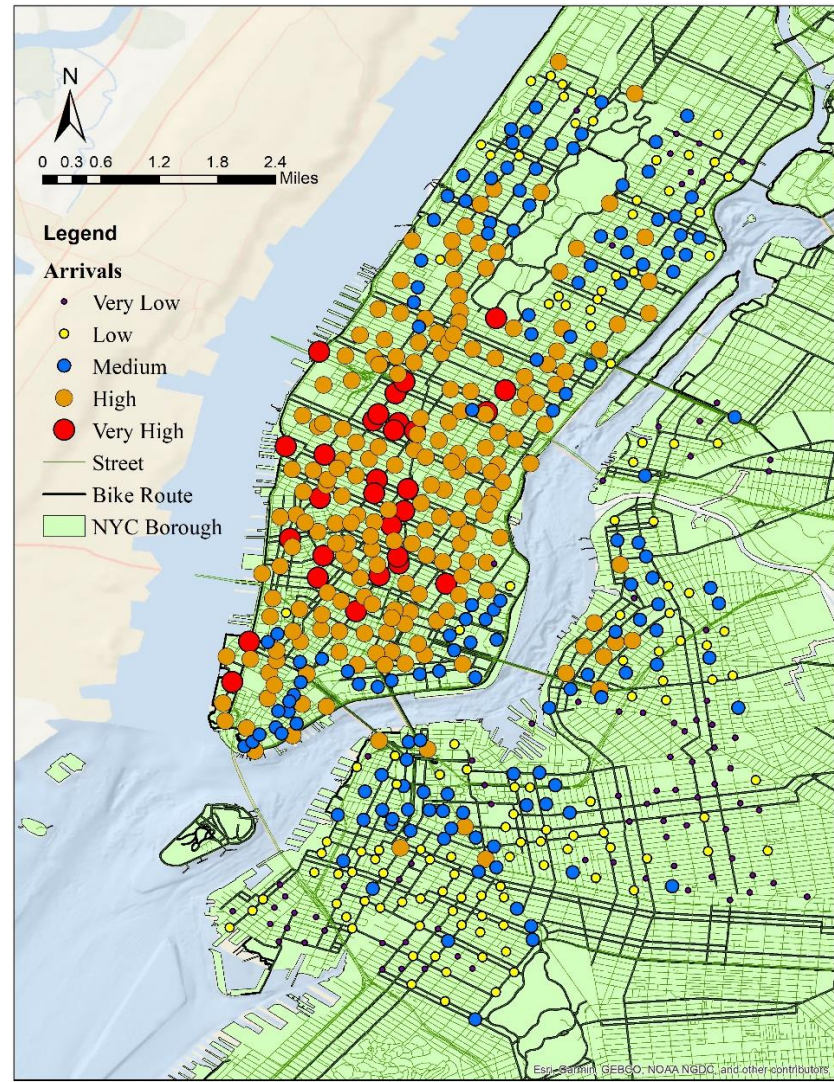


**Figure 2** Data formation flow chart



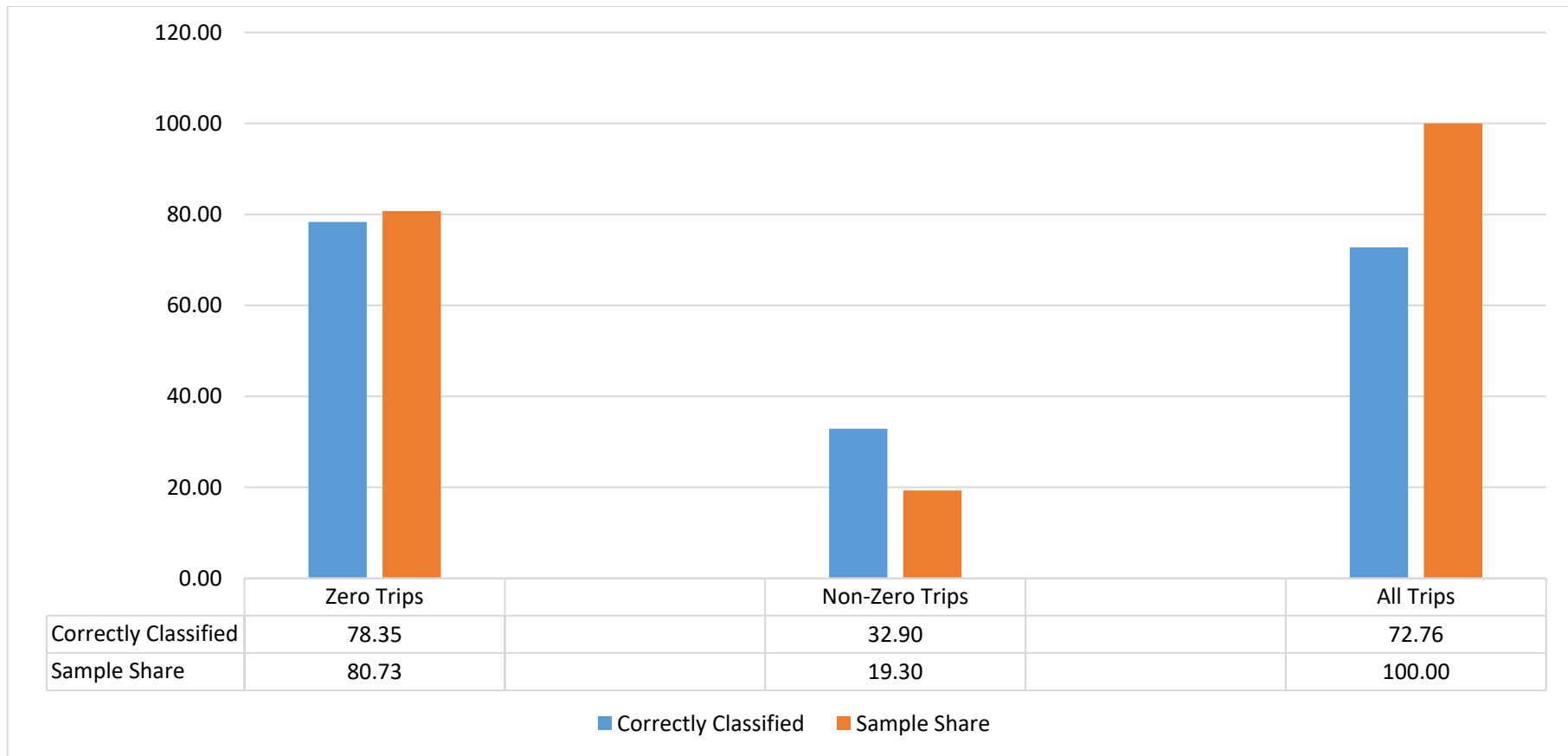


**(a) Trip generation at origin stations**

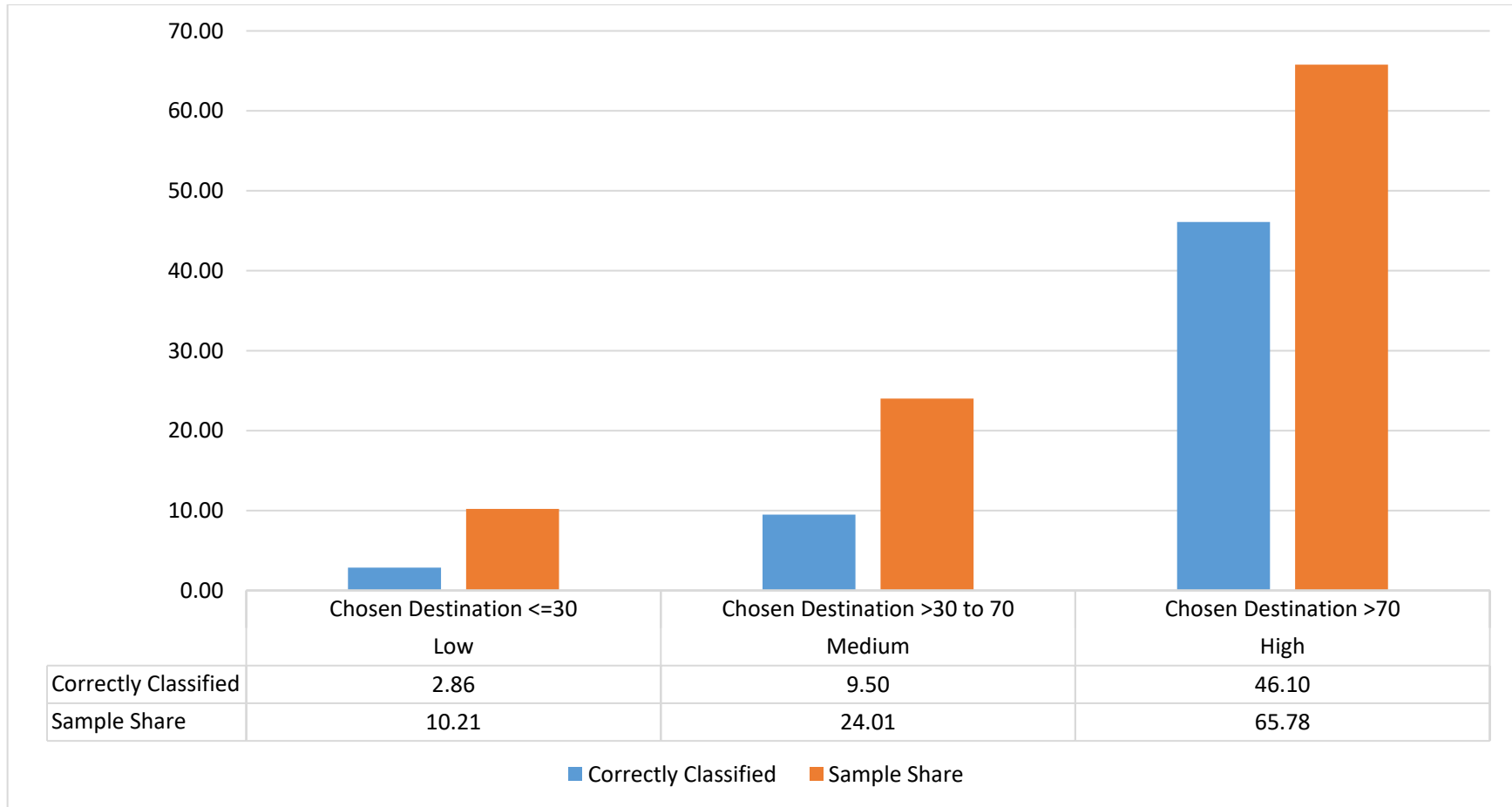


**(b) Trip attraction at destination stations**

**Figure 3** Bicycle-sharing trips in NYC's CitiBike system

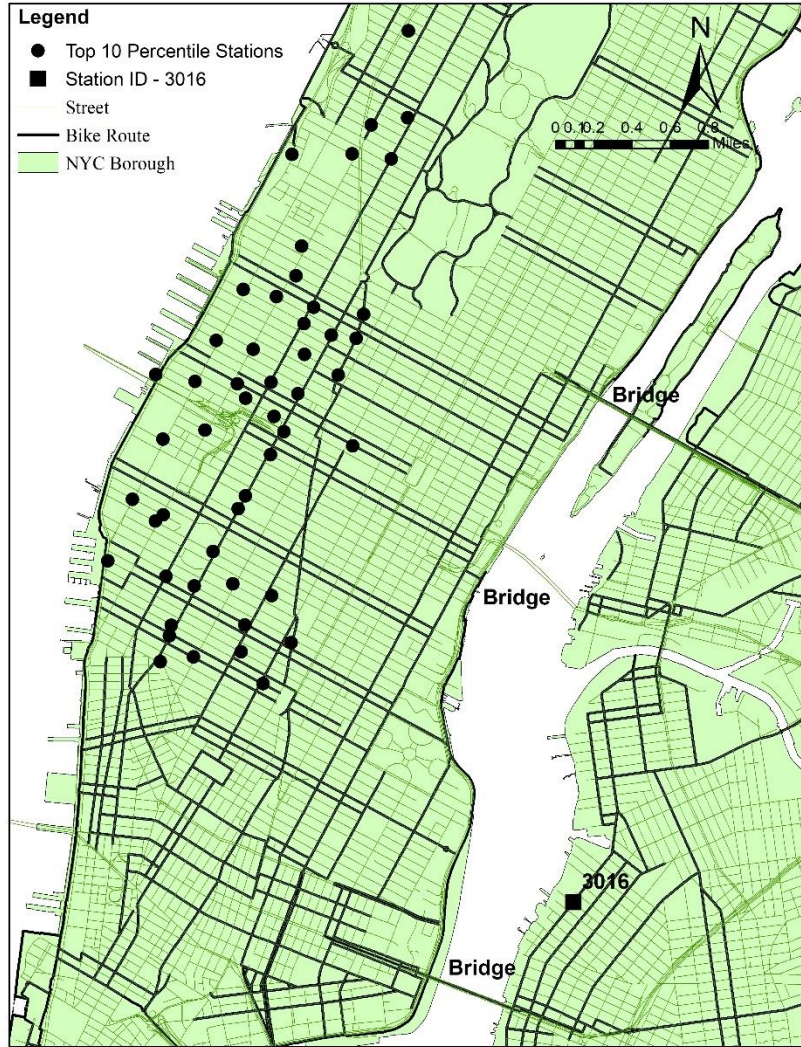


**Figure 4** Prediction measure of bicycle-sharing trips in NYC’s CitiBike system

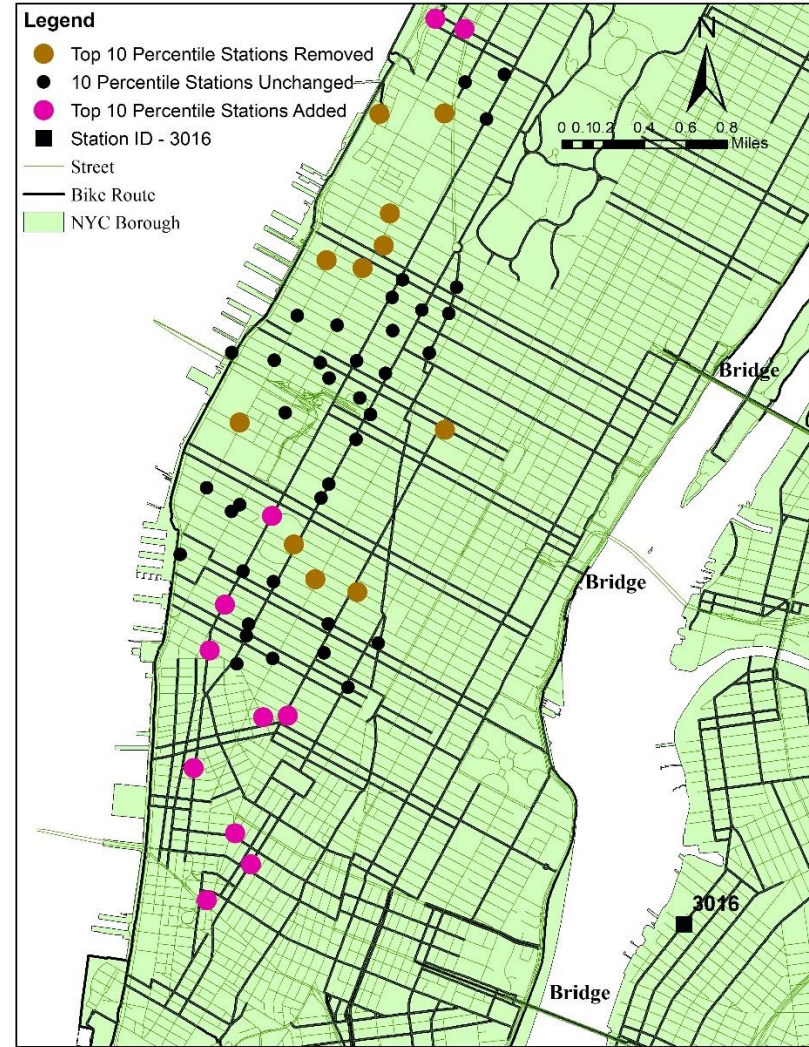


**Figure 5** Variations of prediction measure of bicycle-sharing trips with chosen alternative stations



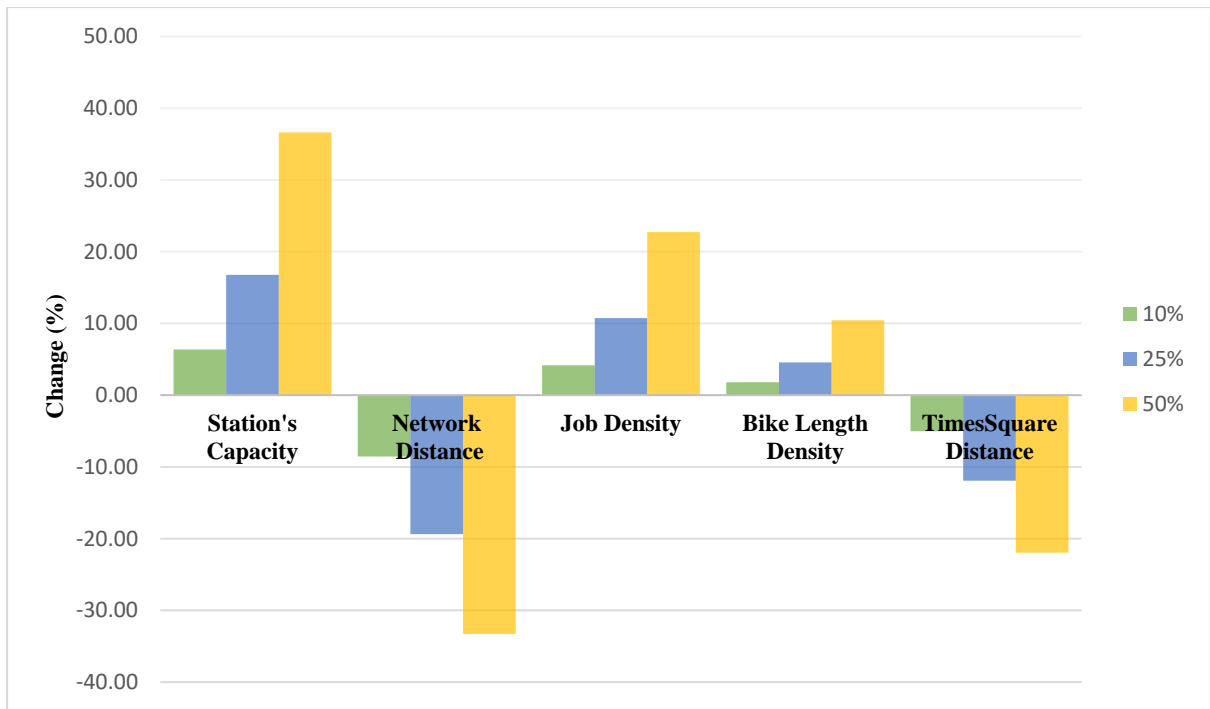


(a) Top ten percentile destination stations before increase



(b) Top ten percentile destination stations after increase

Figure 6 Top ten percentile destination stations



**Figure 7** Elasticity effects considering utility changes

**Table 1** Descriptive summary of sample characteristics

<b>Continuous Variables</b>	<b>Min</b>	<b>Max</b>	<b>Mean</b>	<b>Std. Deviation</b>
<b>Dependent Variable</b>				
<b>Trip Demand</b>				
Total Trip (Weekly per Origin)	1.00	3726.00	402.17	390.06
<b>Destination Choice</b>				
Alternative Destination Chosen	1.00	354.00	111.69	65.79
Total Trip (Weekly O-D Pair)	1.00	175.00	3.60	5.15
<b>Independent Variables</b>				
<b>Trip Attributes</b>				
Network Distance (m) ( $\times 10^{-5}$ )	0.05	0.41	0.14	0.08
<b>Bicycle Infrastructure and Transportation Attributes</b>				
Length of Bicycle Facility in 250m Buffer ( $m \times 10^{-4}$ )	0.00	0.91	0.24	0.17
Length of Street in 250m Buffer ( $m \times 10^{-4}$ )	0.14	0.84	0.38	0.10
Station Capacity ( $\times 10^{-2}$ )	0.07	0.67	0.32	0.10
Number of Neighboring Station in 250m Buffer ( $\times 10^{-1}$ )	0.00	0.50	0.11	0.10
Capacity of Neighboring Station in 250m Buffer ( $\times 10^{-3}$ )	0.00	0.27	0.04	0.04
Number of Subway Stations in 250m Buffer ( $\times 10^{-1}$ )	0.00	0.70	0.06	0.09
Number of Bus Stops in 250m Buffer ( $\times 10^{-1}$ )	0.00	1.10	0.22	0.22
<b>Weather Attributes</b>				
Temperature ( $^{\circ}$ F)	19	84	50.06	13.56
Precipitation (in)	0	3.02	0.16	0.44
Humidity (%)	26	98	61.44	17.5
<b>Land Use and Built Environment Attributes</b>				
Population Density (People per $m^2 \times 10^{-4}$ )	0.00	0.87	0.26	0.17
Job Density (Number of Jobs per Person)	0.00	0.90	0.66	0.17
Number of Establishment (per $m^2 \times 10^{-4}$ )	0.00	1.20	0.09	0.14
Walk Score ( $\times 10^{-2}$ )	0.69	1.00	0.97	0.05
Transit Score ( $\times 10^{-2}$ )	0.61	1.00	0.96	0.07
Bike Score ( $\times 10^{-2}$ )	0.45	0.95	0.85	0.09
Number of Facilities in 250m Buffer ( $\times 10^{-3}$ )	0.00	0.16	0.03	0.02
Number of Recreational Facilities in 250m Buffer ( $\times 10^{-3}$ )	0.00	0.002	0.08	0.30
Number of Restaurants in 250m Buffer ( $\times 10^{-3}$ )	0.00	0.55	0.04	0.08
Number of Sidewalk café in 250m Buffer ( $\times 10^{-3}$ )	0.00	0.14	0.02	0.02
Area of Parks in 250m Buffer ( $m^2 \times 10^{-6}$ )	0.00	0.18	0.09	0.05
Commercial Area in 250m Buffer ( $m^2 \times 10^{-6}$ )	0.00	0.55	0.26	0.14
Elevation ( $m \times 10^{-3}$ )	0.00	0.16	0.04	0.03
Distance to Times Square ( $m \times 10^{-3}$ )	0.58	1.32	0.52	0.28
<b>Categorical Variables</b>				
<b>Temporal Attributes</b>	<b>Percentage</b>			
Winter	48.90			
Spring	51.10			

**Table 2** Linear mixed model results

<b>Parameter</b>	<b>Estimates</b>	<b>t-stats</b>
Intercept ( $\times 10^{-3}$ )	-0.949	-6.914
<b>Bicycle Infrastructure and Transportation Attributes</b>		
Station's Capacity ( $\times 10^{-2}$ )	.370	6.474
Number of Subway Stations in 250m Buffer ( $\times 10^{-1}$ )	0.341	3.308
Length of Bicycle Facility in 250m Buffer ( $\text{m} \times 10^{-4}$ )	0.288	4.525
Standard Deviation ( $\text{m} \times 10^{-4}$ )	0.158	2.891
<b>Temporal Attributes</b>		
Season: Winter (Base: Spring)	-0.268	-41.847
<b>Land Use and Built Environment Attributes</b>		
Job Density	0.180	3.166
Non-motorized vehicle score ( $\times 10^{-2}$ )	1.423	9.982
Number of Facilities and Recreational Point in 250m Buffer ( $\times 10^{-3}$ )	1.316	4.453
Distance to Times Square ( $\text{m} \times 10^{-5}$ )	-5.886	-14.170
<b>Correlation Parameters</b>		
$\sigma$	0.247	33.875
$\sigma_1$	0.314	7.923
<b>Restricted Log-Likelihood</b>	-1863.186	
<b>Number of Observations</b>	574	

**Table 3 MDCEV model results**

Parameter	Estimates	t-stats
<b>Trip Attributes</b>		
Network Distance (m x 10 <sup>-5</sup> )	-0.132	-275.023
Network Distance x Winter (m x 10 <sup>-5</sup> )	-0.806	-10.509
<b>Socio-demographic Attributes</b>		
Population Density (People per m <sup>2</sup> x 10 <sup>-4</sup> )	0.106	7.928
Job Density (Number of Jobs per Person)	0.592	36.649
Establishment Density (per m <sup>2</sup> x 10 <sup>-4</sup> )	0.170	13.306
<b>Bicycle Infrastructure and Transportation Attributes</b>		
Station's Capacity	1.621	51.437
Length of Bicycle Facility in 250m Buffer (m x 10 <sup>-4</sup> )	0.586	54.288
Length of Street in 250m Buffer (m x 10 <sup>-4</sup> )	-0.042	-2.758
Number of Neighboring Stations in 250m Buffer (x10 <sup>-1</sup> )	0.319	4.895
Capacity of Neighboring Stations in 250m Buffer (x10 <sup>-3</sup> )	-2.702	-15.370
Number of Subway Stations and Bus Stops in 250m Buffer	0.076	9.023
<b>Land Use and Built Environment Attributes</b>		
Transit Score (x10 <sup>-2</sup> )	1.780	39.039
Non-motorized vehicle score (x10 <sup>-2</sup> )	5.088	104.465
Number of Restaurants and sidewalk cafe in 250m Buffer	0.231	11.962
Park Area in 250m Buffer (m <sup>2</sup> x 10 <sup>-6</sup> )	0.135	3.467
Number of Facilities in 250m Buffer (x10 <sup>-3</sup> )	3.318	35.678
Number of Recreational Facilities in 250m Buffer (x10 <sup>-3</sup> )	1.245	15.281
Distance to Times Square (m x 10 <sup>-5</sup> )	-16.636	-168.059
Elevation (m x10 <sup>-3</sup> )	-4.675	-49.331
Commercial Area in 250m Buffer (m <sup>2</sup> x 10 <sup>-6</sup> )	0.195	9.853
<b>Satiation Parameters</b>		
Distance to Times Square (m x 10 <sup>-5</sup> )	7.723	143.405
<b>Log-Likelihood at Convergence</b>	-1376961.379	
<b>Number of Observations</b>	2870	