The Link Between For-Hire Service Pickups and Built Environment Characteristics: Evidence from New York City

Jina Mahmoudi, P.E.[1] and Lei Zhang, Ph.D.[2]
[1]Ph.D. Candidate, Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn Martin Hall, College Park, MD 20742; email: zhina@umd.edu
[2]National Transportation Center, Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn Martin Hall, College Park, MD 20742; email: lei@umd.edu

ABSTRACT
This study employs Structural Equation Modeling techniques to investigate the relationship between demand for taxi, Uber and Lyft services, socioeconomic and built environment characteristics as well as access to transit and bike-sharing modes, for taxi zones within New York City. The results show that income and car ownership levels influence demand for these for-hire modes. Additionally, higher activity density and higher extent of mixed land-use are associated with increased demand for for-hire modes, while pedestrian-friendly street networks are linked with lower demand levels. Also, temporal destination accessibility as well as accessibility to transit and bike-sharing significantly influence demand for taxi, Uber and Lyft. The findings provide a better understanding of the link between for-hire modes and built environment as well as accessibility to other modes, which can be used to improve demand forecasting of taxi, Uber, and Lyft services in large cities.

INTRODUCTION
A sound understanding of the factors that influence demand for each mode of travel is essential for planning an effective multimodal transportation system in urban settings. A crucial component of such a complex multimodal system is the for-hire mode, which has a potential to either compete with or complement other travel modes. Demand levels for for-hire service can be very high in big cities. A study for the case of New York City (NYC) estimated that taxicab ridership was nearly 241 million in NYC, which translated into 11% of all fare-paying riders and 25% of all riders traveling within Manhattan *(1)*. However, the taxi industry in NYC is facing serious competition from emerging on-demand ride-hailing services such as Uber and Lyft *(2, 3)*. These modern ride-hailing services are utilizing cutting edge smart-phone-based Intelligent Transportation System (ITS) technologies to reshape the urban transportation landscape. A recent study reported that from 2014 to 2015, the number of Uber trips had a sizable increase of 10 million, while taxi trips decreased by 0.8 million *(3)*. These statistics provide an example for the dramatic increase in demand for modern ride-hailing services and their potential impact on taxi demand. Nevertheless, traditional taxicab remains a key category of the for-hire mode in urban areas despite the tremendous growth of modern ride-hailing services (i.e., Uber and Lyft).

Recognizing the notable impact of taxis on urban travel, several studies have probed to determine the factors that influence demand for taxis in major urban areas *(2-5)*. However, few studies have rigorously examined land use and built environment factors that influence taxi demand, and the link between taxi demand and accessibility to other modes. Specifically, little is known about the role of access to bike-sharing on demand for taxis versus demand for Uber and Lyft. In addition, while Uber and Lyft are rapidly growing in the number of trips, several questions remain related to the perceived competition of these modern ride-hailing services with traditional taxicabs. To address these gaps in knowledge, the present study develops models for estimating the annual number of picks-ups for traditional taxis versus those of Uber and Lyft. Structural Equation Modeling techniques are employed to develop models specified by variables representing various attributes of taxi zones within New York City (NYC) for the year of 2015.

LITERATURE REVIEW

Existing literature on demand for the for-hire mode mainly focuses on demand for taxis. Schaller *(4)* modeled the number of taxicabs using data from 118 U.S. cities and found that the number of taxicabs in U.S. cities was strongly associated with the number of subway commuters, the number of airport taxi trips as well as the number of households that owned no private vehicles. Gonzales et al. *(5)* modeled taxi pick-ups and drop-offs for census tracts in NYC by hour of the day. The study utilized Global Positioning System (GPS) data and found that median income, the number of employment opportunities, and transit accessibility were significant factors in estimating taxi demand. Traditionally, taxicab companies dominated the for-hire market. However, emerging ride-hailing services such as Uber and Lyft are growing in popularity—presenting an unprecedent level of competition with taxicabs *(2)*. Considering the potential impacts of Uber and Lyft, studies were conducted to analyze the impact of these emerging services on traditional for-hire services *(2)*, and on local economies *(6)*. These studies found that Uber has shown a tremendous growth in popularity in major U.S. metropolitan areas. Later, Correa et al. *(3)* examined the impact of Uber on taxi market by analyzing demand for Uber and taxi in NYC. Results indicated that higher demands for taxi and Uber within NYC neighborhood are associated with lower transit access time, higher length of roadways, lower levels of car ownership, higher levels of income and more employment opportunities.

More recent papers tend to involve Uber and Lyft data for demand analysis of for-hire modes; however, little is known about the impact of socioeconomic, built environment characteristics as well as accessibility to alternative travel modes on demand levels for these ride-hailing services. The present study aims to fill that gap by examining the link between the above-mentioned factors and demand for taxi, Uber and Lyft modes in NYC.

DATA

For-Hire Mode Trip Data were obtained from the New York Taxi & Limousine Commission (NYC TLC). The data contain individual pickup times and locations specified by coordinates (for taxis) and by taxi zones (for Uber and Lyft). Comparing the general demand patterns in 2015, of over the 200 million for-hire mode trips made in NYC, yellow taxi covered over 70 percent of the total trips followed by Uber, green taxi, and Lyft (Figure 1).

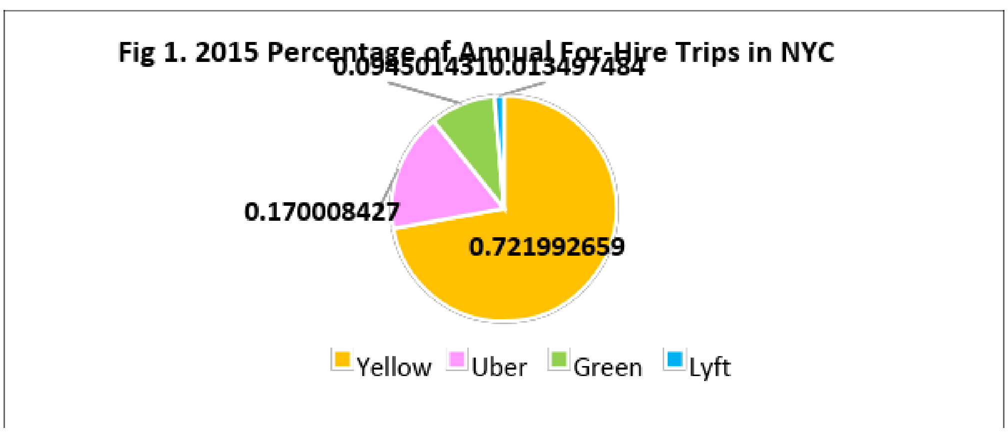

Fig 1. 2015 Percentage of Annual For-Hire Trips in NYC

Smart Location Database (SLD) is a product of the U.S. Environmental Protection Agency (EPA). This dataset provides information on land use and built environment attributes such as population and employment density, mixed land use, neighborhood design, destination and transit accessibility as well as socioeconomic characteristics at the census block group level.

National Transit-Oriented Database (TOD) provides location information for transit stations in U.S. cities. The TOD database is used to obtain the location of transit stations within NYC.

Citi Bike Data are utilized to obtain bike sharing station locations within NYC. Citi Bike is the official bike sharing system in NYC, and the largest bike sharing program in the U.S.
Walk Score and Bike Score Data provide information on walkability/bikeability of locations.

METHODOLOGY

Modeling technique. This study employs Structural Equation Modeling (SEM) techniques to develop models for estimating the annual number of picks-ups for traditional taxis versus those of Uber and Lyft. SEM techniques are the most advanced tools to address endogeneity issues in models with interdependent variables. SEM not only captures correlations between the dependent and independent variables, but also provides explanation of why these variables are related *(7)*. Thus, SEM has an advantage over the Ordinary Least Squares (OLS) techniques.
Model variables. The geographic unit of analysis for the present study is taxi zone. The dependent (i.e., endogenous) variables are defined as the annual number of taxi, Uber and Lyft pickups for each taxi zone within NYC during the year 2015. The model also includes various independent (i.e., exogenous) variables representing socioeconomic and built environment characteristics as well as accessibility to transit and bike sharing within NYC taxi zones. Table 1 provides descriptions for model variables.

To obtain the values for socioeconomic and built environment attributes at the taxi zone level, census block group attributes from SLD were averaged for each NYC taxi zone. Transit accessibility of each taxi zone is represented by average distance from the centroid to the nearest transit station, aggregate frequency of transit service as well as the total number of transit-oriented developments within each taxi zone in NYC. Bike share accessibility of each taxi zone is represented by the total number of bike sharing stations within the taxi zone. The Walk Score and Bike Score for the centroid of each taxi zone are included to account for walkability and bikeability of taxi zones. Also, the well-established entropy formula used in previous research *(8)* is employed to compute entropy values (i.e., extent of mixed land use) for taxi zones.

Table 1. Variable Descriptions, Descriptive Statistics and Data Sources

Variable	Variable Description	Computation/Units	Mean	S.D.	Source
Dependent (Endogenous) Variables					
AnnualTaxiPickups	Annual Taxi pickups	Annual number of Taxi pickups (10^3) [a]	505.1	965.3	NYC TLC
AnnualUberPickups	Annual Uber pickups	Annual number of Uber pickups (10^3) [a]	130.6	208.3	NYC TLC
AnnualLyftPickups	Annual Lyft pickups	Annual number of Lyft pickups (10^3) [a]	10.4	14.1	NYC TLC
Independent (Exogenous) Variables					
PercentNoCarHHs	Percentage of HHs with No Cars	% of HHs with zero cars [b]	47.3	26.8	SLD
PercentLowWage	Percentage of Low-Wage Workers	% of workers earning ≤ $1250/month [b]	21.9	5.5	SLD
ActivityDen	Activity Density	Activity density [(jobs + housing units)/acre] [a, b]	105.9	184.8	SLD
MixedLandUse	Mixed Land Use	Employment and household entropy [b]	0.40	0.2	SLD
RoadDen	Road density	Road network miles/ mi^2 [b, c]	28.8	8.9	SLD
IntDen	Intersection Density	Auto-oriented intersections/mi^2 [a, b]	5.0	9.4	SLD
PedNetDen	Pedestrian-friendly Network Density	Facility miles of pedestrian-oriented links/mi^2 [a, b]	21.7	6.4	SLD
DisTransit	Distance to Transit	Distance from centroid to nearest transit stop (meters) [a, c]	391.9	222.5	SLD
TransitFreq	Transit Frequency	Aggregate frequency of transit service/mi^2 (10^3) [a, b]	2.8	3.7	SLD
TOD	Transit Accessibility	No. of Transit-oriented Developments	2.0	2.2	TOD

DesAcc	Destination Accessibility (Auto. + Transit)	Number of jobs within a 45-minute auto or transit commute (10^3) [a, b]	810.6	424.9	SLD
BikeShare	Bike-shareAccessibility	Number of CitiBike stations	2.3	4.4	CitiBike
WalkScore	Walk Score	Walk Score of the centroid of taxi zone	84.9	20.2	WalkScore®
BikeScore	Bike Score	Bike Score of the centroid of taxi zone	65.7	14.7	WalkScore®
Number of Observations (Taxi Zones)		***263***			

[a] variable log-transformed before inclusion in model (to normalize its distribution); [b] average of values for all census block groups within each taxi zone; [c] variable not included in the final models due to high correlations (>0.5) with other independent variables; S.D. = standard deviation; HH = household; mi^2 = square mile; No. = Number.

Pearson correlation matrices were developed to examine the correlation coefficients between original independent variables. A correlation coefficient greater than 0.5 or less than -0.5 was considered a high correlation based on past studies *(9)*. Highly-correlated variables were eliminated to reduce the risk of multicollinearity in the model.

Model specification. The SEM model structure used in this study represents the following three regression equations for taxi demand, Uber demand and Lyft demand for each taxi zone:

1) AnnualUberPickups = α_1AnnualTaxiPickups + α_2SES + α_3BE + ε_1
2) AnnualTaxiPickups = β_1AnnualUberPickups + β_2AnnualLyftPickups + β_3SES + β_4BE + ε_2
3) AnnualLyftPickups = σ_1AnnualTaxiPickups + σ_2SES + σ_3BE + ε_3

Where, SES = vector of socioeconomic attributes; BE = vector of built environment attributes; α_i, β_i and σ_i = regression coefficients; and ε_i = regression error terms. The SEM estimates these coefficients and error terms for all three equations simultaneously.

RESULTS AND DISCUSSION

Table 2 summarizes the SEM results which provide direct, indirect and total effects of each independent variable on each dependent variable. Variables whose total effects were not significant in the model are not included in Table 2.

Estimation results. The negative direction of the total coefficients for the *PercentLowWage* variable indicates that the annual number of pickups by taxi and Lyft services decrease as the average percentage of workers with lower incomes increases in a taxi zone. This result is consistent with previous findings *(3)* and may be explained by the propensity of lower income earners to opt for low-cost modes such as transit, walking and bicycling. The total effects of the variable representing car ownership (*PercentNoCarHHs*) is significant in the taxi and Lyft equations with an expected positive effect meaning not owning a personal vehicle is associated with higher demand for taxi and Lyft. This result is also consistent with previous findings *(3)*. Further, land use and built environment factors significantly impact the annual number of pickups for all three modes. Higher activity density (*ActivityDen*) is associated with increased annual taxi, Uber and Lyft pickups. A higher extent of mixed land use (entropy) is related to higher levels of annual pickup activity by taxi and Uber.

In addition, the results show that automobile-oriented designs are associated with increased numbers of annual pickups by taxi, Uber and Lyft services as exhibited by the positive total effects of the automobile-oriented intersection density variable (*IntDen*). By contrast, pedestrian-friendly street networks (*PedNetDen*) is associated with lower numbers of annual pickups by taxi and Uber, whereas location walkability (*WalkScore*) is associated with lower numbers of annual pickups by Lyft. These results suggest that where urban design is more conducive to walking, individuals may walk to destinations instead of using ride-hailing services. Increased accessibility to transit shows a positive effect as increased frequency of local transit service (*TransitFreq*) is associated with more taxi, Uber and Lyft pickups. Additionally, as the total number of transit-oriented developments (*TOD*s) within the taxi zone increases, demand for all three modes increases. These findings suggest the use of for-hire modes for first/last mile connections to transit stations. Increased accessibility to bike sharing is also correlated with higher ride-hailing demand levels. The variable for total number of bike sharing stations (*BikeShare*) within a taxi zone exhibits a positive total coefficient in all three equations meaning that as the number of bike sharing stations increases in a taxi zone, so do the number of annual taxi, Uber and Lyft pickups. This finding suggests that bike-share customers may use bike sharing for short-range local trips, but rely on for-hire modes for longer distance trips. The variable representing temporal destination accessibility by both automobile or transit (*DesAcc*) shows positive relationships with annual pickups by taxi, Uber and Lyft services.

Table 2. Structural Equation Model Results for For-Hire Modes in NYC

Independent (Exogenous) Variables	Direct Coefficients	*p-value*	Indirect Coefficients	*p-value*	**Total Coefficients**	***p-value***
Annual Uber Pickup[a] Equation						
ActivityDen[a]	0.340	0.006	0.013	0.654	**0.353*****	**0.009**
MixedLandUse	0.008	0.084	0.0005	0.671	**0.008***	**0.085**
IntDen[a]	0.091	0.056	0.009	0.646	**0.100***	**0.082**
PedNetDen[a]	-0.403	0.066	0.028	0.672	**-0.375***	**0.078**
TransitFreq[a]	0.076	0.042	0.013	0.612	**0.089*****	**0.004**
TOD	0.085	0.006	0.006	0.612	**0.091*****	**0.001**
DesAcc[a]	0.889	0.169	0.197	0.603	**1.086****	**0.023**
BikeShare	0.083	0.000	0.005	0.619	**0.088*****	**0.000**
Annual Taxi Pickup[a] Equation						
AnnualUberPickups[a]	0.762	0.000	-0.083	0.655	**0.679*****	**0.000**
PercentNoCarHHs	0.016	0.000	0.0006	0.889	**0.016*****	**0.004**
PercentLowWage	-0.031	0.061	-0.008	0.639	**-0.039***	**0.076**
ActivityDen[a]	-0.026	0.768	0.218	0.032	**0.192***	**0.093**
MixedLandUse	0.0005	0.924	0.007	0.082	**0.007***	**0.093**
IntDen[a]	0.070	0.075	0.067	0.164	**0.137****	**0.047**
PedNetDen[a]	-0.682	0.001	0.260	0.243	**-0.422***	**0.066**
TransitFreq[a]	0.141	0.000	0.054	0.022	**0.195*****	**0.000**
TOD	0.036	0.186	0.056	0.005	**0.092*****	**0.009**
DesAcc[a]	2.191	0.000	0.710	0.037	**2.901*****	**0.000**
BikeShare	0.024	0.106	0.056	0.000	**0.080*****	**0.000**
Annual Lyft Pickup[a] Equation						
AnnualTaxiPickups[a]	1.124	0.000	-0.124	0.625	**1.000*****	**0.000**
AnnualUberPickups[a]	--	--	0.763	0.000	**0.763*****	**0.000**
PercentNoCarHHs	-0.008	0.107	0.018	0.007	**0.010***	**0.080**
PercentLowWage	-0.037	0.070	0.030	0.084	**-0.007***	**0.076**
ActivityDen[a]	0.115	0.128	0.216	0.157	**0.331*****	**0.006**
IntDen[a]	-0.095	0.051	0.154	0.036	**0.059***	**0.066**
TransitFreq[a]	-0.128	0.000	0.219	0.000	**0.091*****	**0.008**
TOD	-0.019	0.568	0.104	0.009	**0.085****	**0.013**
DesAcc[a]	-2.508	0.000	3.262	0.000	**0.754****	**0.049**
BikeShare	-0.017	0.332	0.090	0.000	**0.073*****	**0.000**
WalkScore	-0.003	0.050	-0.013	0.370	**-0.016***	**0.084**

[a] variable log-transformed before inclusion in model to normalize its distribution; *, **, *** = coefficient significant at the 10%, 5% and 1% significance level, respectively; -- = no direct path assumed in the model.

Reciprocal effects of dependent (endogenous) variables. The results also indicate that demand levels for Uber, taxi and Lyft are interrelated and influence each other. The total coefficient of the Uber demand variable (*AnnualUberPickups*) in the taxi pick up equation shows that higher levels of taxi demand are associated with higher levels of Uber demand. Also, total coefficient estimates in the Lyft equation suggest that higher demand for Lyft in a taxi zone is correlated with higher demand for taxi and Uber even after controlling for built environment factors. These findings seem counter-intuitive as one would expect that higher demand for Uber and Lyft in a taxi zone would lead to lower demand for taxi. However, that expectation assumes substitution of the Uber and Lyft mode for taxi mode which needs further analysis, and perhaps the availability of for-hire mode demand data across several years to determine if any substitution trends exist. The results of the SEM model presented in this study suggest that demand for taxi and Uber/Lyft within a NYC taxi zone may coexist without substituting for each other even after controlling for activity density and socioeconomic factors. This means that taxis and Uber/Lyft services may have their own separate demand due to factors unrelated to socioeconomic and built environment characteristics, and may not always be competing with each other in the real world as assumed in past studies *(3)*. It should be noted that for instance, the models in this study did not include a variable for taxi, Uber or Lyft fares due to unavailability of such data. Little or no differences in service fares may have been a factor in the counter-intuitive results obtained here. Therefore, further research is needed regarding the issue of competition between Uber/Lyft and taxi services, and any substitutions of Uber or Lyft demand for taxi demand.

CONCLUSIONS

This empirical analysis develops an SEM model for estimating the annual number of pickups for traditional taxis versus those of Uber and Lyft for taxi zones within New York City. The study sheds light on key socioeconomic, built environment and land use attributes which are significantly correlated with demand for ride-hailing services. These effects should be considered in demand forecasting analysis for taxi, Uber and Lyft services. The SEM model employed in this study allows for estimation of bidirectional relationships between demand for taxi, Uber and Lyft. Results indicate that demand levels for these ride-hailing services are interrelated and influence each other. However, the possibility of competition and substitution of these modes for each other should be addressed in future work considering additional data (perhaps data for several years) and additional variables such as service fare. The findings from this study can assist transportation professionals and decision-makers to more comprehensively understand factors that influence demand for ride-hailing modes, and to make policy decisions accordingly.

ACKNOWLEDGMENT

This research is partially funded by FHWA and the NTC at the University of Maryland. Findings presented in the paper do not necessarily represent the official views of the sponsoring agencies.

REFERENCES

1. Schaller Consulting. (2006). *"The New York City Taxicab Fact Book"*. www.schallerconsult.com/taxi/taxifb.pdf. Accessed June 8, 2018.
2. Cramer, J. and Krueger, A.B. (2016). "Disruptive change in the taxi business: The case of Uber." *The American Economic Review*, *106*(5), pp.177-182.
3. Correa, D., Xie, K. and Ozbay, K. (2017*). "Exploring the Taxi and Uber Demand in New York City: Empirical Analysis and Spatial Modeling."* The 96th TRB Annual Meeting.
4. Schaller, B. (2005). "A regression model of the number of taxicabs in US cities." *Journal of Public Transportation*, *8*(5), p.4.
5. Gonzales, E.J., Yang, C., Morgul, E.F., Ozbay, K. (2014). *Modeling Taxi Demand with GPS Data from Taxis and Transit. Mineta National Transit Research Consortium*.
6. Rogers, B. (2015). "The social costs of Uber." *University of Chicago Law Review Dialogue*, *82*, p.85.
7. Kelloway, E.K. (1998). "*Using LISREL for structural equation modeling*".Sage.
8. Mahmoudi, J. and Zhang, L., 2018. Impact of County-Level Built Environment and Regional Accessibility on Walking: A Washington, DC–Baltimore Case Study. *Journal of Urban Planning and Development*, *144*(3).

9. Rose, J.M. and Hensher, D.A. (2014). “Demand for taxi services: new elasticity evidence.” *Transportation*, *41*(4), pp.717-743.