Transport Networking Companies Demand and Flow Estimation in New York City

Bibhas Kumar Dey*
Transportation Modeler
Connetics Transportation Group
Tel: 1-407-927-6588; Fax: 1-407-823-3315
Email: bibhas.ce@knights.ucf.edu

Sudipta Dey Tirtha
Doctoral Student
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-543-7521
Email: sudiptadeytirtha2018@knights.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

Date: February 11, 2021

*Corresponding author

Word count: 222 abstract + 6600 texts + 992 references + 3 tables x 250 words (each) = 8,494 equivalent words

ABSTRACT
Given the burgeoning growth in transport networking companies (TNC) based ride hailing systems and their growing adoption for trip making, it is important to develop modeling frameworks to understand TNC ride hailing demand flows at the system level. We identify two choice dimensions: (1) a demand component that estimates origin level TNC demand at the taxi zone level and (2) a distribution component that analyzes how these trips from an origin are distributed across the region. The origin level demand is analyzed using linear mixed models while flows from origin to multiple destinations is analyzed using a multiple discrete continuous model system (MDCEV). The data for our analysis is drawn from New York City Taxi & Limousine Commission (NYTLC) for twelve months from January through December 2018. For our analysis, we examine weekday morning peak hour demand and distribution patterns. The model components are developed using a comprehensive set of independent variables. The model estimation results offer very intuitive results for origin demand and distribution of flows across destinations. We validated the model by predicting trips to destination taxi zones and found that predicted model performs well in identifying high preference destination zones. In addition, elasticity effects are computed by evaluating the percentage change in baseline marginal utility in response to increasing the value of exogenous variables by 10%, 25% and 50% respectively.

Keywords: Ride hailing, Taxi zone level demand, Flow distributions, Linear mixed model, Multiple discrete continuous extreme value

BACKGROUND AND MOTIVATION

Ride hailing services have been available as a mode of transportation since the early 17th century in the form of horse-drawn hackney carriages in Europe. With the advent of the automobile, taxis for hire have been the most common ride hailing transportation alternative. However, ride hailing has undergone a rapid transformation in the recent few years in response to the transformative technological changes including smart mobile availability, ease of hailing a ride using mobile applications, integration of seamless payment systems and real-time driver and user reviews. In fact, the convenience offered by transport networking companies (TNC) such as Uber, Lyft, and Via has allowed for a tremendous growth in ride hailing demand. For example, in New York City, the average daily trips by taxi (yellow taxi) was varying between 400 thousand and 500 thousand for the years 2010 and 2014 (*1*). However, since 2014, with the advent TNC services in the city, the total number of trips have increased. Specifically in 2018, the daily trips have increased to more than a million trips with traditional taxi accounting for nearly 300 thousand trips, and TNC services accounting for 700 thousand trips. These trends are not specific to New York City. A recent report analyzing reimbursed travel in the US has found that the share of Uber and Lyft has increased from 8% to 72.5% within 2014-2018 at the cost of taxi and rental car business share (*2*). The prevalence of TNC services is also not restricted to US. Uber operates in over 60 countries, while Didi Express in China, Ola in India currently capture a large share of the ride hailing market in these countries. The immense growth in market share and the spread of these services across the world illustrate how the ride hailing market has undergone a rapid transformation in a short time frame.

The rapid transformation of the ride hailing market coupled with emerging shared mobility service expansions (such as Carshare, Bikeshare, and Scooter share) offers an unprecedented opportunity to address the existing mobility shortcomings in urban regions (as highlighted in a recent TCRP report (*3*). 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 to provide holistic mobility options in denser urban regions. Specifically, dense urban regions with well-connected public transit systems can strategically target reducing the reliance on private automobile ownership (and use) by incorporating ride-hailing alternatives in trip planning tools. Further, by examining the spatio-temporal ride hailing data, transit agencies and shared mobility platforms can identify urban pockets with service needs to provide last mile connectivity. Towards understanding these patterns it would be beneficial to understand TNC demand and its spatial distribution in the region.

The current research effort contributes to this goal by developing quantitative models of TNC demand and flow distribution patterns. Using data from the NYC Taxi and Limousine commission, we conduct a comprehensive analysis of morning peak hour ride hailing data from Uber, Lyft, Juno and Via from 2018. The study develops (1) a demand component that estimates origin level TNC demand at the taxi zone level and (2) a distribution component that analyzes how these trips from an origin are distributed across the region. The former component is analyzed using linear mixed models and the latter component is analyzed using a multiple discrete continuous model system. The model components are developed using a comprehensive set of independent variables including aggregate trip attributes, transportation infrastructure variables, land use and built environment variables, weather attributes, and temporal attributes. The model estimates are validated using a hold out sample. Further, a policy exercise is conducted to illustrate how the proposed model system can be utilized for evaluating the impact of changes to independent variables.

EARLIER RESEARCH AND CURRENT STUDY

Ride hailing in its traditional form has received attention from various researchers (for example see (*4*) for detailed literature review of traditional taxi services). The research on TNC services is an emerging topic of interest in several fields including computer science, transportation, economics, and social sciences. In our analysis, we restrict ourselves to literature on TNC systems that are directly relevant from a transportation perspective.

Earlier research efforts focused on TNC ride hailing can be grouped into two streams. The first stream of studies explored TNC evolution, factors that affected usage, licensing and policy formulation, system level interaction frameworks, pricing mechanisms, and comparison across ride hailing services (with taxis or between various smart phone based ride hailing companies). These studies typically rely on questionnaire interviews, and online surveys for data collection. TNC evolution studies focused on the definition of ride hailing systems, how ride hailing services have evolved over time (*5-7*), investigated the challenges and opportunities presented by real-time services and highlighted various opportunities for future (*8; 9*). Djavadian and Chow (*10*) developed an agent based framework to identify social optimum considering a pricing criterion within a two-sided market system. Zha et al. (*11*) conducted an aggregate economic analysis of ride-sourcing markets where customers and drivers are matched using an exogenous function. The authors provide guidance to regulators on the mechanisms to improve social welfare. A TCRP report (*3*) examining shared modes of travel (such as bikesharing, carsharing, and TNC systems) by conducting surveys and interviews across seven urban regions (Austin, Boston, Chicago, Los Angeles, San Francisco, Seattle, and Washington, DC). The study concluded that individuals who adopt shared modes for their travel needs are more open to public transit alternatives. Further, these shared modes can serve as complementary modes to public transit. A set of studies explored the influence of various factors affecting TNC usage. For example, Cramer and Krueger (*12*) analyzed passenger service times for Uber and taxi across five major cities in the US. The authors concluded that availability of driver-passenger reviews, Uber's flexible labor supply model coupled with inefficient taxi regulations for passenger safety contributed to higher Uber utilization rates. Nie (*13*) also examined the competition between taxi industry and TNC and interestingly found that taxi industry in Shenzhen, China survived the emergence of ridesourcing. Rayle et al. (*14*) conducted a trip intercept survey to understand the source of TNC demand and concluded that nearly 50% of the demand is transferred from public transit and driving. Multiple studies explored pricing strategies employed by various ride hailing companies (*15-17*). Studies examining Uber surge pricing strategies, concluded that surge pricing has a negative impact on demand. Smart et al. (*18*) compared the performance of Uber and taxi services in terms of waiting time and cost using survey of riders in low income neighborhoods in Los Angeles. The data analysis found that Uber offered lower waiting times and provided service at a lower cost (even under surge pricing).

A second stream of studies conducted quantitative analysis using TNC usage data exploring trip patterns (a) to identify factors influencing TNC demand, (b) to understand TNC demand and its relationship with existing transportation modes. Earlier research has found that Uber demand is affected by temporal and weather patterns (*19; 20*). Other factors that were found to affect ride hailing demand include land use attributes such as lower transit access time (TAT), higher length of roadways, lower vehicle ownership, higher income and more job opportunities (*21-23*). Studies comparing the emerging ride hailing services with existing services such as public transit and bicycle sharing offer interesting results. Gerte et al. (*24*) found evidence for shifting taxi demand to smart phone based ride hailing services in New York City. Further, the study also found evidence of substitution relationship between ride hailing and bicycle share systems. Dey et al. (*25*) also studied the impact of various factors on shifting NYC's TNC services demand from traditional (yellow and green) taxi services.

Komaduri et al. (*26*) analyzed data from RideAustin, to examine the trip length and temporal distribution of the trips. A comparison of the adoption of RideAustin relative to public transit alternatives illustrated that individuals were choosing RideAustin to minimize travel time (highlighting the higher value of time for these travelers). Lavieri et al. (*27*) employed the same data to develop a two stage framework for TNC demand analysis. The study employs averaged daily TNC origin and distribution flows within a two step procedure. The model components developed include a spatially lagged multivariate count model for TNC demand and fractional split model for trip distribution. Poulsen et al. (*28*) examined how the two systems that were introduced in the same time performed - Uber and Green taxis - in Manhattan area and found that the growth rate for Uber was substantially higher. Babar and Burtch (*29*) compared the utilization rate of transit service in the US after the introduction of TNC services and found that utilization rate of bus service dropped while long-haul transit services (such as subway and commuter rail) experienced increasing utilization. The spectrum of quantitative methodologies employed in earlier studies include descriptive analysis, linear regression, logistic regression, difference in difference model and panel based random effects multinomial logit model.

Current Study in Context

The review highlights the burgeoning literature on TNC services across the world. However, given that TNCs are a very recent development several dimensions remain uninvestigated. While TNC demand has been examined in earlier research, the temporal and spatial aggregation employed in the past have not allowed for easy integration of these approaches into existing planning frameworks. Further, earlier studies have rarely examined how the TNC demand is distributed across the study region. Of the earlier research efforts Lavieri et al. (*27*) developed a two stage framework for understanding TNC flow distribution. However, the authors focused on an average model where the two dependent variables were averaged over the study period to conduct the analysis. The averaging process, while simplifying the analysis avoiding repeated measures of data, does not process the rich distributional differences across the data and thus might not be suitable for prediction applications on a daily basis.

The primary objective of our research is to develop TNC demand based planning models that can be integrated within existing frameworks or used to augment the outputs from existing demand frameworks. With this primary objective, the current study makes the following contributions. First, the current study develops a TNC demand model at the Taxi zone level for the morning peak hour (represented as pickups in the data). The demand variable is approximated as a continuous variable and a linear mixed model framework is employed to analyze the data. Second, conditional on the origin taxi zone demand, we develop a distribution model to determine TNC flows from the origin to all destinations in the study region. There are two major challenges associated with modeling the TNC flow distribution. First, the destinations for TNC 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 taxi zones in the system. The multiple discrete continuous approaches that follow Kuhn-Tucker (KT) approaches developed in literature can be adapted to address this choice dimension. In a recent study, Dey et al. (*30*) developed a similar framework for studying bicycle sharing system flows. MDCEV framework employed in this study has several advantages over the alternative approaches (such as fractional split model or the traditional trip based model). First, the MDCEV model allows us to capture for satiation effects – i.e. as more trips are destined to a zone, there is a drop in the value gained for subsequent trips. The accommodation for such zones can account for potential challenges with high demand to a zone such as unavailability of TNC services. Second, the fractional split model allocates a proportion to an alternative as a function of exogenous variables. Given the functional form, it is theoretically possible that some probability is allocated to each alternative (however small). However, in the presence of a large number of alternatives – as is the case in our context – the proportion allocated to these potentially unchosen alternatives could amount to be a significant value. Thus, it might be necessary to adopt a two-step model where a binary model determines whether a zone is chosen or not and then for these chosen alternatives, a proportion is assigned. The reader would note that destination preferences can also be modeled employing a disaggregate trip level model (such as Faghih-Imani and Eluru (*31*) for bikeshare). However, in the absence of any individual specific characteristics an aggregate model reduces the data burden while offering similar insights.

The data for our analysis from January 2018 through December 2018 is drawn from NYC Taxi & Limousine Commission (NYTLC). The data provides taxi zonal level daily origin demand and the corresponding flow patterns from the origin to all destinations across the system. The two model components were developed using a host of independent variables including trip attribute, transportation infrastructure variables, land use and built environment variables, weather attributes, and temporal attributes. The model estimation results for the proposed model offers intuitive results. The proposed model was also validated using a hold-out sample and prediction exercise is undertaken.

DATA

Data Source

New York City with high residential density and large tourist population is an ideal market for ride hailing systems. The NYC Taxi and Limousine Commission (TLC) provides spatially aggregated trip data from all ride hailing companies (taxi, Uber, Lyft, Juno and Via) for public use (https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page). The trip itinerary dataset for 2018 for Uber, Lyft, Juno and Via was processed to obtain daily morning peak hour TNC usage patterns. The dataset provides information on start and end time of trips, origin and destination defined as taxi zone ID, trip distance and vehicle license number. 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).

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 500 minutes in duration (around 0.5% of all trips) were deleted considering that these trips are not typical ride-sharing trips. These trips could also be a result of two possibilities; either destination of those trips could be outside NYC or due to technical issues the trip information was recorded incorrectly. Third, trips that had the origin and destination outside of NYC taxi zone were also eliminated. Therefore, we focus on trips that originated and were destined within NYC taxi zone region only.

For the given study period (January 2018 to December 2018), the total number of available taxi zones in NYC was 260. Initially, we aggregated morning peak (6.30 am-9.30am) trip data for each day for each week (total 52 weeks) from each origin taxi zone ID to every possible destination taxi zone ID (260). The average number of daily trips generated and attracted at each taxi zone is presented in Figure 1. In Figure 1, the number of trips generated (Figure 1a) and attracted (Figure 1b) to each taxi zone is categorized into multiple classes from very low to very high. The figures clearly highlight the high TNC usage in Manhattan and airport locations (LaGuardia, John F. Kennedy International Airport and Newark airport).

For our analysis, to ensure that holiday weekends that are likely to have a different user patterns do not influence our analysis, we selected morning peak period trip data for 43 weeks without any holidays. The processing of the large sample of trip data is substantially time-consuming and significantly increases the model run times. To obtain a reasonable sample size for model estimation, we sampled following two steps; 1) 150 taxi zones were selected randomly from the total 260 taxi zones and 2) for each taxi zone one weekday was randomly selected for each week.

Thus, the data sampled had 150 taxi zone with 43 weekday morning peak trip data during 2018. While the data considered is a sample, the consideration of a reasonably large sample size (6450) allows for robust model estimations. We organized the dataset into two components for our analysis; 1) For zonal level origin demand (aggregating total daily morning peak trip at the origin level) and 2) Trip distribution from origin to destination (aggregating daily morning peak trip at the O-D pair level).

Independent Variable Generation

Several independent variables were generated in our study (see Table 1). These can be grouped into five categories: 1) Trip attribute, 2) Transportation infrastructure variables, 3) Land use and built environment variables, 4) Weather attributes, and 5) Temporal attribute.

Trip attribute includes the network distance between each origin-destination taxi zone 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. The variable will serve as a surrogate for travel time. As all the data is for morning peak, the impact of congestion is likely to be affecting all records similarly.

Transportation infrastructure attributes created at the taxi zone level include bike route length density (capturing the effect of availability of bicycle facilities on system usage), number of bikeshare stations, length of streets (minor and major streets). Number of subway stations and bus stops in the taxi zone were generated to examine the influence of public transit on rider’s preference of destination station.

Several land use and built environment variables were considered including population density, job density and establishment density, the number of institutional 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 each taxi zones. Distance of destination from Times Square and airport were estimated by using the shortest path algorithm tool of ArcGIS software. Airport indicator variable for the taxi zone was generated to examine the additional impact of airport destination. Population, job density and median income information was collected from US Census for 2014-2017 and extrapolated for 2018 at the census tract level considering average yearly population change from 2014-2017. Household car ownership information for 2018 was used to generate proportion of zero car ownership at taxi zone level to examine the impact of car ownership on riders’ destination preferences. Non-motorized vehicle score (average of walk score and bike score) and transit score associated with each taxi zone was considered at the census tract level. Further, crime density and accident density were also generated at taxi zone level. Total number of crimes of all types for previous year (2017) was aggregated at census tract level and crime density was estimated by dividing with the corresponding year’s population. In a similar manner, total number of accidents of all kind for each day of 2018 was considered to generate accident density.

Weather variables include average temperature, precipitation, and snow for that particular day. Several interaction variables were also created. Seasonality is the only temporal variable considered. We consider winter (December-February), Spring (March-May), Summer (June-August) and Fall (September-November) as dummy variables.

Descriptive Analysis

The data at an aggregate system level in the form of average number of trips by taxi zone for each week is presented in Figure 2. The various weeks with lower demand correspond to the weeks with holidays supporting our hypothesis that these weeks have a different demand pattern. The dependent variable distribution is generated to understand origin level demand and distribution of these flows across the study region. On average, 384 trips depart from each origin taxi zone in the morning peak hour and are destined to about 67 alternative taxi zones. The sample characteristics of the independent variables generated are presented in Table 1.

ECONOMETRIC FRAMEWORKS

Linear Mixed Model for Station Level Weekly Origin Demand

The taxi zonal level daily pick up 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 daily peak hour demand at the same taxi zone for fourty-three 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 $w = 1, 2, \dots, W$ be an index to represent each taxi zone ($W = 150$), $M = 1, 2, \dots, 43$ be an index to represent the various day of weeks of data compiled for each pick up taxi zone. The dependent variable (daily peak hour demand) is modeled using a linear regression equation which, in its most general form, has the following structure:

$$y_{mw} = \beta X_{mw} + \varepsilon_{mw} \tag{1}$$

where y_{mw} is the natural logarithm of weekly demand, X is an $K \times 1$ column vector of attributes and the model coefficients, β, is an $K \times 1$ column vector. The random error term, ε_{mw}, is assumed to be normally distributed across the dataset. In our analysis, the repetitions over days 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 43 random days from a sample of 43 weeks for each taxi zone, we decided to employ a simpler covariance structure. The exact functional form of the covariance structure assumed is shown below:

$$\Omega = \begin{pmatrix} \Omega^2 + \Omega_1^2 & \Omega_1 & \dots & \Omega_1 \\ \Omega_1 & \Omega^2 + \Omega_1^2 & \dots & \Omega_1 \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_1 & \Omega_1 & \dots & \Omega^2 + \Omega_1^2 \end{pmatrix} \tag{2}$$

The covariance structure restricts the covariance across all fourty-three records to be the same. The parameters estimated in this correlation structure are Ω and Ω_1 . The parameter Ω represents the error variance of ε, Ω_1 represents the common correlation factor across daily 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 (*32*).

MDCEV Model for Destination Choice

According to Bhat et al. (*33*), we consider the following functional form for utility in this paper, based on a generalized variant of the translated Constant Elasticity of Substitution (CES) utility function:

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

where $U(x)$ is a quasi-concave, increasing, and continuously differentiable function with respect to the consumption quantity (Ix1)-vector ($x_i \geq 0$ for all i), and λ_i associated with drop off taxi zone i. λ represents the baseline marginal utility ($\lambda_i > 0$ for all i), γ is a translation parameter (γ should be greater than zero) which enables corner solutions while simultaneously influencing satiation and α influences satiation ($\alpha \leq 1$).

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

$$\lambda\,(y_{iw}, \rho_{iw}) = \exp\,(\delta y_{iw} + \rho_{iw}) \quad (4)$$

where y_{iwq} is a set of attributes characterizing drop off taxi zone i during day w, δ corresponds to a column vector of coefficients, and ρ_{iw} captures idiosyncratic (unobserved) characteristics that impact the baseline utility for destination stations. The overall random utility function of Equation (3) then takes the following form:

$$U(x) = \sum_{i=1}^{I} \frac{\gamma}{\alpha} \exp\,(\delta y_{iw} + \rho_{iw}) \left\{ \left(\frac{x_i}{\gamma} + 1 \right)^{\alpha} - 1 \right\} \quad (5)$$

Following (*34; 35*), consider a generalized extreme value distribution for ρ_i and assume that ρ_{iw} is independent of y_{iw} $(i = 1,2, \ldots, I)$. The ρ_{iw}'s are also assumed to be independently distributed across alternatives with a scale parameter normalized to 1. Due to the common role of γ and α, it is very challenging to identify both γ and α in empirical application (see (*35*) for detailed discussion). Hence, either γ or α parameter is estimated. When the α - profile is used, the utility simplifies to:

$$U(x) = \sum_{i=1}^{I} \frac{1}{\alpha} \exp\,(\delta y_i + \rho_i)\{(x_i + 1)^{\alpha} - 1\} \quad (6)$$

When the γ - profile is used, the utility simplifies to:

$$U(x) = \sum_{i=1}^{I} \gamma \exp(\delta y_i + \rho_i)\, ln\left(\frac{x_i}{\gamma} + 1 \right) \quad (7)$$

In this study, γ - profile is used. Finally, the probability that an pick up taxi zone has flows to the first D drop-off taxi zones $D \geq 1$ is:

$$P(e_1^*, e_2^*, e_3^*, \ldots, e_D^*, 0,0, \ldots, 0) = \left[\sum_{n=1}^{D} d_n \right] \left[\sum_{n=1}^{D} \frac{1}{d_n} \right] \left[\frac{\prod_{n=1}^{D} e^{U_n}}{(\prod_{d=1}^{K} e^{U_i})^{D}} \right] (D-1)! \quad (8)$$

where $(\sum_{n=1}^{D} m_n)(\sum_{n=1}^{D} {}^{1}/_{m_n})$ is defined as Jacobian form for the case of equal unit prices across goods (*35*) where, $m_n = \left(\frac{1-\alpha}{e_n^* + \gamma} \right)$.

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).

ESTIMATION RESULTS

The model estimation results from the two models are discussed – TNC demand model followed by the trip distribution model results. The reader would note that variable selection was guided by earlier research on emerging modes of transportation such as bikeshare and TNC. From the universal set of variables we tested, variables that were significant at 95% confidence interval and offered intuitive results were included in the models.

Trip Demand Model

Model Fit Measures

A linear regression model was estimated at first as benchmark for evaluating the linear mixed model. To compare these two models, a Log-likelihood ratio (LR) test was computed. The LR value was found to be 1915 which was higher than any corresponding chi-square value for 2 degrees of freedom. 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.

Linear Mixed Model Results

The linear mixed model estimation results for morning peak hour TNC origin demand are presented in Table 2. The model estimation results offer intuitive findings. TNC demand, as expected is positively associated with population density. Increased median income of households within the taxi zones is found to increase demand for TNC trips (see (*18; 22*) for similar results). The presence of airport in the taxi zone also contributes to increased TNC demand. Higher number of trips are likely to be generated from taxi zones with higher population than lower populated zones. The presence of different institutional facilities (such as schools, colleges, hospitals, and office) in the taxi zones increases the zonal demand. The presence of discretionary opportunities such as a higher presence of restaurants and sidewalk café also drives TNC demand. Taxi zones with higher proportion of residential area is positively associated with Peak hour morning TNC flows. The result illustrates the adoption of TNC service for morning commute activities from these zones. The results for precipitation variables highlight that in the presence of precipitation individuals are likely to make a trip via TNC services (see (*19*) for similar result). The results also indicate a positive influence of summer and fall season compared to winter and spring season. The finding is in line with earlier research (*19*). The result is also possibly reflecting the increased tourist activity during these seasons.

Correlation Parameters

In the linear mixed model, we estimate a parameter that recognizes the repeated measures of data for each taxi zone. The correlation parameter is statistically significant highlighting the role of common unobserved factors influencing the demand from taxi zones.

TNC Distribution Model

Model Fit Measures

The final log-likelihood values for the estimated MDCEV model and equal probability MDCEV model are -1531122.801 and -1712633.216 respectively. The log-likelihood ratio (LR) test-statistic of comparison between the final model and the equal probability model is 363020.830. The LR test-statistic value is significantly higher than the corresponding chi-square value for 22 additional degrees of freedom highlighting that the MDCEV distribution model offers a reasonable fit.

MDCEV Model Results

The model results of TNC morning peak hour distribution model are presented in Table 3. The presentation of results is organized by the various variable categories. The reader would note that a single utility equation is estimated for all the destination zones (analogous to location choice model estimation for large number of alternatives). A positive (negative) coefficient indicates an increase (decrease) in the variable results in increasing the utility of the alternative destination.

<u>Land Use and Built Environment Attributes</u>

Zones located in census tracts with higher population density are more likely to be chosen as destination locations. Similarly, job density also impacts destination preference positively. The results together point towards the adoption of TNC services for daily commute trips (see (Correa et al. (*22*) for similar result). Taxi zones with high income are preferred destination zones for TNC services. The model parameter for taxi zone level zero car household proportion highlights the increased adoption of TNC services among these zones (Correa et al. (*22*) found similar association with lower vehicle ownership households).

As expected, increased transit accessibility within a taxi zone increases the likelihood of the zone being chosen as a destination. On the other hand, the results indicate that zones with higher non-motorized score are less preferred destinations. While the result seems counterintuitive, it might be alluding to potential competition between TNC ride hailing and bicycle sharing systems in these zones. The presence of activity opportunities in the forms of restaurants and cafes, institutional facilities, and recreational centers and point of interests (POI) are positively associated with the destination zone preference. Taxi zone with higher commercial area serves as an attraction for TNC demand. The increase in land use mix value (range between 0 and 1) has a positive impact on destination zone preference.

The presence of airport in the destination taxi zone, as expected, increases the preference for the zone. The model also considers the influence of another major landmark in the region - Times Square. The parameter indicates that as the taxi zone is further from Times Square the preference of the zone as a destination reduces. The result illustrates how Times Square and its proximal zones serve as attraction centers for regular and tourist travel.

<u>Trip Attributes</u>

In the current research context, a negative coefficient was obtained for network distance of O-D pair. With the increasing distance to the destination, TNC demand distribution propensity reduces.

<u>Transportation Infrastructure and Attributes</u>

Several transportation infrastructure variables were considered in the demand distribution models. Of these variables, bike lane density, bikeshare stations, street length, bus stops and subway stations presented significant impacts on destination preferences. Taxi zones with higher bike length density (defined as ratio of bike length to overall roadway length) reduce the preference for the destination zone. The negative association with number of bikeshare stations within a taxi zone highlights that TNC demand is likely to be lower for a destination zone with more bikeshare stations. It is possible that the result alludes to potential competition between these modes for the last mile connectivity. It would be interesting to explore these differences further in future studies. An increase in the street length within the destination zones results in an increased likelihood of the zone being chosen as destination (similar to findings of Correa et al. (*22*)). As the number of bus stops and subway stations in the taxi zone increases, we observe increased preference for that destination. The coefficient actually indicates a potential complementarity between TNC flows and transit flows. TNC users might use public transit for large portions of their trip and then use TNC for their final travel to the destination.

<u>Temporal and Weather Attributes</u>

The reader would note that temporal and weather attributes cannot be considered directly in destination distribution model. Hence, we interacted these variables with destination specific variables such as network distance and distance to Times Square. The results offer interesting results. In Winter, the negative influence of network distance increases further indicating that shorter trips are preferred (relative to other moths). The temperature variable interacted with network distance indicates that the influence of network distance is moderated by higher temperature i.e. as temperature increases the negative impact of network distance reduces. The precipitation variable interacted with network distance and distance to Times Square highlights the increase in sensitivity to travel time under precipitation conditions. The weather variables as a whole highlight how TNC distance impact is lower in good weather relative to poor weather.

<u>Satiation Parameter</u>

We used distance to Times Square from taxi zones as a satiation parameter. In MDCEV model, the satiation parameter captures the extent of decrease in marginal utility across different destination zones. The satiation parameter is statistically significant at 95% confidence level, thereby implying that there are clear satiation effects in destination choice as distance of destination from Times Square increases. To elaborate, as the zone is further away from Times Square, the satiation impacts are higher indicating fewer trips will be made to the zone.

VALIDATION ANALYSIS RESULTS

For validation purpose, a hold-out sample was prepared following the same procedure used to extract the estimation sample. After extracting 150 taxi zones for our base dataset, the remaining 110 taxi pick up zones were set aside for validation. Then we randomly chose 43 days from 43 corresponding weeks throughout the year for these 110 zones. The same approach of data preparation employed for estimation sample is exercised for validation sample (110 origins x 43 days x 260 destinations). Using the validation data, the model results from the estimation sample were used to generate a prediction measure in the form of predictive log-likelihood. The difference in the log-likelihood for the predicted and equal probability model is 3626720.830 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 day at disaggregate level. Note that, zero trips to any destination for a week was also considered. To identify the preferred destination zones, top 10 percentile of preferred destination zones was captured for each pickup zone and validated with the top 10 percentile predicted destination zones. For the performance evaluation, we compute the correctly classified predicted trips for top 10 percentile destined zones for each taxi zone considering the total trips throughout the year. The reader would note that about 71% of the top destination zones were correctly classified. To provide a visual representation, we selected 5 random taxi zones from 5 NYC boroughs and predicted the top 10 percentile destination zones for them considering average daily morning peak hour trips throughout the year and compared them with observed top destination zones for that particular zone (See Figure 3). Across the five boroughs, based on the observed and predicted measures from the Figure, taxi zones situated in Brooklyn offered the best prediction performance while taxi zone from Staten Island has inferior prediction performance. Overall, the two validation exercises, highlight the applicability of the proposed approach for TNC demand and distribution prediction.

POLICY ILLUSTRATION

The model results from Table 3 provide an indication of how the exogenous variables affect the network flows considering destination choice. However, they cannot provide the exact magnitude of the effect of these exogenous variables. Hence, elasticity effects computation considering changes of baseline marginal utility was used to evaluate the impact of exogenous variables on destination choice. The elasticity effects are computed by evaluating the percentage change in marginal utility of an alternative in response to increasing the value of exogenous variables from best fit model by 10%, 25% and 50% respectively. We selected five independent variables for presentation including job density, median income, network distance, institutional facilities and bus stops and subway stations. The computed elasticities are presented in Figure 4. Based on elasticity effects results in Figure 4, following observations can be made. *First,* the elasticity estimate for job density variable indicates that about 6.5, 17 and 37% increase in utility happens due to 10, 25 and 50% change in the independent variable. All the other results can be interpreted similarly. *Second*, rank order of the top three significant variable in terms of changes for the utility without considering positive or negative impact include network distance, job density and median income. *Third*, network distance between O-D can be considered as a proxy for travel time. The increasing value of this variable provides a snapshot of the impact of additional travel time due to traffic congestion or other safety incidents. Overall, the elasticity analysis results provide an illustration on how the proposed model can be applied to determine the critical factors contributing to increase in utility to choose a taxi zone as destination.

CONCLUSIONS

Given the burgeoning growth in ride hailing systems and their growing adoption for trip making, it is important to develop modeling frameworks to understand ride hailing demand flows at the zonal level. The current research effort contributes to this goal by developing quantitative models of TNC demand and flow distribution patterns. We identify two choice dimensions: (1) a demand component that estimates origin level TNC demand at the taxi zone level and (2) a distribution component that analyzes how these trips from an origin are distributed across the region. The origin level demand is analyzed using linear mixed models while flows from origin to multiple destinations is analyzed using a multiple discrete continuous model system (MDCEV).

The data for our analysis is drawn from New York City Taxi & Limousine Commission (NYTLC) for twelve months from January through December 2018. For our analysis, we examine weekday morning peak hour demand and distribution patterns. The model components are developed using comprehensive set of independent variables including aggregate trip attributes, transportation infrastructure variables, land use and built environment variables, weather attributes, and temporal attributes. The model estimation results provide intuitive findings for both zonal level demand and flow distribution behavior. The model estimates are validated using a holdout sample set aside. The data fit relative to the equal probability MDCEV model highlighted the significant improvement in data fit for the estimated model. Several prediction exercises were also conducted to illustrate the value of the proposed model framework including identifying the top 10 percentile destinations and elasticity effect of changes to independent variables. The policy analysis results offer intuitive results and provide a mechanism for transportation planners to evaluate the impact of various changes on TNC demand and distribution.

The framework developed can be employed by planning agencies to evaluate how TNC originate in the region and their distribution. The model framework can be employed to evaluate how TNC flows in the future evolve as a function of various attributes. The future TNC flow prediction can be used to explore TNC flow inequity and potential mobility impacts of transportation infrastructure. The reader would note that while estimating the distribution component model (MDCEV) might be involved for practitioners, its application for prediction is not as involved and is relatively easier with tools available in open source platforms such as R.

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 this context, another potential avenue for future research is the consideration of sampling for MDCEV models (similar to sampling in MNL models). It might also be interesting to evaluate the proposed approach with the approach proposed in Lavieri et al. (*27*), and/or the traditional trip distribution approaches. Empirically, several improvements can be considered in future research. It would be useful to examine the bikeshare flows between various zones and their impact on TNC flows. The model developed might also benefit from the consideration of transit connectivity between taxi zones in the region.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: Naveen Eluru, Bibhas Kumar Dey; data collection: Bibhas Kumar Dey; model estimation and validation: Bibhas Kumar Dey; analysis and interpretation of results: Bibhas Kumar Dey, Sudipta Dey Tirtha, Naveen Eluru; draft manuscript preparation: Bibhas Kumar Dey, Sudipta Dey Tirtha, Naveen Eluru. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] NYC Taxi & Limousine Commission. (2019, March). Taxi and Ride hailing Usage in New York City. Retrieved July 2019, from https://toddwschneider.com/dashboards/nyc-taxi-ridehailing-uber-lyft-data/.

[2] Richter W. (2018, July 30). Uber and Lyft are gaining even more market share over taxis and rentals. Retrieved July 2019, from https://www.businessinsider.com/uber-lyft-are-gaining-even-more-market-share-over-taxis-and-rentals-2018-7.

[3] Feigon S, Murphy C. Shared mobility and the transformation of public transit. Transportation Research Board, Washington, D.C. 2016. http://www.trb.org/Main/Blurbs/174653.aspx.

[4] Faghih-Imani A, Anowar S, Miller EJ, Eluru N. Hail a cab or ride a bike? A travel time comparison of taxi and bicycle-sharing systems in New York City. Transportation Research Part A: Policy and Practice. 2017 Jul 1;101:11-21.

[5] Chan ND, Shaheen SA. Ridesharing in North America: Past, present, and future. Transport reviews. 2012 Jan 1;32(1):93-112.

[6] Sun C, Edara P. Is getting an Uber-Lyft from a sidecar different from hailing a taxi? Current dynamic ridesharing controversy. Transportation Research Record. 2015 Jan;2536(1):60-6.

[7] Furuhata M, Dessouky M, Ordóñez F, Brunet ME, Wang X, Koenig S. Ridesharing: The state-of-the-art and future directions. Transportation Research Part B: Methodological. 2013 Nov 1;57:28-46.
[8] Agatz N, Erera A, Savelsbergh M, Wang X. Optimization for dynamic ride-sharing: A review. European Journal of Operational Research. 2012 Dec 1;223(2):295-303.
[9] Amey A, Attanucci J, Mishalani R. Real-time ridesharing: opportunities and challenges in using mobile phone technology to improve rideshare services. Transportation Research Record. 2011;2217(1):103-10.
[10] Djavadian S, Chow JY. An agent-based day-to-day adjustment process for modeling 'Mobility as a Service'with a two-sided flexible transport market. Transportation research part B: methodological. 2017 Oct 1;104:36-57.
[11] Zha L, Yin Y, Yang H. Economic analysis of ride-sourcing markets. Transportation Research Part C: Emerging Technologies. 2016 Oct 1;71:249-66.
[12] Cramer J, Krueger AB. Disruptive change in the taxi business: The case of Uber. American Economic Review. 2016 May;106(5):177-82.
[13] Nie YM. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. Transportation Research Part C: Emerging Technologies. 2017 Jun 1;79:242-56.
[14] Rayle L, Dai D, Chan N, Cervero R, Shaheen S. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. Transport Policy. 2016 Jan 1;45:168-78.
[15] Chen L, Mislove A, Wilson C. Peeking beneath the hood of uber. InProceedings of the 2015 internet measurement conference 2015 Oct 28 (pp. 495-508).
[16] Chen MK, Sheldon M. Dynamic pricing in a labor market: Surge pricing and the supply of Uber driver-partners. University of California (Los Angeles) Working Paper URL http://citeseerx. ist. psu. edu/viewdoc/download. 2015 Nov 3.
[17] Guo S, Liu Y, Xu K, Chiu DM. Understanding ride-on-demand service: Demand and dynamic pricing. In2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) 2017 Mar 13 (pp. 509-514). IEEE.
[18] Smart R, Rowe B, Hawken A. Faster and cheaper: How ride-sourcing fills a gap in low-income Los Angeles neighborhoods.
[19] Brodeur A, Nield K. Has uber made it easier to get a ride in the rain?
[20] Gerte R, Konduri KC, Eluru N. Is there a limit to adoption of dynamic ridesharing systems? Evidence from analysis of Uber demand data from New York City. Transportation Research Record. 2018 Dec;2672(42):127-36.
[21] Alemi F, Circella G, Handy S, Mokhtarian P. What influences travelers to use Uber? Exploring the factors affecting the adoption of on-demand ride services in California. Travel Behaviour and Society. 2018 Oct 1;13:88-104.
[22] Correa D, Xie K, Ozbay K. Exploring the taxi and Uber demand in New York City: An empirical analysis and spatial modeling. In96th Annual Meeting of the Transportation Research Board, Washington, DC 2017.
[23] Davidson A, Peters J, Brakewood C. Interactive travel modes: Uber, transit, and mobility in New York City. 2017.
[24] Gerte R, Konduri KC, Ravishanker N, Mondal A, Eluru N. Understanding the relationships between demand for shared ride modes: case study using open data from New York City. Transportation research record. 2019 Dec;2673(12):30-9.
[25] Dey BK, Tirtha SD, Eluru N, Konduri KC. Transformation of Ridehailing in New York City: A Quantitative Assessment. Working paper, Department of Civil, Construction and Environmental Engineering, University of Central Florida. 2021.

[26] Komanduri A, Wafa Z, Proussaloglou K, Jacobs S. Assessing the impact of app-based ride share systems in an urban context: Findings from Austin. Transportation Research Record. 2018 Dec;2672(7):34-46.
[27] Lavieri PS, Dias FF, Juri NR, Kuhr J, Bhat CR. A model of ridesourcing demand generation and distribution. Transportation Research Record. 2018 Dec;2672(46):31-40.
[28] Poulsen LK, Dekkers D, Wagenaar N, Snijders W, Lewinsky B, Mukkamala RR, Vatrapu R. Green cabs vs. uber in new york city. In2016 IEEE International Congress on Big Data (BigData Congress) 2016 Jun 27 (pp. 222-229). IEEE.
[29] Babar Y, Burtch G. Examining the impact of ridehailing services on public transit use. September. 2017;25:2017.
[30] Dey BK, Anowar S, Eluru N. A framework for estimating bikeshare origin destination flows using a multiple discrete continuous system. Transportation Research Part A: Policy and Practice. 2019;144:119-33.
[31] Faghih-Imani A, Eluru N. Analysing bicycle-sharing system user destination choice preferences: Chicago's Divvy system. Journal of transport geography. 2015 Apr 1;44:53-64.
[32] Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. Journal of the American statistical association. 1977 Jun 1;72(358):320-38.
[33] Bhat CR, Sen S, Eluru N. The impact of demographics, built environment attributes, vehicle characteristics, and gasoline prices on household vehicle holdings and use. Transportation Research Part B: Methodological. 2009 Jan 1;43(1):1-8.
[34] Bhat CR. A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. Transportation Research Part B: Methodological. 2005 Sep 1;39(8):679-707.
[35] Bhat CR. The multiple discrete-continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions. Transportation Research Part B: Methodological. 2008 Mar 1;42(3):274-303.

Figure 1(a) Trip generation at taxi zones
Figure 1(b) Trip attracted at destined taxi zones
Figure 1 Ride hailing trips in NYC's taxi zone level.

Figure 2 Trip Rates of TNC demand by week.

Figure 3(a) Manhattan
Figure 3(b) Brooklyn
Figure 3(c) Bronx
Figure 3(d) Queens
Figure 3(e) Staten Island
Figure 3 Top 10 percentile destined zones for randomly selected pickup zones from 5 NYC borough.

Figure 4 Elasticity effects considering change in marginal utilities.

Table 1 Descriptive Summary of Sample Characteristics

Variable Names	Definition	Descriptive Minimum	Descriptive Maximum	Descriptive Mean
Dependent Variables				
Trip Demand				
Total Trip (Daily per Origin)	Total number of daily morning peak hour trips made in an origin taxi zone	1.0	1983.0	384.0
Destination Choice				
Alternative Destination Chosen	Total number of alternative taxi zone chosen as destination	1.0	171.0	66.0
Total Trip (Daily O-D Pair)	Total number of daily morning peak hour trips destined to a taxi zone	0.0	542.0	1.5
Independent Variables				
Trip Attributes				
Network Distance (m x 10^{-6})	Shortest distance between taxi zones	0.0	55.5	2.43
Land Use and Built Environment Attributes				
Population Density	Population in the taxi zone /Total area of the taxi zone in square meters	0.0	0.6	0.1
Employment Density	Total number of jobs in taxi zone /Total number of populations in taxi zone	0.0	1.0	0.6
Median Income (10^{-6})	Median person income in taxi zone	0.0	1.6	0.7
Proportion of Zero Car HH	Zero Car Ownership HH in the taxi zone /All HH in the taxi zone	0.0	0.9	0.5
Facilities	Total number of institutional facilities in taxi zone	0.0	660.0	210.9
Point of Interests	Number of point of interests in the taxi zone	0.0	487.0	120.3
Park and Recreational Centers	Total number of park and recreational centers in the taxi zone	0.0	8.0	0.9
Restaurants	Total number of restaurants in the taxi zone	0.0	1287.0	146.3
Sidewalk Cafe	Total number of sidewalk café in the taxi zone	0.0	491.0	113.4
Theaters	Total number of theaters in the taxi zone	0.0	23.0	0.1
Commercial Area (m^2 x 10^{-6})	Total commercial area of the taxi zone in square meters	0.0	73.8	11.7
Residential Area (m^2 x 10^{-6})	Total residential area of the taxi zone in square meters	0.0	56.1	18.9
Office Area (m^2 x10^{-6})	Total office area of the taxi zone in square meters	0.0	62.3	42.9
Park Area (m^2 x 10^{-6})	Total park area of the taxi zone in square meters	0.0	57.2	5.9
Land use mix	Land use mix = $\left[\frac{-\sum_{k}(p_k(lnp_k))}{lnN}\right]$, where k is the category of land-use, p is the proportion of the developed land area devoted to a specific land-use, N is the number of land-use categories in a Taxi zone	0.0	0.9	0.3
Crime Density	Total number of per capita crimes that occurred in the previous year in the taxi zone	0.0	113.5	89.6
Accidents Density	Total number of per-capita accidents in the taxi zone	0.0	0.1	0.0

Street Length (m x 10^{-6})	Street length of all type in meter per taxi zone	0.0	0.8	0.1
Bike Lane Density	Ratio of bike length to street length	0.0	0.5	0.1
Walk Score	Walk Score in the taxi zone	0.0	100.0	90.1
Bike Score	Bike Score in the taxi zone	0.0	95.0	72.1
Transit Score	Transit Score in the taxi zone	0.0	100.0	88.1
Non-motorized Vehicle Score	Non-motorized (Walk and Bike) Score in the taxi zone	0.0	97.5	81.6
Distance to Times Square (m x 10^{-3})	Shortest Distance to Times Square in miles	0.0	43.6	2.7
Distance to Airport (m x 10^{6})	Distance to the nearest airport from each taxi zone	0.0	11.2	2.2
Transportation Infrastructure Attributes				
Bike Share Station	Total number of bikeshare stations in the taxi zone	0.0	27.0	2.2
Bus Stops	Total number of bus stops in the taxi zone	0.0	55.0	20.0
Subway Stations	Total number of subway stations in the taxi zone	0.0	14.0	2.9
Weather Attributes				
Temperature (°F)	Average temperature in a day	24.0	86.0	60.1
Precipitation	Average precipitation in a day	0.0	3.0	0.2
Snow (inch)	Average snow depth in a day	0.0	8.0	0.2
Categorical Variable	**Definition**	**Frequency (%)**		
Temporal Attributes				
Season	Spring (March-May)	30.2		
	Summer (June-August)	27.9		
	Fall (September-November)	23.3		
	Winter (December-February)	18.6		
Built Environment and Land Use Attributes				
Historic District	Presence of origin on historic district or not	29.7		
Airport Indicator	Airport within the Taxi Zone	1.2		

Table 2 Linear Mixed Model Results for TNC Origin Demand

Parameter	Estimates	t-stats
Intercept	-1.679	-3.903
Land Use and Built Environment Attributes		
Population Density	1.261	8.869
Median Income ($x10^{-6}$)	8.035	4.079
Airport as an Indicator	0.804	4.079
Number of Institutional Facilities in a Taxi Zone ($x10^{-3}$)	0.195	1.655
Number of Restaurants and Side cafe in a Taxi Zone ($x10^{-3}$)	0.316	2.803
Residential Area (m^2 $x10^{-6}$)	0.316	2.803
Temporal Attributes		
Precipitation (cm)	3.740	26.106
Season: Summer and Fall (Base: Winter and Spring)	1.548	8.574
Correlation Parameters		
Ω	5.253	56.116
Ω_1	3.776	8.429
Restricted Log-Likelihood	37161.892	
Sample Size	6450	

Table 3 MDCEV Model Results

Parameter	Estimates	t-stats
Land Use and Built Environment Attributes		
Population Density	0.462	22.824
Job Density	1.122	45.023
Median Income ($x10^{-6}$)	5.445	67.210
Proportion of Zero Car HH	1.376	78.465
Transit Score ($x10^{-2}$)	0.958	30.103
Non-motorized vehicle score ($x10^{-2}$)	-1.807	-51.698
Number of Restaurants and sidewalk café in Taxi Zone ($x10^{-3}$)	0.438	42.622
Number of Institutional Facilities in Taxi Zone ($x10^{-3}$)	0.194	8.528
Number of Point of Interests and Recreational Points in Taxi Zone ($x10^{-3}$)	1.401	41.801
Commercial Area (m^2 $x10^{-6}$)	1.641	87.265
LU Mix	0.723	35.999
Airport Indicator	3.702	335.179
Times Square Distance (m x 10^{-3})	-0.378	-66.091
Trip Attributes		
Network Distance (m x 10^{-3})	-2.547	-174.790
Transportation Infrastructure and Attributes		
Bike Lane Density in Taxi Zone	-0.730	-22.787
Number of Bikeshare Stations in Taxi Zone ($x10^{-2}$)	-0.108	-26.258
Street Length in Taxi Zone (m x 10^{-3})	0.106	3.348
Number of Bus Stops and subway stations in Taxi Zone ($x10^{-3}$)	1.174	62.354
Temporal and Weather Attributes		
Network Distance (m x 10^{-3}) x Winter	-0.577	-5.659
Network Distance (m x 10^{-3}) x Temperature (°F x 10^{-2})	2.460	10.983
Times Square Distance (m x 10^{-3}) x Precipitation (cm)	-0.031	-7.267
Network Distance (m x 10^{-3}) x Precipitation (cm)	-0.721	-13.517
Satiation Parameters		
Times Square Distance (m x 10^{-3})	0.087	42.497
Log-Likelihood at Convergence	-1531122.801	
Sample Size	6450	