Ridesharing and the Use of Public Transportation

Research-in-Progress

Katherine Hoffmann
Stern School of Business, NYU
44 W. 4th Street, New York NY 10012
khoffmann@stern.nyu.edu

Panagiotis G. Ipeirotis
Stern School of Business, NYU
44 W. 4th Street, New York NY 10012
panos@stern.nyu.edu

Arun Sundararajan
Stern School of Business, NYU
44 W. 4th Street, New York NY 10012
digitalarun@nyu.edu

Abstract

We investigate the effects of mobile-sourced ridesharing via platforms like Uber, Lyft, and Didi Chuxing on the use of public transit systems. Our study uses trip-level data about Uber usage in New York City, turnstile data about subway usage, and trip-level data about taxicab and shared bike usage. We find that on the surface, ridesharing and subway usage are positively correlated. Exploiting a series of exogenous shocks to the system – the closing of subway stations – to better isolate substitution effects, our preliminary results suggest that the average shock results in an increase of over 30% in the use of ridesharing, highlighting the potential for crowd-based systems to serve as infrastructure that helps smooth unexpected supply and demand surges. Our ongoing work studies how these substitution patterns vary with neighborhood socioeconomic indicators, and how substitution towards mobile-hailed ridesharing compares to traditional taxi and bike sharing. We hope to lay a data-driven foundation to better understand how sharing economy alternatives substitute and complement existing and future capital-intensive transit systems, and to provide a more judicious basis for assessing impacts on different population segments.

Keywords: Data science and business analytics, sharing economy, ridesharing

Introduction

As the sharing economy has gained prominence in the world economy, mobile-hailed ridesharing companies like Uber and Lyft in the US, Didi Chuxing in China, and Ola in India have begun to challenge traditional transportation providers. The growing popularity and exponential growth of such services has led to concerns that they will lower the economic viability of public goods provision, perhaps even eventually replacing existing mass transit systems with private alternatives that are less inclusive.

Ridesharing companies, in contrast, highlight their ability to provide services where public transit has failed. Upon its arrival in Boston, Uber ran promotions when the Red Line metro was scheduled to undergo repairs, citing "slow, inefficient buses" and "35,000 people left out in the cold waiting for shuttle service" (Uber 2011). For April Fool's day in New York City, Uber announced a premature "expansion" of Manhattan's long-awaited Second-Avenue subway line, offering rides along Second Avenue for the price of

a subway fare (Ninomiya 2014). And in San Diego, Uber explicitly claims: "gaps in public transportation become hubs for Uber...we complement public transit" (Donahue 2015).

In this paper, we describe an empirical study that takes a first step at quantifying the relationship between ridesharing and the use of public transportation, and the influence of new sharing economy alternatives on individual choices of transport mode at the system level. Specifically, we use data on Uber and subway ridership from the city of New York to study the extent to which ridesharing services act as "invisible infrastructure" (Sundararajan 2016) that absorbs the demand surplus caused by subway service disruptions. We find that subway service disruptions are associated with increases in the use of ridesharing. The magnitude of these increases is economically significant relative to average Uber ridership but small relative to average subway ridership.

Our findings suggest that Uber and Lyft's business strategy of competing for market share among consumers inconvenienced by public transit makes sense; such services may benefit from subway service disruptions. In doing so, this work also illustrates the potential of applying large-scale Open Data at the city level in a business analytics context. Although the increase in Uber rides during periods of subway service disruptions matches our expectations, the magnitude of this increase is harder to predict. The fact that we were unable to establish a similar relationship between subway disruptions and green taxi or Citi Bike ridership further suggests the need for empirical validation of such intuitive trends.

Background and Literature Review

As ridesharing services have grown in popularity, researchers and practitioners have begun to explore how they influence transport mode choice at the individual level. These studies generally suggest that ridesharing is a substitute for driving. For example, Rayle et al. (2014) conducted an intercept survey of ridesharing users and matched this survey with existing data on taxi users and taxi trips from the same time period. The authors found that many ridesharing customers reported less actual use of their cars, although the authors found no relationship between ridesharing and self-reported changes in car ownership. Similarly, a recent report from the American Public Transportation Association (APTA) used survey data to suggest that ridesharing is more likely to replace a trip in a private car than a trip with public transportation (Shared-Use Mobility Center 2016).

Interestingly, there is also evidence of this substitution among individuals whose consumption choices have impaired their ability to drive. Greenwood and Wattal (2015) use variation in the timing of Uber's market entry across California to estimate the impact of Uber availability on the number of drivers killed in alcohol-related incidents. Encouragingly, they find that UberX reduces traffic fatalities nine to fifteen months after it arrives in a new city, as drunk drivers may substitute away from their own cars when an affordable alternative is available.

These studies also agree that ridesharing appeals to those who typically rely on the public transportation system. For example, Rayle et al. (2014) found that ridesharing users were less likely to have a car than taxi customers, while the APTA report argued that the use of transport modes like ridesharing is associated with less car ownership and more use of public transportation. The authors differ on their attitudes towards whether ridesharing is a complement or substitute for public transportation; while Rayle et al. (2014) note that users saved approximately 10 minutes on average by choosing ridesharing over public transportation alternatives, APTA observes that ridesharing is most popular at times when public transportation operates less (e.g. late nights). The report concludes that public transportation systems would do well to leverage complementarities with such private-sector alternatives.

There are fewer studies of the systemic impact of ridesharing on local transportation markets. Most recently, a number of journalists and citizen scientists have begun to explore this question. For example, FiveThirtyEight has analyzed ridership data to assert that Uber is cutting into the market share of New York City yellow cabs, even as it is increasing door-to-door service in New York's outer boroughs (Bialik et al. 2015; Fischer-Baum and Bialik 2015). Nevertheless, this analysis is based on correlations rather than any sort of causal claim. Although such analyses form an important first step towards understanding ridesharing's impacts, a more thorough investigation of this relationship is needed. The research described below is an attempt to address this gap.

Figure 1. Weekly Uber Pickups and Subway Entries (April 2014 - August 2015)

Research Questions

Our research studies the extent to which ridesharing acts as a short-term substitute for public transportation. Although short-run substitution seems likely, it is possible that ridesharing and public transportation target *different individuals*. For example, the cost of ridesharing may be prohibitive for individuals who normally travel with public transportation; or, the inconvenience and safety risks associated with public transportation may lead ridesharing users to avoid public transportation altogether. On the other hand, ridesharing and public transportation may be associated with *different trip types*. For example, individuals may prefer ridesharing for social trips, while they may rely on public transportation for commuting. We are interested in establishing whether such substitution does occur, and if it does, estimating its magnitude and the extent to which it is correlated with demographic and geographic characteristics of displaced commuters such as average income, baseline reliance on public transportation, and the level of transit service available.

Data

We use publicly available data from the city of New York, a natural setting for this study: the tri-state area is home to the largest public transportation system in the United States, and New York's Metropolitan Transportation Authority (MTA) provides approximately one third of all public transit rides nationwide (MTA 2016a). As of January 2016, the MTA reported an average of 6.57 million heavy rail trips a day (APTA 2016).

New York City is also home to a growing number of ridesharing users. Our data on Uber pickups spans the period from April – September 2014 and January – June 2015. During this period, ridesharing operated at a fraction of the scale of the subway system; the maximum number of rides per day that we observed was just under 134,000. Nevertheless, during this period the number of Uber pickups grew dramatically (see Figure 1), rising from 0.55 million per month in April 2014 t0 2.82 million in June 2015.

For context, we extend our analysis of subway and Uber data to include three other popular mode choices: yellow taxis (approximately 428,000 rides per day during our period of observation), green taxis (approximately 49,000 rides per day during our period of observation), and Citi Bike (approximately 22,000 rides per day during our period of observation). Yellow taxis are the classic mode of for-hire transportation in New York, with a service area that is heavily concentrated in mid-town Manhattan. Green cabs were introduced in 2013 to offset this spatial disparity; they serve uptown Manhattan and the outer boroughs, and are prohibited from picking up passengers south of West 110th Street or East 96th Street. Finally, Citi Bike has acted as New York's bike share alternative since 2013; it is worth noting that Citi Bike ridership is highly seasonal. Citi Bike's docks are concentrated primarily in central Manhattan and western Brooklyn, though its service area continues to expand.
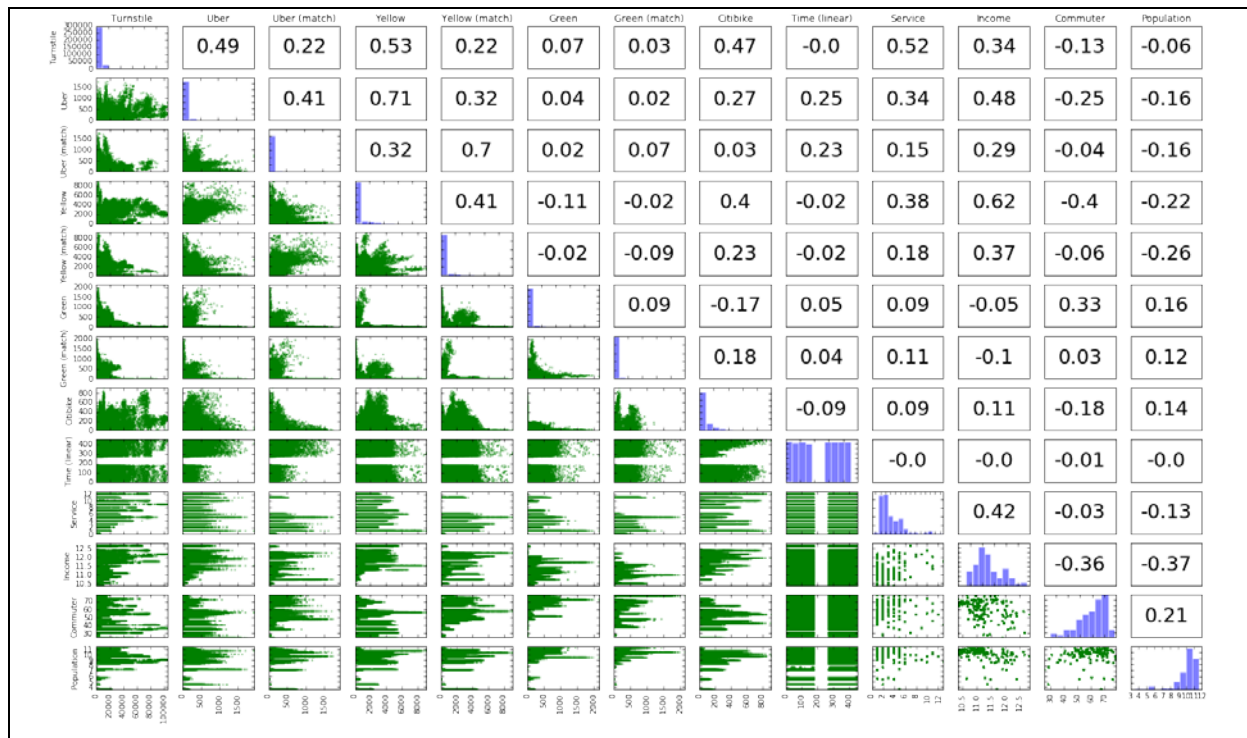
Figure 2. Correlograms for Key Variables in the Aggregated Dataset

Cells show histograms (diagonal), scatterplots (below diagonal) and correlations (above diagonal). From top left, the variables shown are: turnstile entries, Uber pickups, Uber pickups in the matched zone, yellow taxi pickups, yellow taxi pickups in the matched zone, green taxi pickups, green taxi pickups in the matched zone, Citi Bike pickups, a linear measure of time, the number of subway lines serving the zone, log mean income of the zone, the percent of workers who commute with public transportation in the zone, and log population in the zone.

Data sources

The MTA has published weekly data on the number of turnstile entries and exits since 2010 (MTA 2016b). The data consists of cumulative readings for stations across the city, collected in approximately four-hour intervals at the *subunit channel position* level (which appears to correspond to a turnstile or a group of turnstiles). The data represents 4,635 unique subunit channel positions, which are associated with 732 *control areas* (roughly analogous to station entrances or exits) and 468 *remote units* (analogous to stations).

Several steps were taken to normalize the data for analysis. First, although cumulative entry and exit readings were generally recorded in four-hour intervals starting at midnight, the timing of readings was not always consistent. Therefore, readings were resampled at the hourly level, and missing values were filled through linear interpolation. Second, because these readings were cumulative, each reading was subtracted from the previous hour's reading to obtain the differential number of entries and exits. Finally, occasional unexplained jumps in the cumulative readings occurred (possibly corresponding to when a counter was reset). To address these anomalies, negative entry or exit counts and entry or exit counts exceeding 10,000 per hour for a single subunit channel position were discarded.

New York City's Taxi and Limousine Commission (TLC) collects data on passenger pickups for Uber and other for-hire vehicle companies. Through Freedom of Information Law (FOIL) requests, journalists have made this data publicly available for April – September 2014 and January – June 2015 (FiveThirtyEight 2016). The 2014 Uber data contains trip-level pickup data for 5 Uber bases with GPS coordinates, whereas the 2015 data contains trip-level pickup data from 6 Uber bases with location identified according to one of New York's 263 taxi zones.

Yellow and green taxi data was accessed from the TLC, which publishes geotagged trip-level data for its cabs (NYC Taxi & Limousine Commission 2016). Similarly, geotagged trip-level bike share data was obtained from the Citi Bike website (Citi Bike NYC 2016). Complementary demographic data from the 2010 census and the 2014 American Community Survey (ACS) was downloaded for New York City census tracts from the American FactFinder data portal (U.S. Census Bureau 2016a, 2016b).
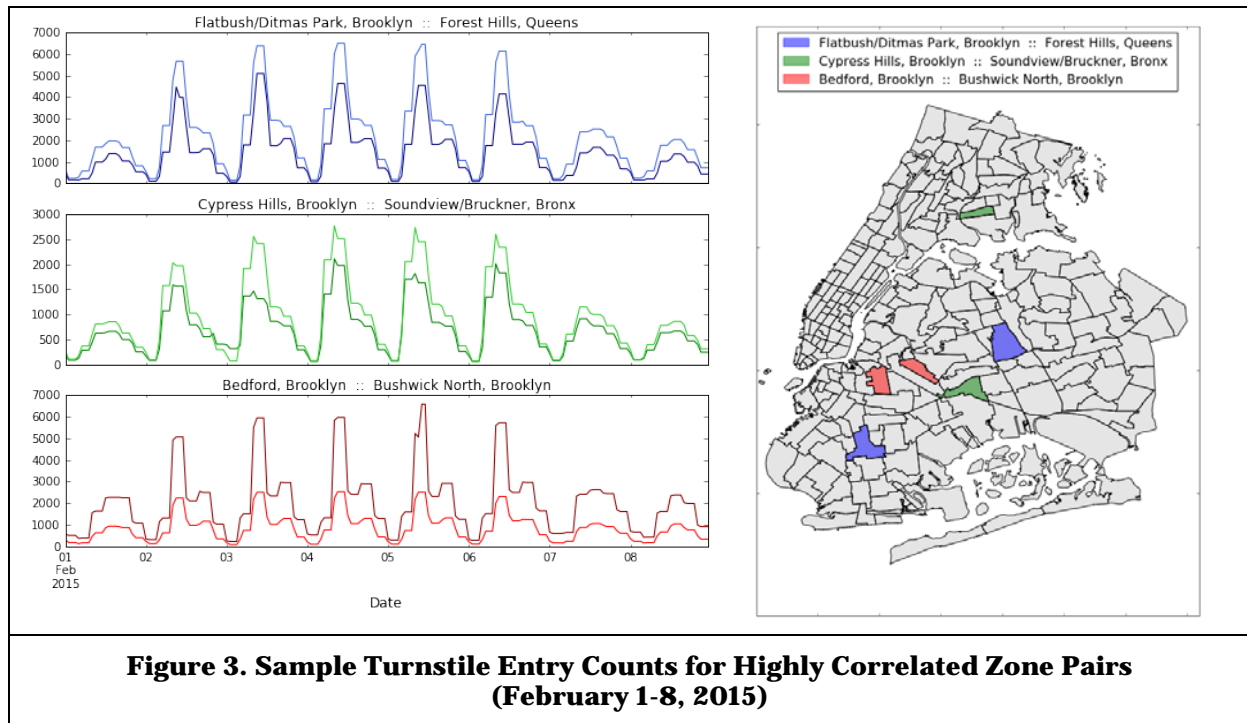
Aggregation

To create a compiled dataset for analysis, all datasets were restricted to the range of dates for which Uber pickup counts were available: April – September 2014 and January – June 2015. To match the taxi zone level of aggregation in the 2015 Uber data, all ride-level data (Uber pickups from 2014, taxi rides, and Citi Bike pickups) was assigned to a taxi zone using GPS coordinates and aggregated. Similarly, each turnstile control area was assigned to a taxi zone, and turnstile entry counts were aggregated by taxi zone. Finally, because the turnstile data was primarily read in four-hour intervals starting at midnight, all zone-level ride counts were aggregated over the four-hour intervals leading up to 12am, 4am, 8am, 12pm, 4pm, and 8pm. The final dataset consists of 334,052 observations representing 2,160 four-hour intervals and 156 taxi zones. Summary data for key variables is shown in Figure 2.

Associating the census data with taxi zones required extrapolating from tract-level statistics. The 2,168 census tracts associated with New York's five boroughs did not all fit neatly into taxi zone boundaries; therefore, we used the taxi zone boundaries to "slice" the census tract shapefile, resulting in 4,992 unique tract segments. To calculate aggregate statistics at the taxi zone level s_z, we computed:

$$s_z = \frac{1}{\sum_t n_t \frac{a_{tz}}{a_t}} \sum_t s_t * n_t \frac{a_{tz}}{a_t} \quad,$$

where n_t is population of tract t; a_t is the area of tract t; a_{tz} is the area of tract t that falls in zone z; and s_t is the relevant statistic for tract t (e.g. the percent of workers commuting with public transit). In other words, we used the proportion of the tract *area* that fell in the taxi zone, a_{tz}/a_t, to estimate the proportion of the tract *population* that fell in the zone. We then used this population to weight the tract-level statistic s_t when aggregating at the zone level.

Figure 3. Sample Turnstile Entry Counts for Highly Correlated Zone Pairs (February 1-8, 2015)

Model

Following prior work, our model design decisions included: (1) *the unit of analysis*, typically framed as a one-way trip, a round-trip tour, or a daily activity schedule; (2) *the level of aggregation*, typically aggregated by region or market segment, or disaggregate at the individual or household level; and (3) *the type of data used*, which may consist of revealed or stated preference data, longitudinal or cross-sectional data, and observational or attitudinal data (Ben-Akiva 2008; Ben-Akiva and Lerman 1985).

Because New York's subway data does not associate turnstile entries with exits at the individual level, and because only pickup data was available for Uber rides, our unit of analysis is effectively a *trip leg start*; we observe the point at which individuals begin to use a particular transport mode. Although restrictive, this should suffice for our purposes; the assumption is that an individual, attempting to use the subway and finding that it is out of service, makes a choice among the other available alternatives. Similarly, the constraints in the Uber data described above required aggregation at the taxi zone level. Fortunately, the richness of the available data allows both a longitudinal and observational analysis, with preferences revealed by actual behavior.

Disruptions

To test the hypothesis that ridesharing may act as a short-run substitute for public transportation, we first examined the impact of subway service disruptions on the use of Uber. A taxi zone was considered "disrupted" if one or more of its subway remote units (stations) experienced no turnstile entries or exits within a four-hour period. Almost all disruptions (98%) lasted for two days or less, with 89% lasting one day or less and 38% lasting only four hours. Disruptions most commonly started on Saturday (37%) and ended on Sunday (33%), a trend which is consistent with the MTA's frequent use of weekends for planned repairs. Overall, taxi zones experienced disruptions approximately 0.88% of the time, yielding 2,912 four-hour disruption incidents in our dataset. The model was estimated as follows:

$$uber_{it} = \alpha + \beta_1\ disruption_{it} + \beta_2\ subway_{it} + \gamma_1\ time_t + \gamma_2 zone_i + \varepsilon_{it}\ , \tag{1}$$

where $uber_{it}$ represents the number of Uber pickups for zone i in period t; α is a constant; $disruption_{it}$ is a binary variable indicating a subway disruption for zone i in period t, $subway_{it}$ represents the number of turnstile entries for zone i in period t; $time_t$ is a vector of time controls that includes a linear variable indicating the passage of time in days, binary indicators for day of week, and binary indicators for hour of day; $zone_i$ is a set of taxi zone fixed effects; and ε_{it} is an error term clustered at the zone level. The coefficient of interest is β_1, the estimated change in the number of Uber rides associated with a disruption in a given taxi zone.

Zone Matching

To address the concern that Uber pickups and subway service disruptions may be correlated with a third omitted variable, we also create pairs of "matched" geographical areas with similar subway ridership patterns and test a specification that controls for Uber ridership in the matched areas. In other words, for each taxi zone, our goal is to identify a zone that can act as a counterfactual to the disrupted zone.

Specifically, we created a list of all possible pairs from the 156 zones for which turnstile entry counts existed. We removed all pairs of zones that were served by the same subway lines, due to concerns that disruptions may travel along a given subway line. We also removed all pairs of zones that were spatially contiguous, due to concerns that turnstile entries in one zone might be negatively correlated with turnstile entries in neighboring zones when disruptions occur. For each of the remaining eligible pairs, we calculated the Pearson correlation between the time series of four-hour turnstile entry counts, and matched zones in descending order of correlation. Of the 156 zones for which turnstile entry counts existed, we were able to pair 124 zones, with correlations in turnstile entries ranging from 0.98 to 0.43. Figure 3 contains a sample plot with one week's worth of turnstile entry data from the 3 zone pairs with the highest correlation, along with a map of these zones. We then estimated the following model:

$$uber_{it} = \alpha + \beta_1\ disruption_{it} + \beta_2\ subway_{it} + \beta_3 uber_{mt} + \gamma_1\ time_t + \gamma_2 zone_i + \varepsilon_{it}\ , \tag{2}$$

where $uber_{mt}$ represents the number of Uber pickups for matched zone m in period t, and $uber_{it}$, α, $disruption_{it}$, $subway_{it}$, $time_t$, $zone_i$, and ε_{it} are as described above. Once again, the coefficient of interest is β_1, the estimated change in the number of Uber rides associated with a disruption in a given taxi zone. Results of fixed-effects regressions for both models are presented in Table 1.

Other Transport Modes

In order to add context to our estimates of the relationship between subway disruptions and Uber ridership, we also fit models (1) and (2) using yellow taxi, green taxi, and Citi Bike pickups as the dependent variable. Note that while the zone *pairings* remained the same regardless of the dependent variable (because zone matching was completed using correlations in the turnstile entry data), the actual *count* of rides in the matched zone reflect yellow and green taxi ridership, respectively. Unfortunately, the Citi Bike station data covered only 39 of New York's 263 taxi zones; consequently, the number of zone pairings with data for both zones was too low to estimate specification (2). Results are shown in Table 1.

Demographic Characteristics

In models (1) and (2), we rely on fixed effects to capture zone-level variation in mode preferences. Nevertheless, transport models often quantify such variation by framing mode choice as a function of individual attributes – such as the demographic characteristics of the traveler or the land use of surrounding areas – as well as the available mode alternatives (Ben-Akiva and Lerman 1985). In their review of the literature, Taylor and Fink (2003) further highlight transportation system characteristics as potentially important predictors of ridership, such as the availability of service and the level of coordination between different transit lines.

As described above, we estimated key demographic characteristics at the taxi zone level from New York's census and ACS data, including: mean income, total population, and the percentage of workers who commute with public transit. Using data on subway entrance locations, we are also able to measure the number of subway lines serving each taxi zone, which we adopt as a rough proxy for the availability of public transportation alternatives (assuming that neighborhoods with many intersecting lines represent more central transit hubs). To investigate how demographic characteristics correlate with transport mode choice, we replace zone fixed effects with zone-level characteristics to estimate the following model:

$$uber_{it} = \alpha + \beta_1 commuters_i + \beta_2 income_i + \beta_3\ service_i + \beta_4\ population_i + \gamma_1\ time_t + \varepsilon_{it}\ , \qquad (3)$$

where $commuters_i$ is the estimated percentage of workers in zone i who commute using public transportation; $income_i$ is the log estimated mean income in zone i; $service_i$ is the number of subway lines serving zone i; $population_i$ is the log estimated population of zone i; and $uber_{it}$, α, $time_t$, and ε_{it} are as above. We also estimate this model with yellow taxi, green taxi, and Citi Bike ridership as the dependent variable; results of Ordinary Least Squares regressions using these specifications are shown in Table 2.

Results

In general, turnstile entries and Uber pickups were positively correlated, with each turnstile entry corresponding to 0.005 additional Uber pickups. This is unsurprising, given that there are likely to be overall fluctuations in the total number of riders for different four-hour intervals and zones that are not captured by time- or zone-level fixed effects.

After controlling for these variations in zone-level turnstile traffic, we nevertheless see evidence of increased Uber ridership during periods of subway service disruption. Zones that experienced a disruption were associated with 14.17 additional Uber rides in a four-hour period, a 31% increase relative to the average of 45.61 rides per zone. This supports our hypothesis that Uber may serve as a short-run replacement for public transit for some riders. Our results also indicate a successful matching process: each additional Uber pickup in a matched zone was associated with 0.24 additional pickups in the zone of interest. However, controlling for Uber pickups in the matched zone did not influence our estimate of the impact of subway disruptions.

Yellow taxi pickups followed a similar pattern. Yellow taxi pickups were positively correlated with turnstile entries; each entry was associated with 0.032 additional pickups. Furthermore, disruptions were associated

with 44.16 additional yellow taxi rides, a 12% increase over the baseline average of 379 pickups per taxi zone in a four-hour period. Including controls for pickups in the matched zone only increased our estimate of the absolute and relative magnitude of substitution. We were unable to support or reject any relationship between green taxi rides and turnstile entries, or to identify any measurable variation in these rides with subway service disruptions, perhaps on account of the newness of this mode. While Citi Bike rides are positively correlated with turnstile entry volume, we were also unable to establish a relationship between these rides and subway closures.

Table 1. Uber, Taxi, and Citi Bike Pickups as a Function of Subway Service Disruptions

Pickups	Uber (1)	Uber (2)	Yellow Taxi (3)	Yellow Taxi (4)	Green Taxi (5)	Green Taxi (6)	Citi Bike (7)
Turnstile Entries	0.005*** (0.001)	0.005*** (0.001)	0.032*** (0.005)	0.036*** (0.007)	0.000 (0.000)	0.000 (0.001)	0.003*** (0.001)
Disruption	14.173*** (5.071)	14.173* (7.188)	44.158** (17.042)	54.655** (21.048)	0.099 (3.709)	-0.830 (3.840)	4.859 (9.574)
Pickups in Matched Zone		0.240*** (0.082)		0.127** (0.050)		0.043 (0.030)	
Time (linear)	0.170*** (0.018)	0.110*** (0.018)	-0.079*** (0.013)	-0.059*** (0.014)	0.037*** (0.006)	0.032*** (0.006)	-0.063*** (0.009)
Taxi Zone Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	334,052	263,056	334,052	265,238	334,052	265,238	84,140
R-squared	0.309	0.310	0.253	0.220	0.099	0.093	0.394
Number of Taxi Zones	156	124	156	124	156	124	39
Mean LHS	45.61	38.22	379	311.9	49.74	47.88	75.13

Table 2. Uber, Taxi, and Citi Bike Pickups as a Function of Zone Characteristics

Pickups	Uber (1)	Yellow Taxi (2)	Green Taxi (3)	Citi Bike (4)
Log Mean Income	66.386*** (7.872)	714.477*** (109.740)	16.897 (14.438)	23.789 (19.255)
Workers Commuting with Public Transit (%)	-1.038** (0.430)	-16.719*** (4.189)	3.287*** (0.681)	-1.625** (0.654)
# Subway Lines Serving Zone	7.480*** (1.635)	61.572*** (22.519)	3.576 (2.820)	2.065 (2.416)
Log Population	1.682 (3.968)	12.850 (34.227)	10.953** (4.465)	14.512** (6.056)
Time (linear)	0.168*** (0.017)	-0.098*** (0.019)	0.038*** (0.006)	-0.064*** (0.008)
Taxi Zone Controls	No	No	No	No
Observations	334,052	334,052	334,052	84,140
R-squared	0.363	0.472	0.168	0.308
Number of Taxi Zones	156	156	156	39
Mean LHS	45.61	379	49.74	75.13

*** $p<0.01$ ** $p<0.05$ * $p<0.1$. Standard errors are shown in parentheses and were clustered at the taxi zone level. Dataset has one observation per four-hour period, by taxi zone. Fixed effects for the day of the week and four-hour period of the day, as well as a constant, were included but are not shown.

In examining demographic characteristics, we find a strong positive correlation between Uber and yellow taxi pickups and the log mean income in a given taxi zone, though this effect is stronger in relative terms for yellow taxis. Interestingly, we find that Uber and yellow taxi ridership is higher in taxi zones served by more subway lines, but it is negatively correlated with the percent of workers commuting with public transportation in the zone. This may reflect the fact that Uber and yellow taxi pickups are concentrated in zones with a large number of competing transport options.

Conclusion and Next Steps

To our knowledge, this research represents the first attempt to pair ridesharing and subway ridership data from New York City to investigate substitution between the two transport modes. Our preliminary findings act as a proof-of-concept for this approach. They support the hypothesis that a decrease in public transportation use can be partially offset by an increase in ridesharing, at least in the short run and in response to subway system shocks. Although the magnitude of this response is high relative to average ridesharing levels, it is a small fraction of subway usage. It remains to be seen how much this substitution grows as mobile-hailed ridesharing becomes increasingly mainstream.

In extending our analysis to alternative transit modes, we find a parallel increase in yellow taxi rides in the face of service disruptions, though this increase appears to be smaller in relative terms. We find no evidence of an impact on green taxi or Citi Bike ridership. Perhaps, since many subway service disruptions occur in the evenings (when it may be difficult or dangerous to street-hail a green taxi outside of central Manhattan, and when riding a bike is less appealing), there is not substitution of significance. In either case, our findings suggest that while subway disruptions should intuitively drive riders to other forms of transportation, this effect is not equal across all transport modes.

As research-in-progress, our work has many limitations. Although subway disruptions provide a natural experiment for the study of short-term substitution behavior, disruptions are not a wholly exogenous shock. Planned disruptions may be optimally scheduled during periods when subway ridership is low (and Uber ridership is high) so as to minimize system impact, while unplanned disruptions may be correlated with omitted variables such as crime rates (e.g. the closure of stations due to police activity), which also drive Uber use. Furthermore, Uber's own competitive behavior may introduce distortions into observed substitution patterns; for example, in the past Uber has run promotions to coincide with periods when disruptions occur (Ninomiya 2015).

In addition, while we have tried to control for obvious determinants of ridership and disruptions using time- and zone-level fixed effects, we have not captured the full range of variation in factors that jointly determine these variables. An exploratory analysis of demographic variation in ridership suggests considerable differences across modes. Notably, Uber and yellow taxi ridership is highest in zones with higher mean income and fewer residents who commute with public transportation, despite the fact that these zones receive better levels of subway service.

One of our next steps will be to estimate a discrete choice model that combines our data on subway, Uber, taxi, and Citi Bike ridership into a single framework. This approach will enable a more careful analysis of the tradeoffs between different modes, and will allow us to model these tradeoffs as a direct function of rider demographics. We also plan to extend our analysis beyond subway service disruptions to examine other situations where substitution between the two modes is likely. For example, by collecting more data on time- and location-specific Uber price promotions, we hope to study the impact of other types of shocks on the distribution of mode choice. Finally, as the city publishes additional data on for-hire vehicles, we plan to extend our analysis to other ridesharing providers, building a more complete picture of New York's transit system as a whole.

We hope to lay a data-driven foundation to better understand how sharing economy alternatives substitute and complement existing and future capital-intensive transit systems, and to provide a more judicious basis for assessing impacts on different population segments. Taken together, our results will provide more granular insights into how ridesharing affects mode choice in the most heavily-used municipal transportation system in the US – a valuable asset to any policymaker looking to understand and manage the sharing economy's new competitors.

References

APTA. 2016. "Individual Agency Data Collected by the FTA National Transit Database," Public Transportation Ridership Report, American Public Transportation Association (available at http://www.ntdprogram.gov/ntdprogram/pubs/MonthlyData/January%202016%20Raw%20Database.xls).

Ben-Akiva, M. E. 2008. *Travel Demand Modeling*, Presented at the Transportation Systems Analysis: Demand & Economics, Massachusetts Institute of Technology, Fall (available at http://ocw.mit.edu/courses/civil-and-environmental-engineering/1-201j-transportation-systems-analysis-demand-and-economics-fall-2008/lecture-notes/MIT1_201JF08_lec05.pdf).

Ben-Akiva, M. E., and Lerman, S. R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press.

Bialik, C., Flowers, A., Fischer-Baum, R., and Mehta, D. 2015. "Uber Is Serving New York's Outer Boroughs More Than Taxis Are," *FiveThirtyEight*, August 10 (available at https://fivethirtyeight.com/features/uber-is-serving-new-yorks-outer-boroughs-more-than-taxis-are/).

Citi Bike NYC. 2016. "Citi Bike System Data," (available at http://www.citibikenyc.com/system-data).

Donahue, K. 2015. "Uber and Public Transit Working Hand in Hand in San Diego," *Uber Newsroom*, February 3 (available at http://newsroom.uber.com/sd/2015/02/uber-and-public-transit-working-hand-in-hand-in-san-diego/).

Fischer-Baum, R., and Bialik, C. 2015. "Uber Is Taking Millions of Manhattan Rides Away From Taxis," *FiveThirtyEight*, October 13 (available at http://fivethirtyeight.com/features/uber-is-taking-millions-of-manhattan-rides-away-from-taxis/).

FiveThirtyEight. 2016. "Uber TLC FOIL Response," *GitHub* (available at https://github.com/fivethirtyeight/uber-tlc-foil-response).

Flynn, M. 2015a. "Ride in Manhattan for $10 or Less," *Uber Newsroom*, July 10 (available at https://newsroom.uber.com/us-new-york/ride-in-manhattan-for-10-or-less/).

Flynn, M. 2015b. "Ride Free to Bed-Stuy and Crown Heights," *Uber Newsroom*, June 18 (available at https://newsroom.uber.com/us-new-york/ride-free-to-bed-stuy-and-crown-heights/).

Flynn, M. 2015c. "Share Your Ride, Reduce Emissions," *Uber Newsroom*, April 22 (available at https://newsroom.uber.com/us-new-york/share-your-ride-reduce-emissions/).

Greenwood, B. N., and Wattal, S. 2015. "Show Me the Way to Go Home: An Empirical Investigation of Ride Sharing and Alcohol Related Motor Vehicle Homicide," SSRN Scholarly Paper No. 2557612, Rochester, NY: Social Science Research Network (available at http://papers.ssrn.com/abstract=2557612).

Metropolitan Transportation Authority. 2016a. "The MTA Network: Public Transportation for the New York Region," (available at http://web.mta.info/mta/network.htm).

Metropolitan Transportation Authority. 2016b. "Turnstile Data," (available at http://web.mta.info/developers/turnstile.html).

Ninomiya, K. 2014. "Uber Announces Second Avenue Subway Expansion," *Uber Newsroom*, April 1 (available at https://newsroom.uber.com/uber-announces-second-avenue-subway-expansion-2/).

Ninomiya, K. 2015. "$5 for 5 Weekends," *Uber Newsroom*, April 17 (available at https://newsroom.uber.com/us-new-york/5-for-5-weekends/).

NYC Taxi & Limousine Commission. 2016. "TLC Trip Record Data," (available at http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml).

Rayle, L., Shaheen, S., Chan, N., Dai, D., and Cervero, R. 2014. "App-Based, On-Demand Ride Services: Comparing Taxi and Ridesharing Trips and User Characteristics in San Francisco," Working Paper No. UCTC-FR-2014-08, University of California Transportation Center (available at http://www.uctc.net/research/papers/UCTC-FR-2014-08.pdf).

Shared-Use Mobility Center. 2016. "Shared Mobility and the Transformation of Public Transit," Research Analysis No. TCRP J-11/TASK 21, American Public Transportation Association (available at http://www.apta.com/resources/reportsandpublications/Documents/APTA-Shared-Mobility.pdf).

Sundararajan, A. 2016. *The Sharing Economy: The End of Employment and the Rise of Crowd-Based Capitalism.* Boston: MIT Press.

Taylor, B. D., and Fink, C. N. Y. 2003. "The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature," Working Paper, UCLA Institute of Transportation Studies (available at http://escholarship.org/uc/item/3xk9j8m2).

Uber. 2011. "Red Line Fail? Ride for Free on Boston's Uber Line," *Uber Newsroom*, November 10 (available at https://newsroom.uber.com/us-massachusetts/red-line-fail-introducing-the-u-line/).

Uber. 2015. "Explore the Bronx This Weekend," *Uber Newsroom*, May 15 (available at https://newsroom.uber.com/us-new-york/explore-the-bronx-this-weekend/).

U.S. Census Bureau. 2016a. "DP-03. Selected Economic Characteristics. 2010-2014 American Community Survey 5-Year Estimates." (available at https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_14_5YR_DP03&src=pt).

U.S. Census Bureau. 2016b. "Profile of General Population and Housing Characteristics: 2010," (available at https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_DP_DPDP1&src=pt).

Wadhwa, T. 2014. "Could Lyft and Uber Put Public Transit Out of Business?" *Forbes* (available at http://www.forbes.com/sites/tarunwadhwa/2014/11/13/will-lyft-and-ubers-shared-ride-service-put-public-transit-out-of-business/).