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

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable Descriptions</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td><strong>DEPENDENT VARIABLES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trip fare model</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trip fare</td>
<td>Ln (Trip fare)</td>
<td>2.079</td>
</tr>
<tr>
<td><strong>INDEPENDENT VARIABLES (CONTINUOUS)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trip distance</td>
<td>Distance traveled in each trip</td>
<td>4.149</td>
</tr>
<tr>
<td>Network distance</td>
<td>Ln (Shortest distance between census tracts)</td>
<td>1.895</td>
</tr>
<tr>
<td>Demand in last 15 minutes at origin</td>
<td>Ln (Demand in last 15 minutes in each origin census tract)</td>
<td>2.006</td>
</tr>
<tr>
<td>Demand in last 15 minutes at destination</td>
<td>Ln (Demand in last 15 minutes in each destination census tract)</td>
<td>2.032</td>
</tr>
<tr>
<td><strong>Land Use and Built Environment Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network distance from CBD</td>
<td>Ln (Network distance to census tract from Central Business District (CBD))</td>
<td>1.871</td>
</tr>
<tr>
<td>Residential area</td>
<td>Total residential area in each census tract (area/100) in acre</td>
<td>0.602</td>
</tr>
<tr>
<td>Commercial area</td>
<td>Total commercial area in each census tract (area/100) in acre</td>
<td>0.115</td>
</tr>
<tr>
<td>Institutional area</td>
<td>Total institutional area in each census tract (area/100) in acre</td>
<td>0.113</td>
</tr>
<tr>
<td>Recreational area</td>
<td>Total recreational area in each census tract (area/100) in acre</td>
<td>0.074</td>
</tr>
<tr>
<td>Land use mix</td>
<td>Land use mix = $\left(\sum_{k} p_{k} \ln p_{k}\right) / \ln N$, 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</td>
<td>0.134</td>
</tr>
<tr>
<td><strong>Transportation Infrastructure Attributes</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable Descriptions</th>
<th>Descriptive Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bike lane density</td>
<td>Length of bike lane in each census tract per acre (Density*100) (mi/acre)</td>
<td>0.321 0.352</td>
</tr>
<tr>
<td>Length of street</td>
<td>Length of street in each census tract</td>
<td>5.597 4.953</td>
</tr>
<tr>
<td>Number of bus stops</td>
<td>Number of bus stops in each census tract</td>
<td>12.486 8.330</td>
</tr>
<tr>
<td>Number of L stations</td>
<td>Number of stations of L transit system in each census tract</td>
<td>0.156 0.529</td>
</tr>
<tr>
<td>Number of divvy stations</td>
<td>Number of divvy stations in each census tract</td>
<td>1.029 1.441</td>
</tr>
<tr>
<td>Walk score</td>
<td>Walk score (a measure of serviceability of walkability) in each census tract</td>
<td>82.397 26.225</td>
</tr>
<tr>
<td>Transit score</td>
<td>Transit score (a measure of serviceability of public transit) in each census tract</td>
<td>8.260 0.996</td>
</tr>
</tbody>
</table>

### Sociodemographic Attributes

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable Descriptions</th>
<th>Freq.</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment density</td>
<td>Number of employments in each census tract per acre (Density/100)</td>
<td>0.236</td>
<td>0.381</td>
</tr>
</tbody>
</table>

### Weather Attributes

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable Descriptions</th>
<th>Mean</th>
<th>Std. dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snow depth</td>
<td>Standard score ($\frac{x-\mu}{\sigma}$) of snow depth in each census tract. Where x is the observed value of snow depth, $\mu$ is the mean of the distribution of the values of snow depth and $\sigma$ is the standard deviation of the distribution of the values of snow depth</td>
<td>0.004</td>
<td>1.034</td>
</tr>
</tbody>
</table>

### INDEPENDENT VARIABLES (CATEGORICAL)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Variable Descriptions</th>
<th>Freq.</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trip Attributes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trip starts at AM peak</td>
<td>Trip starts within AM peak period</td>
<td>1965.000</td>
<td>19.650</td>
</tr>
<tr>
<td>Trip starts at PM peak</td>
<td>Trip starts within PM peak period</td>
<td>2560.000</td>
<td>25.600</td>
</tr>
<tr>
<td>Trip starts at other time</td>
<td>Trip starts in other time period</td>
<td>5475.000</td>
<td>54.750</td>
</tr>
<tr>
<td>Trip ends at AM peak</td>
<td>Trip ends within AM peak period</td>
<td>1876.000</td>
<td>18.760</td>
</tr>
<tr>
<td>Trip ends at PM peak</td>
<td>Trip ends within PM peak period</td>
<td>2481.000</td>
<td>24.810</td>
</tr>
<tr>
<td>Trip ends at other time</td>
<td>Trip ends in other time period</td>
<td>5643.000</td>
<td>56.430</td>
</tr>
<tr>
<td>Shared trip indicator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Trip authorized as shared</td>
<td>1507.000</td>
<td>15.070</td>
</tr>
<tr>
<td>No</td>
<td>Trip is not authorized as shared</td>
<td>8493.000</td>
<td>84.930</td>
</tr>
<tr>
<td>Sociodemographic Attributes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low income indicator</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Census tract with median income under $58 thousand USD (15th percentile)</td>
<td>466.000</td>
<td>58.543</td>
</tr>
<tr>
<td>No</td>
<td>Census tract with median income over $58 thousand USD (15th percentile)</td>
<td>330.000</td>
<td>41.457</td>
</tr>
</tbody>
</table>

---

**Figure 1** Total number of weekday trips (a) originated; (b) destined
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, \( q (=1, 2, 3, \ldots, Q=10,000) \) be an index to represent each individual trip, \( y_q \) be an index to represent the fare associated with a trip \( q \), and \( s (=1, 2, \ldots, S=30) \) 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:

\[
y_q = (\alpha' + \eta')z_q + \varepsilon_q
\]

where \( \alpha' \) is a vector of coefficients to be estimated, \( \eta \) 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 \( \varepsilon_q \) is an idiosyncratic random error term assumed independently normally distributed with variance \( \gamma^2 \). Now, we can express the probability of a trip, \( q \) having fare, \( y_q \) as follows:

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

where \( \phi(.) \) 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:

\[
u_{qs} = (\beta' + \eta')x_{qs} + \varepsilon_{qs}
\]

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 \( \beta \) is a vector of model coefficients to be estimated. The random error term, \( \varepsilon_{qs} \), 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:

\[
P(s_q) = \frac{\exp \left( (\beta' + \eta')x_{qs} \right)}{\sum_{s=1}^{S} \exp \left( (\beta' + \eta')x_{qs} \right)}
\]

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 $\eta$. 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 $\eta$ the probability for a trip, $q$ to have fare, $y_q$ and destination $s$ can be expressed as follows:

$$P(y_q, s_q|\eta) = P(y_q) \times P(s_q)$$  \hspace{1cm} (5)

The complete set of parameters to be estimated in the model system of Equation (5) are $\alpha, \beta$ and $\gamma$ and standard error term, $\sigma$. Let, $\Omega$ 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:

$$L_q|\Omega = \prod_{s=1}^{S} [P(y_q, s_q|\eta)]^{d_{qs}}$$  \hspace{1cm} (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:

$$L_q = \int_{\Omega} (L_q|\Omega)d\Omega$$  \hspace{1cm} (7)

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

$$\text{LL} = \sum_{q=1}^{Q} \ln L_q$$  \hspace{1cm} (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 $\Omega$. 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:

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

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 illustrate 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 several 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**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Trip Fare Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>1.638</td>
<td>86.286</td>
</tr>
<tr>
<td><strong>Trip Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trip distance</td>
<td>0.104</td>
<td>77.797</td>
</tr>
<tr>
<td>Trip start time (Base: Other periods)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM peak period</td>
<td>0.096</td>
<td>3.546</td>
</tr>
<tr>
<td>Trip end time (Base: Other periods)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM peak period</td>
<td>0.045</td>
<td>2.276</td>
</tr>
<tr>
<td>PM peak period</td>
<td>0.045</td>
<td>2.326</td>
</tr>
<tr>
<td>Shared trip indicator (Base: No)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>-0.371</td>
<td>-37.494</td>
</tr>
<tr>
<td><strong>Origin Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand in last 15 minutes at origin*AM peak period</td>
<td>-0.019</td>
<td>-2.683</td>
</tr>
<tr>
<td>Demand in last 15 minutes at origin*PM peak period</td>
<td>0.021</td>
<td>2.683</td>
</tr>
<tr>
<td>Network distance to origin from CBD</td>
<td>0.010</td>
<td>5.271</td>
</tr>
<tr>
<td>Network distance to origin from CBD*PM peak period</td>
<td>-0.011</td>
<td>-2.713</td>
</tr>
<tr>
<td>Number of L stations*AM peak period</td>
<td>-0.008</td>
<td>-1.643</td>
</tr>
<tr>
<td>Employment density at origin</td>
<td>-0.012</td>
<td>-2.675</td>
</tr>
<tr>
<td>Snow depth</td>
<td>-0.010</td>
<td>-2.838</td>
</tr>
<tr>
<td><strong>Destination Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand in last 15 minutes at destination*AM peak period</td>
<td>0.032</td>
<td>5.157</td>
</tr>
<tr>
<td>Network distance to destination from CBD</td>
<td>-0.007</td>
<td>-3.836</td>
</tr>
<tr>
<td>Network distance to destination from CBD*AM peak period</td>
<td>-0.005</td>
<td>-2.335</td>
</tr>
<tr>
<td>Number of Divvy stations</td>
<td>-0.004</td>
<td>-4.535</td>
</tr>
<tr>
<td>Land use mix</td>
<td>0.183</td>
<td>1.959</td>
</tr>
<tr>
<td>Employment density</td>
<td>0.014</td>
<td>3.096</td>
</tr>
<tr>
<td>Scale</td>
<td>0.229</td>
<td>31.564</td>
</tr>
<tr>
<td><strong>Destination Choice Model</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Land Use and Built Environment Attributes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Network distance between O-D</td>
<td>-1.092</td>
<td>-57.340</td>
</tr>
<tr>
<td>Distance from CBD</td>
<td>-0.539</td>
<td>-16.967</td>
</tr>
<tr>
<td>Distance from CBD*Trip starts at AM peak</td>
<td>-0.254</td>
<td>-4.705</td>
</tr>
<tr>
<td>Residential area</td>
<td>-0.927</td>
<td>-14.191</td>
</tr>
<tr>
<td>Commercial area</td>
<td>0.396</td>
<td>7.397</td>
</tr>
<tr>
<td>Institutional area</td>
<td>-0.168</td>
<td>-2.939</td>
</tr>
<tr>
<td>Recreational area</td>
<td>0.306</td>
<td>7.262</td>
</tr>
</tbody>
</table>
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 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
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
Comparison of Taxis, Transit, and Ridesourcing Services in San Francisco. Transport Policy,


