Transformation of Ridehailing in New York City: A Quantitative Assessment

Bibhas Kumar Dey*
Transportation Modeler
Connetics Transportation Group
Tel: 1-407-823-4815; Fax: 1-407-823-3315
Email: bibhas.ce@knights.ucf.edu

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

Naveen Eluru
Professor
Department of Civil, Environmental and Construction Engineering
University of Central Florida
Tel: 1-407-823-4815; Fax: 1-407-823-3315
Email: naveen.eluru@ucf.edu

Karthik C. Konduri
Assistant Professor
Department of Civil and Environmental Engineering
University of Connecticut
Castleman Building Rm 331, Storrs, CT 06269
Tel: 860-486-2733; Email: karthik.konduri@uconn.edu

*Corresponding author

ABSTRACT

The proposed study contributes to our understanding of the ongoing transformation of ridehailing market by examining the New York City Taxi & Limousine Commission data from a fine spatial and temporal resolution. We examine taxi zone based demand data from NYC for each month and explore the reasons contributing to (a) the increase in ridehailing demand and (b) the shift from traditional taxi services to Transportation Networking Company (TNC) services. The first component – taxi zone ridehailing demand - is analyzed adopting a negative binomial count model. The second component - share of traditional and TNC services demand - is analyzed using a multinomial fractional split model. The two model components are stitched together in a joint framework that allows for the influence of repeated observations as well as for the presence of common unobserved factors affecting the two components. The model estimation considered a comprehensive set of independent variables including transportation infrastructure variables, land use and built environment variables, weather attributes, and temporal attributes. Several performance measures were generated using the joint model for estimation and validation datasets. A prediction exercise is conducted to illustrate how the proposed model system can be utilized for predicting future ridehailing trends. Finally, an elasticity exercise is conducted to estimate the influence of independent variables on the ridehailing market.

Keywords: *Ridehailing demand, NB-MNL Fractional split model, Time elapsed, Correlation*

1 INTRODUCTION

In most urban regions, individuals, who do not have access to or do not prefer to use personal vehicles, have the option of either using public transit, bike/scooter sharing systems (for short distance trips) or a ridehailing service (such as taxi or Uber). While public transit systems are constrained by predefined routes and fixed schedules, bike/scooter sharing systems are limited by small distance range, ridehailing services at a cost provide individuals with convenient door-to-door car trips without the additional challenges associated with driving/bicycling (such as having to find a parking spot, concentrating on driving and physical effort of bicycling). In recent years, ridehailing has undergone a rapid transformation in response to the transformative technological changes including smart mobile availability, ease of hailing a ride using mobile applications, integration of seamless payment systems and real-time driver and user reviews. The convenience offered by transport networking companies (TNC) (such as Uber, Lyft, and Via) has allowed for tremendous growth in ridehailing demand. For example, in New York City, the average daily trips by taxi (Yellow taxi) was varying between 400,000and 500,000 for the years 2010-2014 *(1)*. However, since 2014, with the advent of TNC services in the city, the total number of trips have increased. Based on New York City Taxi & Limousine Commission (NYCTLC) report *(1)*, from 2015 to 2018, TNC daily trips increased from 60,000 to 700,000 while traditional taxi (Yellow and Green together) daily trips declined from 450,000 to 285,000. The trend observed in NYC is not an exception. A recent report analyzing reimbursed travel in the US has found that the share of Uber and Lyft has increased from 8% to 72.5% from 2014-2018 at the cost of taxi and rental car business share *(2)*.

The TNC service induced transformation can be viewed as constituting two major components. The first component is the overall increase in ridehailing demand possibly drawing from population of individuals driving, using public transit, and even inducing newer travel. The second component of the transformation is the shift in the share of traditional taxi service demand toward TNC services *(3)*. In a short timeframe, in NYC, TNC services have increased their market share from 0 to nearly 70% by the end of 2018. While preliminary research has begun to explore the reasons for the transformation, it is safe to assume economists and social scientists will continue to examine the transformation for several years into the future.

The proposed study contributes to our understanding of this transformation by examining the NYC data from a fine spatial and temporal resolution by adopting an innovative joint econometric model system. The study examines two components of the transformation (a) the increase in ridehailing demand and (b) the shift from traditional taxi services to TNC services. The first component – taxi zone ridehailing demand - is analyzed adopting a negative binomial count model. The second component - share of traditional and TNC services demand - is analyzed using a multinomial fractional split model. The two model components are stitched together in a joint framework that allows for the influence of repeated observations as well as for the presence of common unobserved factors affecting the two components. The study employs trip level data from the NYCTLC from January 2015 through December 2018 for the analysis. The data is aggregated by taxi zone for every month in the study period and analyzed by ridehailing alternatives: Yellow taxi, Green taxi and TNC services (including Uber, Lyft, Juno and Via).

2 EARLIER RESEARCH AND CURRENT STUDY

Literature related to ridehailing vehicles can be categorized into three main streams: a) studies investigating various operational and quantitative aspects of taxis, b) studies investigating the evolution and various qualitative aspects of TNC based ridehailing and c) studies examining the relationship between various ridehailing systems and their interaction with public transportation[1].

The _first group of studies_ focused on taxi services from different perspectives, including entry regulation (see Schaller *(4)* for US and Canada regulation and Çetin and Eryigit *(5)* for Istanbul regulation), demand and pricing *(6-8)*, and impact of emerging technologies such as electric and autonomous vehicles *(9-11)*. Several studies analyzed different aspects of taxi operations including taxi passenger search schemes and routing of vacant taxis to improve the efficiency of taxi services *(12-17)*. Crash injury severity and safety issues related to taxi services are also examined by several researchers *(18-22)*.

The _second group of studies_ explored TNC evolution, factors that affected usage, licensing and policy formulation, pricing mechanisms, and comparison across ridehailing services (with taxis or between various smart phone based ridehailing companies). These studies typically rely on questionnaire interviews, and online surveys for data collection. TNC evolution studies focused on the definition of ridehailing systems, how ridehailing services have evolved over time *(23-25)*, investigated the challenges and opportunities presented by real-time services and highlighted various opportunities for the future *(26; 27)*. A set of studies explored the influence of various factors affecting TNC usage. For example, Cramer and Krueger *(28)* analyzed passenger service times for Uber and taxi across five major cities in the US. The authors concluded that availability of driver-passenger reviews, Uber's flexible labor supply model coupled with inefficient taxi regulations for passenger safety contributed to higher Uber utilization rates. Multiple studies explored pricing, and waiting times associated with various ridehailing companies *(29-32)*. Another subset of studies conducted quantitative analysis using TNC usage data exploring trip patterns (a) to identify factors influencing TNC demand, (b) to understand TNC demand and its relationship with existing transportation modes. Factors that were found to affect ridehailing demand include temporal and weather patterns, land use attributes such as lower transit access time, higher length of roadways, lower vehicle ownership, higher income and more job opportunities *(33-35)*.

The _third group of studies_ is comprised of research conducting comparative analysis using ridehailing usage data. The research conducted in this paper falls into this third category. A group of studies investigate the new age ridehailing demand considering relationship between ridehailing services with public transit system *(3; 36-38)*. Rayle et al. *(37)* conducted a trip intercept survey to understand the source of TNC demand and concluded that nearly 50% of the demand is transferred from public transit and driving. Studies comparing the emerging ridehailing services with existing services such as public transit and bicycle sharing offer interesting results. Gerte et al. *(3)* found evidence for shifting taxi demand to smart phone based ridehailing services in New York City. Further, the study also found evidence of substitution relationship between ridehailing and bicycle share systems. Komanduri et al. *(38)* analyzed data from RideAustin, to examine the trip length and temporal distribution of the trips. A comparison of the adoption of RideAustin relative to public transit alternatives illustrated that riders were choosing RideAustin to minimize travel time (highlighting the higher value of time for these travelers). Using the same data, Yu and Peng *(39)* and Lavieri et al. *(40)* studied TNC trip demand and found that population characteristics, household characteristics, built environment and transit supply influence TNC trip demand. Poulsen et al. *(41)* examined how the two systems that were introduced in the same time performed - Uber and Green taxis - in Manhattan area and found that the growth rate for Uber was substantially higher. Babar and Burtch *(42)* compared the utilization rate of transit service in the

[1] For a recent detailed literature review on emerging ridehailing alternatives see Wang and Yang *(43)*.

US after the introduction of TNC services and found that utilization rate of bus service dropped while long-haul transit services (such as subway and commuter rail) experienced increasing utilization. Nie *(44)* examined the competition between taxi industry and TNC and interestingly found that taxi industry in Shenzhen, China survived the emergence of ridesourcing. Wang et al. *(45)* analyzed the market equilibrium under TNC and traditional taxi services and found that changes in platform charges can alter the equilibrium. More recently multiple studies have conducted economic analysis of ridehailing markets focusing on passenger and driver matching and studying the impact of ride splitting among passengers on TNC demand *(46-48)*.

2.1 Current Study in Context

The proposed study contributes to our understanding of the ongoing transformation of ridehailing market by examining the NYC data from a fine spatial and temporal resolution using an innovative joint econometric model. Specifically, as opposed to considering the transformation at a regional scale and in a 4 year period, we examine taxi zone based demand data from NYC for each month and explore the reasons contributing to (a) the increase in ridehailing demand and (b) the shift from traditional taxi services to TNC services. A negative binomial count model and a multinomial fractional split model are used to analyze ridehailing demand and proportion of traditional and TNC demand respectively. As the data for the two components is obtained for the same spatial record, there are several common unobserved factors influencing the two variables. The database generated also has multiple data points for each spatial unit. Thus, a joint econometric model that accommodates for repeated measures (panel) and common unobserved factors across the two dependent variables is developed. Specifically, we build on the cross-sectional joint negative binomial and multinomial fractional split model developed in Bhowmik et al. *(49)* for a different empirical context. The reader would note that traditional and TNC services demand can also be analyzed using a multivariate random parameter negative binomial model system. In the multivariate model system, the main interaction across different demand variables is sought through unobserved effects. On the other hand, the fractional split approach directly relates an exogenous variable to demand proportions simultaneously and allows for the estimation of a parsimonious specification. Further, as illustrated in Bhowmik et al. *(49)*, the fractional split model system outperforms the traditional multivariate count system in model prediction.

The econometric model will explore how various attributes typically considered within the transportation planning process affect the transformation. The comprehensive set of independent variables considered include transportation infrastructure variables, land use and built environment variables, weather attributes, and temporal attributes (as has been considered in other studies such as Yu and Peng *(39)* and Lavieri et al. *(40)*). The study recognizes that technology adoption cannot be explained by simply considering the variables described. To quantify the impact of time, we explicitly consider time elapsed since the beginning of TNC data collection in NYC as a surrogate variable of technological adoption impact. To conduct our research analysis, the data is drawn from New York City Taxi & Limousine Commission (NYCTLC) website from January 2015 through December 2018. The data was processed to obtain monthly pickup demand for three ridehailing alternatives (Yellow taxi, Green taxi and TNCs) at the taxi zone level. The model estimates are validated using a holdout sample. Further, a policy exercise is conducted to illustrate how the proposed model system can be utilized for predicting future ridehailing trends.

3 DATA

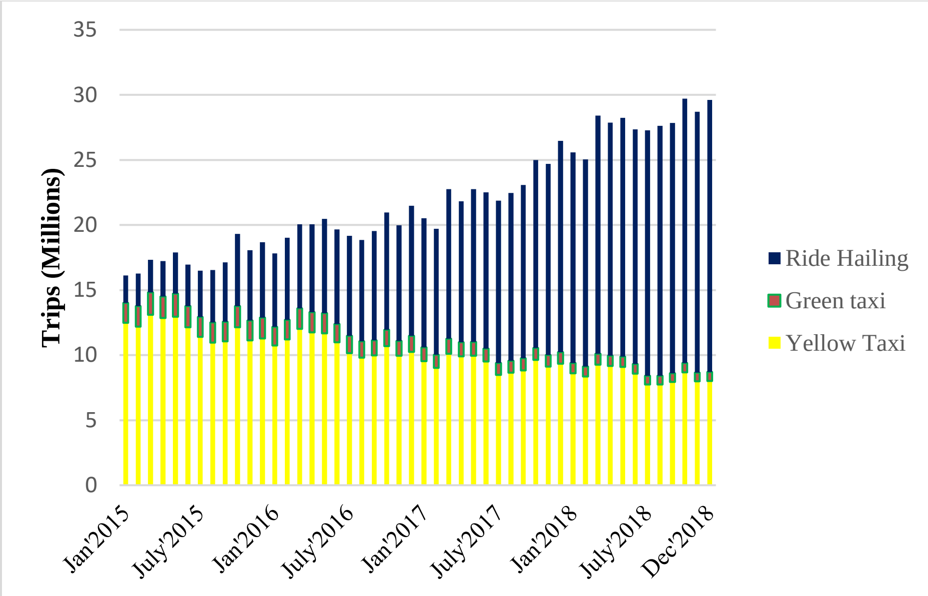
3.1 Data Source

The NYCTLC provides spatially aggregated trip data from all transportation networking companies (taxi, Uber, Lyft, Juno and Via) for public use (https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page). Yellow taxis are traditional and iconic ridehailing service in NYC while Green taxis known as boro taxis and street-hail liveries started operation in August 2013 and operate pickups in northern Manhattan, Bronx, Brooklyn, Staten Island and Queens with the ability to drop off anywhere in NYC. TNCs became operation at around a similar time frame. Thus, it is informative to examine how the share of Green taxi and TNCs has evolved with time. The trip itinerary dataset was collected from 2015-2018 for Yellow taxi, Green taxi and TNC (Uber, Lyft, Juno and Via) for our analysis. The dataset provides information on start and end time of trips, origin and destination defined as taxi zone ID, trip distance and vehicle license number. The trip data was augmented with other sources including: (1) built environment attributes derived from New York City open data (https://nycopendata.socrata.com); (2) socio-demographic characteristics at the census tract/zip code level gathered from US 2010 census data; (3) the weather information corresponding to the Central Park station retrieved from the National Climatic Data Center (http://www.ncdc.noaa.gov/data-access).

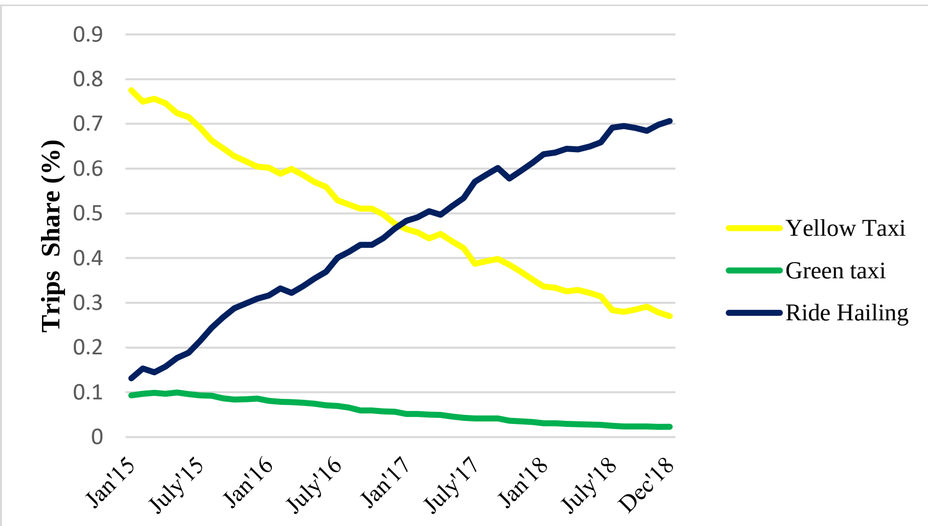
3.2 Sample Formation and Dependent Variable

A series of data cleaning and compilation exercises were undertaken for generating the sample data for estimation purposes. *First*, trips with missing or inconsistent information were removed. *Second*, trips longer than 500 minutes in duration (around 0.08% of all trips) were deleted considering that these trips are not typical ride-sharing trips. These trips could also be a result of two possibilities; either destination of those trips could be outside NYC or due to technical issues the trip information was recorded incorrectly. *Third*, trips that had the origin and destination outside of NYC taxi zone were also eliminated. Therefore, we focus on trips that originated and were destined within NYC taxi zone region only.

For the given study period (January 2015 to December 2018), the total number of available taxi zones in NYC was 259. Initially, we aggregated pickup data for each month from January 2015 to December 2018 for each origin taxi zone ID. Figure 1(a) represents the total trips generated in each month from January 2015 to December 2018 by each ridehailing alternatives while Figure 1(b) represents the proportion of total trips shared by Yellow taxi, Green taxi and TNC services. The evolving number of trips by ridehailing type offers clear depiction of how demand has increased as well as how TNC demand has surpassed traditional taxi demand. TNC service share crossed the share of Yellow taxi in February 2017. Figure 1(b) represents the trips proportion shared by the three ridehailing alternatives from 2015 to 2018. The Figure highlights TNC's trip share increased from 13% to 70% from 2015-2018 while Yellow taxis share declined from 77% to 27%. It is important to note that the share of Green taxi dropped consistently to become almost negligible in 2018. The main reason we still retained Green taxi as a separate alternative is to contrast two services (Green taxi and TNCs) that started operation in the same time frame. For our analysis, we aggregated trip data for 48 months from January 2015 to December 2018. To obtain a reasonable sample size for model estimation, 24 months were randomly selected for each taxi zone for analysis.



(a) Total monthly trips of all ridehailing alternatives.



(b) Monthly trips share between three ridehailing alternatives.

Figure 1 Dependent variable distribution.

3.3 Exogenous Variables

Several independent variables generated in our study are described below:

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

Several land use and built environment variables were considered including population density, job density and establishment density, the number of institutional facilities (schools, colleges, hospitals), the number of point of interests (museums, shopping malls), and the number of restaurants (including coffee shops and bars), total area of parks and commercial space (office, industry, retail) within each taxi zones. Distance of destination from Times Square and airport were estimated by using the shortest path algorithm tool of ArcGIS software. Airport indicator variable for the taxi zone was generated to examine the additional impact of airport destination. Population, job density and median income information was collected from US Census for 2015-2017 and extrapolated for 2018. Household car ownership information for 2018 was used to generate proportion of zero car ownership at taxi zone level to examine the impact of car ownership on riders' trip count and mode choice preferences. Non-motorized vehicle score (average of walk score[2] and bike score) and transit score associated with each taxi zone was considered at the census tract level. Further, crime density and accident density were also generated at taxi zone level. Total number of crimes of all types for previous year was aggregated at census tract level and crime density was estimated by dividing corresponding year's population. In a similar manner, total number of accidents for each month was considered to generate accident density.

Weather variables include average temperature, precipitation, and snow for that particular month of the year. Several interaction variables were also created. Seasonality is the one of the temporal variables considered. We consider winter (December-February), Spring (March-May), Summer (June-August) and Fall (September-November) as dummy variables. Finally, we recognize that technology adoption cannot be explained by simply considering the variables described. To quantify the impact of time, we explicitly consider time elapsed since the beginning of TNC data collection (and other functional forms of the variable) as a temporal variable.

4 METHODOLOGY

The proposed joint econometric system jointly models "total number of trips" and "proportion of trips by type of ridehailing". The first variable is modeled using a Negative Binomial (NB) model and the second variable is analyzed using the multinomial logit fractional split (MNLFS) model. The mathematical details of the Joint NB-MNLFS model follows.

4.1 NB Component

Let i be the index for taxi zone ($i = 1,2,3, \dots, N$) and y_{it} be the ridehailing demand for a taxi zone i in time period $(t = 1,2,3, \dots, T)$. The NB probability expression for random variable y_{it} can be written as *(50)*:

$$P_{it}(y_{it}) = \frac{\Gamma(y_{it} + \alpha^{-1})}{\Gamma(y_{it} + 1)\Gamma(\alpha^{-1})} \left(\frac{1}{1 + \alpha\mu_{it}}\right)^{\frac{1}{\alpha}} \left(1 - \frac{1}{1 + \alpha\mu_{it}}\right)^{y_{it}} \tag{1}$$

where, P_{it} is the probability that taxi zone i has y_{it} number of trips over time period of t. $\Gamma(\cdot)$ is the Gamma function, α is the NB dispersion parameter and μ_{it} is the expected number of trips listed in taxi zone i for time period t and can be expressed using a log-link function as:

[2] Walk (bike) score provide a numerical value assessing the walkability (bikability) in the community. The exact methodology employed is detailed in their webpage https://www.walkscore.com/.

$$\mu_{it} = E(y_{it}|\boldsymbol{x}_{it}) = exp\left((\boldsymbol{\partial} + \aleph_i)\boldsymbol{x}_{it} + \delta_{itj} + \varphi_{it}\right) \tag{2}$$

where, $\boldsymbol{x}_{it}$ is a vector of explanatory variables associated with taxi zone i for time period t. $\boldsymbol{\partial}$ is a vector of coefficients to be estimated. $\aleph_i$ is a vector of unobserved factors on ridehailing demand propensity and its associated zonal characteristics assumed to be a realization from standard normal distribution: $\aleph_i \sim N(0, \boldsymbol{\varsigma}^2)$. δ_{itj} captures unobserved factors that simultaneously impact total number of trips and proportion of trips by ridehailing type j $(j = 1, 2{,}3; \mathrm{J} = 3)$ for taxi zone i and time period t. φ_{it} is a gamma distributed error term with mean 1 and variance α.

4.2 MNLFS Component

Let z_{itj} be the fraction of trips by ridehailing type j in taxi zone i and time period t.

$$0 \leq z_{itj} \leq 1, \qquad \sum_{j=1}^{J} z_{itj} = 1 \tag{3}$$

Let the fraction z_{itj} be a function of a vector w_{itj} of relevant explanatory variables associated with attributes of taxi zone i and time period j.

$$\begin{gathered} E[z_{itj}|w_{itj}] = Q_{itj}(\cdot) \\ 0 < Q_{itj}(\cdot) < 1, \;\; \textstyle\sum_{j=1}^{J} Q_{itj}(\cdot) = 1 \end{gathered} \tag{4}$$

where $Q_{itj}(\cdot)$ is a predetermined function. The properties specified in equation (4) for $Q_{itj}(\cdot)$ warrant that the predicted fractional ridehailing types will range between 0 and 1 and will add up to 1 for each zone. In this study, a MNL functional form for Q_{itj} in the fractional split model of equation (4). Then equation (4) is rewritten as:

$$E(z_{itj}|w_{itj}) = Q_{itj}(\cdot) = \frac{\exp(\left(\boldsymbol{\beta'_j} + \boldsymbol{\sigma_{ij}}\right)w_{itj} \pm \delta_{itj} + \xi_{itj})}{\sum_{j=1}^{J}\exp(\left(\boldsymbol{\beta'_j} + \boldsymbol{\sigma_{ij}}\right)w_{ij} \pm \delta_{itj} + \xi_{itj})}, j = 1,2,3, \dots, \tag{5}$$

where, $\boldsymbol{w}_{itj}$ is a vector of attributes, $\boldsymbol{\beta}'_j$ is the corresponding vector of coefficients to be estimated for ridehailing type j. $\boldsymbol{\sigma}_{ij}$ is a vector of unobserved factors assumed to be a realization from standard normal distribution: $\boldsymbol{\sigma} \sim N(0, \boldsymbol{\nu}_j^2)$. ξ_{itj} is the random component assumed to follow a Gumbel type 1 distribution. δ_{itj} term generates the correlation between equations for total number of trips and trip proportions by ridehailing types. The $\pm$ sign in front of δ_{itj} in equation (5) indicates that the correlation in unobserved zonal factors between total trips and trip proportions by ridehailing type may be positive or negative. A positive sign implies that taxi zones with higher number of trips are intrinsically more likely to incur higher proportions for the corresponding ridehailing types. On the other hand, negative sign implies that taxi zones with higher number of trips intrinsically incur lower proportions for different ridehailing types. To determine the appropriate sign, we empirically test the models with both $'+'$ and $'-'$ signs independently. The model structure that offers the superior data fit is considered as the final model.

It is important to note here that the unobserved heterogeneity between total number of trips and trip proportions by ridehailing types can vary across taxi zones. Therefore, in the current study, the correlation parameter δ_{itj} is parameterized as a function of observed attributes as follows (see Rahman et al. *(51)*; Tirtha et al. *(52)* and Faghih-Imani and Eluru *(53)* for similar specification):

$$\delta_{itj} = \boldsymbol{\pi_j} \boldsymbol{\tau}_{itj} \tag{6}$$

where, $\boldsymbol{\tau}_{itj}$ is a vector of exogenous variables, $\boldsymbol{\pi_j}$ is a vector of unknown parameters to be estimated (including a constant).

In examining the model structure of total trip count and proportion of trips by ridehailing types, it is necessary to specify the structure for the unobserved vectors $\boldsymbol{\varsigma}, \boldsymbol{\sigma}$ and $\boldsymbol{\pi}$ represented by Ω. In this paper, it is assumed that these elements are drawn from independent realization from normal population: $\Omega \sim N(0, (\boldsymbol{\varsigma^2}, \boldsymbol{\nu}_j{}^2, \beth_j^2))$. Thus, conditional on Ω, the likelihood function for the joint probability can be expressed as:

$$\mathcal{L}_i = \int_{\Omega} P(y_{it}) \times \prod_{t=1}^{T} \prod_{j=1}^{J} \left(E\left(z_{itj} \middle| w_{itj}\right) \right)^{z_{itj}} f(\Omega) d\Omega \tag{7}$$

Finally, the log-likelihood function is:

$$\mathcal{LL} = \sum_i Ln(L_i) \tag{8}$$

All the parameters in the model are estimated by maximizing the logarithmic function $\mathcal{LL}$ presented in equation (8). The parameters to be estimated in the joint model are: $\boldsymbol{\partial}, \boldsymbol{\alpha}, \boldsymbol{\beta'_j}, \boldsymbol{\nu}_j$ and $\boldsymbol{\beth_j}$**.** To estimate the proposed joint model, 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 across individuals (see *54-56* for examples of Quasi-Monte Carlo approaches in literature).

5 ESTIMATION RESULTS

5.1 NB-MNL Fractional Split Joint Model

The reader would note that the proposed methodology is flexible to allow for unobserved heterogeneity. The unobserved parameters tested in our analysis include: (1) correlation between demand component and ridehailing proportion components, (2) correlation across ridehailing proportion components and (3) random parameters in demand and proportion components. However, running the joint model with a large number of unobserved parameters is computationally burdensome. Hence, we adopt a sequential approach testing for sets of unobserved parameters in the initial estimations and then re-estimating the models with variables that offer significant parameters in the initial estimations. Table 1 presents the model estimation results of the joint NB-MNL fractional split model. The second column provides the results of the NB component while columns 3 through 5 present the results of the MNL fractional split model.

The model results are discussed separately for total ridership demand and proportion by ridehailing alternatives.

5.1.1 Total Ridership Demand (NB Component)

A positive (negative) sign for a variable in the ridehailing demand component of Table 1 indicates that an increase in the variable is likely to result in more (less) ridehailing trips.

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

As expected, zones located in census tracts with higher population density are more likely to be associated with higher number of trips. Similarly, increased job density and median income of in taxi zones is found to increase demand for ridehailing trips (see Correa et al. *(34),* Smart et al. *(32)* for similar results). The increased proportion of zero car households in urban areas increases demand for ridehailing (Correa et al. *(34)* found similar association with lower vehicle ownership households). As expected, increased transit accessibility within a taxi zone increases the propensity for higher ridehailing demand while taxi zones with higher non-motorized score reduce the appeal towards use ridehailing. It is possible that the presence of bicycle sharing serves as a competitive alternative for shorter trips (see Faghih-Imani et al. *(57)* for analysis in the context of short trips).

Several variables associated with travel generation are found to affect ridehailing demand. The presence of activity opportunities in the form of restaurants and cafes, recreational centers and point of interests (POI) is positively associated with demand (see Li et al. *(58)* and Wenzel et al. *(59)* for similar results). Taxi zones with higher residential area are positively associated with ridehailing demand. The result potentially alludes to the adoption of ridehailing service for commute activities from residential zones. As expected, availability of airport in taxi zones increases demand for ridehailing. The total area of recreational parks in the taxi zone has a positive influence on ridehailing demand. The result highlights the role of recreational parks serving as generators of ridehailing demand.

The study also considered the impact of landmarks such as Airports and Times Square[3] on ridehailing demand. The presence of an airport in the taxi zone, as expected, contributes to higher ridehailing demand. Interestingly, as the distance of taxi zone from airports increases, the model indicates an increase in ridehailing demand. On the other hand, as the distance from Times Square increases, ridehailing demand is expected to reduce. The result is intuitive as Times Square and the proximal zones serve as attraction centers for regular and tourist travel.

[3] Times Square represents an iconic destination in New York representing the center of the business district. Several earlier papers modeling bikeshare and ridesourcing have considered Times Square as a significant point of interest (see Dey et al. *(60, 61),* and Liu et al. *(62)*).

Table 1 Joint NB-MNLFS Model Estimation Results

Joint Component	**NB Model (Counts)**		**MNLFS Model (Proportions)**					
Ridehailing Type	**Total Trips**		**Yellow Taxi**		**Green Taxi**		**TNC**	
Variable Name	**Estimate**	**t-stat**	**Estimate**	**t-stat**	**Estimate**	**t-stat**	**Estimate**	**t-stat**
Constant	-1.426	-10.40	2.688	9.44	0.639	1.42	---	---
Land Use and Built Environment Attributes								
Population Density	0.245	2.12	2.069	4.35	-3.813	-3.55	---	---
Job Density	2.553	19.02	---	---	---	---	1.968	4.07
Median Income ($x10^{-3}$)	0.651	17.08	1.366	7.33	---	---	---	---
Proportion of Zero Car HH	1.003	9.70	---	---	3.508	5.28	0.830	1.85
Transit Score ($x10^{-2}$)	1.478	8.51	---	---	---	---	---	---
Non-motorized vehicle score ($x10^{-2}$)	-1.189	-6.34	---	---	---	---	---	---
Number of Restaurants and sidewalk café in Taxi Zone ($x10^{-3}$)	0.655	10.66	---	---	---	---	-2.975	-4.84
Number of Point of Interests and Recreational Points in Taxi Zone ($x10^{-3}$)	0.194	8.52	4.459	5.04	---	---	---	---
Residential Area (m^2 x 10^{-6})	1.570	8.94	---	---	---	---	---	---
Park Area (m^2 x 10^{-6})	1.484	10.22	16.665	4.89	-5.302	-2.43	---	---
Airport Indicator	0.723	35.99	3.511	9.47	---	---	---	---
Airport Distance (m x 10^{-3})	4.089	60.66	---	---	---	---	0.313	2.63
Times Square Distance (m x 10^{-3})	-1.047	-35.77	-2.384	-14.33	-0.511	-2.65	---	---
Accident Density ($x10^{-3}$)	---	---	---	---	-1.684	-2.53	---	---
Transportation Infrastructure and Attributes								
Bike Lane Density in Taxi Zone	-1.522	-8.97	-2.111	-2.22	---	---	---	---
Number of Bikeshare Stations in Taxi Zone ($x10^{-2}$)	-0.059	-2.65	---	---	-0.322	-1.97	---	---
Street Length in Taxi Zone (m x 10^{-3})	0.401	2.30	-10.183	-4.15	---	---	---	---
Number of Bus Stops and Subway Stations in Taxi Zone ($x10^{-3}$)	1.174	62.35	-3.815	-4.84	---	---	---	---
Temporal and Weather Attributes								
Times Square Distance (m x 10^{-3}) x Summer (Season)	-0.577	-5.65	---	---	---	---	---	---
Time Elapsed as Month Sequel	2.194	33.96	-0.054	-14.35	-0.083	-18.84	---	---
Snow Depth (cm)	-0.031	-7.26	0.281	2.97	---	---	---	---
Dispersion Parameters	0.160	27.45	---	---	---	---	---	---
Correlation	---	---	0.785	10.20	---	---	0.785	10.20
Model Fitness								
No. of observations	6216							
Quasi Log-likelihood	-16673.58							

5.1.1.2 Transportation Infrastructure and Attributes

Several transportation infrastructure variables such as bike lane density, bikeshare stations, street length, bus stops and subway stations were considered in the demand model. The parameter estimates for bike length indicate that probability of ridehailing trips decreases with increasing bike length density in the taxi zone. The negative association with number of bikeshare stations within a taxi zone highlights that ridehailing trip demand is likely in competition with bikeshare demand (for shorter distance share). For shorter distance travel, the segment of ridehailing users may shift to bikeshare services especially under congested conditions (see Faghih-Imani et al. *(57)* for competition between bikeshare and Taxi). The reader would note that the impact of bikeshare stations on ridehailing demand while significant is quite small in magnitude. An increase in the street length within a taxi zone has a positive impact on demand. (similar to findings of Correa et al. *(34)*). The number of bus stops and subway stations in the taxi zone has a positive coefficient indicating an increment in ridehailing demand. This result highlights the complementarity between ridehailing and public transit alternatives (please see Sadowsky & Nelson *(63)*; Tirachini & del Río *(64); Yu & Peng (39)* for similar results).

5.1.1.3 Temporal and Weather Attributes

An interaction variable of summer season with Times Square distance from each taxi zone was used and the results highlight an interesting result. The results indicate that the ridehailing demand in summer reduces faster than rest of the year as we move away from Times Square. The result clearly highlights the attraction of Times Square during summer months for visitors and their plausible adoption of ridehailing. Time elapsed variable that counts the month from January 2015 to December 2018 was used to find the impact of temporal trend attribute on ridehailing trip count. The result highlights the positive association with ridehailing representing how with time overall demand has increased. Finally, as the depth of snow in the taxi zone increases, ridehailing demand reduces. This is expected as trip generation across all modes is likely to reduce under snowy conditions (please see Brodeur & Nield *(65)* for similar results).

5.1.2 Trip Proportion (MNL Fractional Split Component Model)

In the MNL fractional split model, a positive (negative) sign for a variable indicates that an increase in the variable is likely to result in higher proportion of trips for the corresponding alternative relative to the base alternative for that variable.

5.1.2.1 Constant parameters

The constant parameters have no substantive interpretation after introducing independent variables.

5.1.2.2 Land Use and Built Environment Attributes

In the context of land use and built environment attributes, population density in a census tract has significant impact on trip proportions. Increasing population has a positive impact on Yellow taxi proportion and negative impact on Green taxi proportion. The result seems reasonable since Green taxi has regulations restricting on-street pickup. In a similar vein, with higher job density, the proportion of TNC increases. The result potentially indicates preference among employed individuals for TNC. Taxi zones with high median income have positive association with Yellow taxi proportion. The result probably reflects the indifference to typically higher fares of Yellow taxi relative to TNCs. With increasing zero car ownership households, the likelihood of Green taxi and TNC services trips proportion increases. Zero car households are inclined to adopting TNC services that are usually less expensive compared to taxis.

A negative association is observed for the presence of restaurants and cafes with TNC trip proportions while recreational centers and point of interests (POI) have an increased likelihood for the Yellow taxi proportions. While this might appear counter-intuitive, the finding potentially indicates lower wait times for taxi services in these locations. It is also important to recognize that recreational centers and POIs result in an overall increase in the number of ridehailing trips. Thus, the proportion reduction might simply indicate a higher share of taxi trips in these taxi zones (relative to other taxi zones). The exact magnitude of the impact is a complex interaction across the two components and must be carefully evaluated. In terms of land use type, only proportion of park area variable has significant impact on trip proportions. The likelihood of Yellow taxi trips increases for a high percentage of park area in a taxi zone while Green taxi trip proportion reduces. The result highlights how Yellow taxi is more preferred in these taxi zones relative to the other alternatives. The result is possibly a manifestation of the differences in service regions for Yellow and Green taxis. The reader would note that an increase in park area is associated with an overall ridehailing demand increase. Thus, even with this increase in share for Yellow taxi, the overall share of TNC is likely to be higher than other two alternatives. As expected, availability of airport in taxi zones increases the inclination of choosing Yellow taxis (See similar results for Yellow taxi share for airport originated trips *(66)*). At airports, taxi services receive preferential treatment with pickup bays and are hence likely to have a larger share. As the distance between taxi zone and airport increases, the share of TNC alternative increases. It is possible that TNC services are more readily available in these locations. Further, the pricing of TNC services is cheaper and are hence preferred away from airport locations. As taxi zones are further from Times Square, trip proportions for both taxis reduce reflecting their low availability as we move further away from Times Square. The results for accident density from the previous year reveal that taxi zones with higher accident density is likely to reduce Green taxi proportion (no impact on total ridership). It is possible that accident density is potentially a surrogate for roadway infrastructure challenges in locations served by Green taxi. This is an interesting result and warrants additional examination (in the future).

5.1.2.3 Transportation Infrastructure and Attributes

Several transportation infrastructure characteristics considered are found to be significant determinants of trip proportions by various ridehailing alternatives. Yellow taxi trip proportions are negatively associated with higher bike length density. The result might be a reflection of the service region differences between Yellow and Green taxis. The result warrants further attention in future research. Among transportation attributes, trip proportion of Green taxi trips is found to be lower for taxi zones with higher bike sharing stations in vicinity while Yellow taxi trip proportions are negatively associated with higher number of bus stops in taxi zones. An increase in the street length within a taxi zone results in a decreased of Yellow taxi proportions.

5.1.2.4 Temporal and Weather Attributes

Elapsed time considering month is negatively associated with Yellow and Green taxi trips proportions. The result suggests that Yellow and Green taxi trips number reduces with the time elapsed from January 2015 (as expected). The estimated snow depth variable implies a positive effect on Yellow taxi trip proportions. It is possible that, under snowy conditions, the inventory of Yellow taxi fleet is unchanged while the number of TNC services reduce.

5.1.2.5 Common Unobserved Parameters

Several unobserved parameters were tested including: (1) correlation between demand component and ridehailing proportion components, (2) correlation across ridehailing proportion components and (3) random parameters in demand and proportion components. Of these tested parameters only common correlation between trip proportions of Yellow taxi and TNC services was significant. The correlation between the two components could be either positive or negative. In our analysis, we found the positive sign to offer better fit. The results indicate that unobserved factors that increase the proportion of Yellow taxi also increase the proportion of TNC services.

6 PERFORMANCE EVALUATION

The estimated models were used to predict the expected ridership at the taxi zone level and the proportion of the three ridehailing alternatives. These generated values were used to estimate the predicted number of trips by each ridehailing alternative. These estimated values are compared to the observed values to evaluate model performance. Three different measures: mean percentage error (MPE), mean absolute percentage error (MAPE) and root mean square error (RMSE) were computed based on the estimates from the joint model. A description of the measures follows:

MPE measures the prediction accuracy and is defined as:

$$\text{MPE} = mean(\frac{\hat{y}_{it} - y_{it}}{y_{it}}) \tag{9}$$

where, i represents the taxi zones and takes the value of i = 1, 2, ….., N (=259), t represents the months and takes the value of t = 1, 2, …., T (=20). The smaller the MPE, the better the model predicts observed data.

MAPE measure the error in terms of percentage and is defined as:

$$\text{MAPE} = mean\left|\frac{\hat{y}_{it} - y_{it}}{y_{it}}\right| \tag{10}$$

The smaller the MAPE, the better the model predicts observed data. These measures of fit are generated at disaggregate level: across all crash types and across all observations.

Root Mean Square Error (RMSE) is basically the standard deviation of the residuals (prediction errors). It highlights how much data is concentrated around the best fit line.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{y}_{it} - y_{it})^2}{N \times T}} \tag{11}$$

The measures were generated for the estimation sample as well as for the hold out sample. The hold-out sample was prepared following the same procedure used to extract the estimation sample. We used a sample of 20 months per taxi zone for validation. Figure 2 presents the values of these measures for joint NB-MNLFS model for estimation and validation datasets. The results highlight that the joint NB-MNLFS model gives quite intuitive result across the various measures computed. The results also highlight the relatively small range of errors for estimation and validation datasets. The model performance does not worsen for validation dataset highlighting the appropriateness of the developed model for analyzing the data.

7 PREDICTION ANALYSIS

To illustrate how the proposed model can be adopted for future demand prediction, we conduct a hypothetical policy analysis. We consider the independent variables from 2018 to remain constant for the first 6 months of 2019 and examine the number of trips by ridehailing alternative. The model prediction values, thus generated are compared with the observed trips by ride alternative for the corresponding time period. The comparison of the observed and predicted trips by ride alternative are presented in Figure 3. The predicted TNC trips increased from 20 million to 25 million from December 2018 through June 2019 while Yellow taxi trip reduced from 7.4 million to 6.4 million. Overall, the results clearly indicate a good match between observed and predicted trips by ride alternative. For Yellow taxi, the results compare favorably with slightly larger error in March 2019. From the figures, the reader would note that trips by Green taxi have the largest deviation. However, this is an artifact of the small share of Green taxi magnifying any shifts in number of trips. For TNC, the observed and predicted trips follow closely except for March 2019. To evaluate the exact mis-match in trip number by ridehailing alternative, we computed percentage error in prediction normalized to total number of trips. The estimated average percentage error for the three ridehailing alternatives (Yellow taxi, Green taxi and TNC) is 1.29, 0.59 and 1.80% respectively with the range of these errors varying from a minimum of 0.53% through a maximum of 2.11% for Yellow taxi, 0.42 through 1.13% for Green taxi and 0.02 through 6.90% for TNC. These results also indicate that the maximum error for Yellow taxi and TNC was for the month of March. We observed an anomaly in the data for the total number of ridehailing trips in March and this could be the reason for the slightly larger error. In spite of this discrepancy, the proposed model performs adequately. The comparison presented only documents the overall system level performance. Further, we also track the two components of the transformation – new ridehailing demand and shift from taxi trips to TNC services. The results presented in Table 2 provide a percentage change measure (relative to the preceding month) for three types of trips: TNC trips, Taxi trips and Total trips. From the results, it is evident that there is an increase in TNC trips while there is a reduction in taxi trips. Further, the TNC trips increase is of a larger magnitude than a simple shift of taxi trips illustrating how TNC trips are contributing to increased ridehailing demand (except for February). The model outputs are provided at a fine spatial resolution that can be employed by city planners and ridehailing operators to effectively plan and manage for changing ridehailing patterns.

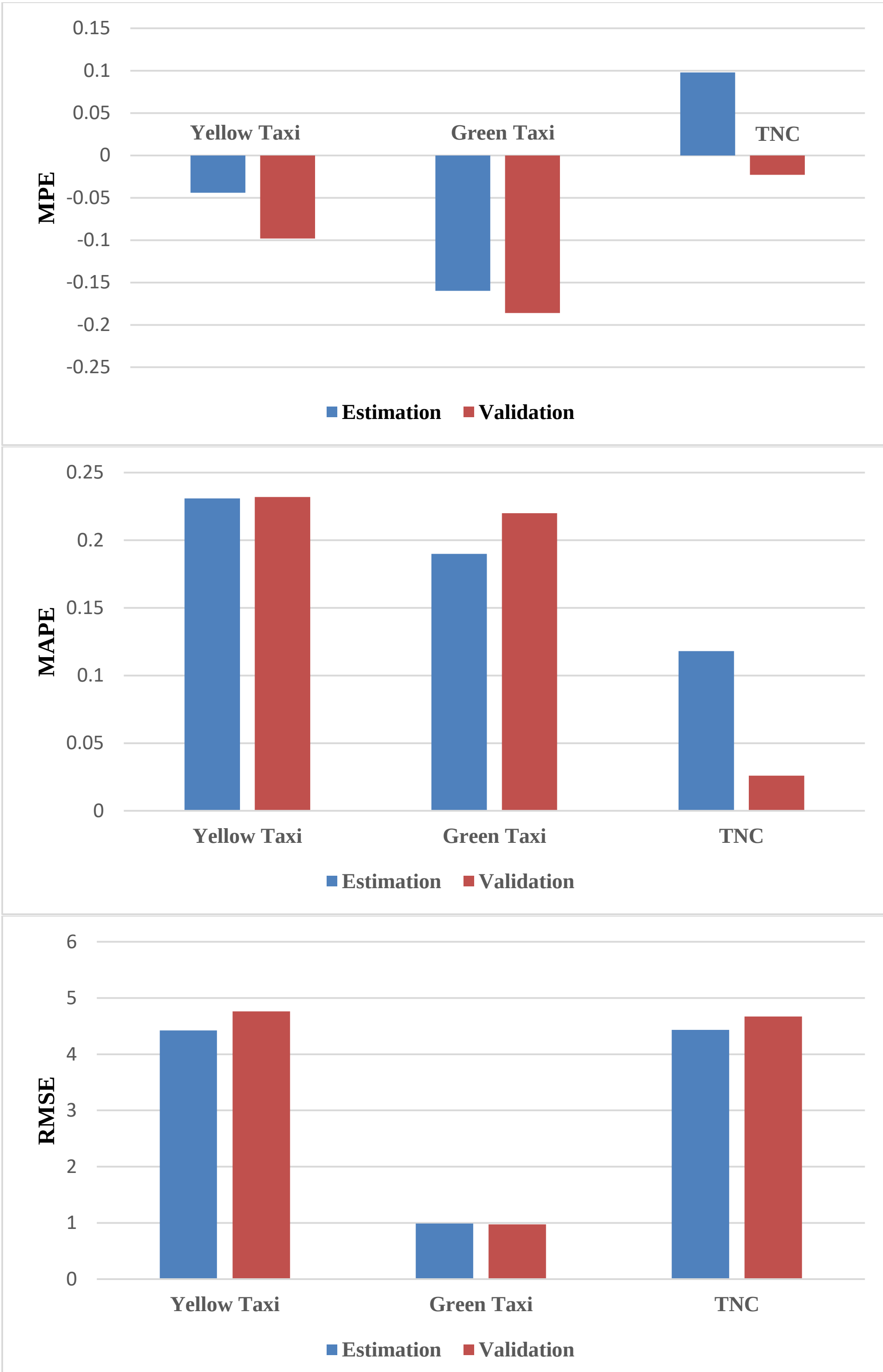


Figure 2 Sample predictive performance measure.

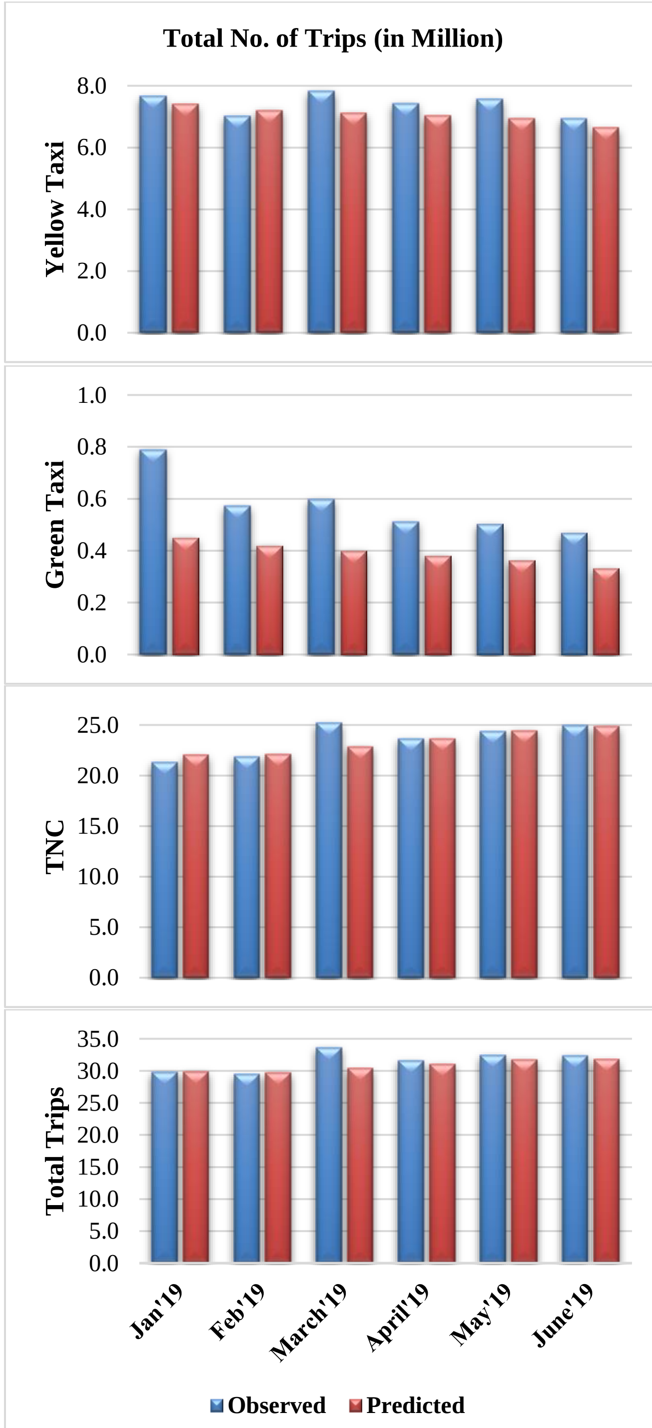


Figure 3 Predicted trip comparison

Table 2 Percentage changes in demand by months

Ridehailing Modes	Feb-19	Mar-19	Apr-19	May-19	Jun-19
TNC Trips	0.390	3.446	3.406	3.366	1.595
Taxi Trips	-2.928	-1.338	-1.396	-1.429	-4.651

Total Trips	-0.482	2.220	2.218	2.218	0.157

8 ELASTICITY ANALYSIS

The parameters of the independent variables in Table 1 do not directly provide the exact magnitude of the effects of variables on the taxi zone level trip demand and proportion of the trips. In order to highlight the effect of various attributes on total trips and proportion across ridehailing alternatives, an elasticity analysis is conducted (see Eluru & Bhat *(67)* for a discussion on the methodology for computing elasticities). For this purpose, we identify a subset of variables including population density, job density, median income, proportion of zero car household, transit score, non-motorized vehicle score, residential area, park area, airport indicator, bike lane length, number of bike stations and street length. Table 3 presents elasticity analysis results for both model components. For the <u>first component</u>, we investigate the percentage change in the expected taxi zone level trip counts in response to a 10% change in the explanatory variable. From the results, it is found that job density in the taxi zone has the highest impact on total ridehailing trips made in that zone. It is particularly interesting to note the significantly larger influence of job density relative to population density on ridehailing demand potentially highlighting how ride hailing demand is closely associated with employment opportunities. In their study, Correa et al. *(34)* found that job density affected ridehailing demand while population density was not significant. The median household income and share of zero car households also have a considerable influence on ridership (Correa et al. *(34)* observed the negative impact of higher vehicle ownership and positive influence of income on ridehailing demand). Further, transit score and non-motorized vehicle score highly affect the total ridehailing trips (see Correa et al. *(34)* for similar results). Of course, it is important to recognize that altering these variables by 10% within a taxi zone is not a simple process and could potentially take several years. For the <u>second component</u>, we investigate the percentage change in the predicted proportion of trips for each mode considered in response to a 10% in the explanatory variable. The reader should note that these elasticity effects are directly associated with the proportion of trips (not actual trips by these alternatives). Hence, while the proportion of an alternative might increase or reduce, the actual change in the alternative trips will depend on the net effect of the two components. From the analysis, it is found that Job density, bike lane density, and street length variables positively influence TNC proportion. Proportion of zero car households in the taxi zone has a significant positive impact on Green Taxi proportion. Median Income and Park area variables affect Yellow Taxi mode proportion positively.

Table 3 Elasticity analysis results

Ridehailing Alternatives	**Total trips**	**Yellow Taxi**	**Green Taxi**	**TNC**
Population Density	0.551[1]	2.519	-4.637	-0.075
Job Density	17.835	-4.351	-5.879	1.677
Median Income	6.442	5.222	-1.732	-1.034
Zero Car HH	6.932	-4.121	12.493	-0.424
Transit Score	15.439	0.000	0.000	0.000
Non-motorized Score	-10.127	0.000	0.000	0.000
Residential Area	3.591	0.000	0.000	0.000
Park Area	2.371	2.965	-1.315	-0.550
Bike Lane Density	-1.647	-1.056	0.480	0.195

Bikeshare Stations	-0.401	0.102	-0.409	0.022
Street Length	0.371	-4.532	1.949	0.848

[1] = percentage change in demand due to change in the independent variables

9 CONCLUSIONS

We develop an innovative joint econometric model system to study two components of the ridehailing demand transformation; (a) the increase in ridehailing demand and (b) the shift from traditional taxi services to TNC services. The first component is analyzed adopting a negative binomial count model while the second component is analyzed using a multinomial fractional split model. The two model components are stitched together in a joint framework that allows for the influence of repeated observations as well as for the presence of common unobserved factors affecting the two components.

The data for our analysis is drawn from New York City Taxi & Limousine Commission (NYTLC) for four years from January 2015 through December 2018. The model estimation considered a comprehensive set of independent variables. Among those tested, F for the total trip component, land use and built environment variables (such as population density, job density, median income, transit score, non-motorized vehicle score, restaurants and cafes, points of interest and presence of airport), transportation infrastructure variables (such as bicycle and public transit infrastructure), temporal and weather attributes (such as season and snow depth) offer significant impacts. For the proportion model component, we observe that only a subset of these variable offer significant parameters. Several variables such as population density and job density offer significant impacts. However, the impact is specific to a subset of ridehailing alternatives. In terms of the influence of common unobserved factors in the joint model, only one variable capturing the interconnectedness between Yellow Taxi and TNC proportions was found to be significant.

The model estimation effort was augmented via a comprehensive estimation sample prediction, hold-out sample prediction and elasticity analysis exercises. First, we compared the performance of the model on the estimation sample and validation sample (for the same time frame). Second, we examined the performance of the model on data from the model future for examining prediction performance in the future. Finally, to illustrate the impact of various independent variables an elasticity exercise is conducted. The comparison exercise is conducted using a host of traditional metrics used for model evaluation. It is encouraging to observe that the proposed model provides excellent match with estimation and validation datasets for the estimation time frame. Further, the performance of the model is quite good even for prediction into the future. The results indicate that the predicted model tracks the evolving trends by ridehailing alternatives very closely. Finally an elasticity exercise was conducted to identify the important variables affecting the two components of the joint model. From our analysis, job density, employment density, households with zero vehicle ownership, bike infrastructure and transit score significantly affect ridehailing demand and the proportion across various ridehailing alternatives.

The proposed framework can be employed by urban transportation agencies to understand the influence of emerging TNC demand at a fine spatial and temporal resolution while accommodating for the influence of a host of variables. The high resolution framework is sensitive enough to accommodate for various scenario impacts such as future developments in the urban region, incentive/penalty structures on TNC and evolving land use and built environment patterns. Further, the model outputs can be useful for long range planning exercises. The ridehailing demand estimates from the joint model can be combined with passenger demand from a four step or activity based model to conduct an assessment of urban travel patterns and spatio-temporal congestion profiles.

To be sure, the study is not without limitations. It might be interesting to enhance the study methodology by accounting for unobserved temporal effects (heteroscedasticity) across the multiple years of data. In future efforts, it might also be useful to include monthly economic indicators (such as employment and wages) in the model to control for macroeconomic conditions. Finally, access to more details on trip pricing and travel time can allow the development of more advanced and realistic demand frameworks.

AUTHOR CONTRIBUTION STATEMENT

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

REFERENCES

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

[2] Rajagopalan, B. S., and K. K. Srinivasan. Integrating household-level mode choice and modal expenditure decisions in a developing country: multiple discrete–continuous extreme value model. Transportation Research Record, Vol. 2076, No. 1, 2008, pp. 41-51.

[3] Gerte, R., K. C. Konduri, N. Ravishanker, A. Mondal, and N. Eluru. Understanding the Relationships between Demand for Shared Ride Modes: Case Study using Open Data from New York City. Transportation Research Record, Vol. 2673, No. 12, 2019, pp. 30-39.

[4] Schaller, B. Entry controls in taxi regulation: Implications of US and Canadian experience for taxi regulation and deregulation. Transport policy, Vol. 14, No. 6, 2007, pp. 490-506.

[5] Çetin, T., and K. Y. Eryigit. Estimating the effects of entry regulation in the Istanbul taxicab market. Transportation Research Part A: Policy and Practice, Vol. 45, No. 6, 2011, pp. 476-484.

[6] Chang, S. J., and C.-H. Chu. Taxi vacancy rate, fare, and subsidy with maximum social willingness-to-pay under log-linear demand function. Transportation Research Record, Vol. 2111, No. 1, 2009, pp. 90-99.

[7] Milioti, C. P., M. G. Karlaftis, and D. Spyropoulou. Impact of the Financial Recession on the Taxi Market: Perceptions of Travelers and Taxi Drivers. Transportation Research Record, Vol. 2536, No. 1, 2015, pp. 67-75.

[8] Zhang, W., and S. V. Ukkusuri. Optimal fleet size and fare setting in emerging taxi markets with stochastic demand. Computer-Aided Civil and Infrastructure Engineering, Vol. 31, No. 9, 2016, pp. 647-660.

[9] Burghout, W., P. J. Rigole, and I. Andreasson. Impacts of shared autonomous taxis in a metropolitan area. In Proceedings of the 94th annual meeting of the Transportation Research Board, Washington DC, 2015.

[10] Chrysostomou, K., A. Georgakis, M. Morfoulaki, K. Kotoula, and G. Myrovali. Using Big Taxi GPS Data to Investigate Feasibility of Electric Taxis in Thessaloniki, Greece. 2016.

[11] Jung, J., J. Y. Chow, R. Jayakrishnan, and J. Y. Park. Stochastic dynamic itinerary interception refueling location problem with queue delay for electric taxi charging stations. Transportation Research Part C: Emerging Technologies, Vol. 40, 2014, pp. 123-142.

[12] Wong, K., S. Wong, H. Yang, and J. Wu. Modeling urban taxi services with multiple user classes and vehicle modes. Transportation Research Part B: Methodological, Vol. 42, No. 10, 2008, pp. 985-1007.
[13] Wong, R., W. Szeto, and S. Wong. Bi-level decisions of vacant taxi drivers traveling towards taxi stands in customer-search: Modeling methodology and policy implications. Transport policy, Vol. 33, 2014, pp. 73-81.
[14] Wong, R., W. Szeto, and S. Wong. A two-stage approach to modeling vacant taxi movements. Transportation Research Procedia, Vol. 7, 2015, pp. 254-275.
[15] Yang, H., and S. Wong. A network model of urban taxi services. Transportation Research Part B: Methodological, Vol. 32, No. 4, 1998, pp. 235-246.
[16] Zhang, W., S. V. Ukkusuri, and J. J. Lu. Impacts of urban built environment on empty taxi trips using limited geolocation data. Transportation, Vol. 44, No. 6, 2017, pp. 1445-1473.
[17] Zhan, X., and S. V. Ukkusuri. Probabilistic urban link travel time estimation model using large-scale taxi trip data. 2015.
[18] Dalziel, J. R., and R. S. Job. Motor vehicle accidents, fatigue and optimism bias in taxi drivers. Accident Analysis & Prevention, Vol. 29, No. 4, 1997, pp. 489-494.
[19] Lam, L. T. Environmental factors associated with crash-related mortality and injury among taxi drivers in New South Wales, Australia. Accident Analysis & Prevention, Vol. 36, No. 5, 2004, pp. 905-908.
[20] Peltzer, K., and W. Renner. Superstition, risk-taking and risk perception of accidents among South African taxi drivers. Accident Analysis & Prevention, Vol. 35, No. 4, 2003, pp. 619-623.
[21] Tay, R., and J. Choi. Factors associated with crashes involving taxi owners and non-owners: A case of moral hazard and adverse selection? Accident Analysis & Prevention, Vol. 87, 2016, pp. 78-82.
[22] Tseng, C.-M. Operating styles, working time and daily driving distance in relation to a taxi driver's speeding offenses in Taiwan. Accident Analysis & Prevention, Vol. 52, 2013, pp. 1-8.
[23] Chan, N. D., and S. A. Shaheen. Ridesharing in North America: Past, present, and future. Transport Reviews, Vol. 32, No. 1, 2012, pp. 93-112.
[24] Sun, C., and P. Edara. Is getting an Uber-Lyft from a sidecar different from hailing a taxi? Current dynamic ridesharing controversy. Transportation Research Record, Vol. 2536, No. 1, 2015, pp. 60-66.
[25] Furuhata, M., M. Dessouky, F. Ordóñez, M.-E. Brunet, X. Wang, and S. Koenig. Ridesharing: The state-of-the-art and future directions. Transportation Research Part B: Methodological, Vol. 57, 2013, pp. 28-46.
[26] Agatz, N., A. Erera, M. Savelsbergh, and X. Wang. Optimization for dynamic ride-sharing: A review. European Journal of Operational Research, Vol. 223, No. 2, 2012, pp. 295-303.
[27] Amey, A., J. Attanucci, and R. Mishalani. Real-time ridesharing: opportunities and challenges in using mobile phone technology to improve rideshare services. Transportation Research Record, Vol. 2217, No. 1, 2011, pp. 103-110.
[28] Cramer, J., and A. B. Krueger. Disruptive change in the taxi business: The case of Uber. American Economic Review, Vol. 106, No. 5, 2016, pp. 177-182.
[29] Chen, L., A. Mislove, and C. Wilson. Peeking beneath the hood of uber.In Proceedings of the 2015 Internet Measurement Conference, ACM, 2015. pp. 495-508.
[30] Chen, M. K., and M. Sheldon. Dynamic pricing in a labor market: Surge pricing and the supply of Uber driver-partners. University of California (Los Angeles) Working Paper. URL http://citeseerx. ist. psu. edu/viewdoc/download, 2015.

[31] Guo, S., Y. Liu, K. Xu, and D. M. Chiu. Understanding ride-on-demand service: Demand and dynamic pricing.In 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), IEEE, 2017, pp. 509-514.
[32] Smart, R., B. Rowe, and A. Hawken. Faster and cheaper: How ride-sourcing fills a gap in low-income Los Angeles neighborhoods. 2015.
[33] 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, Vol. 13, 2018, pp. 88-104.
[34] Correa, D., K. Xie, and K. Ozbay. Exploring the Taxi and Uber Demand in New York City: An Empirical Analysis and Spatial Modeling. Presented at 96th Annual Meeting of the Transportation Research Board, Washington, D.C., 2017.
[35] Davidson, A., J. Peters, and C. Brakewood. Interactive travel modes: Uber, transit, and mobility in New York City. 2017.
[36] Feigon, S., and C. Murphy. TCRP Report 188: Shared Mobility and the Transformation of Public Transit. Transportation Research Board of the National Academies, Washington, D.C., 2016. http://www.trb.org/Main/Blurbs/174653.aspx.
[37] 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, Vol. 45, 2016, pp. 168-178.
[38] Komanduri, A., Z. Wafa, K. Proussaloglou, and S. Jacobs. Assessing the impact of app-based ride share systems in an urban context: Findings from Austin. Transportation Research Record, Vol. 2672, No. 7, 2018, pp. 34-46.
[39] Yu, H., and Z.R. Peng. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. Journal of Transport Geography, Vol. 75, 2019, pp. 147-163.
[40] 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, Vol. 2672, No. 46, 2018, pp. 31-40.
[41] Poulsen, L. K., D. Dekkers, N. Wagenaar, W. Snijders, B. Lewinsky, R. R. Mukkamala, and R. Vatrapu. Green Cabs vs. Uber in New York City.In 2016 IEEE International Congress on Big Data (BigData Congress), IEEE, 2016. pp. 222-229.
[42] Babar, Y., and G. Burtch. Examining the impact of ridehailing services on public transit use. Available at SSRN 3042805, 2017.
[43] Wang, H., and H. Yang. Ridesourcing systems: A framework and review. Transportation Research Part B: Methodological, Vol. 129, 2019, pp. 122-155.
[44] Nie, Y.M. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. Transportation Research Part C: Emerging Technologies, Vol. 79, 2017, pp. 242-256.
[45] Wang, X., F. He, H. Yang, and H.O. Gao. Pricing strategies for a taxi-hailing platform. Transportation Research Part E: Logistics and Transportation Review, Vol. 93, 2016, pp. 212-231.
[46] Zha, L., Y. Yin, and H. Yang. Economic analysis of ride-sourcing markets. Transportation Research Part C: Emerging Technologies, Vol. 71, 2016, pp. 249-266.
[47] Li, W., Z. Pu, Y. Li, and X.J. Ban. Characterization of ridesplitting based on observed data: A case study of Chengdu, China. Transportation Research Part C: Emerging Technologies, Vol. 100, 2019, pp. 330-353.

[48] Zhu, Z., X. Qin, J. Ke, Z. Zheng, and H. Yang. Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. Transportation Research Part A: Policy and Practice, Vol. 132, 2020, pp. 713-727.
[49] Bhowmik, T., S. Yasmin, and N. Eluru. A joint econometric approach for modeling crash counts by collision type. Analytic methods in accident research, Vol. 19, 2018, pp. 16-32.
[50] Cameron, A. C., T. Li, P. K. Trivedi, and D. M. Zimmer. Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. The Econometrics Journal, Vol. 7, No. 2, 2004, pp. 566-584.
[51] Rahman M., S. Yasmin, and N. Eluru. A Joint Panel Binary Logit and Fractional Split Model for Converting Route-Level Transit Ridership Data to Stop-Level Boarding and Alighting Data. Transportation Research Part A, Vol. 139, 2020, pp. 1-16.
[52] Tirtha, S.D., S. Yasmin, and N. Eluru. Modeling of Incident Type and Incident Duration Using Data from Multiple Years. Analytic Methods in Accident Research, Vol. 28, 2020, p. 100132.
[53] Faghih-Imani, A., and N. Eluru. Determining the Role of Bicycle Sharing System Infrastructure Installation Decision on Usage: Case Study of Montreal BIXI System. Transportation Research Part , Vol. 94, 2016, pp. 685-698.
[54] Bhat, C. R. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transportation Research Part B: Methodological, Vol. 35, No. 7, 2001, pp. 677-693.
[55] Eluru, N., C. R. Bhat, and D. A. Hensher. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention, Vol. 40, No. 3, 2008, pp. 1033-1054.
[56] Yasmin, S., and N. Eluru. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. Accident Analysis & Prevention, Vol. 59, 2013, pp. 506-521.
[57] Faghih-Imani A., S. Anowar, E. Miller, and N. Eluru, Hail a Cab or Ride a Bike? A Travel Time Comparison of Taxi and Bicycle-Sharing Systems in New York City. Transportation Research Part A: Policy and Practice, Vol. 101, July 2017, pp.11-21
[58] Li, T., P. Jing, L. Li, D. Sun, and W. Yan. Revealing the Varying Impact of Urban Built Environment on Online Car-Hailing Travel in Spatio-Temporal Dimension: An Exploratory Analysis in Chengdu, China. Sustainability, Vol. 11, No. 5, 2019, p. 1336.
[59] Wenzel, T., C. Rames, E. Kontou, and A. Henao. Travel and energy implications of ridesourcing service in Austin, Texas. Transportation Research Part D: Transport and Environment, Vol. 70, 2019, pp. 18-34.
[60] Dey, B. K., S. Anowar, and N. Eluru. A Framework for Estimating Bikeshare Origin Destination Flows Using a Multiple Discrete Continuous System. Transportation Research Part A, Vol. 144, 2021, pp.119-133.
[61] Dey, B.K., S.D. Tirtha, and N. Eluru. Transport Networking Companies Demand and Flow Estimation in New York City. Transportation Research Record, 2021, p. 03611981211000752.
[62] Liu, Y., A. Singleton, D. Arribas-bel, and M. Chen. Identifying and understanding road-constrained areas of interest (AOIs) through spatiotemporal taxi GPS data: A case study in New York City. Computers, Environment and Urban Systems, Vol. 86, 2021, p. 101592.
[63] Sadowsky, N., and E. Nelson. The Impact of Ride-Hailing Services on Public Transportation Use: A Discontinuity Regression Analysis. Economics Department, Bowdoin College, 2017. https://digitalcommons.bowdoin.edu/econpapers/13/
[64] Tirachini, A., and M. del Río. Ride-hailing in Santiago de Chile: Users' characterisation and effects on travel behaviour. Transport Policy, Vol. 82, 2019, pp. 46-57.

[65] 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, Vol. 152, 2018, pp. 1-16.
[66] Schneider, W. T. (2018, March). Analyzing 1.1 Billion NYC Taxi and Uber Trips, with a Vengeance. Retrieved July 2019, from https://toddwschneider.com/posts/analyzing-1-1-billion-nyc-taxi-and-uber-trips-with-a-vengeance/#airport-traffic.
[67] Eluru, N., and C.R. Bhat. A joint econometric analysis of seat belt use and crash-related injury severity. Accident Analysis and Prevention, Vol. 39, No. 5, 2007, pp. 1037–1049.