A Joint Econometric Model Framework for Transportation Network Companies (TNC) Users' Trip Fare and Destination Choice Analysis

Dewan Ashraful Parvez
Doctoral Student
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-437-2587
Email: parvez@knights.ucf.edu
ORCiD number: 0000-0002-8663-7808

Sudipta Dey Tirtha*
Post-doctoral Scholar
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 407-543-7521
Email: sudiptadeytirtha2018@knights.ucf.edu
ORCiD number: 0000-0002-6228-0904

Tanmoy Bhowmik
Post-doctoral Scholar/Graduate Faculty Scholar
Department of Civil, Environmental & Construction Engineering
University of Central Florida
Tel: 1-407-927-6574
Email: tanmoy78@knights.ucf.edu
ORCiD number: 0000-0002-0258-1692

Naveen Eluru
Professor
Department of Civil, Environmental and Construction Engineering
University of Central Florida
Tel: 1-407-823-4815
Email: naveen.eluru@ucf.edu
ORCiD number: 0000-0003-1221-4113

Submitted To: Transportation Research Record: Journal of the Transportation Research Board

Submission Date: January 17, 2023

Funding Source: None
Data Accessibility Statement: Data will be shared upon request

---

*Corresponding author

ABSTRACT
In this study, we examine the factors affecting Chicago Transportation Networking Companies (TNC) pricing and destination choice behavior. While trip fare has been examined from various perspectives, earlier fare models have not considered an exhaustive set of independent variables. Further, trip fare decisions are significantly influenced by trip destination. Hence, in our study a joint model of trip fare and destination choice is proposed. The joint model system – linear regression for fare and multinomial logit model for destination - is developed based on Chicago TNC weekday trip data from January 2019 to December 2019. A wide range of origin and destination specific land use and built environment factors, transportation infrastructure attributes, and weather attributes were found to be significant in the model system. Based on log-likelihood (LL) and Bayesian Information Criterion (BIC) measures, the model performance of the proposed joint model is found to be superior compared to independent fare and destination models. The applicability of our proposed fare and destination choice model is illustrated through fare prediction and destination elasticity analysis. The framework can potentially be employed to generate TNC fare for inclusion in Level of Service measures for TNC model in the mode choice model.

Keywords: Transportation Networking Companies (TNC), Joint Linear Regression (LR) and Multinomial Logit (MNL) Model, Prediction, Elasticity Analysis

INTRODUCTION

Transportation Networking Companies (TNCs) are reshaping the transportation sector with operations in more than 10,000 cities across the world (*1*). As of 2021, the global ride share market is valued at 85.8 billion and is predicted to be valued at 185.1 billion by the end of 2026 (*2*). In a recent report, TNC heavyweight Uber (*3*) indicated that about 118 million users have used Uber service at least once a month in 2021. The magnitude of the user base, considering the ongoing COVID-19 pandemic in 2021, is illustrative of the major influence of TNC on mobility in urban regions. TNCs are emerging as makeshift public transport options across many urban regions across the world including Chicago (*4*), San Francisco (*5*), Boston (*6*), Santiago, Chille (*7*), Chengdu, China (*8*), and Hanover, Germany (*9*). As TNCs become an increasingly significant transportation mobility alternative across the world, there is growing literature examining TNC impact on the various facets of the transportation system.

An important consideration with the growing adoption of TNC alternatives is the inclusion of these systems within urban travel demand modeling frameworks. Several research efforts have examined the impact of TNCs in the context of demand generation and distribution across the urban region. However, incorporating TNCs within the current mode choice frameworks across urban regions is not typically explored. The main reason it is challenging to develop mode choice frameworks for TNCs is the lack of an easy to adopt framework for generating Level of Service (LOS) measures. The generation of travel time measure is relatively easy as automobile travel times can be directly applied for TNC travel times (for solo passengers). However, generating the cost measure is not straightforward. The proposed study is geared towards tackling this challenge of predicting trip level TNC fare that can be incorporated within travel demand model frameworks for generating travel cost measure for TNC alternatives in mode choice. In modeling trip fare, the current study postulates that TNC user's selection of a trip is closely linked with destination and the associated fare. For example, the destination attractiveness of a location with high density of hospitality venues (hotels/motels) is quite high. At the same time, the fare to such destinations might also be higher due to the demand. This is an example of how a destination attribute affects fare and destination choice. These can be readily considered in fare and destination models. However, it is also likely that factors such as local events (such as a concert) occurring in a destination might affect fare and demand. The information on such events might not be available for modeling. Hence, the influence of such unobserved information can be considered in the form of common unobserved factors affecting fare and destination. Further, given the ease with which TNC rides can be selected on smartphone apps, it is possible that TNC users can revisit their choice of destination in response to the fare levels shown in the app. With these considerations, in our study, we develop a joint model of fare and destination choice where trip fare is modelled using linear regression model (LR) and destination choice is modelled using a multinomial logit model (MNL). The model estimation exercise is conducted using TNC data from Chicago region. Specifically, weekday trip data spanning January 2019 through December 2019 is employed for our analysis. Trip fare and destination data are further augmented with a host of independent variables including trip attributes, origin attributes, destination attributes, land use and built environment attributes, socio-demographic attributes, and weather attributes. The model estimation process is augmented by elasticity analysis to illustrate how the proposed model can be employed to understand the influence of various independent variables on fare and destination selection.

The rest of the paper is organized as follows: Literature review section summarizes relevant literature and positions the current study. Data section documents the data processing procedures and provides an overview of the data used in our analysis. The mathematical details of the models are described in the following section. Model Estimation Results section describes the results from the models. An elasticity analysis illustrating the impact of independent variables is documented in the next section. Conclusions section presents an overview of the paper and identifies potential directions for future research.

LITERATURE REVIEW AND CURRENT STUDY IN CONTEXT

We present an overview of earlier research efforts on the two TNC dimensions of interest in our research – trip fare and destination.

TNC fare is evaluated in two ways in earlier research. *First*, fare is considered as an independent variable affecting the decision to use TNC alternatives. In these studies, various TNC associated decisions such as solo or pooled trip (*10*, *11*), competition between transit and TNC (*5*, *12*, *13*), role of income in affecting TNC usage (*12*, *14*, *15*), driver economics and turnover (*16*–*18*) and satisfaction with TNC (*12*, *19*, *20*) are examined. Important findings from these studies include: (a) high income individuals prefer TNC to transit (*12*, *15*), (b) higher TNC pricing power is observed in highly walkable areas (*21*, *22*), (c) turnover for ridehailing services is significantly high (*16*), (d) sharing TNC demand and supply information with drivers may lead to higher satisfaction level among drivers (*20*), and (e) a higher inclination among younger individuals for using TNC(*5*, *12*, *23*). *Second*, studies examined dynamic pricing policy (or surge pricing) in their analysis of TNC systems. In these studies, fare is modeled as a continuous variable within an optimization framework (*24*–*26*). The approaches provide elegant mathematical formulations for profit maximization or demand imbalance minimization in the context of a equilibrium based optimization models to estimate price and/or demand. The mathematical formulations are applicable under a host of assumptions such as restricted number of TNCs (*25*), neglecting spatial variations (*27*), the distances in the network are equidistant (*24*), and limits on the number of modal alternatives (for example only Drive vs TNC in Afifah and Guo (*25*)). The demand, price and model choice equations in these approaches are simplified and focus on a small set of variables such as trip length (*11*). While these approaches are very helpful, applying these methods for large urban regions with temporal and spatial variations are not readily practical. In our review, we found only 3 studies that developed direct fare models using TNC data (*11*, *28*, *29*) where a small set of variables such as trip distance, trip time, tolls and additional charges were considered.

Destination selection behavior has been examined in multiple ride sharing domains including bicycle-sharing system (*30*–*32*), taxi (*33*–*35*), TNC and Shared Autonomous Vehicle (SAV) (*36*, *37*). The preferred approach employed at the disaggregate level is the Multinomial Logit Model (MNL) based on the random utility maximization approach (*30*). Other model structures employed for analysis of destination dimensions such as Traffic Analysis Zone (TAZ) (*38*) includes a Generalized Spatially Correlated Logit (GSCL) Model. In some studies, aggregate destination allocations are analyzed using Multiple Extreme Continuous Extreme Value (MDCEV) models (*32*). Important findings on destination choice preferences include: (a) destination choice is highly correlated with employment status (*39*), (b) presence of high demand in the neighborhood is a strong contributor of demand (*32*), (c)lower fare price increases the utility of a destination (*40*), (d) duration of stay and home location prior to the activity affect destination choice (*41*), and (e) destination choice behavior is influenced by the perceived destination image from individual's social network (*42*).

Contributions of the Current Study

Several studies have recognized that pricing algorithms are influenced by spatio-temporal demand (such as demand at origin in preceding 15 minutes), origin and destination land use and built environment factors, transportation infrastructure attributes, and weather attributes (*24*, *43*). However, none of the earlier research studies have incorporated a wide range of attributes in modeling TNC fare. The first contribution of our study is to develop a comprehensive trip fare model while accounting for a host of independent variables. In this study, we recognize that trip fare values are closely aligned with trip destination. Hence, the second contribution of our study is to develop a joint model system that accounts for common unobserved factors affecting fare and destination. The study develops a joint linear regression (LR) for fare and multinomial logit (MNL) model for destination labelled as the LR-MNL model. The model system is developed using TNC trip data from Chicago for the year 2019. Chicago data has been employed in the literature to study various TNC dimensions including spatial demand variations and willingness to use pool alternative (*10*, *21*, *44*). Finally, the current study contributes empirically by allowing us to understand Chicago TNC pricing model and destination choice behavior. The framework can potentially allow us to generate TNC fare for mode choice model. In application, the model developed can be employed in a sequence – destination choice outcome followed by trip fare prediction. The model framework can also allow us to identify systemic differences across the Chicago city in pricing (if any) and how various destination attributes influence destination preferences.

DATA PREPARATION

Data Source

City of Chicago has made TNC data available for analysis beginning in November 2018. As of 2019, three TNCs were operating in the Chicago area: Uber, Lyft and Via (*45*). For this current study, daily weekday trip data of more than 50 million records for 12 months starting from January 2019 to December 2019 was compiled for our analysis(*45*). Origin and destination for each of these trips have been aggregated at the census tract level while trip times (start time & end time), trip fare are rounded to nearest 15 minutes and 2.50 USD respectively. The trip dataset is further augmented by trip attributes such as trip start & end time, trip distance, shared trip indicator provided by Transportation Network Providers-Chicago Data Portal (*45*), land use and built environment variables including distance from Central Business District (CBD), residential area, commercial area, institutional area, recreational area accessed from Chicago Data portal and Chicago Metropolitan Agency for Planning (CMAP) (44, 46), Transportation infrastructure attributes including bike lane density, street length, number of bus stops, number of transit stations, number of divvy stations walk score, transit score compiled from Chicago Data portal and Chicago Metropolitan Agency for Planning(CMAP) (*45*, *46*) and sociodemographic attributes such as low income indicator, employment density drawn from US Census Bureau (*47*) and weather attributes such as snow depth obtained from National Climatic Data Center (NCDC) (*48*). A summary of the independent variables is provided in Table 1.

Sample Formation

The data processing procedures were implemented in the following sequence. First, records with missing and inconsistent information were dropped from the dataset. Second, trips that originated or destined outside of Chicago city area were removed from the dataset. Finally, weekday trips were retained amounting to more than 44 million of records. The spatial distribution of weekday trips by origin and destination census tract are presented in **Figure 1(a)** and **Figure 1(b)** respectively. Employing the full set of records (44 million) would increase computational time for modeling exercise significantly. Further, using such large datasets in econometric models might lead to overfitting. To address these issues, we randomly select 25 samples of 10,000 records for our model estimation exercise. These samples will allow us to ensure that the parameters estimated using one sample are not significantly different from other samples of data. Towards this end, we conduct a rigorous statistically valid comparison of model estimates across all 25 samples prior to selecting a sample for further analysis.

For the destination choice models, all census tracts in the region are potential alternatives. In our data for Chicago we identified 801 census tracts (*49*). From this broad set of alternatives, destination choice models are developed employing a random sample of 30 alternatives (inclusive of the chosen alternative). Similar random sampling process has been adopted in earlier literature for destination choice models(see 57–60 for details).

TABLE 1 Descriptive Statistics of Variables

Variables	**Variable Descriptions**	**Descriptive Statistics**	
		Mean	**Std. dev.**
DEPENDENT VARIABLES			
Trip fare model			
Trip fare	Ln (Trip fare)	2.079	0.577
INDEPENDENT VARIABLES (CONTINUOUS)			
Trip Attributes			
Trip distance	Distance traveled in each trip	4.149	4.129
Network distance	Ln (Shortest distance between census tracts)	1.895	0.494
Demand in last 15 minutes at origin	Ln (Demand in last 15 minutes in each origin census tract)	2.006	1.482
Demand in last 15 minutes at destination	Ln (Demand in last 15 minutes in each destination census tract)	2.032	1.537
Land Use and Built Environment Attributes			
Network distance from CBD	Ln (Network distance to census tract from Central Business District (CBD))	1.871	0.560
Residential area	Total residential area in each census tract (area/100) in acre	0.602	0.519
Commercial area	Total commercial area in each census tract (area/100) in acre	0.115	0.174
Institutional area	Total institutional area in each census tract (area/100) in acre	0.113	0.280
Recreational area	Total recreational area in each census tract (area/100) in acre	0.074	0.232
Land use mix	Land use mix = $[\frac{-\sum_k(P_k(lnP_k))}{lnN}]$, where k is the category of land-use, p is the proportion of the developed land area for specific land-use, N is the number of land-use categories	0.134	0.045
Transportation Infrastructure Attributes			

Variables	Variable Descriptions	Descriptive Statistics	
		Mean	**Std. dev.**
Bike lane density	Length of bike lane in each census tract per acre (Density*100) (mi/acre)	0.321	0.352
Length of street	Length of street in each census tract	5.597	4.953
Number of bus stops	Number of bus stops in each census tract	12.486	8.330
Number of L stations	Number of stations of L transit system in each census tract	0.156	0.529
Number of divvy stations	Number of divvy stations in each census tract	1.029	1.441
Walk score	Walk score (a measure of serviceability of walkability) in each census tract	82.397	26.225
Transit score	Transit score (a measure of serviceability of public transit) in each census tract	8.260	0.996
Sociodemographic Attributes			
Employment density	Number of employments in each census tract per acre (Density/100)	0.236	0.381
Weather Attributes			
Snow depth	Standard score ($\frac{x-\mu}{\sigma}$) of snow depth in each census tract. Where x is the observed value of snow depth, μ is the mean of the distribution of the values of snow depth and σ is the standard deviation of the distribution of the values of snow depth	0.004	1.034
INDEPENDENT VARIABLES (CATEGORICAL)			
Variables	**Variable Descriptions**	**Freq.**	**Percentage**
Trip Attributes			
Trip starts at AM peak	Trip starts within AM peak period	1965.000	19.650
Trip starts at PM peak	Trip starts within PM peak period	2560.000	25.600
Trip starts at other time	Trip starts in other time period	5475.000	54.750
Trip ends at AM peak	Trip ends within AM peak period	1876.000	18.760
Trip ends at PM peak	Trip ends within PM peak period	2481.000	24.810
Trip ends at other time	Trip ends in other time period	5643.000	56.430
Shared trip indicator			
Yes	Trip authorized as shared	1507.000	15.070
No	Trip is not authorized as shared	8493.000	84.930
Sociodemographic Attributes			
Low income indicator			
Yes	Census tract with median income under $58 thousand USD (15th percentile)	466.000	58.543
No	Census tract with median income over $58 thousand USD (15th percentile)	330.000	41.457

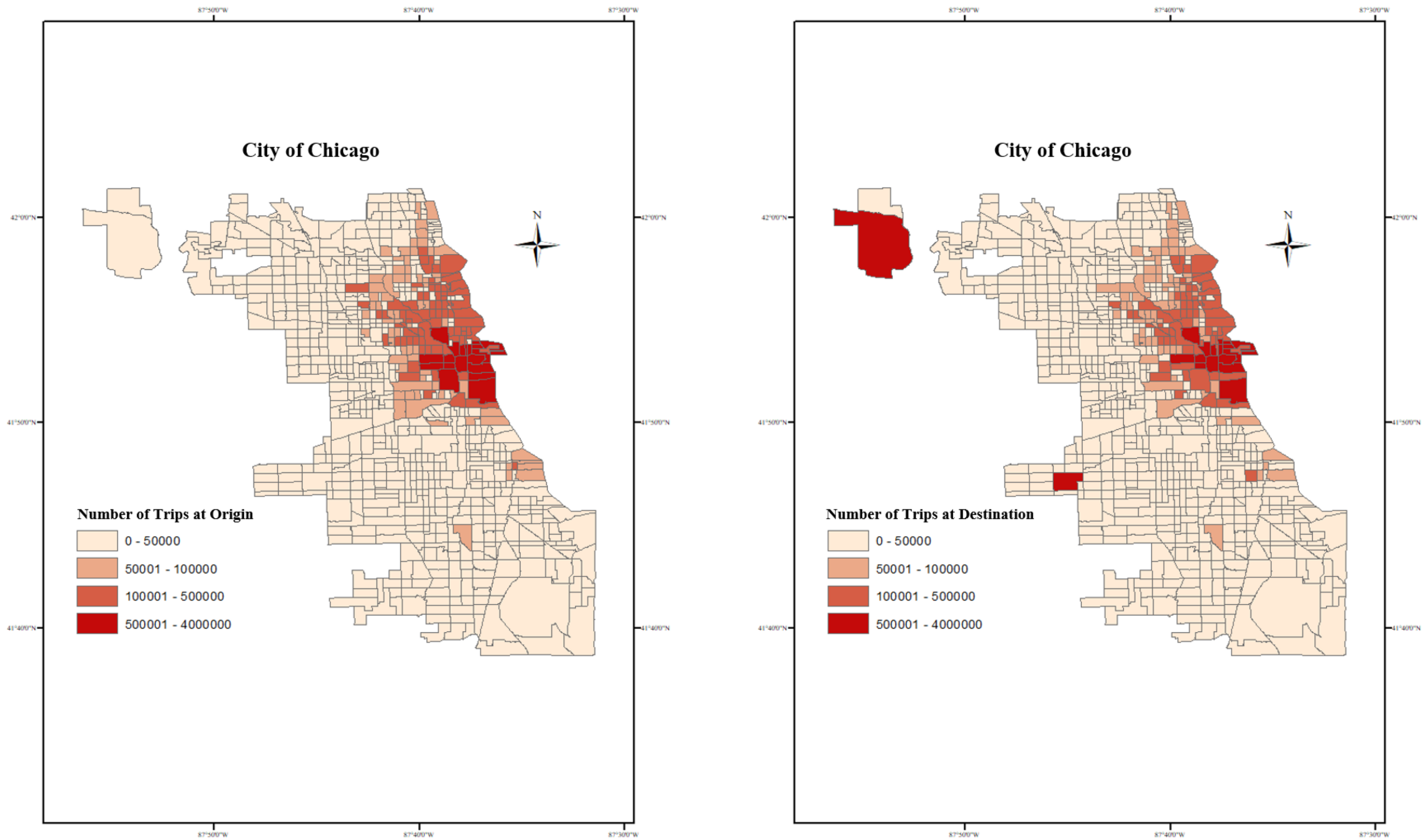

Figure 1 Total number of weekday trips (a) originated; (b) destined

ECONOMETRIC METHODOLOGY

In this study, we develop a joint trip fare and trip destination model where trip fare is modelled using a linear regression model and trip destination is modelled using a multinomial logit model. Let, q (=1, 2, 3,…., Q=10,000) be an index to represent each individual trip, y_q be an index to represent the fare associated with a trip q, and s (= 1, 2, …, S=30) be an index to represent destination alternatives (census tracts). In the following sections, we describe two model components and then present estimation procedure for the joint model.

Trip Fare Model

In the linear regression formulation, we express y_q as a function of independent variables z_q as follows:

$$y_q = (\alpha' + \eta')z_q + \varepsilon_q \tag{1}$$

where α' is a vector of coefficients to be estimated, η represents the effect of common unobserved factors modifying the impact of z_q in the trip fare and trip destination models (see Equation 2) and ε_q is an idiosyncratic random error term assumed independently normally distributed with variance γ^2. Now, we can express the probability of a trip, q having fare, y_q as follows:

$$P(y_q) = \frac{\phi\left[\frac{y_q - (\alpha' + \eta')z_q}{\gamma}\right]}{\gamma} \tag{2}$$

where ϕ(.) is the standard normal probability distribution function.

Trip Destination Model

In the MNL model, the random utility of an alternative s for trip q takes the following form:

$$u_{qs} = (\beta' + \eta')x_{qs} + \epsilon_{qs} \tag{3}$$

where u_{qs} is the utility obtained by user q by choosing census tract s as the destination from a choice set of 30 census tracts. x_{qs} is a vector of attributes and β is a vector of model coefficients to be estimated. The random error term, ϵ_{qs}, is assumed to be independent and Gumbel-distributed identically across the dataset. In random utility maximization (RUM) approach, a user making the trip, q will choose a census tract as the destination that offers the highest utility. Therefore, the probability expression takes the following multinomial logit form:

$$P(s_q) = \frac{exp\,((\beta' + \eta')x_{qs})}{\sum_{s=1}^{S} exp\,((\beta' + \eta')x_{qs})} \tag{4}$$

The destination alternatives in our study context are not labelled (i.e., they are not typical categorical alternatives such as travel mode (car, bike)). Hence, our model estimation approach considers a generic parameter structure across all alternatives. The approach will allow for parameter estimation for variables that vary across destination alternatives such as destination employment or destination land use mix. In the model structure, accounting for variables at the trip level such as trip start time or origin destination can be considered as an interaction term with variables varying across the destination (such as Trip starts in AM peak x Number of divvy stations in CT).

Estimation Procedure

To complete the model structure of the Equations (1) and (3), it is necessary to define the structure for the unobserved vector η. In this paper, we assume that this vector is independent realizations from normal distributions as follows: $\eta \sim N(0, \sigma^2)$. With this assumption, the joint probability expression for trip fare and trip destination may be derived. Conditional on η the probability for a trip, q to have fare, y_q and destination s can be expressed as follows:

$$P(y_q, s_q)|\eta = P(y_q) \times P(s_q) \tag{5}$$

The complete set of parameters to be estimated in the model system of Equation (5) are α, β and γ and standard error term, σ. Let, Ω represents a vector that includes all the standard error parameters to be estimated. Given this assumption, the joint likelihood for trip fare and trip destination is provided as follows:

$$L_q|\Omega = \prod_{s=1}^{S} \left[P(y_q, s_q)|\eta\right]^{d_{qs}} \tag{6}$$

where d_{qs} is a dummy variable taking a value of 1 if a user making the trip, q chooses the destination, s and 0 otherwise. Finally, the unconditional likelihood function may be computed for a trip, q as follows:

$$L_q = \int_{\Omega} (L_q|\Omega)d\Omega \tag{7}$$

Now, we can express the log-likelihood function of the final joint model as follows:

$$\mathrm{LL} = \sum_{q=1}^{Q} \ln L_q \tag{8}$$

The log-likelihood function in Equation (8) involves the evaluation of a multi-dimensional integral of size equal to the number of rows in Ω. We apply Quasi-Monte Carlo simulation techniques based on the scrambled Halton sequence to approximate this integral in the likelihood function and maximize the logarithm of the resulting simulated likelihood function (See Bhat, (*54*); Yasmin and Eluru, (*55*) for more details).

MODEL DEVELOPMENT

As described earlier, we estimate the model components employing a randomly chosen dataset of 10,000 records for computational efficiency and avoiding overfitting. Given the possibility that the random sample might not represent the population, we draw 25 samples of 10,000 and examine the role of randomness in the parameter stability across the samples for linear regression and multinomial logit models. To examine parameters stability, we employ the following revised Wald test statistic approach across 25 samples:

$$\text{Parameter test statistic} = abs\left[\frac{(sample\ parameter - population\ benchmark)}{\sqrt{{SE_{sample}}^2 + {SE_{population}}^2}}\right]$$

The population benchmark is computed as the average value of the parameter across the 25 samples. If any parameter for a sample is significantly different from the population benchmark, the Wald statistics will be larger than the 90% t-statistic value of 1.65. **Figure 2** and **Figure 3** illustrates the range of revised Wald test statistic for 25 samples for trip fare model and destination choice model respectively in a box plot. It is evident from **Figure 2** and **Figure 3** that means of the revised Wald test statistic of all the exogenous variables (and majority of the realizations) are well within 90% test statistic. To be precise, in case of the trip fare model (destination choice model), only two (six) test statistic values among 450 (500) values generated were found to be greater than 90% test statistic. Therefore, we can conclude that the parameters estimated across the random samples are stable and there is no significant difference in parameters estimated across samples.

After establishing that the random sample based models are stable, we estimate a joint LR-MNL model which accounts for common unobserved heterogeneity between trip fare and trip destination for one sample.

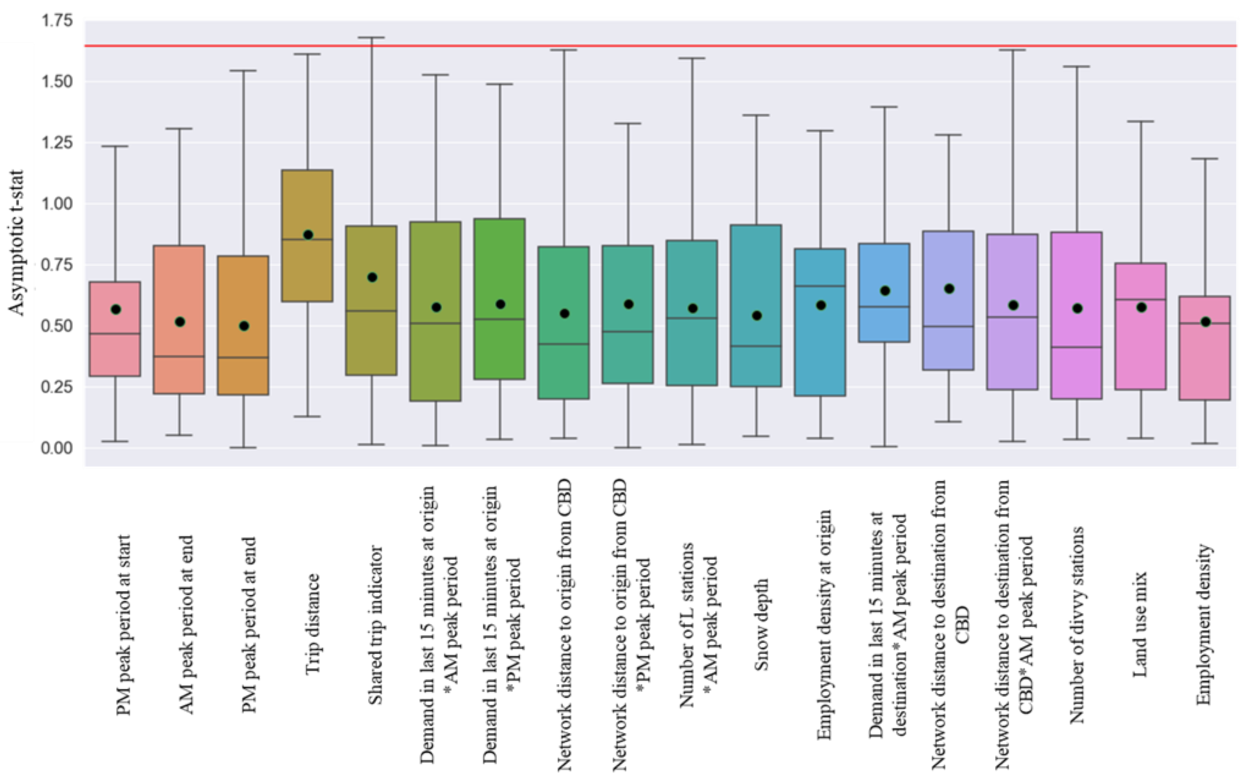

Figure 2 Asymptotic t-statistic for the parameters estimated of trip fare model

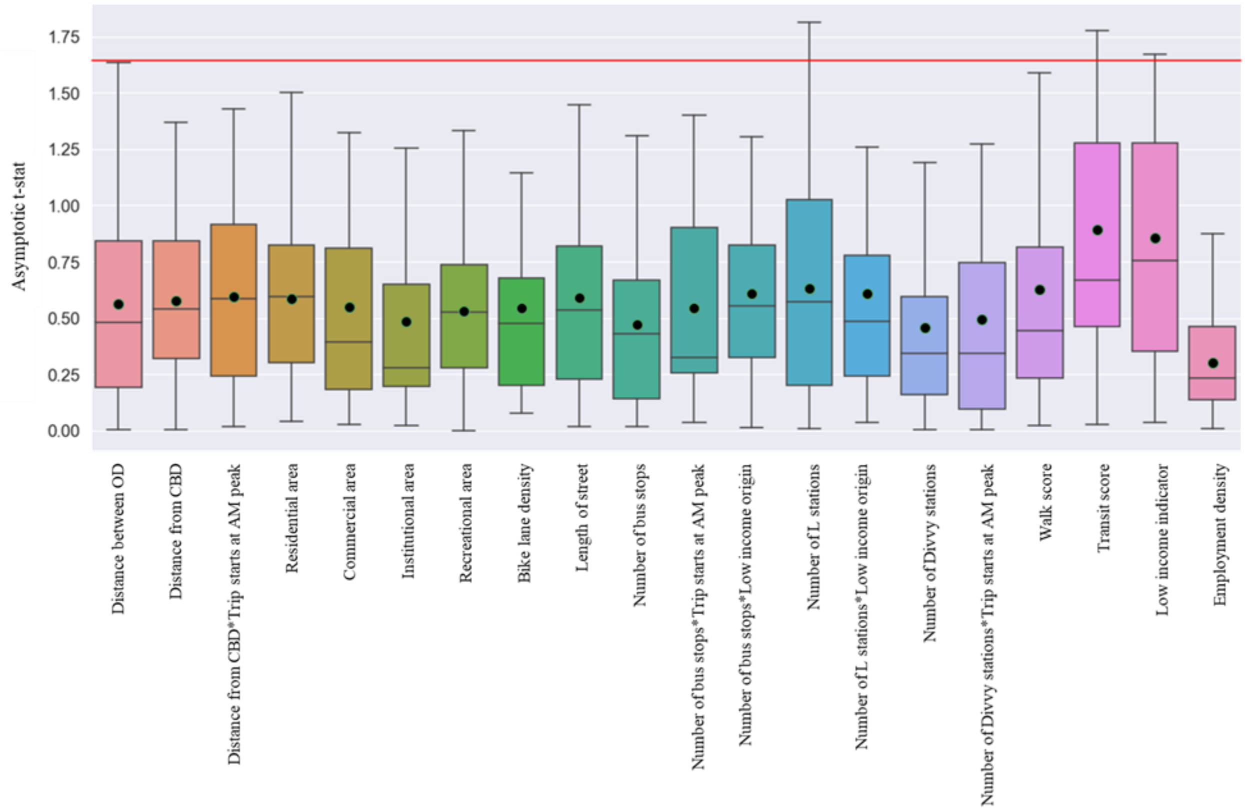

Figure 3 Asymptotic t-statistic for the parameters estimated of destination choice model

MODEL ESTIMATION RESULTS

The model performance of the proposed joint model is compared to the independent fare and destination models using log-likelihood (LL) and Bayesian Information Criterion (BIC) measures. The LL (BIC) values of the independent LR and MNL model are -22857.920 (46075.043). For the joint LR-MNL model system, LL (BIC) values were found to be -22717.000 (45793.203). Hence, the joint model system clearly outperforms the independent model system. For the sake of brevity, the results from the Joint LR-MNL model estimation results are discussed (see **Table 2).** The discussion is organized by variable group.

Trip Fare Model

Trip Attributes

In **Table 2 s**everal trip attributes are found to have significant impact on TNC fare. Trip distance, as expected, has a positive impact on trip fare. Controlling for everything else, longer trips have higher fares. If the trip starts and ends in PM peak period, fare is likely to increase for the corresponding trip. Similarly, when a trip ends in AM peak period, an increase in fare is observed. The results are along expected lines and suggest that during peak periods a higher fare is levied. Finally, we also find that shared trips are likely to have a lower fare as expected.

TABLE 2 Joint LR-MNL Model Result

Variable	Estimate	t-stat
Trip Fare Model		
Constant	1.638	86.286
Trip Attributes		
Trip distance	0.104	77.797
Trip start time (Base: Other periods)		
PM peak period	0.096	3.546
Trip end time (Base: Other periods)		
AM peak period	0.045	2.276
PM peak period	0.045	2.326
Shared trip indicator (Base: No)		
Yes	-0.371	-37.494
Origin Attributes		
Demand in last 15 minutes at origin*AM peak period	-0.019	-2.683
Demand in last 15 minutes at origin*PM peak period	0.021	2.683
Network distance to origin from CBD	0.010	5.271
Network distance to origin from CBD*PM peak period	-0.011	-2.713
Number of L stations*AM peak period	-0.008	-1.643
Employment density at origin	-0.012	-2.675
Snow depth	-0.010	-2.838
Destination Attributes		
Demand in last 15 minutes at destination*AM peak period	0.032	5.157
Network distance to destination from CBD	-0.007	-3.836
Network distance to destination from CBD*AM peak period	-0.005	-2.335
Number of Divvy stations	-0.004	-4.535
Land use mix	0.183	1.959
Employment density	0.014	3.096
Scale	0.229	31.564
Destination Choice Model		
Land Use and Built Environment Attributes		
Network distance between O-D	-1.092	-57.340
Distance from CBD	-0.539	-16.967
Distance from CBD*Trip starts at AM peak	-0.254	-4.705
Residential area	-0.927	-14.191
Commercial area	0.396	7.397
Institutional area	-0.168	-2.939
Recreational area	0.306	7.262
Transportation Infrastructure Attributes		
Bike lane density	0.160	4.685
Street Length	0.073	25.453
Number of bus stops	0.007	3.064

Variable	Estimate	t-stat
Number of bus stops*Trip starts at AM peak	0.020	5.408
Number of bus stops*Low income origin	0.017	4.692
Number of L stations	-0.117	-7.473
Number of L stations*Low income origin	-0.120	-3.726
Number of Divvy stations	0.037	4.465
Number of Divvy stations*Trip starts at AM peak	0.039	2.524
Walk score	0.004	4.158
Transit score	0.089	3.597
Demographic Attributes		
Low income indicator (Base: Median income over 15th percentile)		
Yes	-0.926	-24.680
Employment density	0.061	2.242
Unobserved heterogeneity		
Constant in LR and Distance between O-D in MNL	0.266	36.494
Constant in LR and Street Length in MNL	0.039	7.148

Origin Attributes

In our analysis, we wanted to consider the influence of demand in preceding time intervals on trip fare. For this purpose, origin demand in the last 15 minutes in AM and PM peak periods was considered in the model. The model estimates offer interesting results. In the AM peak period, higher demand has a negative coefficient. While this might appear counter-intuitive on first glance, the reader will recognize that the demand variable interacts with the AM peak main effect thus, the net effect is still likely to be positive. For PM peak period, the impact on fare is more pronounced clearly highlighting that higher demand at the origin contributes to a higher fare.

From **Table 2**, it is evident that trip fare is likely to increase as distance between origin of the trip and CBD increases. The result represents the supply side challenge (or rerouting costs) for drivers to pick up riders away from CBD (see (*56*) for similar findings). The negative coefficient for interaction of distance variable and PM peak period indicates that during PM peak the impact of distance from CBD is moderated potentially due to increase expected supply for TNC. Chicago L, a rapid transit system, operates inside the city of Chicago. The number of L stations in the AM peak period has negative impact on TNC fare highlighting potential competition between Chicago L and TNC (*57*). Interestingly, higher employment density at the origin is negatively associated with TNC fare potentially reflecting the presence of infrastructure for non-motorized modes and improved land use (*58*, *59*). The results indicate that in adverse weather conditions such as higher level of snow depth, TNC fares are likely to be lower possibly due to supply demand imbalance (*60*, *61*).

Destination Attributes

The demand in the last minutes at the destination also offers interesting results. We find that interaction of destination demand with AM peak is positive indicating that higher fares are likely to destinations with higher demand in AM peak (similar findings in 11). As the distance of the destination census tract increases from CBD, TNC fare is likely to be lower. The result is expected because with all else same, travel away from CBD is typically faster and thus trip fare is expected to be lower. The effect is more pronounced in the AM peak period as congestion is likely to be lower away from CBD during AM peak.

Chicago bike sharing system (Divvy) and TNC appear to have competitive relationship as highlighted by the negative coefficient on the number of divvy stations (see (*30*) for evidence of how individuals use divvy system to make commuting trips in CBD). The results also indicate that destination with diverse land use is likely to have higher fares. TNC travel in these locations will be slower and hence require longer travel time resulting in higher fares. Finally, destinations with higher employment density will contribute to higher TNC fare as expected.

Destination Choice Model

Land use and Built Environment Attributes

Several land use and built environment variables offer significant and expected results. As the distance between origin and destination and distance of the destination from CBD increases, the likelihood of the alternative being selected reduces. The impact of distance to CBD is significantly higher in the AM peak period as users are unlikely to travel away from the CBD in the AM peak. The various built-up areas also offer expected results. Census tracts with residential and institutional areas are less likely to be destination. On the other hand, census tracts with higher areas of commercial and recreational areas have a higher likelihood of being chosen (see (*32*, *63*) for similar results).

Transportation Infrastructure Attributes

The results for transportation infrastructure attributes offer multiple significant and nuanced relationships with destination preferences. Destination attributes that represent non-motorized and transit infrastructure such as bike lanes, bus stops, divvy stations, walk score and transit offer positive association with destination choice. Several earlier studies have documented these some or all of these relationships (*21*, *22*, *60*, *64–66*). For bus stops and divvy stations, the impact on destination selection is even higher during the AM peak period. An exception to this is the parameter for L stations. The result clearly highlights that in census tracts with L stations, TNC users are less likely to choose these destinations. The income of origin census tract also offers a conflicting interaction with bus stops and L stations. The users starting their travel from low-income census tracts have higher affinity to travel to destinations with higher number of bus stops. However, the result is exactly opposite in the context of L stations. The variation might be reflecting the different neighborhood characteristics of census tracts with higher number of buses vis-à-vis census tracts with higher number of L stations (*62*, *67*, *68*).

Demographic Attributes

Census tracts with lower income are less likely to be chosen as TNC destinations. The result indicates to income inequity in the adoption of TNC for mobility needs in Chicago and other urban regions (see similar findings in (*23*, *69*, *70*). As expected, on weekdays, a census tract with higher employment density is likely to attract more TNC trips (*31*, *69*).

Unobserved Heterogeneity

The proposed LR-MNL joint model system accommodates for common unobserved heterogeneity between trip fare and destination choices. Several unobserved factors were tested in the joint model. The variables that offered significant unobserved correlation are reported in the last row panel of **Table 2**. The two parameters represent interaction of a constant in fare model with origin -destination distance and street length. These significant correlations reinforce our hypothesis that trip fare and destination choices are influenced by shared factors and incorporating such correlation is important.

PREDICTION AND ELASTICITY ANALYSIS

To illustrate the applicability of the proposed model, we employ the model results for understanding the influence of independent variables on fare and destination choice models. We employ the model results for the fare model to generate trip cost predictions for five randomly chosen trips in the PM peak and off-peak periods. These trips are plotted in **Figures 4 (a) and (b).** The prediction illustrates how the proposed model can be employed for generating trip fares across the region. The trip fares presented in **Figure 4** illustrate the higher cost of TNC during PM peak (relative to off-peak period). The procedure can be readily applied to generate travel cost schemes for a mode choice model in the region with TNC alternative.

For the destination model, an elasticity analysis has been undertaken in an effort to capture the changes in dependent variables (destination) in response to changes in independent variables. **Figure 5** illustrate the percent change in fare and aggregate probability of the chosen destination alternative respectively due to change in independent variables by 10%. The results summarized in **Figure 5** offer interesting results. We notice that distance between origin destination, transit score and street length variables exhibit the highest impact on destination preferences. We also observe that walk score, divvy stations, bus stops, residential area and distance form CBD affect destination preferences reasonably. In summary, the elasticity effect highlights how transportation planners and TNC owners can examine trends influencing destination choice behavior.

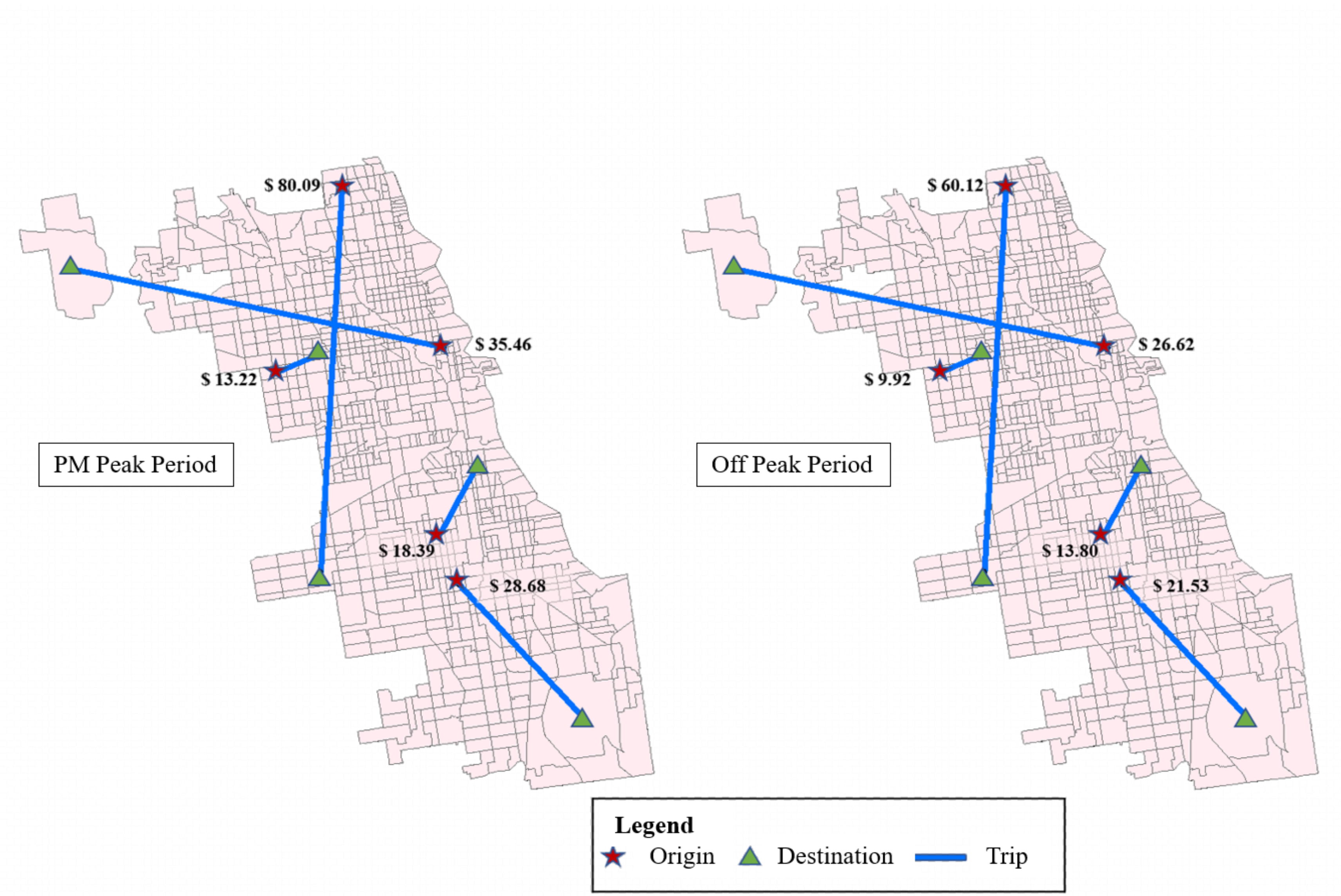

Figure 4 Trip fare prediction across (a) PM peak period; (b) Off peak period

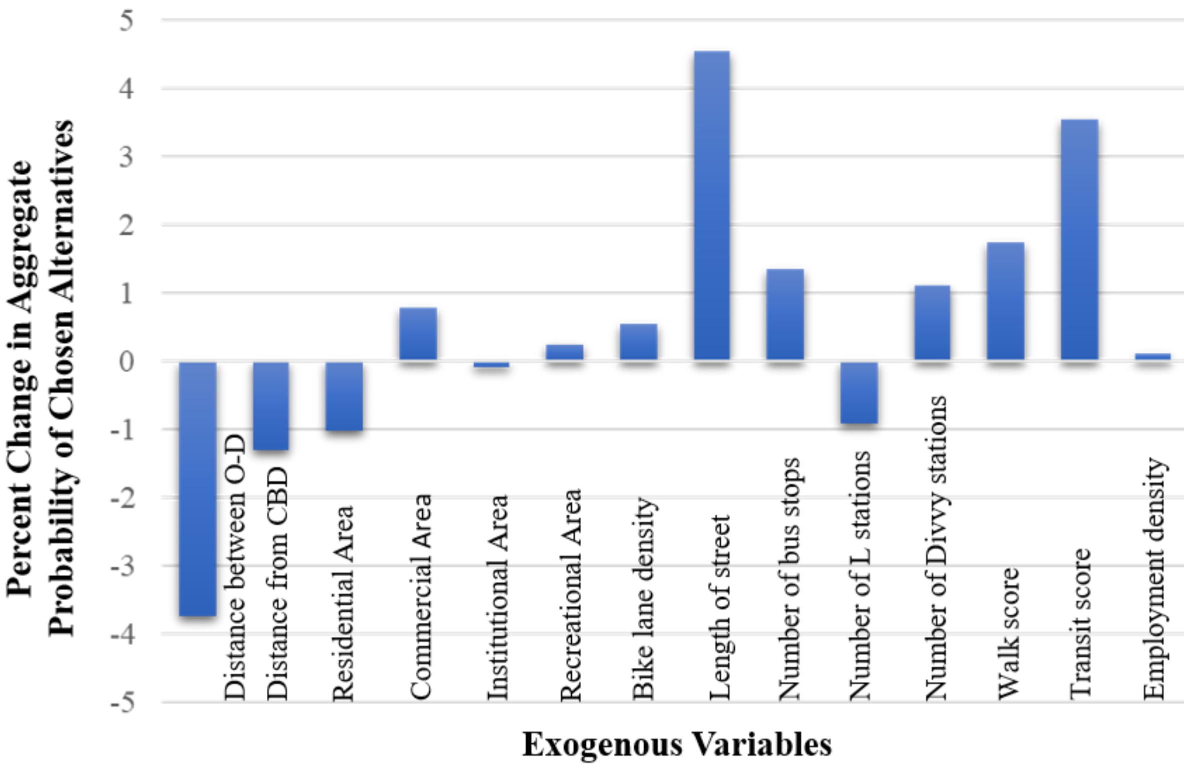

Figure 5 Elasticity analysis

CONCLUSIONS

Given the prevalence of Transportation Networking Companies (TNCs) across the world, there is growing literature dedicated to TNC usage analysis. However, there is limited research on comprehensively examining the influence of independent variables on TNC fare. In this study, we postulate that TNC trip fare is closely linked to TNC trip destination and develop a joint econometric model linking the two outcomes. A wide range of origin and destination specific land use and built environment factors, transportation infrastructure attributes, and weather attributes were found to be significant in the joint the model system. Based on log-likelihood (LL) and Bayesian Information Criterion (BIC) measures, the model performance of the proposed joint model is found to be superior compared to independent fare and destination models. The model results were augmented with fare prediction exercise and destination model elasticity analysis. The fare prediction exercise illustrated how the proposed model can be employed to generate TNC travel costs for use in a mode choice model with TNC alternative. The destination elasticity analysis highlighted the important factors affecting destination preferences.

The study is not without limitations. TNC trip data does not provide any user related information. Access to sociodemographic, socioeconomic, and other relevant information can significantly enhance the models developed in our analysis. Trip level TNC data employed in this study provides trip origin and destination aggregated at the census tract level potentially to preserve user and operator privacy. The aggregated destination information can result in large differences in travel distances for short trips within the census tracts. The model developed can be further refined in the presence of more disaggregate data. It is also important to recognize that TNC trip fare can be influenced by business strategies of TNCs that are not readily declared publicly. Understanding the effect of TNC business strategies might be an avenue for future research.

ACKNOWLEDGEMENT

The authors are grateful to Chicago Transportation Network Providers (TNP), Chicago Data portal, Chicago Metropolitan Agency for Planning (CMAP), National Climatic Data Center (NCDC) and United States Census Bureau for providing access to Chicago TNC trip data, land use, transportation infrastructures, weather, and sociodemographic data. The authors are also thankful to Dr. Bibhas Kumar Dey for introducing Chicago TNC trip data to the authors.

AUTHOR CONTRIBUTION STATEMENT

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

CONFLICT OF INTEREST STATEMENTS

The authors do not have any conflicts of interest to declare.

REFERENCES

1. How Many Uber Drivers Are There in 2022? | Ridester.Com. https://www.ridester.com/how-many-uber-drivers-are-there/. Accessed Jul. 6, 2022.
2. Forecast of Ride-Sharing Market Size 2021 | Statista. https://www.statista.com/statistics/1155981/ride-sharing-market-size-worldwide/. Accessed Jul. 6, 2022.
3. Uber Technologies, Inc. - Uber Announces Results for Fourth Quarter and Full Year 2021. https://investor.uber.com/news-events/news/press-release-details/2022/Uber-Announces-Results-for-Fourth-Quarter-and-Full-Year-2021/. Accessed Jul. 6, 2022.
4. Schwieterman, J., and C. S. Smith. Sharing the Ride: A Paired-Trip Analysis of UberPool and Chicago Transit Authority Services in Chicago, Illinois. *Research in Transportation Economics*, 2018. 71:9–16.
5. Rayle, L., D. Dai, N. Chan, R. Cervero, and S. Shaheen. Just a Better Taxi? A Survey-Based Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco. *Transport Policy*, 2016. 45: 168–178.
6. Gehrke, S. R., M. P. Huff, and T. G. Reardon. Social and Trip-Level Predictors of Pooled Ride-Hailing Service Adoption in the Greater Boston Region. *Case Studies on Transport Policy*, 2021. 9: 1026–1034.
7. Tirachini, A., and M. del Río. Ride-Hailing in Santiago de Chile: Users' Characterisation and Effects on Travel Behaviour. *Transport Policy*, 2019. 82:46–57.
8. Kong, H., X. Zhang, and J. Zhao. How Does Ridesourcing Substitute for Public Transit? A Geospatial Perspective in Chengdu, China. *Journal of Transport Geography*, 2020. 86: 102769.
9. Gilibert, M., I. Ribas, N. Maslekar, C. Rosen, and A. Siebeneich. Mapping of Service

Deployment Use Cases and User Requirements for an On-Demand Shared Ride-Hailing Service: MOIA Test Service Case Study. *Case Studies on Transport Policy*, 2019. 7:598–606.
10. Hou, Y., V. Garikapati, D. Weigl, A. Henao, M. Moniot, and J. Sperling. Factors Influencing Willingness to Pool in Ride-Hailing Trips. *Transportation Research Record: Journal of the Transportation Research Board*, 2020. 2674:419–429.
11. Jacob, J., and R. Roet-Green. Ride Solo or Pool: Designing Price-Service Menus for a Ride-Sharing Platform. *European Journal of Operational Research*, 2021. 295:1008–1024.
12. Alemi, F., G. Circella, S. Handy, and P. Mokhtarian. What Influences Travelers to Use Uber? Exploring the Factors Affecting the Adoption of on-Demand Ride Services in California. *Travel Behaviour and Society*, 2018. 13:88–104.
13. Schwieterman, J. P. Uber Economics: Evaluating the Monetary and Travel Time Trade-Offs of Transportation Network Companies and Transit Service in Chicago, Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673:295–304.
14. Dong, X. Linking TNC with Passengers: Investigating TNC Use among Lower-Income Residents with Limited Access to Cars. *Travel Behaviour and Society*, 2022. 27:184–191.
15. Young, M., and S. Farber. The Who, Why, and When of Uber and Other Ride-Hailing Trips: An Examination of a Large Sample Household Travel Survey. *Transportation Research Part A: Policy and Practice*, 2019. 119:383–392.
16. SherpaShare. https://www.sherpashare.com/share/the-top-demographic-trends-of-the-on-demand-workforce/. Accessed Jun. 30, 2022.
17. Opinion | When Your Boss Is an Algorithm - The New York Times. https://www.nytimes.com/2018/10/12/opinion/sunday/uber-driver-life.html. Accessed Jul. 2, 2022.
18. Henao, A., and W. E. Marshall. An Analysis of the Individual Economics of Ride-Hailing Drivers. *Transportation Research Part A: Policy and Practice*, 2019. 130:440–451.
19. Hall, J. V., and A. B. Krueger. An Analysis of the Labor Market for Uber's Driver-Partners in the United States. *Ilr Review*, 2017. 71(3):705–732.
20. Hu, B., M. Hu, and H. Zhu. Surge Pricing and Two-Sided Temporal Responses in Ride Hailing. *Manufacturing and Service Operations Management*, 2021. 24 (1):91–109.
21. Dean, M. D., and K. M. Kockelman. Spatial Variation in Shared Ride-Hail Trip Demand and Factors Contributing to Sharing: Lessons from Chicago. *Journal of Transport Geography*, 2021. 91:102944.
22. Marquet, O. Spatial Distribution of Ride-Hailing Trip Demand and Its Association with Walkability and Neighborhood Characteristics. *Cities*, 2020. 106:102926.
23. Grahn, R., C. D. Harper, C. Hendrickson, Z. Qian, and H. S. Matthews. Socioeconomic and Usage Characteristics of Transportation Network Company (TNC) Riders. *Transportation*, 2020. 47(6):3047–3067.
24. Bimpikis, K., O. Candogan, and D. Saban. Spatial Pricing in Ride-Sharing Networks. *Operations Research, 2019.* 67(3): 744–769.
25. Afifah, F., and Z. Guo. Spatial Pricing of Ride-Sourcing Services in a Congested Transportation Network. *Transportation Research Part C: Emerging Technologies, 2022. 142:103777*
26. Guda, H., and U. Subramaniana. Your Uber Is Arriving: Managing On-Demand Workers Through Surge Pricing, Forecast Communication, and Worker Incentives. *Management Science*, 2019. 65(5):1995–2014.
27. Zha, L., Y. Yin, and Y. Du. Surge Pricing and Labor Supply in the Ride-Sourcing Market.

Transportation Research Procedia, 2017. 23:2–21.
28. Liu, M., E. Brynjolfsson, and J. Dowlatabadi. Do Digital Platforms Reduce Moral Hazard? The Case of Uber and Taxis. *Management Science*, 2021. 67(8):4665–4685.
29. Pepić, L. The Sharing Economy: Uber and Its e?Ect on Taxi Companies. *Acta Economica*, Vol. 16, No. 28, 2018, pp. 123–136. https://doi.org/10.7251/ACE1828123P.
30. Faghih-Imani, A., and N. Eluru. Analysing Bicycle-Sharing System User Destination Choice Preferences: Chicago's Divvy System. *Journal of Transport Geography*, 2015. 44:53–64.
31. Mehadil Orvin, M., and M. Rahman Fatmi. Modeling Destination Choice Behavior of the Dockless Bike Sharing Service Users. *Transportation Research Record: Journal of the Transportation Research Board*, 2020. 2674(11):875–887.
32. Kumar Dey, B., S. Anowar, and N. Eluru. A Framework for Estimating Bikeshare Origin Destination Flows Using a Multiple Discrete Continuous System. *Transportation Research Part A: Policy and Practice*, 2021. 144:119–133.
33. Zhu, J., and X. Ye. Development of Destination Choice Model for Taxi Passengers in Shanghai, China. *CICTP 2017: Transportation Reform and Change - Equity, Inclusiveness, Sharing, and Innovation - Proceedings of the 17th COTA International Conference of Transportation Professionals*, 2018. 2018-January:3989–3997.
34. Zhang, S., X. Liu, J. Tang, S. Cheng, Y. Qi, and Y. Wang. Spatio-Temporal Modeling of Destination Choice Behavior through the Bayesian Hierarchical Approach. *Physica A: Statistical Mechanics and its Applications*, 2018. 512:537–551.
35. Zhu, J., and X. Ye. Development of Destination Choice Model with Pairwise District-Level Constants Using Taxi GPS Data. *Transportation Research Part C: Emerging Technologies*, 2018. 93:410–424.
36. El Zarwi, F., A. Vij, and J. L. Walker. A Discrete Choice Framework for Modeling and Forecasting the Adoption and Diffusion of New Transportation Services. *Transportation Research Part C: Emerging Technologies*, 2017. 79:207–223.
37. Perrine, K. A., K. M. Kockelman, and Y. Huang. Anticipating Long-Distance Travel Shifts Due to Self-Driving Vehicles. *Journal of Transport Geography*, 2020. 82:102547.
38. Sener, I. N., R. M. Pendyala, and C. R. Bhat. Accommodating Spatial Correlation across Choice Alternatives in Discrete Choice Models: An Application to Modeling Residential Location Choice Behavior. *Journal of Transport Geography*, 2011. 19:294–303.
39. Schönfelder, S. ;, H. ; Li, R. ; Guensler, J. ; Ogle, and K. W. Axhausen. Analysis of Commute Atlanta Instrumented Vehicle GPS Data. *Arbeitsberichte Verkehrs- und Raumplanung*, 2006. 350
40. Nicolau, J. L., and F. J. Más. Sequential Choice Behavior: Going on Vacation and Type of Destination. *Tourism Management*, 2008. 29(5):1023–1034.
41. Kitamura, R., C. Chen, and R. Narayanan. Traveler Destination Choice Behavior: Effects of Time of Day, Activity Duration, and Home Location. *Transportation Research Record: Journal of the Transportation Research Board*, 1998. 1645:76–81.
42. Pan, X., S. Rasouli, and H. Timmermans. Investigating Tourist Destination Choice: Effect of Destination Image from Social Network Members. *Tourism Management*, 2021. 83:104217.
43. Zha, L., Y. Yin, and Z. Xu. Geometric Matching and Spatial Pricing in Ride-Sourcing Markets. *Transportation Research Part C: Emerging Technologies*, 2018. 92:58–75.
44. Du, J., H. A. Rakha, and H. Breuer. An In-Depth Spatiotemporal Analysis of Ride-Hailing Travel: The Chicago Case Study. *Case Studies on Transport Policy*, 2022. 10(1):118–129.
45. Transportation Network Providers - Trips | City of Chicago | Data Portal.

https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p. Accessed Jul. 5, 2022.
46. Land Use Inventory - CMAP. https://www.cmap.illinois.gov/data/land-use/inventory. Accessed Jul. 5, 2022.
47. Census Bureau Data. https://data.census.gov/cedsci/. Accessed Jul. 10, 2022.
48. Datasets | Climate Data Online (CDO) | National Climatic Data Center (NCDC). https://www.ncei.noaa.gov/cdo-web/datasets. Accessed Jul. 5, 2022.
49. Boundaries - Census Tracts - 2010 | City of Chicago | Data Portal. https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Census-Tracts-2010/5jrd-6zik. Accessed Jul. 5, 2022.
50. Pozsgay, M. A., and C. R. Bhat. Destination Choice Modeling for Home-Based Recreational Trips: Analysis and Implications for Land Use, Transportation, and Air Quality Planning. *Transportation Research Record: Journal of the Transportation Research Board, 2001*. 1777:47–54.
51. Scott, D. M., P. A. Coomes, and A. I. Izyumov. The Location Choice of Employment-Based Immigrants among U.S. Metro Areas*. *Journal of Regional Science*, 2005. 45(1):113–145.
52. Scott, D. M., and S. Y. He. Modeling Constrained Destination Choice for Shopping: A GIS-Based, Time-Geographic Approach. *Journal of Transport Geography*, 2012. 23:60–71.
53. Mcfadden, D. Modelling the Choice of Residential Location. 1977.
54. Bhat, C. R. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research Part B: Methodological*, 2001. 35(7):677–693.
55. Yasmin, S., and N. Eluru. Evaluating Alternate Discrete Outcome Frameworks for Modeling Crash Injury Severity. *Accident Analysis and Prevention*, 2013. 59: 506–521.
56. Gurumurthy, K. M., Kockelman, K. M., & Auld, J. A system of shared autonomous vehicles for Chicago. *Journal of Transport and Land Use*, 2021. 14(1):933-948.
57. Joewono, T. B., Matthew, A., & Rizki, M. Loyalty of Paratransit Users in the Era of Competition with Ride Sourcing. *Sustainability*, 2021. 13(22):12719.
58. Ryan, S., and L. F. Frank. Pedestrian Environments and Transit Ridership. *Journal of Public Transportation*, 2009. 12(1):3.
59. Seong, E. Y., N. H. Lee, and C. G. Choi. Relationship between Land Use Mix and Walking Choice in High-Density Cities: A Review of Walking in Seoul, South Korea. *Sustainability 2021. 13(2):810*.
60. Dey, B. K., S. D. Tirtha, N. Eluru, and K. C. Konduri. Transformation of Ridehailing in New York City: A Quantitative Assessment. *Transportation Research Part C: Emerging Technologies*, 2021. 129:103235.
61. Brodeur, A., and K. Nield. An Empirical Analysis of Taxi, Lyft and Uber Rides: Evidence from Weather Shocks in NYC. *Journal of Economic Behavior & Organization*, 2018. 152:1–16.
62. Grahn, R., S. Qian, H. S. Matthews, and C. Hendrickson. Are Travelers Substituting between Transportation Network Companies (TNC) and Public Buses? A Case Study in Pittsburgh. *Transportation*, 2021. 48(2):977–1005.
63. Zhang, W., F. Gao, S. Sun, Q. Yu, J. Tang, B. Liu, and D. Ma. A Distribution Model for Shared Parking in Residential Zones That Considers the Utilization Rate and the Walking Distance. *Journal of Advanced Transportation*, 2020. 2020.
64. Lyft Becomes America’s Largest Bikeshare Service - Lyft Blog.

https://www.lyft.com/blog/posts/lyft-becomes-americas-largest-bikeshare-service. Accessed Jul. 7, 2022.

65. King, D. A., M. W. Conway, and D. Salon. Do For-Hire Vehicles Provide First Mile/Last Mile Access to Transit? *Transp. Find*, 2020. 14:1-7
66. Is Ridehailing Equitably Available Across Chicago? | Center for Neighborhood Technology. https://cnt.org/blog/is-ridehailing-equitably-available-across-chicago. Accessed Jul. 7, 2022.
67. Barajas, J. M., and A. Brown. Not Minding the Gap: Does Ride-Hailing Serve Transit Deserts? *Journal of Transport Geography*, 2021. 90:102918.
68. Dong, X. Trade Uber for the Bus?: An Investigation of Individual Willingness to Use Ride-Hail Versus Transit. *Journal of the American Planning Association*, 2020. 86(2):222–235.
69. Dey, B. K., S. D. Tirtha, and N. Eluru. Transport Networking Companies Demand and Flow Estimation in New York City: *Transportation Research Record: Journal of the Transportation Research Board, 2021*. 2675(9):139–153.
70. Lavieri, P. S., F. F. Dias, N. R. Juri, J. Kuhr, and C. R. Bhat. A Model of Ridesourcing Demand Generation and Distribution. *Transportation Research Record: Journal of the Transportation Research Board, 2018.* 2672(46):31–40.