Ridesharing and the Use of Public Transportation

Completed Research Paper

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

Panos 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 how the digital business model of on-demand ridesharing platforms like Uber and Lyft interacts with an established, centralized public mass transit system. Our study uses data on ridesharing, taxi, shared bike, and subway usage in New York City and exploits a series of exogenous shocks to the system – the closing of subway stations – to isolate substitution effects. We find that the average shock is associated with a 2.8 - 3.3% increase in the use of ridesharing, which translates into 5.5 additional Uber rides and 1.5 additional Lyft rides per taxi zone and four-hour period. Although this suggests that on-demand ridesharing acts as infrastructure that helps smooth unexpected transportation supply and demand surges, the estimated effect is small relative to the average number of subway rides displaced. Our results indicate that the flexibility inherent in ridesharing's crowd-based business model could be further exploited to support capital-intensive transit systems in the future.

Keywords: ridesharing; sharing economy; cities; transportation; difference-in-differences; digital business models

Introduction

Digital technologies are changing mobility business models, increasingly influencing how people move in physical space. As sharing economy business models transform the transportation sector, digitally-enabled platforms like Uber and Lyft in the US, Didi Chuxing in China and Latin America, and Ola in India have begun to challenge traditional urban transportation providers, expanding the options available to city residents and offering the promise of an increasingly multi-modal urban transport future.

Ridesharing companies highlight how their flexible digital business model allows service provision where public transit has failed. Lyft runs a "friends with transit" campaign and advertises "one gajillion new stops," with services that take riders "the rest of the way" (Lyft 2015). 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 investigate the extent to which ridesharing supplements public transportation at the system level. Specifically, we treat subway station closures in New York City as a natural experiment, using open data on Uber, Lyft, taxi, Citi Bike, and subway ridership to study how crowd-based capacity harnessed via a platform may serve as "invisible infrastructure" that absorbs the demand spikes caused by disruptions to centralized public mass transit. This is an important instance of a broader question – how digitally-enabled, decentralized systems can work alongside centralized infrastructure – whose answer is key to understanding the value of platform-based business models.

We answer three focused questions. *First, do riders substitute to taxi, platform-based ridesharing, or Citi Bike when faced with subway disruptions?* We find that subway service disruptions are associated with statistically significant increases in the use of Uber, Lyft, and taxi services, but at the city level, we find no evidence of a significant increase in Citi Bike use.

Second, do riders prefer the digital alternative – platform-based ridesharing – over the physical-world alternative – traditional street-hail taxis – when faced with a public transit disruption? In percentage terms, we find no evidence of a citywide preference for Uber and Lyft relative to taxis. While subway service disruptions are associated with 2.8% and 3.3% increases in the use of Uber and Lyft, respectively, they are associated with 8.2% and 7.0% increases in the use of yellow and green taxi services.

Third, how much of displaced subway ridership do the digitally-based and physical-world modes absorb, and which modes absorb the greatest number of displaced riders? We find that the digitally-enabled modes absorb only a small fraction of displaced riders, and that they likely have the ability to play a far greater role in the future. On average, we estimate that each disruption displaces over 1,500 rides, but that the corresponding increases in taxi, platform-based ridesharing, and Citi Bike use together account for less than 40 additional rides. In terms of the absolute number of riders accommodated, we estimate that ridesharing plays a more important public transit role relative to taxis in the "outer boroughs" of Brooklyn, Queens, and the Bronx (relative to Manhattan), and that the significance of this role is growing over time.

Our findings suggest that the strategy adopted by digital transit players – i.e. competing for market share among consumers inconvenienced by public transit – makes sense; platform-based demand is increased by subway disruptions. However, these modes are not currently absorbing the majority of displaced riders. This suggests that the flexibility inherent in their crowd-based business model could be exploited further. Finally, although an increase in Uber, Lyft, and taxi rides during periods of subway service disruptions matches our expectations, the magnitude of this increase – and the variation in effects across modes, space, and time – is harder to predict, highlighting the value of quantitatively estimating the impact of digital platforms on centralized infrastructure systems.

Literature Review

Using national panel datasets, a number of recent studies have investigated substitution and complementarity between ridesharing and public transit in cities across the US. For example, using variation in the timing of Uber's market entry across cities to fit a difference-in-differences model, Hall et al. (2018) find that ridesharing is a complement for public transportation, leading to a 5% increase in ridership over a two-year period. Nelson and Sadowsky (2017) add nuance to this narrative by exploring how competition influences the effect of ridesharing on public transportation. They find that ridesharing is initially a complement to public transportation after the entry of the first ridesharing provider, but – presumably as a result of competitive pressure to reduce prices – becomes a substitute once a second ridesharing provider arrives.

Also using a difference-in-differences approach, Babar and Burtch (2017) find differential effects by mode, arguing that Uber has decreased the use of city bus and increased the use of subways and commuter rail. This is consistent with Clewlow and Mishra (2017), who use a survey of seven cities to conclude that ridesharing use was associated with a 6% decline in the use of public transportation at the individual level; they find that the negative effects of ridesharing were concentrated on bus and light rail services, whereas ridesharing was a complement for commuter rail.

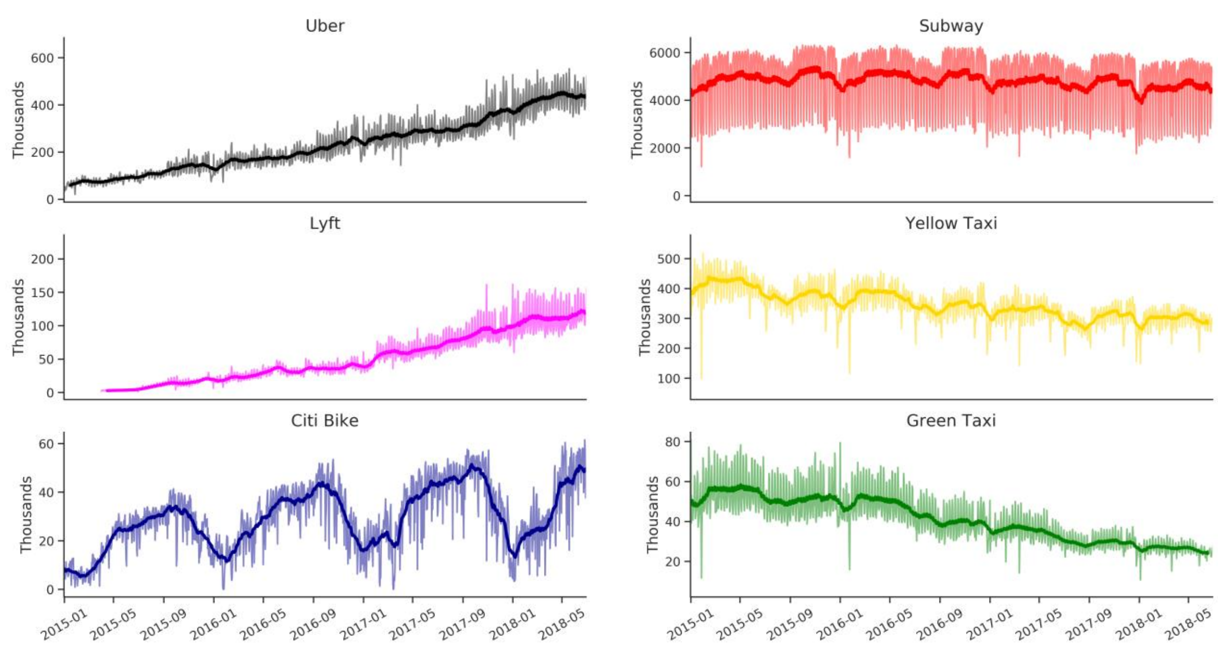
These studies estimate the effect of Uber and Lyft across a large number of US cities. However, a key takeaway from these studies is that the impact of ridesharing on public transit is not uniform; for example, Babar and Burtch (2017) find that the quality of service provision moderates the effect of ridesharing on transport use, while Hall et al. (2018) find that complementarity is strongest for small transit agencies and big cities.

A better understanding of such heterogeneous impacts is crucial because the *typical* US city can be quite different from a *large* one. Most American cities are characterized by low urban density ("sprawl"), high levels of car ownership, and low transit coverage. A focus on such cities overlooks the experience of many of the nation's transit users, which are located in dense urban areas: the Federal Transit Administration's most recent ridership data indicates that 42% of nationwide transit rides taken in June 2019 were affiliated with New York's Metropolitan Transportation Agency (Federal Transit Administration 2019).

There are three ways in which the role of ridesharing platforms might differ in a city like New York. First, ridesharing platforms may face more competition as a driving alternative. There are multiple good options, including a strong taxi industry and a widely-used public transportation system; for example, Walk Score has ranked New York #1 in the nation for walkability and transit scores (Walk Score 2019).

Second, large cities may face greater challenges given systematic substitution to ridesharing alternatives. For example, New York struggles with worsening traffic congestion and the need to reduce the number of vehicles in Central Manhattan. Given that the subway saw almost 5 million daily rides on average in our sample, replacing even 10% of these trips with ridesharing would more than double the number of daily Uber trips as of June 2018.

Despite these potential constraints, one key promise of ridesharing platforms in this context is their ability to provide additional, dynamically responsive capacity to supplement existing transit systems. For example, hailing a car using an electronic app may enable riders to react quickly to unexpected route changes, and (with the help of surge pricing) may attract drivers to areas in need. From this perspective, ridesharing platforms may offer superior alternatives to taxi or bikeshare services, which do not have this rapid response capacity enabled by platform algorithms in combination with flexible supply. Nevertheless, we are aware of no existing studies that examine this "invisible infrastructure" functionality.



Background shows daily fluctuations in ridership; foreground shows centered 30-day moving averages.

Figure 1. Daily Ridership (January 2015 – June 2018)

Relative to prior research, our work provides three key contributions. First, we formally document the use of ridesharing in response to subway disruptions, quantifying the extent to which New Yorkers use ridesharing as an emergency alternative. Second, we provide evidence on the role of ridesharing relative to other modes in the context of disruptions, generating insights into the comparative value provided by these new sharing economy entrants. Finally, we explore the relationship between ridesharing and public transportation in granular detail: rather than comparing monthly ridership across aggregate categories of transit modes in metropolitan areas, we are able to study ride choices by mode at four-hour intervals across different regions of the city. Our approach builds on existing results by contributing to a more holistic understanding of how ridesharing alternatives integrate into public transit systems.

Context

Mode choice in NYC

To examine these questions, we use open 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 in the course of our study period, the subway recorded an average of 4.8 million rides per day (Metropolitan Transportation Authority 2018a).

New York City is also home to a growing number of ridesharing users. For example, from January 2015 to June 2018, Uber pickups grew more than 7-fold, from under 60,000 daily rides to over 440,000 daily rides. Similarly, from April 2015 – June 2018, Lyft pickups grew more than 40-fold, from an average of under 3,000 daily rides to over 120,000 daily rides (see Figure 1). Our analysis also includes three other popular mode choices: yellow taxis (approximately 345,000 rides per day during our period of observation), green taxis (approximately 41,000 rides per day during our period of observation), and Citi Bike (approximately 29,000 rides per day during our period of observation) (Citi Bike NYC 2018; Taxi and Limousine Commission 2018a). Yellow taxis are the classic mode of for-hire transportation in New York, with a service area that is heavily concentrated in lower Manhattan. Green cabs, or “boro taxis”, were introduced in 2013 to offset this spatial disparity; they serve uptown Manhattan and the outer boroughs, and are prohibited from picking up street passengers in Manhattan south of West 110th Street or south of East 96th Street, or at the airport (Taxi and Limousine Commission 2018b). Finally, Citi Bike has operated as New York’s bike share alternative since 2013; as shown in Figure 1, its ridership is highly seasonal. Citi Bike’s docks are concentrated primarily in central Manhattan and western Brooklyn, though its service area continues to expand.

Although the growth in ridesharing has affected all parts of the city, the shift in mode share has been most dramatic in the outer boroughs. According to our data, by 2018 both Uber and Lyft were providing more average daily rides in the outer boroughs (combined) than within Manhattan. In non-central zones – such as in the Bronx and the far reaches of Queens and Brooklyn – daily ridesharing use can be one or even two orders of magnitude higher than taxi use. This reflects the expansion of ridesharing in virtually every taxi zone in the city over time as the distribution of taxi ridership has remained geographically concentrated.

Subway Disruptions as an Exogenous Shock

While there is a popular consensus that ridesharing has significantly damaged the taxi industry (Fischer-Baum and Bialik 2015), there is an ongoing debate on the extent to which ridesharing has influenced the use of the subway. Among monthly ridesharing users surveyed in New York’s 2017 Citywide Mobility Survey, 50% reported that ridesharing replaced public transport trips, while 43% reported that it replaced taxi or car services (NYC Department of Transportation 2017). However, only 35% of New Yorkers surveyed were ridesharing app members. Furthermore, when the survey sample as a whole was asked to characterize ridesharing and the subway, ridesharing was less often characterized as “convenient” (47 vs. 35%), “reliable” (23 vs. 19%), “fast” (23 vs. 15%), and “inexpensive” (23 vs. 8%). Both were perceived as comparably “safe” (15 vs. 13%), and ridesharing was more “comfortable” (14 vs. 20%).

Even with rich mode choice data, the causal identification of substitution between ridesharing and public transportation is challenging for two key reasons. First, ridesharing and public transit use may be correlated due to external factors, which may be measurable (e.g. average income, population density, or weather), or unmeasurable (e.g. neighborhood popularity). Therefore, estimating the impact of one mode on another requires controlling for these *omitted variables*. Second, there may be feedback between the two modes of transport. For example, ridesharing could attract riders away from the subway due to its convenience. However, this might lead to less crowding on the subway while street congestion worsens; as a result, riders might substitute back to public transportation. In this case, there is also a need to control for *simultaneity* between the two transport modes.

To address these problems, we sought to identify a source of exogenous variation affecting either subway or ridesharing use. The analysis described below focuses on subway service disruptions, examining cases in which no passengers enter or leave a station for at least four hours. Such disruptions clearly reduce subway ridership, and it seems credible that they affect the use of ridesharing only through their immediate impact on the usability of the subway. Since disruptions affect different parts of the subway system at different times, we are able to use panel data analysis to control for neighborhood- and time-specific variables which may drive ridership trends.

While subway disruptions offer an opportunity to explore the relationship between modes, they are also of critical policy relevance to cities like New York. Officially opened in 1904, the New York City subway system is one of the oldest in the world, and the decline of its antiquated infrastructure is a contentious political issue (Metropolitan Transportation Authority 2019). Currently, policy discussions are focused on the system's falling performance, and the decline in annual ridership that started in 2016.

Disruptions are caused by a number of different problems, including: crowding and passenger behavior, scheduled maintenance, and emergency repairs. Leader and Cafiero (2015) estimate that these factors accounted for 40%, 26%, and 22% of delays in 2014, respectively. For example, trash thrown on the tracks by passengers can cause fires, and trains can be held in the station due to sick passengers or police activity. Some conductors believe that the signals installed to control train speed may be mis-calibrated and over-sensitive, triggering a train's brakes even when it is not speeding; this leads them to slow down pre-emptively (Gordon 2018). Finally, some planned closures result from required upgrades to track infrastructure or the need to repair flood damage. Understanding whether taxi, ridesharing, and Citi Bike provide viable alternatives – and how much of the displaced ridership they are able to absorb – is of immediate concern in this context.

Although our approach of using subway disruptions to study ridesharing is new, there is a broader precedent for the use of natural experiments in the economics of transportation literature. For example, Davis (2008) studies the impact of alternate-day restrictions on license plates in Mexico city. Using a regression discontinuity design, he finds that the net effect of the policy on a variety of pollution indicators is negligible, and that the policy actually increases the number of vehicles in circulation. Anderson (2014) studies the effect of a public transit strike in Los Angeles. Using a regression discontinuity design, he finds that the effect of the strike on highway congestion is larger than predicted under previous models. Although neither is a direct analogy to our approach, these studies underscore the potential of natural experiments to answer policy questions involving large-scale, established, and dynamic systems in which true experimentation is difficult.

Hypotheses

We test three core hypotheses.

H1: In response to subway disruptions, ridership of taxi, ridesharing, and Citi Bike will increase.

While this hypothesis is straightforward, it is useful for confirming our basic intuition. There are three alternative possibilities: riders could forego travel altogether; riders could substitute to alternative modes such as walking, bus, and driving; and riders could substitute to *some* of the modes above, but not *all* of them.

H2: In response to subway disruptions, riders prefer ridesharing over taxi or Citi Bike.

There are two reasons to believe that ridesharing is well-suited to addressing short-term disruptions in public transit. First, Uber and Lyft's mobile apps may make it easier to connect with drivers during unexpected demand peaks — particularly in the outer boroughs, during rush hour, or late at night. Second, these platforms have access to centralized, real-time information on demand, making it easier for them to address unanticipated peaks by encouraging additional drivers to move to the area, for example through the use of surge pricing incentives.

However, ridesharing and public transportation may appeal to different individuals. For example, the cost of ridesharing may be prohibitive for people who rely most on the public system, a challenge that may only be exacerbated by the use of surge pricing. Ridesharing and public transportation may also be associated with different trip types: for example, individuals may prefer ridesharing for social trips, while they may rely on public transportation for commuting.

H3: The rate of substitution to alternative modes is limited by capacity constraints.

One concern with the use of alternative modes as described by H1 and H2 above is the possibility that these modes cannot fully absorb displaced demand. As discussed below, there are over 5,000 subway rides on average recorded per zone and four-hour period; none of the alternative modes described here surpasses even 10% of this ridership. The ability of taxis, ridesharing, and Citi Bike to respond to disruptions requires either that vehicles be available in disrupted zones, or that they can be quickly allocated to disrupted zones. Therefore, we examine how the excess demand resulting from subway disruptions compares to typical ridership in each zone.

Data and Analysis

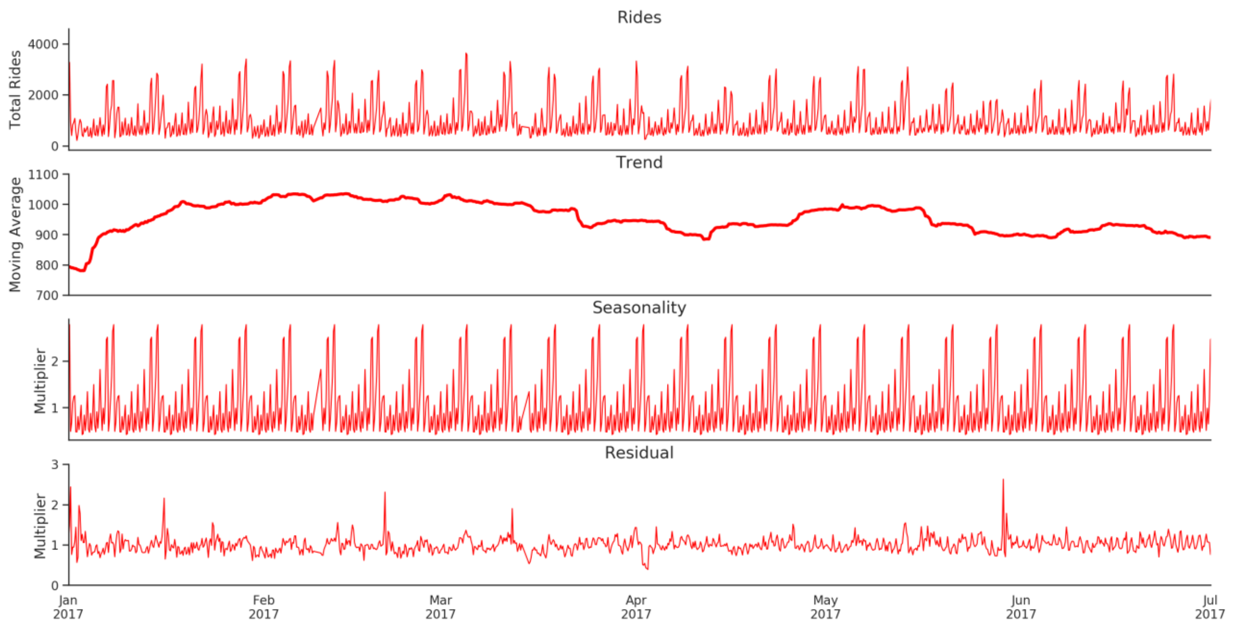
We compile a dataset of ridership for the six transport modes described above, covering New York City for a period of three and a half years. Raw data on subway entries and exits by turnstile and four-hour period was collected from the Metropolitan Transportation Authority (Metropolitan Transportation Authority 2018b). Data on taxi and ridesharing pickups by GPS location or taxi zone was collected from the Taxi and Limousine Commission (Taxi and Limousine Commission 2018a). Finally, data on bikeshare pickups by dock was collected from Citi Bike (Citi Bike NYC 2018).

To create a compiled dataset for analysis, we restrict all datasets to the range of dates for which Uber pickup counts were consistently available: January 2015 – June 2018. To match the level of aggregation in the Taxi and Limousine Commission (TLC) data, subway and Citi Bike stations are assigned to a taxi zone using their GPS coordinates, and rides are aggregated over these taxi zones. The dataset is then restricted to taxi zones served by the subway system. Furthermore, because the subway turnstile data is primarily read in four-hour intervals starting at midnight, all zone-level ride counts are aggregated over the intervals preceding 12am, 4am, 8am, 12pm, 4pm, and 8pm (inclusive). The final panel dataset consists of $n = 1{,}164{,}023$ observations representing the count of rides by mode in $i = 153$ taxi zones and $t = 7{,}608$ four-hour intervals.

Dependent Variables

Aggregate Ridership

Modeling transportation choices typically involves design decisions regarding: the level of aggregation, the type of data used, and the unit of analysis (Ben-Akiva 2008; Ben-Akiva and Lerman 1985). As described above, we model ridership using count data, aggregated into markets defined by geography and time. Our data reflects preferences revealed by the actual behavior of transport users. Our unit of analysis is effectively trip starts or pickups; we observe the location and time at which riders begin to use a given transport mode. We focus on trip starts due to two constraints in the data. First, the New York City subway system uses electronic Metro Cards that charge passengers upon entry, but are not required upon exit; as a result, subway entries and exits cannot be matched at the individual level. Similarly, our Uber and Lyft data does not contain comprehensive information on drop-off locations until mid-2017. Consequently, we are unable to systematically identify passenger destinations, or to distinguish a multimodal trip from several independent unimodal trips. Instead, we assume a simple underlying model of behavior in which a rider, attempting to use the subway and finding that it is out of service, makes a choice among the available alternatives for completing the current trip leg; we study this choice.



An example decomposition, showing Uber rides in the East Village (zone 79) over four-hour intervals in early 2017.

Figure 2. Example Time Series Decomposition

Adjusting Ridership Data for Trends and Seasonality

Raw ridership counts vary substantially across taxi zones and time. Seasonal fluctuations in ridership occur by hour of day, day of week, and month of year. Furthermore, the nature of seasonality may be different for different taxi zones; for example, due to commuting patterns, remote zones may see a large number of rides initiated in the morning, while central zones may see a large number of rides initiated in the evening. As shown in Figure 1, there are also long-term trends in the use of different modes over time, and short-term citywide shocks due to factors like holidays or bad weather.

To address this variability, we use time series decomposition to remove trends and seasonality from the ridership data for each taxi zone and mode. Since variance often increases with absolute ridership levels, we use a multiplicative decomposition. Specifically, for each for each taxi zone i and mode m, we separate ridership into three components:

$$\text{rides}_t \quad = \quad \text{trend}_t \times \text{seasonality}_{h(t),d(t)} \times \text{residual}_t \tag{1}$$

That is, we first estimate and remove a centered three-week moving average trend about time t. Using the detrended data, we next estimate and remove seasonality by calculating the average detrended ridership for the corresponding hour of day $h(t)$ and day of week $d(t)$ for the mode and zone over all periods. The result of this calculation is a series of residual changes in ridership that are not explained by trend or seasonal variations, which we refer to as *adjusted ridership*. An example of this decomposition can be found in Figure 2.

Adjusted ridership has a simple interpretation. If the value of adjusted ridership is 1 (i.e. $\text{residual}_t = 1$), then the number of rides in a zone is perfectly predicted using the zone's time trend and seasonality: $\text{rides}_t = \text{trend}_t \times \text{seasonality}_{h(t),d(t)}$. If adjusted ridership is 1.2, then the zone's ridership is 120% of the ridership expected based on trend and seasonality alone: $\text{rides}_t = \text{trend}_t \times \text{seasonality}_{h(t),d(t)} \times 1.2$. Similarly, if adjusted ridership is 0.8, then the zone's ridership is 80% of the ridership expected based on trend and seasonality alone: $\text{rides}_t = \text{trend}_t \times \text{seasonality}_{h(t),d(t)} \times 0.8$. Therefore, changes in adjusted ridership capture a given period's deviation from "normal" ridership in percentage terms.

We note briefly that adjusted ridership may be missing if there have been no rides for a given mode and zone in a three-week window ($\text{trend}_t = 0$), or if there are never any rides for a mode and zone in a given hour-of-day day-of-week bin ($\text{seasonality}_{h(t),d(t)} = 0$). In these cases, computing the residual involves dividing by zero and we drop such observations. We also note that the distribution of the adjusted ridership variable, while having a mean value of 1, is not symmetric; we therefore also drop occasional outliers with high adjusted ridership values (>10) to prevent these observations from having undue influence on our estimates. Since such cases are rare, we do not consider this a major limitation.

Independent Variables

Inferring Subway Disruptions from Activity Data

To test the hypothesis that ridesharing may act as a short-term substitute for public transportation, we next seek to identify subway service disruptions. A taxi zone is considered "disrupted" if one or more of its subway remote units (~ stations) experiences no turnstile entries or exits within a four-hour period (note that disrupted zones can still experience turnstile entries, if other stations in the zone are open).[1] When computing disruptions, we ignore any remote units that were continuously disrupted for more than two weeks, since we are interested in the impact of short-term changes in service availability. We also ignore any remote units that are disrupted only between 12:00am and 4:00am, to avoid accidentally labeling low-volume subway stations as "disrupted" in the early morning hours.

Using this definition, we identify 3,331 "incidents" in which a taxi zone was continuously disrupted. These incidents correspond to 12,914 disrupted zone-period observations (1.1% of all observations), and affect 110 of the 153 taxi zones. Brooklyn saw the highest average number of disrupted periods per zone, followed by Queens, the Bronx, and then Manhattan. Most disruption incidents (97%) lasted for two days or less, with 83% lasting one day or less, 67% lasting twelve hours or less, and 31% lasting for only four hours. Disrupted periods were most likely to fall on Saturday (26%) or Sunday (48%), which is consistent with the MTA's frequent use of weekends for planned repairs.

It is worth noting that this is a conservative definition of disruptions. As discussed, we have intentionally ignored some isolated late-night disruptions and long-term closures, although these types of service changes may impose a lot of hardship on riders. Assuming that these changes result in substitution towards alternative modes, failing to label these periods as disruptions should introduce an upward bias in our baseline ridership estimates for alternative modes. This should reduce our estimated impact of disruptions, making it *less* likely that we will find an effect.

At the same time, this definition of disruptions does restrict our findings to incidents that are severe enough to appear in the data. Because of the limitations in the resolution of the subway data, we ignore disruptions that do not span an entire four-hour period. Furthermore, we have not captured disruptions that affect only one of the trains serving a station – e.g. express lines that are disrupted even if local trains are running – or that affect only one direction of traffic – e.g. northbound trains that skip stations even when southbound trains provide full service. The impact of lower-level disruptions is an interesting area for future work.

Results

Summary statistics

An overview of the key variables in the dataset is shown in Table 1. On average, each zone experienced 308 yellow taxi rides, 54 green taxi rides, 202 Uber rides, 45 Lyft rides, 72 Citi Bike rides, and 5,223 subway rides in a four-hour period. As the median and maximum columns demonstrate, the distribution of ridership counts is highly skewed. For example, the median number of taxi rides per zone is 7 while the maximum is over 9,000. The median number of subway rides is just over 2,400, while the maximum exceeds 100,000. This motivates our use of adjusted and log dependent variables as described above.

---

[1] A turnstile refers to a set of spinning metal bars, or a revolving metal gate, which regulates entry into the subway system and ensures that only one rider can pass at a time. Riders swipe their fare card to unlock the turnstile and enter the system.

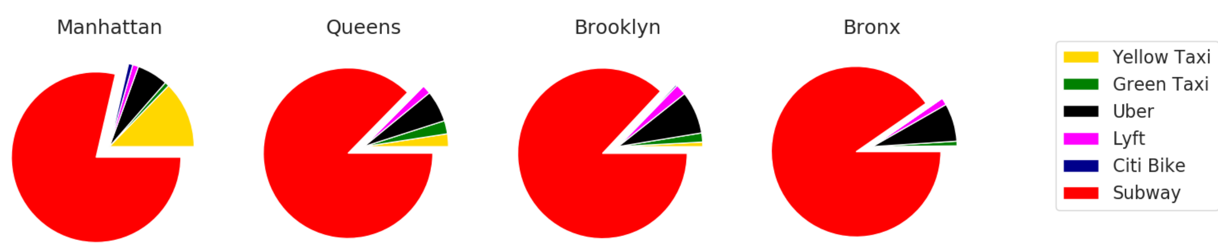
The right panel of Table 1 illustrates borough-level differences in ridership. Manhattan taxi zones experience the highest average volume of rides across all observed modes, while zones in the Bronx experience the lowest volume. There are also notable differences in mode preferences across boroughs, illustrated graphically in Figure 3. Specifically, we calculate ridership by mode as the share of *total* observed rides for a given zone and period, averaged over all records in our dataset. We can see that yellow taxi and Citi Bike have a higher average market share in Manhattan than in the outer boroughs (12.7% vs. 0.1–2.4% for yellow taxi, and 0.8% vs. 0.1%–0.3% for Citi Bike). In contrast, the remaining modes – green taxi, Uber, Lyft, and the subway – have equal or higher market shares in the outer boroughs (0.9—2.5% vs. 0.8% for green taxi; 6.0—8.1% vs. 5.9% for Uber; 1.4—2.1% vs. 1.1% for Lyft; and 87.0–90.4% vs. 78.7% for the subway).

Note also that the count of subway rides in our sample is significantly higher than the count of rides with all other modes combined. The remaining modes (taxi, ridesharing, and Citi Bike) together make up only 10–21% of observed rides per zone and four-hour period. Finally, we note that due to the absence of high-quality data, we exclude walking, buses, and driving from our definition of the "market". According to estimates from the Citywide Mobility Survey, walking, buses, and cars together account for 68% of all trips citywide (NYC Department of Transportation 2017). Incorporating these modes would further reduce the estimated shares of taxi, ridesharing, and Citi Bike as subway alternatives.

Table 1. Summary Statistics

	Pooled Sample				Manhattan		Queens		Brooklyn		Bronx	
	Mean	Std. Dev.	Median	Max	Mean	Share	Mean	Share	Mean	Share	Mean	Share
Yellow Taxi	308	690	7	9,109	904	12.7%	69	2.4%	18	0.9%	1	0.1%
Green Taxi	54	110	10	2,154	161	0.8%	68	2.5%	48	1.7%	10	0.9%
Uber	202	271	103	4,428	347	5.9%	125	6.0%	178	8.1%	70	7.2%
Lyft	45	69	21	1,598	61	1.1%	37	1.7%	53	2.1%	15	1.4%
Citi Bike	72	106	31	1,656	94	0.8%	13	0.1%	36	0.3%	.	.
Subway	5,223	8,227	2,462	107,589	8,830	78.7%	4,104	87.4%	3,975	87.0%	2,295	90.4%
Disruption	0.011	0.1	0	1	0.006	.	0.010	.	0.020	.	0.007	.

Mean ridership is calculated over all zones and four-hour periods in the dataset. Market share was first *calculated for each zone and four-hour period, and* then *averaged across zones and four-hour periods. For this reason the ordering of borough means does not always match the ordering of borough shares.*



The values corresponding to the slices in the chart are included in the "share" columns of Table 1 *above.*

Figure 3. Average Market Share by Mode in the Compiled Dataset

Baseline specification

Our baseline specification is a difference-in-differences model that predicts adjusted ridership using subway service disruptions and a set of period fixed effects. As described in the previous section, we first adjust our ridership counts to account for the substantial heterogeneity in ridership patterns across geographic regions, modes, and times. We explored numerous alternative approaches – including the use of a log dependent variable with a comprehensive set of fixed effects (discussed in further detail below), a ranked dependent variable, a count dependent variable, and the use of matching strategies. We have chosen to focus on our current approach because of its simplicity and flexibility. With the use of a de-trended and de-seasonalized dependent variable, we are better able to interpret our estimated effects of disruptions as deviations from "normal" ridership patterns.

Specifically, for each mode m we (separately) estimate the following model:

$$\text{adjusted}_{it} \quad = \quad \beta \, \text{disruption}_{it} \quad + \quad \gamma_t \quad + \quad \epsilon_{it} \tag{2}$$

where adjusted_{it} represents adjusted ridership for zone i in period t; disruption_{it} is a binary indicator of whether zone i suffered a subway disruption in period t (independent of the mode); γ_t is a set of citywide fixed effects for each four-hour period in our dataset; and ϵ_{it} is an error term clustered at the taxi zone level. Note that zone-level fixed effects have not been included because they are effectively accounted for in the calculation of the adjusted dependent variable, which removed moving average ridership per zone. The coefficient of interest is β, the estimated impact of a disruption on the adjusted ridership of mode m for a typical taxi zone and four-hour period. We estimate this model using the adjusted ridership of yellow taxi, green taxi, Uber, Lyft, Citi Bike, and the subway as dependent variables.

The results from estimating this model are shown in Table 2. These estimates support the hypothesis of substitution: disruptions are associated with a significant 8.2% increase in yellow taxi ridership, a 7.0% increase in green taxi ridership, a 2.8% increase in Uber ridership, a 3.3% increase in Lyft ridership, and no significant change in Citi Bike ridership. For comparison, we also fit the model to the subway data, and estimate that subway ridership is on average 29% lower than usual when at least one station in a taxi zone is disrupted. This latter regression does not have a causal interpretation, since disruptions are *defined* in terms of low subway ridership in at least one station in the zone. However, it serves as an intuitive check that the model works as expected, and provides an estimate of the extent of displaced ridership.

These coefficients allow us to consider the notion of *proportionate substitution*, i.e. the question of whether displaced subway riders choose alternatives with a probability proportional to these alternatives' pre-existing market shares (Train 2009). If rider behavior is characterized by proportionate substitution, then we would expect the disruption coefficients to be equal across all modes (since the coefficients represent percent change relative to typical ridership). If, instead, displaced subway riders have a specific preference for Uber or Lyft, we would expect to see larger coefficients on the disruption variable for these modes. Our estimates show no evidence of such a preference for Uber and Lyft during periods of disruption; if anything, ridesharing sees a *smaller* estimated percentage increase in ridership than taxis during these periods.

Alternative specification

To ensure that the directions of our results are not specific to our particular baseline specification, we also present results using log ridership counts as the dependent variable. Since we do not de-trend or de-seasonalize the log ride count variable, we include a richer set of fixed effects than in the baseline specification. For each mode m, we (separately) estimate the model:

$$\log(1 + r_{it}) \quad = \quad \beta \, \text{disruption}_{it} \quad + \quad \gamma_t \quad + \quad \alpha^{(1)}_{i,h(t),d(t)} \quad + \quad \alpha^{(2)}_{i,h(t),d(t)} \, \text{time}_t \quad + \quad \epsilon_{it} \tag{3}$$

where β, disruption_{it}, γ_t, and ϵ_{it} are as above; r_{it} represents ridership for zone i in time t; $\alpha^{(1)}_{i,h(t),d(t)}$ is a set of zone-specific fixed effects for each four-hour period $h(t)$ and day of the week $d(t)$;[2] time_t is a linear

[2] For example, one dimension of $\alpha^{(1)}_{i,h(t),d(t)}$ could represent "zone 79 on a Monday from 12-4am."

time trend; and $\alpha^{(2)}_{i,h(t),d(t)}$ is a set of coefficients that allow the slope of the linear trend to vary for each zone, hour, and day-of-week combination. In other words, $\alpha^{(2)}_{i,h(t),d(t)}\text{time}_t$ represents a set of time-varying zone fixed effects with intercepts $\alpha^{(1)}_{i,h(t),d(t)}$. This set of fixed effects was chosen because long-term trends in ridership may differ across zones. For example, a given mode might gain riders in some zones, and lose riders in other zones, with the passage of time.

Table 2. Estimated Impact of Disruptions on Adjusted Ridership

	Adj. Yellow Taxi	Adj. Green Taxi	Adj. Uber	Adj. Lyft	Adj. Citi Bike	Adj. Subway
Disruption	0.082***	0.070***	0.028***	0.033***	0.001	-0.290***
	(0.014)	(0.015)	(0.006)	(0.006)	(0.009)	(0.044)
Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.035	0.08	0.189	0.112	0.281	0.186
N Zones	153	115	153	153	69	153
N Observations	1,136,720	852,531	1,144,732	1,064,103	414,406	1,144,710
Mean Dep. Variable	0.96	0.98	1.00	1.00	1.00	1.00
Average Marginal Effect	25.4	3.9	5.5	1.5	0.0	-1511.8

** p<0.1, ** p<0.05, *** p<0.01. Standard errors in parentheses. Fixed effects were included for each four-hour period in the dataset. Errors were clustered by taxi zone ID.*

Table 3. Estimated Impact of Disruptions on Log Ridership

	Log Yellow Taxi	Log Green Taxi	Log Uber	Log Lyft	Log Citi Bike	Log Subway
Disruption	0.066***	0.076***	0.026***	0.044***	0.045	-1.188***
	(0.014)	(0.02)	(0.009)	(0.011)	(0.037)	(0.219)
Period FE	Yes	Yes	Yes	Yes	Yes	Yes
Zone-day-hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Time × Zone-day-hour FE	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-squared	0.981	0.951	0.974	0.942	0.942	0.91
N Zones	153	115	153	153	70	153
N Observations	1,164,023	874,919	1,164,023	1,084,157	423,314	1,164,023
Mean Dep. Variable	2.87	2.57	4.43	2.96	3.29	7.67
Average Marginal Effect	21.8	4.6	10.3	2.0		-3593.0

** p<0.1, ** p<0.05, *** p<0.01. Standard errors in parentheses. Fixed effects were included for each four-hour period in the dataset; time-varying fixed effects were also included for taxi zone ID/day of week/hour of day bins. Errors were clustered by taxi zone ID.*

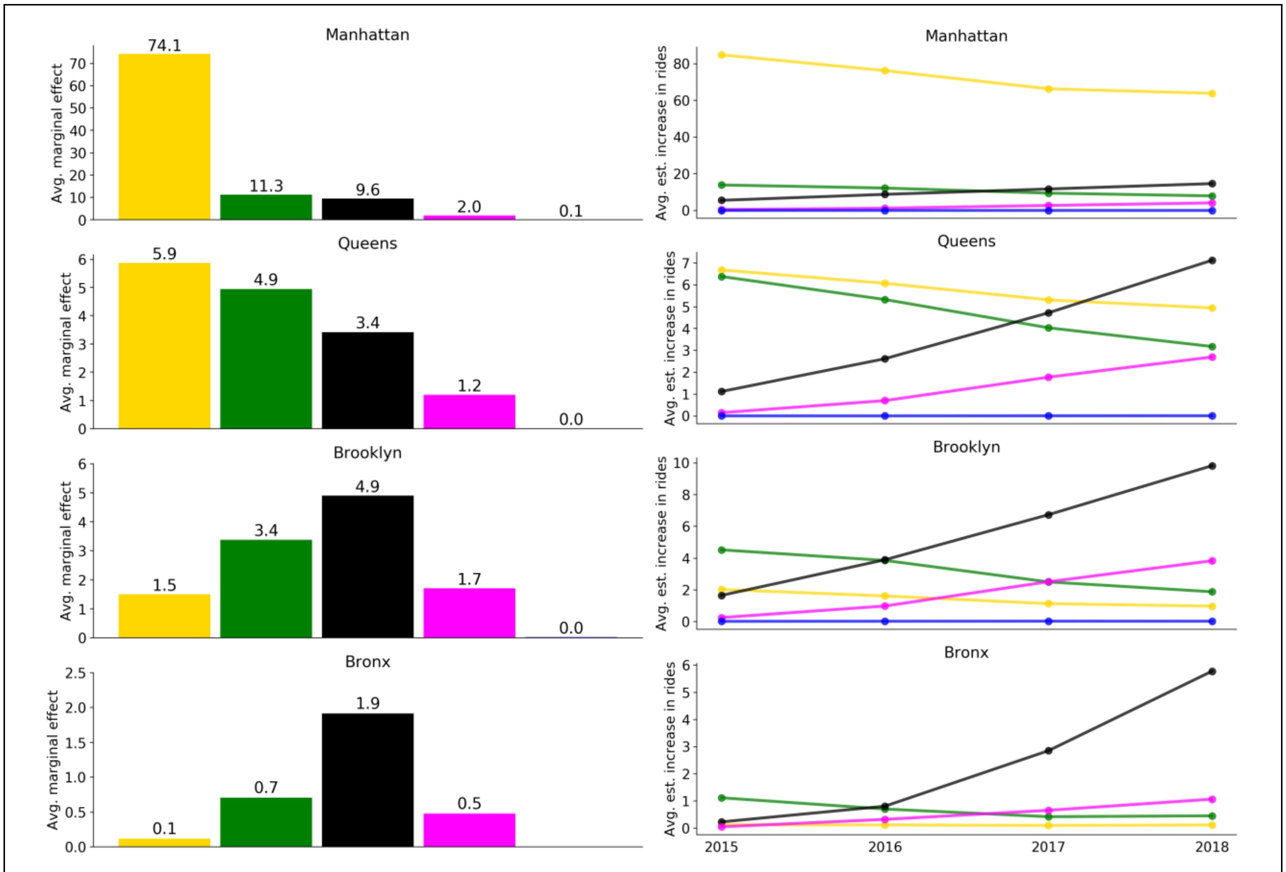


Figure 4. Average Marginal Effects, Disaggregated by Borough and Time

The results, shown in Table 3, are consistent with our baseline specification. We estimate that disruptions are associated with an approximate 6.6% increase in yellow taxi ridership, a 7.6% increase in green taxi ridership, a 2.6% increase in Uber ridership, a 4.4% increase in Lyft ridership, and no significant change in Citi Bike ridership. This model is less flexible than the preferred model, since it assumes a fixed linear time trend in ridership for each zone, day of week, and four-hour period rather than a moving average trend. However, both specifications are similar enough that the alignment of results is encouraging.

Discussion and Policy Implications

Thus far, our results support the hypothesis that taxi and ridesharing act as substitutes for public transportation during periods of subway service disruptions, and contradict the notion that ridesharing might be a "preferred" substitute during these periods. However, the impacts presented in percentage terms do not give a complete picture of the practical significance of this substitution behavior. In this section, we express the magnitude of substitution behavior in three ways. First, we convert percent changes in ridership to average marginal effects in rides. Second, we express the changes in taxi, ridesharing, and Citi Bike ridership relative to the total number of subway rides displaced. Finally, we express the changes in ridership relative to each mode's total estimated capacity to absorb displaced subway riders.

Marginal Effects

We first convert our estimates from relative to absolute changes in ridership, in order to better understand the total number of riders affected. We generate estimates for adjusted ridership by mode with and without disruptions using Equation (2) and the fitted coefficients from Table 2. Then, we multiply these estimates by the trend and seasonality estimates that were previously used to detrend and deseasonalize the raw ridership variable, as described in Equation (1). This allows us to calculate average estimated differences in rides by mode with and without disruptions. That is, for each mode m we calculate:

$$\begin{aligned} \text{marginal effect}_{it} \quad = \quad & (\ \mathbb{E}\ [\ \text{adjusted}_{it} \mid \text{disruption}_{it} = 1\ ,\ \gamma_t\ ] \\ & -\ \mathbb{E}\ [\ \text{adjusted}_{it} \mid \text{disruption}_{it} = 0\ ,\ \gamma_t\ ]\ ) \\ & \times \quad \text{trend}_{it} \quad \times \quad \text{seasonality}_{i,h(t),d(t)} \end{aligned} \tag{4}$$

The bottom row of Table 2 shows the estimated average marginal effect of disruptions citywide. According to our model, disruptions are associated with approximately 25.4 more yellow taxi rides, 3.9 more green taxi rides, 5.5 more Uber rides, 1.5 more Lyft rides, and no more Citi Bike rides, on average. Across the city as a whole, therefore, our estimates suggest that taxis absorb the largest number of displaced rides.

However, if we recalculate these average marginal effects by borough, a more nuanced narrative emerges. As shown in the left panel of Figure 4, in Manhattan alone we estimate that disrupted zones see over 70 additional yellow taxi rides per four-hour period, making yellow taxi the dominant alternative mode. In the outer boroughs, we estimate that green taxi, Uber, and Lyft play much larger roles. In Brooklyn and the Bronx, the estimated number of additional rides due to disruption for each of these three alternative modes surpasses the estimated change in yellow taxi ridership. If we further disaggregate these averages by year, as shown in the right panel of Figure 4, our estimates suggest that Uber and Lyft are playing an increasingly important role in absorbing displaced subway ridership as they gain market share over time, particularly in the outer boroughs.

Effects as a fraction of subway ridership

In addition to interpreting our estimated effects in absolute terms, it is also interesting to consider these effects as a fraction of displaced subway ridership. Therefore, for each mode m and borough b we calculate:

$$\frac{1}{|\mathcal{I}_b||\mathcal{T}|} \sum_{i \in \mathcal{I}_b} \sum_{t \in \mathcal{T}} \frac{\Delta_{imt}}{|\Delta_{ist}|} \tag{5}$$

where $\mathcal{I}_b$ is the set of all taxi zones in borough b, $\mathcal{T}$ is the set of all four-hour time periods, s represents the subway, and Δ represents the estimated change in rides due to disruption. That is, we calculate the average change in rides for mode m as a fraction of the average change in subway rides for the given time period and borough. The results are shown in Figure 5. Comparing across modes and boroughs, we can see that the share of yellow taxi as an alternative falls in the outer boroughs. Aside from yellow taxi, the most important substitutes appear to be Uber, followed by green taxi, Lyft, and finally Citi Bike.

Across all modes and boroughs, however, the estimated proportion of displaced ridership absorbed is small. According to our estimates, yellow taxi absorbs an average of 0–6.9% of rides, green taxi absorbs 0.2–0.8% of rides, Uber absorbs 0.9–1.1% of rides, Lyft absorbs 0.2–0.4% of rides, and Citi Bike absorbs less than 0.1% of rides. Therefore, while disruptions may drive changes in taxi, ridesharing, and Citi Bike ridership that are meaningful relative to each mode's average number of rides, these modes do not appear to be the preferred subway alternative for displaced riders. It seems likely that many riders choose modes that are omitted from our dataset (walking, buses, or cars), substitute to other subway stations in nearby zones, or forego their trips altogether.

Effects Relative to Mode Capacity

Finally, we consider the size of our estimated effects relative to each mode's total capacity. For each mode m and taxi zone i, we define *capacity* to be the maximum ridership for a given year, four-hour period of the day, and day of week. This approach is taken because the availability of different modes may vary over time and space. Intuitively, the number of Uber cars on the road likely changed from 2015 to 2018; the number of Citi Bikes available in a dock might be different at 4-8pm on a Friday relative to 4-8am on a Sunday; and the number of yellow taxis in circulation might be different in Times Square relative to Far Rockaway.

Given these estimates, we define utilization as the ratio between the observed ridership and the total capacity for zone i, mode m, and period t:

$$utilization_{imt} = \frac{r_{imt}}{\max_{\mathcal{T}_t} r_{imt}} \tag{6}$$

where $\mathcal{T}_t$ represents the set of all time periods that share the same hour of day, day of week, and year as time t and r_{imt} represents ridership. Intuitively, we ask: in Flushing, from 8am to 12pm on a *given* Monday in 2016, how does ridership compare to the maximum ridership ever observed in Flushing from 8am to 12pm on *any* Monday in 2016?

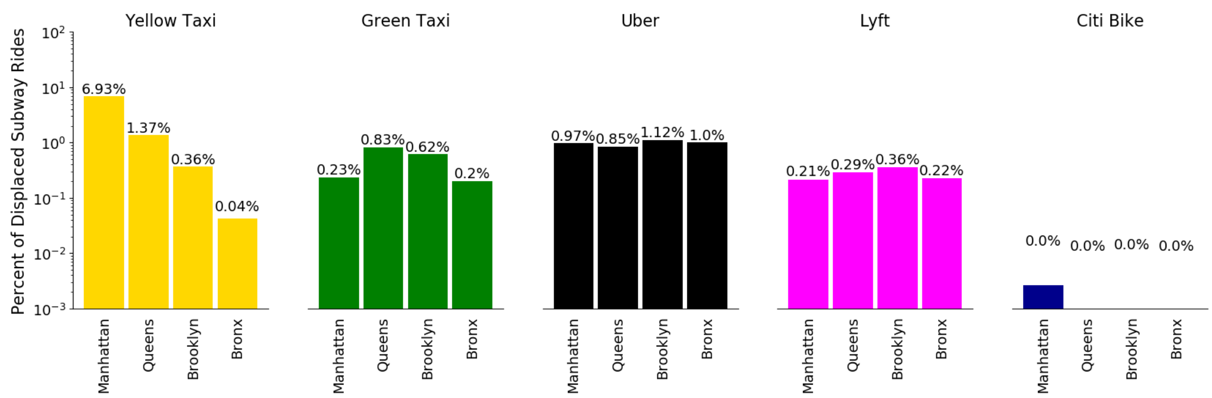
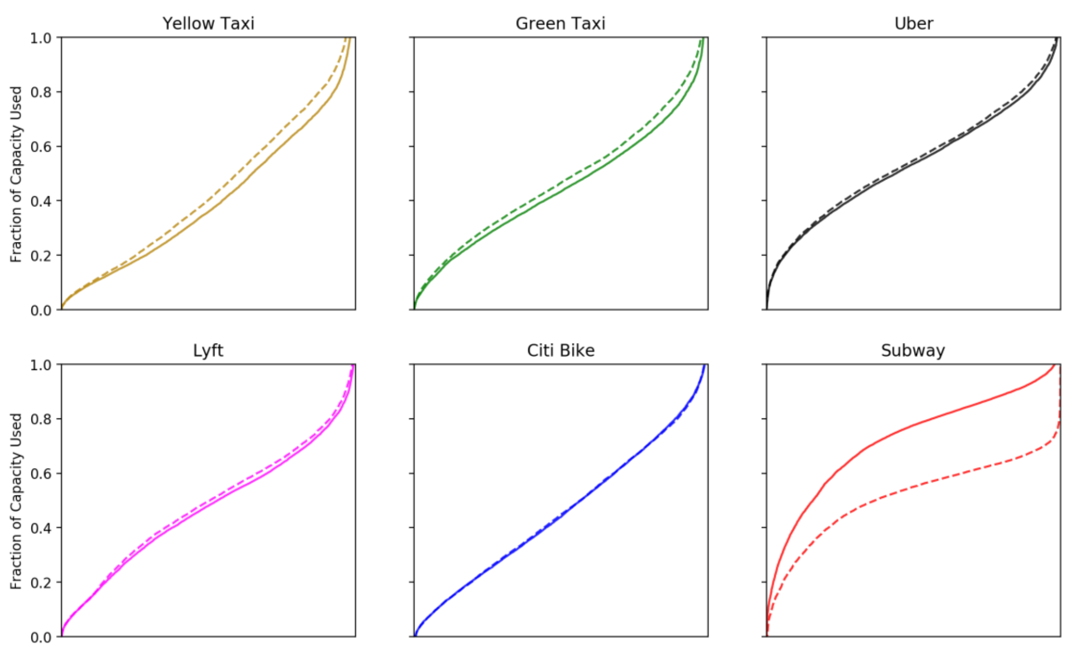


Figure 5. Estimated Effects as a Percentage of Displaced Subway Ridership (Log Scale)



The solid line represents estimates without disruptions, whereas the dotted line represents estimates with disruptions. The x-axis is a sorted random sample from all taxi zones and periods, so that the graphs represent the distribution of utilization in our sample.

Figure 6. Average Utilization With and Without Disruptions

Using this framework, we predict ridership with and without subway disruptions and estimate the corresponding utilization rates for all zones and periods. Figure 6 shows these estimated utilizations, plotted for a sorted random sample of zones and times. The space below the diagonal represents the distribution of utilized capacity across sampled times and zones, whereas the space above the diagonal represents the distribution of remaining capacity across sampled times and zones. The vertical distance between the dotted and horizontal lines represents how much predicted utilization differs for the given zone and time, with and without disruptions.

From this analysis, we can conclude that all modes generally have excess capacity to absorb riders. The subway is the mode operating closest to its capacity limit, whereas it appears that taxis, ridesharing, and Citi Bike have considerable bandwidth to take on additional riders. During periods of disruption, we see that utilization is slightly higher for taxis and ridesharing. However, utilization generally does not approach the estimated capacity limits for a given zone and mode, suggesting that the low levels of observed substitution do not result from a capacity constraint. Furthermore, we can see that for zones and times where utilization is low, we do not estimate larger substitution effects relative to total capacity. Therefore, it does not seem that capacity is a key limiting factor for substitution behavior.

Conclusions

To our knowledge, this research represents the first attempt to pair ridership data from digital platform-based transit and traditional mass transit to investigate substitution at the individual ride level. Using subway disruptions as a natural experiment, we find that a decrease in public transportation availability is partially offset by an increase in taxi and ridesharing use, at least in the short run and in response to subway system shocks. We find no significant citywide effect on Citi Bike use. This finding has implications for cities around the world, as New York is just one of many urban areas struggling with the limited capacity of its existing public transit system and the declining condition of its infrastructure. As cities consider large-scale, long-term investments in transport system upgrades and repairs, platform-based solutions may offer a flexible and responsive strategy for mitigating short-term inefficiencies.

However, our results suggest that ridesharing is not a panacea. By observing behavior in extreme scenarios — i.e. short-run disruptions in which a given subway station sees no entries for four hours or more — we attempt to identify the number of public transport users who, in the limit, can be convinced on short notice to switch from one mode to another. Somewhat surprisingly, our estimated magnitude of substitution is modest. Our analysis of other modes — i.e. yellow and green taxi — suggests that substitution away from the subway does not favor ridesharing in particular, perhaps because of the immediacy of the availability of these street-hail options.

Furthermore, taxi, ridesharing, and Citi Bike absorb only a small fraction of displaced subway riders, and they appear to have excess capacity to absorb more. This suggests that these alternative modes have the potential to play a larger role as “invisible infrastructure” supporting the public transit system. A platform-based approach of this kind could yield benefits across a range of city settings, from short-term accommodation and parking space demand spikes during major events to disaster recovery and coordination. Since multiple platforms compete to provide their services in most major cities globally, a city wishing to encourage or expand these alternatives may have to adopt strategies that enable multiple competing vendors to cooperate and coordinate (Bapna et al., 2010).

Our analysis also suggests that on-demand digital ridesharing platforms may not be perfect substitutes for public transit, perhaps due to differences in price or travel time. While the MTA’s fears that ridesharing is stealing subway ridership may be valid, our results indicate that many riders do not voluntarily substitute, even when the subway is no longer available.

While these findings lay the groundwork for further study, our work has several potential limitations. Although we treat subway disruptions as a natural experiment for the study of short-term substitution behavior, disruptions are not a wholly exogenous shock. Planned disruptions may be optimally scheduled when subway ridership is low (and on-demand ridesharing use is high) in order to minimize system impact, a concern we have tried to overcome through the use of period fixed effects. Disruptions also affect some zones more than others, a concern we have tried to overcome by normalizing the dependent variable using zone-level average ridership. Nevertheless, there is still a risk that unplanned disruptions are correlated with omitted variables (e.g. subway system crowding) that could also drive ridesharing use.

Furthermore, Uber and Lyft'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). Both companies use dynamic pricing that could artificially reduce the viability of ridesharing as a substitute in periods of high demand. Since promotions and dynamic pricing are core components of the ridesharing business model, we do not adjust our estimates to account for them; in order to understand substitution behavior "as-is", we take these practices as given.

Ultimately, we have chosen to examine one facet of the complex and dynamic interaction between transport modes across the city, and future work could make use of additional data to extend this research in three principal directions. First, with access to richer information on subway service changes, a more nuanced definition of subway service disruptions could be constructed. In this case, it might be possible to contrast the impact of planned and unplanned disruptions; estimate varying impacts of disruptions according to their severity; and examine long-run responses to chronic declines in subway service quality. Second, with better estimates of the total number of riders in a taxi zone in any given period, it might be possible to construct more complete market share estimates that could be used as inputs to a discrete choice model. This would allow for an exploration of how factors such as price and travel time affect the choice of alternative modes. Finally, with better information on Uber and Lyft's pricing and promotion strategies, it might be possible to directly account for price endogeneity in such a model.

We hope to lay a data-driven foundation to better understand how on-demand ridesharing alternatives substitute and complement existing and future capital-intensive transit systems. Taken together, our results provide 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

Anderson, M. L. 2014. "Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion," *American Economic Review* (104:9), pp. 2763–2796. (https://doi.org/10.1257/aer.104.9.2763).

Babar, Y., and Burtch, G. 2017. "Examining the Impact of Ridehailing Services on Public Transit Use," SSRN Scholarly Paper No. 3042805, Rochester, NY: Social Science Research Network, September 25. (https://papers.ssrn.com/abstract=3042805).

Ben-Akiva, M. E. 2008. *Travel Demand Modeling*, presented at the Transportation Systems Analysis: Demand & Economics, Massachusetts Institute of Technology. (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.

Citi Bike NYC. 2018. "Citi Bike System Data." (http://www.citibikenyc.com/system-data).

Clewlow, R. R., and Mishra, G. S. 2017. "Disruptive Transportation: The Adoption, Utilization, and Impacts of Ride-Hailing in the United States," No. UCD-ITS-RR-17-07, Davis, CA: UC Davis Institute of Transportation Studies, October.

Davis, L. W. 2008. "The Effect of Driving Restrictions on Air Quality in Mexico City," *Journal of Political Economy* (116:1), pp. 38–81. (https://doi.org/10.1086/529398).

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

Federal Transit Administration. 2019. "Monthly Module Raw Data Release," Text. (https://www.transit.dot.gov/ntd/data-product/monthly-module-raw-data-release).

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

Gordon, A. 2018. "'The Trains Are Slower Because They Slowed the Trains Down,'" *Village Voice*. (https://www.villagevoice.com/2018/03/13/the-trains-are-slower-because-they-slowed-the-trains-down/).

Hall, J. D., Palsson, C., and Price, J. 2018. "Is Uber a Substitute or Complement for Public Transit?" *Journal of Urban Economics* (108), pp. 36–50. (https://doi.org/10.1016/j.jue.2018.09.003).

Leader, J., and Cafiero, P. 2015. "NYCT Subway Performance," May. (http://web.mta.info/mta/news/books/docs/150518_SubwayPerformance.pdf).

Lyft. 2015. "Friends with Transit," *Lyft*. (http://take.lyft.com/friendswithtransit/).

Metropolitan Transportation Authority. 2018a. "The MTA Network: Public Transportation for the New York Region." (http://web.mta.info/mta/network.htm).

Metropolitan Transportation Authority. 2018b. "Turnstile Data." (http://web.mta.info/developers/turnstile.html).

Metropolitan Transportation Authority. 2019. "New York City Transit - History and Chronology." (http://web.mta.info/nyct/facts/ffhist.htm).

Nelson, E., and Sadowsky, N. 2017. *The Impact of Ride-Hailing Services on Public Transportation Use: A Discontinuity Regression Analysis*. (https://digitalcommons.bowdoin.edu/econpapers/13/).

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

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

NYC Department of Transportation. 2017. "Citywide Mobility Survey," NYC Department of Transportation, August. (http://www.nyc.gov/html/dot/downloads/pdf/nycdot-citywide-mobility-survey-report-2017.pdf).

NYC Department of Transportation. 2018. "New York City Mobility Report," NYC Department of Transportation, June. (http://www.nyc.gov/html/dot/downloads/pdf/mobility-report-2018-screen-optimized.pdf).

Taxi and Limousine Commission. 2018a. "TLC Trip Record Data." (http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml).

Taxi and Limousine Commission. 2018b. "Your Guide to Boro Taxis." (http://www.nyc.gov/html/tlc/html/passenger/shl_passenger.shtml).

Train, K. E. 2009. *Discrete Choice Methods with Simulation*, Cambridge University Press. (https://eml.berkeley.edu/books/choice2.html).

Walk Score. 2019. "Cities & Neighborhoods." (http://walkscore.com/cities-and-neighborhoods/).