Uber versus Trains? Worldwide Evidence from Transit Expansions*

Marco Gonzalez-Navarro† Jonathan D. Hall‡ Harrison Wheeler§ Rik Williams¶

October 5, 2022

Abstract

There is a contentious debate on whether ride-hailing complements or substitutes public transportation. We address this question using novel data and an innovative identification strategy. Our identification strategy relies on exogenous variation in local transit availability caused by rail expansions. Using proprietary, anonymized trip data from Uber for 35 countries, we use a dynamic difference-in-differences strategy to estimate how transit expansions affect local Uber ridership in 100 m distance bands centered on the new train station. Our estimates compare Uber ridership within a distance band before and after a train station opens relative to the next further out distance band. Total effects are obtained by aggregating relative effects at all further distance bands. We find that a new rail station opening increases Uber ridership within 100 m of the station by 60%, and that this effect decays to zero for distances beyond 300 m. This sharp test implies Uber and rail transit are complements.

*We are grateful for helpful feedback from Enrico Moretti. This work has received funding from the Social Sciences and Humanities Research Council of Canada [grant number 435-2019-0507], the European Union's Horizon 2020 research and innovation programme [grant number 101022491], the NET Institute, and via an interagency agreement between the US Department of Transportation and the National Science Foundation [award number 1559013].

†Department of Agricultural & Resource Economics, University of California Berkeley, 216 Giannini Hall, Berkeley, California, 94720, USA, marcog@berkeley.edu

‡Department of Economics and Munk School of Global Affairs and Public Policy, University of Toronto, 150 St. George Street, Toronto, Ontario M5S 3G7, Canada, jonathan.hall@utoronto.ca

§Department of Economics, University of California Berkeley, 530 Evans Hall #3880, Berkeley, California, 94720, USA, wheeler@berkeley.edu

¶Uber Technologies, Inc., 1515 3rd St., San Francisco, California 94158, USA, rik@uber.com

1 Introduction

The explosive growth of ride-hailing around the world has sparked an important debate about the repercussions this new mobility option is having on cities. Urban planners and policymakers around the world have grappled with how to regulate ride-hailing companies in their jurisdictions. Indeed, several countries have banned ride-hailing while others have heavily regulated it.[1] Of particular interest to urban planners and policy makers is whether ride-hailing technologies increase or decrease public transit ridership. Understanding the degree of complementarity or substitutability between ride-hailing and public transit is important for at least four reasons. First, reductions in transit ridership can potentially generate major budgetary shortfalls for transit authorities. Second, reductions in transit ridership likely have social welfare costs because transit ridership is inefficiently too low.[2] Third, reductions in transit ridership likely increase congestion and pollution.[3] Fourth, changes in transportation technologies, such as steam railways, the automobile, and limited-access highways, have repeatedly reshaped urban spatial structure.[4]

Because of its importance, determining the impact of ride-hailing on public transportation has attracted significant attention from researchers. In spite of this, there remains great uncertainty as existing estimates vary in sign and magnitude. Estimates range from as high as +5% after two years (Hall et al., 2018) to as low as -16% after four years (Diao et al., 2021).

Our contribution is to address the question of whether ride-hailing and public transportation are complements or substitutes using novel data and an innovative identification strategy. Using proprietary trip data from Uber, we use a dynamic difference-in-differences strategy to estimate how rail transit expansions from around the world affect local Uber ridership. We focus on rail transit due to the difficulty in documenting bus expansions. Our approach has several advantages. First, the timing of rail expansions are plausibly exogenous to underlying trends in Uber ridership. Rail expansions are planned years in advance, and indeed, many of those in our sample were initially planned before the existence of Uber. Second, our detailed and geographically precise data allows us to provide tests for the mechanisms by which transit and ride-hailing impact each other.

[1] Countries banning ride-hailing include Denmark, Hungary, and Bulgaria.

[2] Transit fares are typically above social marginal cost (though below average cost) due to economies of scale and density. Existing research shows that, given the existing set of transportation policies, increasing transit subsidies, and so increasing transit ridership, increases social welfare (e.g., Parry and Small, 2009, Basso and Silva, 2014).

[3] See, for example Anderson (2014) and Gendron-Carrier et al. (forthcoming).

[4] Heblich et al. (2020) shows how steam railways allowed London to double in population and Baum-Snow (2007) estimates that limited-access highways reduced central city population by 8%. Gorback (2019) finds ride-hailing is already affecting urban spatial structure, with UberX doubling restaurant net creation in previously inaccessible locations.

Third, our research design allows us to flexibly control for hyper-local and highly variable time trends in Uber ridership. Fourth, we are able to use data for 35 countries.

We use a dynamic difference-in-differences strategy to control for hyper-local trends in Uber ridership. Our novel approach exploits the high frequency and extremely granular Uber trip data as well as the sharp opening date for new transit stations. We compare the number of Uber trips in two adjacent distance bands around a new train station (for example 0–100 m and 100–200 m from a station) before and after a train station opens for service. While the further distance band plays the role of a local "control group," we expect that it is also affected by the new transit station opening. Thus our estimates at, say, 100—200 m are the effect of a new transit opening on Uber ridership at 100-–200 m relative to the effect of a new transit station opening at 200—300 m. We repeat this estimation strategy for adjacent distance bands up until 1200 m from transit stations. We find that relative treatment effects are indistinguishable from zero beyond 300 m. We obtain the total effect of a new train station on Uber ridership by summing up the relative effects at all distance bands.

Our test shows clear evidence that Uber and train service are highly complementary, as we observe large increases in ride hailing trips after a train station opens. Effects are concentrated within 300 meters of a station, and show no signs of decay after 6 months of train service. We also find that the average length of an Uber trip decreases after train service begins, consistent with the idea that ride-hailing in the presence of a rail station starts being used for last mile trips where both modes of transport are used.

This paper builds on a quickly growing literature seeking to determine whether ride-hailing and public transportation are complements or substitutes. Most papers in this literature use variation across US metropolitan areas in the timing of Uber entry to estimate the effect of ride-hailing on public transit. Hall et al. (2018) finds that ride-hailing complements the average transit agency, while Graehler et al. (2019), Erhardt et al. (2021), and Diao et al. (2021) find ride-hailing is a substitute. Nelson and Sadowsky (2018), Babar and Burtch (2020), and Cairncross et al. (2021) find mixed or statistically insignificant results. This paper takes a completely different approach by exploiting exogenous variation from the timing of train service starts to assess what happens to Uber trips in the face of a new train transit option. The paper also adds to the literature by including data from other countries, allowing for a rich heterogeneity analysis.

There is also a broader literature working to understand the effect of ride-hailing on cities. This includes understanding the impact of ride-hail on traffic safety (Greenwood et al., 2017, Burgdorf et al., 2019, Barrios et al., 2020, Barreto et al., 2020, Anderson and Davis, 2021), and the impacts of surge pricing (Castillo, 2020, Castillo et al., 2021).

2 Data

To investigate the effect of new transit stations on Uber ridership we require data describing new transit station locations and dates of opening as well as panel data on Uber ridership near these stations. We use data on Uber ridership constructed from Uber's trip database. Our data on new transit stations are the result of primary data collection. We describe these datasets and their construction below.

2.1 Uber ridership

We use Uber's trip database to calculate the number of trips starting or stopping within a given distance band from the transit station in a month, for example, all trips within 100–200 meters (m). We do this for 12 different 100 m bands from 0–100 m up to 1,100–1,200 m. To avoid double-counting, each Uber pickup or drop-off is allocated to the station to which it is closest. This means the distance bands are not always perfect circles. Our Uber trip data spans January 2012 to December 2018, but note that Uber was expanding in this time period so that not every city has trip data from 2012. This provides us with between 15 and 79 months of Uber trip data for each transit expansion, with a mean and median of 61 months. We exclude cities where Uber availability began, ended, or paused within six months of the rail expansion. We also excluded New York City and China.

2.2 Transit stations

To collect data on new transit stations, we start with a list of cities Uber operates in worldwide, and for each city, find all rail transit stations that opened after Uber entered the city. We obtain data on subway openings through 2017 from Gendron-Carrier et al. (forthcoming), and extend this dataset to include light rail and commuter rail, and update it through 2018 using online sources such as www.urbanrail.net, www.wikipedia.org, and news sites. We limit attention to rail stations as these are better documented than bus stops. For each station, we record opening date, latitude and longitude, station name, whether it is the terminal station, city, and country. We define the exact latitude and longitude of each station using Google Maps. We also find the locations of the pre-existing stations that had been the terminal stations before the transit expansion. Table 1 reports the 78 cities that expanded their rail transit between the date Uber entered and 2018. The table lists the date Uber entered, the number of expansions, number of new stations, type of service (subway, light rail, etc.), and the first and last date of the expansions. Figure 1 shows the time series of rail expansion events and number of stations opened, showing there is a notable drop in openings in January and February, but they are otherwise fairly uniform over time of year. Figure 2 plots the geographic distribution of cities used in our analysis, showing they are heavily concentrated in India and Southeast Asia, Europe, and North America.

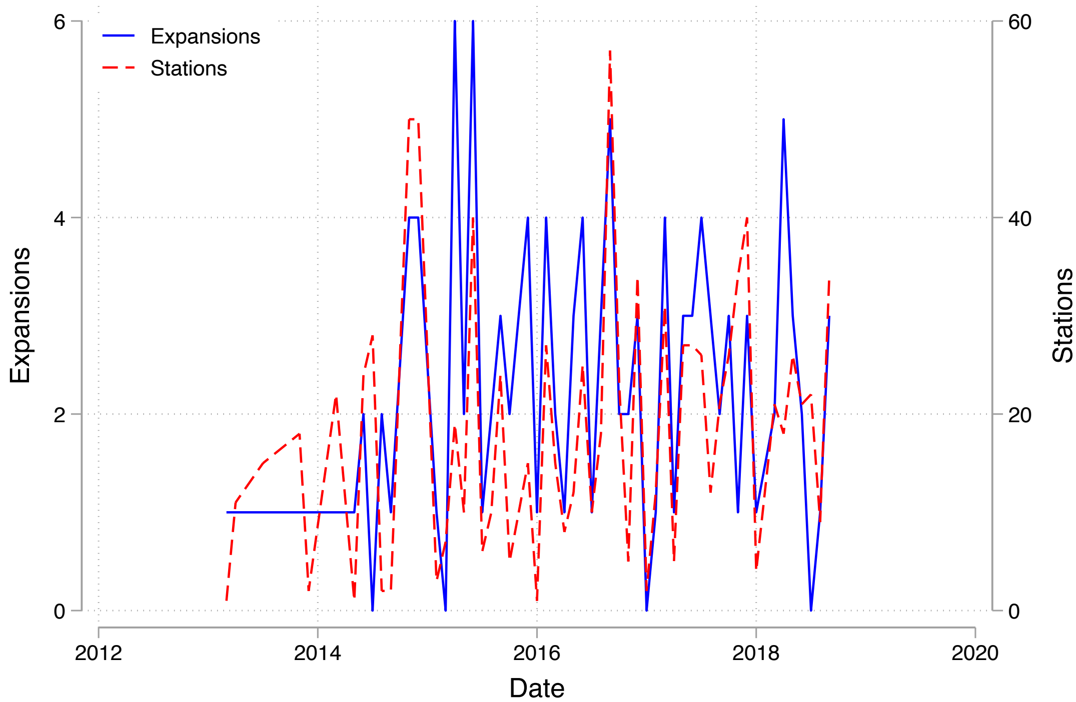

Figure 1: Transit expansions in our data over time

Our analysis requires us to observe Uber ridership at a transit station for sometime before and after the transit station opens. Thus, we face a trade-off between sample size and the length of time we observe each transit station. We focus on a 13-month window, six months before and after the station opening; excluding any transit stations for which we do not have this data.

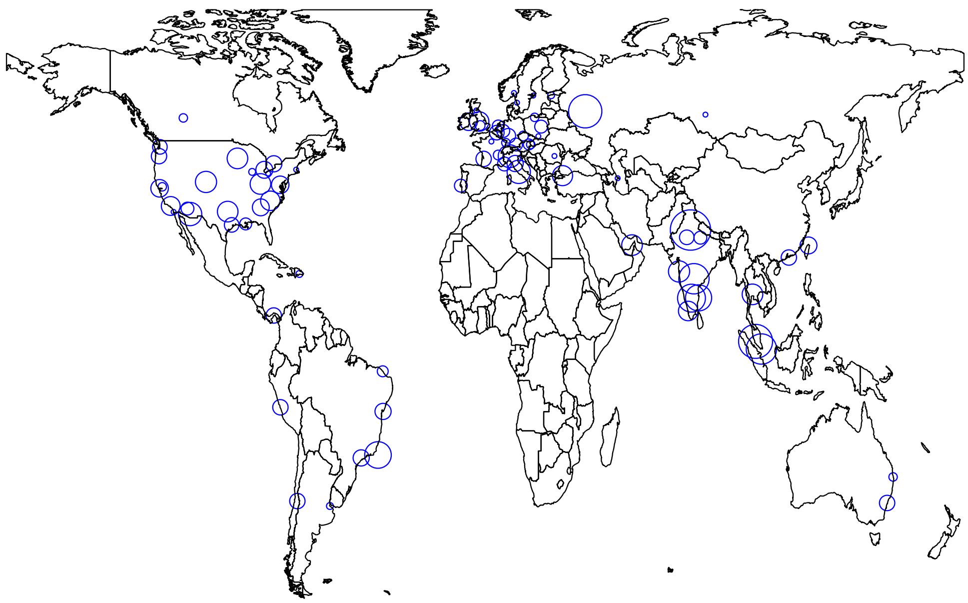

Figure 2: Locations of transit expansions in our data

Notes: This map plots the locations of new transit openings in our data. The size of each circle is proportional to the number of new stations that opened between Uber's entry date and 2018.

Table 1: City-level descriptive statistics

City	# Expansions	# Stations	# Monorail	# Light Rail	# Commuter Rail	# Subway	# Interacts w/ Traffic	Uber Entry	First Exp. Year	Last Exp. Yr
Amsterdam	0	6	0	0	0	6	0	2012	2018	2018
Atlanta	0	12	0	12	0	0	12	2012	2014	2014
Baku	0	1	0	0	0	1	0	2015	2016	2016
Bangalore	4	34	0	0	0	34	0	2013	2014	2017
Bangkok	3	18	0	0	0	18	0	2014	2016	2017
Birmingham, UK	1	4	0	4	0	0	4	2015	2015	2016
Bordeaux	1	9	0	9	0	0	9	2014	2015	2016
Boston	0	1	0	0	0	1	0	2012	2014	2014
Bratislava	0	3	0	3	0	0	2	2015	2016	2016
Brussels	1	11	0	11	0	0	11	2014	2018	2018
Bucharest	0	2	0	0	0	2	0	2015	2017	2017
Buenos Aires	0	1	0	0	0	1	0	2016	2018	2018
Charlotte	1	16	0	16	0	0	5	2013	2015	2018
Chennai	4	29	0	0	0	29	0	2014	2015	2018
Chicago	1	2	0	0	0	2	0	2012	2015	2017
Cincinnati	0	18	0	18	0	0	18	2014	2016	2016
Cleveland	0	1	0	0	0	1	0	2014	2015	2015
Dallas	4	15	0	15	0	0	9	2012	2014	2016
Delhi NCR	12	65	0	0	0	65	0	2013	2014	2018
Denver	1	19	0	19	0	0	4	2012	2013	2017
Detroit	0	12	0	12	0	0	12	2013	2017	2017
Dubai	4	18	0	10	0	8	0	2013	2014	2017
Dublin	0	13	0	13	0	0	9	2014	2017	2017
Dusseldorf	1	9	0	9	0	0	3	2014	2016	2018
Edinburgh	0	1	0	1	0	0	1	2015	2016	2016
Edmonton	0	3	0	3	0	0	0	2014	2015	2015
Florence	0	12	0	12	0	0	12	2015	2018	2018
Fortaleza	2	8	0	7	0	1	0	2016	2017	2018
Frankfurt	0	9	0	0	0	9	0	2014	2016	2016
Gold Coast	0	3	0	3	0	0	3	2014	2017	2017
Gothenburg	0	1	0	1	0	0	1	2014	2015	2015
Hong Kong	3	10	0	0	0	10	0	2014	2014	2016
Houston	2	9	0	9	0	0	9	2014	2015	2017
Hyderabad	1	39	0	0	0	39	0	2014	2017	2018
Istanbul	4	17	0	0	0	17	0	2014	2015	2017
Jaipur	0	9	0	0	0	9	0	2014	2015	2015
Kochi	1	16	0	0	0	16	0	2014	2017	2017
Kuala Lumpur	4	50	0	22	0	28	0	2013	2015	2017
Lisbon	1	11	0	10	0	1	10	2014	2016	2018
Los Angeles	1	13	0	13	0	0	3	2012	2016	2016
Lucknow	0	7	0	0	0	7	0	2016	2017	2017
Lyon	0	2	0	2	0	0	2	2013	2014	2014
Manchester	1	16	0	16	0	0	16	2014	2014	2015
Milan	7	16	0	6	0	10	6	2013	2014	2018
Minneapolis - St. Paul	0	18	0	18	0	0	18	2012	2014	2014
Moscow	12	42	0	0	0	42	0	2013	2014	2017
Munich	0	6	0	6	0	0	6	2013	2016	2016
New Jersey	0	1	0	0	1	0	0	2013	2016	2016
New Orleans	0	6	0	6	0	0	6	2014	2016	2016
Nice	0	12	0	12	0	0	12	2014	2018	2018
Novosibirsk	0	2	0	2	0	0	0	2015	2016	2016
Oslo	0	1	0	0	0	1	0	2014	2016	2016
Panama City, PA	1	2	0	0	0	2	0	2014	2015	2015
Paris	6	76	0	75	0	1	75	2012	2013	2017
Phoenix	1	7	0	7	0	0	7	2012	2015	2016
Portland	0	10	0	10	0	0	7	2014	2015	2015
Prague	0	4	0	0	0	4	0	2014	2015	2015
Rio de Janeiro	7	31	0	26	0	5	26	2014	2016	2017
Rome	2	22	0	0	0	22	0	2013	2014	2015
Sacramento	0	3	0	3	0	0	0	2013	2015	2015
Salvador	3	11	0	0	0	11	0	2016	2016	2018
San Diego	0	1	0	1	0	0	1	2012	2018	2018
San Francisco	4	15	0	0	11	4	0	2012	2014	2018
Santiago	0	10	0	0	0	10	0	2013	2017	2017
Santo Domingo	0	4	0	0	0	4	0	2015	2018	2018
Sao Paulo	8	15	4	0	0	11	0	2014	2017	2018
Seattle	2	14	0	14	0	0	11	2012	2016	2016
Singapore	8	39	0	6	0	33	0	2013	2013	2017
Stockholm	1	2	0	1	1	0	1	2013	2017	2017
Strasbourg	0	2	0	2	0	0	2	2015	2017	2017
Sydney	1	10	0	10	0	0	10	2012	2014	2015
Taipei	2	29	0	0	0	29	0	2013	2014	2017
Tallinn	0	2	0	2	0	0	2	2015	2017	2017
Toronto	2	12	0	2	4	6	2	2012	2015	2017
Tucson	0	23	0	23	0	0	23	2013	2014	2014
Vienna	0	5	0	0	0	5	0	2014	2017	2017
Warsaw	0	7	0	0	0	7	0	2014	2015	2015
Washington D.C.	1	13	0	8	0	5	8	2012	2014	2016

3 Methodology

Our main specification estimates the impact of a new transit station opening on Uber ridership using a dynamic difference-in-differences design that exploits our high-frequency data on Uber trips as well as the sharp opening date of new transit stations. The first difference compares Uber ridership in a given distance band before and after the transit station opens. The second difference adjusts for time trends in Uber ridership. Adjusting for time trends is vital because, over our study period, Uber is growing quickly and is not in a steady state. To address this challenge, we use Uber ridership in the next furthest distance band to adjust for hyper-local time trends in Uber ridership. We expect that Uber ridership in the next furthest distance band is also affected by the new transit station opening, and thus our estimates at, say, 100–200 m are of the effect of a new transit opening on Uber ridership at 100–200 m *relative* to the effect of a new transit station opening at 200–300 m. Should we find a distance where the treatment effect is zero, we can then find the total effect at 100–200 m by summing the relative effects at all further out distances.[5]

Let y_{dit} denote the inverse hyperbolic sine number of Uber trips starting or ending in distance band d around transit station i during month t.[6] Denote the month that station i opens as t_i' and define the time since the station opened ("relative time") as $\tau_{it} = t - t_i'$, noting that τ_{it} is negative before the station opens. In the month a station opens, $\tau_{it} = 0$; for stations that open at the start of the month, nearly the entire month is treated, while for stations that open at the end of the month, nearly the entire month is untreated.

To estimate the relative effect of a new transit station opening on Uber ridership in distance band $\tilde{d}$, we use data on distance band $\tilde{d}$ and the next distance band further out, $\tilde{d}+1$, and estimate the following dynamic difference-in-differences specification:

$$\begin{aligned} y_{dit} = \gamma_{it} + \delta_{di} + \sum_{j \in \{-6,-5,\ldots,6\} \setminus \{-2\}} \alpha_{\tilde{d}j} \times \mathbb{1}_{\tau_{it}=j} \times \mathbb{1}_{d=\tilde{d}} \\ + \left(\beta_{1,\tilde{d}} \times \mathbb{1}_{\tau_{it}<-6} + \beta_{2,\tilde{d}} \times \mathbb{1}_{\tau_{it}>6} \right) \mathbb{1}_{d=\tilde{d}} + \epsilon_{dit}, \\ \forall\ d \in \{\tilde{d}, \tilde{d}+1\}, \quad (1) \end{aligned}$$

where $\mathbb{1}_{\tau_{it}=j}$ is an indicator function for whether the given observation occurs j months

[5]This procedure can be viewed as a discrete approximation to the difference-in-differences style estimator of Diamond and McQuade (2019). Our regression approach estimates the gradient of the treatment effect with respect to distance from the train expansion using finite differences across 100 m bands. Difference-in-differences estimates at farther distances provide tests of whether the treatment effect eventually converges to zero with distance.

[6]We focus on this outcome for two reasons. First, the inverse hyperbolic sine transformation allows us to interpret the regression coefficients as approximately the percentage change in Uber trips. Second, farther rings mechanically have more Uber usage given their larger area. Differences in percents will thus be more relevant than level differences.

after the station opens, $\mathbb{1}_{d=\tilde{d}}$ is an indicator function for whether the given observation is at distance $\tilde{d}$, γ_{it} is a station-time fixed effect, and δ_{di} is a station-distance fixed effect. The coefficients of interest are $\alpha_{\tilde{d}j}$, which are the percentage changes in Uber trips j months after a new transit station opens, relative to distance band $\tilde{d}+1$. The index j runs from -6 to 6, excluding -2, so the second month prior to the transit station openings is the reference category.[7] For $j \geq 0$, α_j is the treatment effect we seek to measure. For $j < 0$, α_j allows us to test whether there are pre-trends in Uber trips before stations open. The coefficients $\beta_{1,d}$ and $\beta_{2,d}$ are event study coefficients for before and after our treatment window. This specification allows us to use all the data for each station (anywhere between 15–79 months) to precisely estimate station-by-distance fixed effects, while only using variation in Uber trips around the station opening date to estimate the effect of a station opening on Uber trips. We cluster our standard errors at the station level.

We often wish to aggregate the treatment effect to a single coefficient, which is helpful when reporting regression results across multiple specifications or distances in a single table or figure. We do so using the following difference-in-difference specification:

$$\begin{aligned} y_{dit} = \gamma_{it} + \delta_{di} + \alpha_{\tilde{d}} \times \mathbb{1}_{\tau_{it}\in\{1,\ldots,6\}} \times \mathbb{1}_{d=\tilde{d}} \\ + \left(\beta_{1,\tilde{d}} \times \mathbb{1}_{\tau_{it}<-6} + \beta_{2,\tilde{d}} \times \mathbb{1}_{\tau_{it}>6} + \beta_{3,\tilde{d}} \times \mathbb{1}_{\tau_{it}\in\{-2,-1,0\}}\right) \mathbb{1}_{d=\tilde{d}} + \epsilon_{dit}, \\ \forall\ d \in \{\tilde{d}, \tilde{d}+1\}, \quad (2) \end{aligned}$$

In this regression, the omitted category is the fourth through sixth months before the station opened and the coefficient of interest is $\alpha_{\tilde{d}}$. $\alpha_{\tilde{d}}$ captures the average percentage effect of a station opening on Uber trips at distance d relative to $d+1$, for months $\tau_{it} \in \{1, ..., 5\}$ relative to months $\tau_{it} \in \{-6, -5, -4\}$. We separately control for the effect two months prior and the month of the station opening. This specification mitigates concerns over the partial treatment of month $\tau_{it} = 0$ and soft openings which may have had an impact on Uber usage prior to the station's official open date.

We have two core identifying assumptions. First, we assume that local governments do not time the opening of new transit stations to coincide with a sharp break in Uber ridership. Indeed, Gendron-Carrier et al. (forthcoming) find that it typically takes 11 years between when the plan is approved for a new subway and the opening date. This means that the vast majority, and maybe even all, of the transit openings in our data were approved before Uber existed. Second, we make the standard difference-in-differences assumptions that Uber ridership in adjacent distance bands would have moved in parallel in the absence of a new transit station opening. A strength of our approach is the granularity of our

[7] We use the second month prior to the opening as the reference category to test for anticipation effects.

data, which uses adjacent distances to provide a more plausible counterfactual. We will present evidence that Uber ridership in adjacent bands moved in parallel prior to the station opening.

4 Results

We start by presenting our estimates of the marginal effect of transit on Uber ridership by distance. We then estimate the total effect, which city or expansion characteristics predict the magnitude of the effect of transit on Uber, and finally, present evidence on the mechanisms by which transit affects Uber.

4.1 Average relative effect

We start with estimating the average effect of a new transit station opening on Uber trips by the distance from a transit station. Figure 3 plots our estimates of the effect on trips within 0–100 m of a station, relative to the effect on trips within 100–200 m. It shows there is a large and statistically significant increase in ridership of 53% within a month of opening (0.43 log points). This effect is persistent for at least six months and does not show signs of decay. There is a smaller increase in the month the station opens (month 0) which could be occurring for two reasons: first, part of this is mechanical, as transit stations do not always open on the first day of the month, and so the month of opening is only partially treated, and second, it likely takes time for travelers to re-optimize their decisions of how to travel.

The figure also shows evidence in favor of our identification assumption. There are no pre-trends in Uber trips before the train station opens, and the increase in Uber ridership we observe only occurs after the train opens and not before.

The sharp increase in Uber ridership when the transit station opens is strong evidence in favor of the transit opening itself increasing Uber ridership, rather than an indirect effect mediated through restaurant openings and other commercial activities. The high frequency of our data allow us to rule out this alternative explanation given that neighborhood or commercial environment changes take years or even decades to occur. While a new restaurant or store may want to time their opening with the transit station opening, this is difficult to achieve.

Figure 4 plots how the average effect during months 1–6 of a new transit station on Uber ridership changes with distance. It shows that the a new transit station opening increases Uber ridership at 100–200 m and 200–300 m (relative to the next further distance band). However, beyond these distances there is no detectable effect. Appendix Figure A1

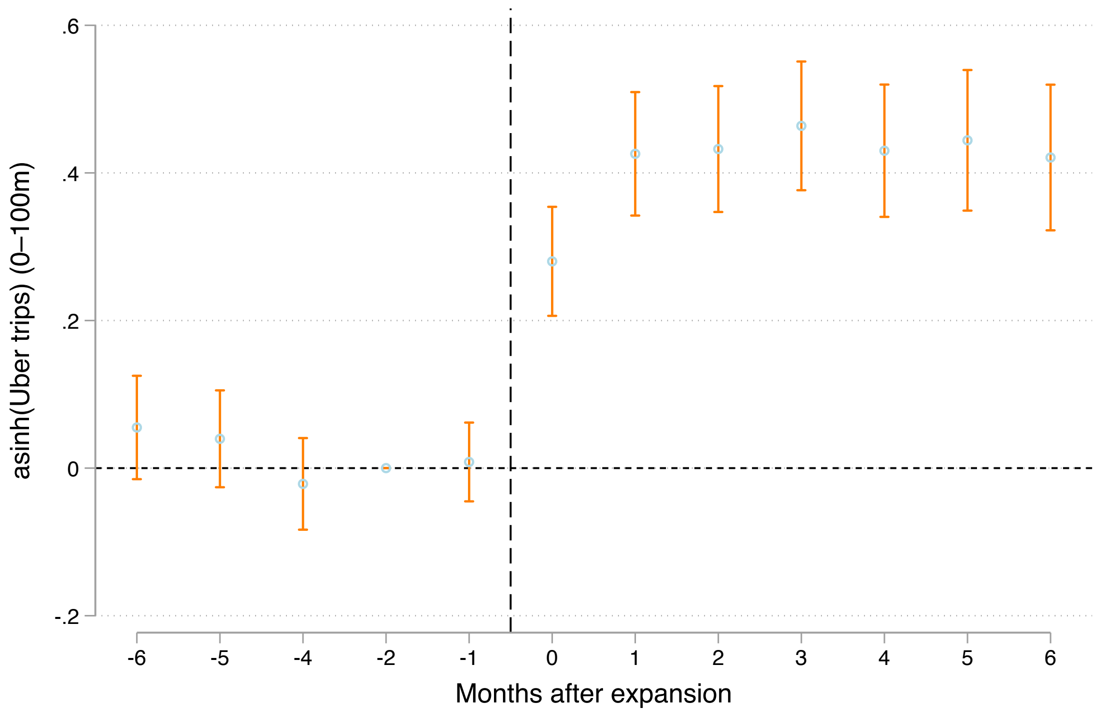

Figure 3: Dynamic difference-in-difference estimates, 0–100 m relative to 100–200 m

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) for the percentage effect of a new train station opening on Uber trips 0–100 m away, relative to 100–200 m away. The coefficients are plotted for each month between six months prior and six months after an expansion. The regression is a dynamic difference-in-differences model of station openings on the inverse hyperbolic sine of Uber trips that either originated or terminated at a given distance band from the opening. The model contains station by distance band and station by time fixed effects; additionally, fixed effects for distance band by more than 6 months before or after an expansion are included. All errors are clustered at the station-level.

plots the coefficients for the full dynamic difference-in-differences specification for each distance.

Our finding of no marginal effect, relative to the next further out distance band, past 300 m from a transit station, implies that the interaction between transit and Uber past 300 m is too small to detect. An alternative hypothesis is that the effect from 300–1100 m is constant, and so our difference-in-differences design nets it out. To test this, we estimate an event study of the effect of a new transit station opening on Uber ridership at 1100–1200 m. This specification does not use the next further out distance band to control for time trends, instead relying on city by month fixed effects. Figure 5 shows that we find no detectable treatment effect.[8]

[8]While this specification controls for city trends, it still does not control for the hyper-local trends captured

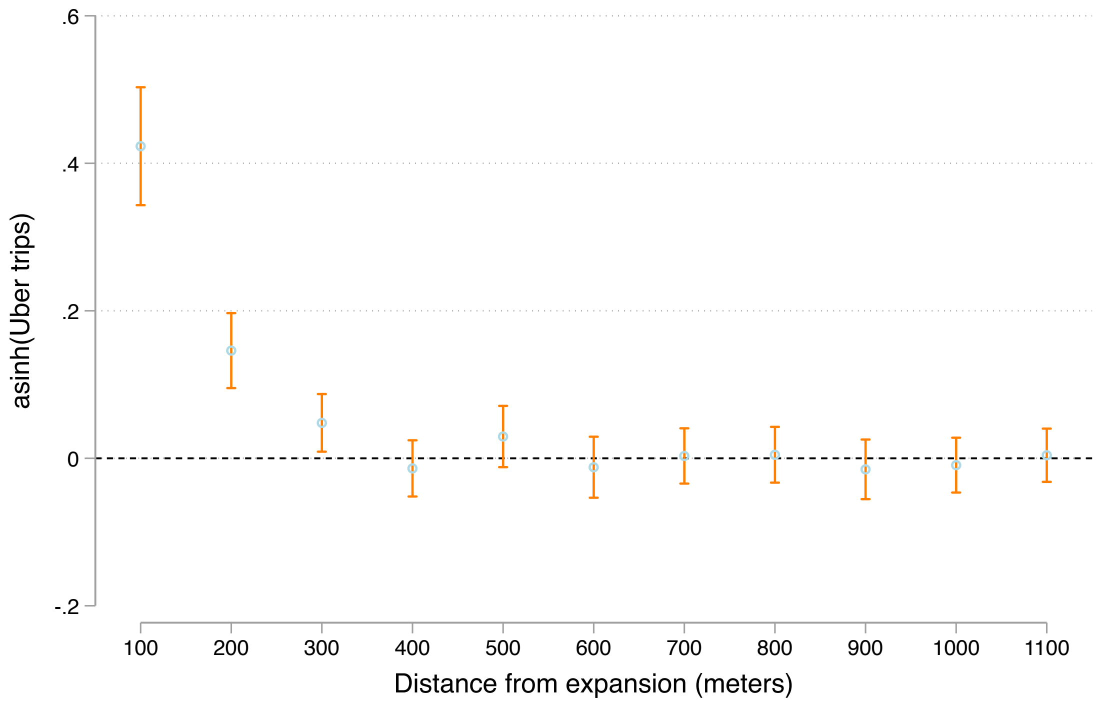

Figure 4: Relative effect by distance of new train stations on Uber trips

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equation (2). These coefficients estimate the effect of a new train station opening on Uber trips at varying distances from the train station during months 1 through 6, relative to the next further distance band, and relative to months -6 through -3. Each distance is measured over a 100 m band ending at the given distance; for example, the coefficient at 400 m reports the change in trips between 300 and 400 m away from a train station.

If we assume the true effect of new transit stations on Uber ridership at 1100–1200 m is zero, then we can sum up the relative treatment effects beyond a given distance to estimate the total treatment effect at that distance band. Figure 6 reports the results from doing so. As expected, the total effects at 0–100 and 100–200 m are larger, the total effect at 200–300 m is relatively unchanged, and the total effect beyond 300 m is indistiguishable from zero. Details of this calculation can be found in subsection B.1.

4.2 Heterogeneity

In this section we explore sources of heterogeneity in the effect of a new transit station opening on Uber ridership. We start by allowing the treatment effect to differ by city. We

in our difference-in-difference analysis. Moreover, the event study specification leads to cities without multiple, differently-timed expansions contributing no variation to the estimation of the main coefficients.

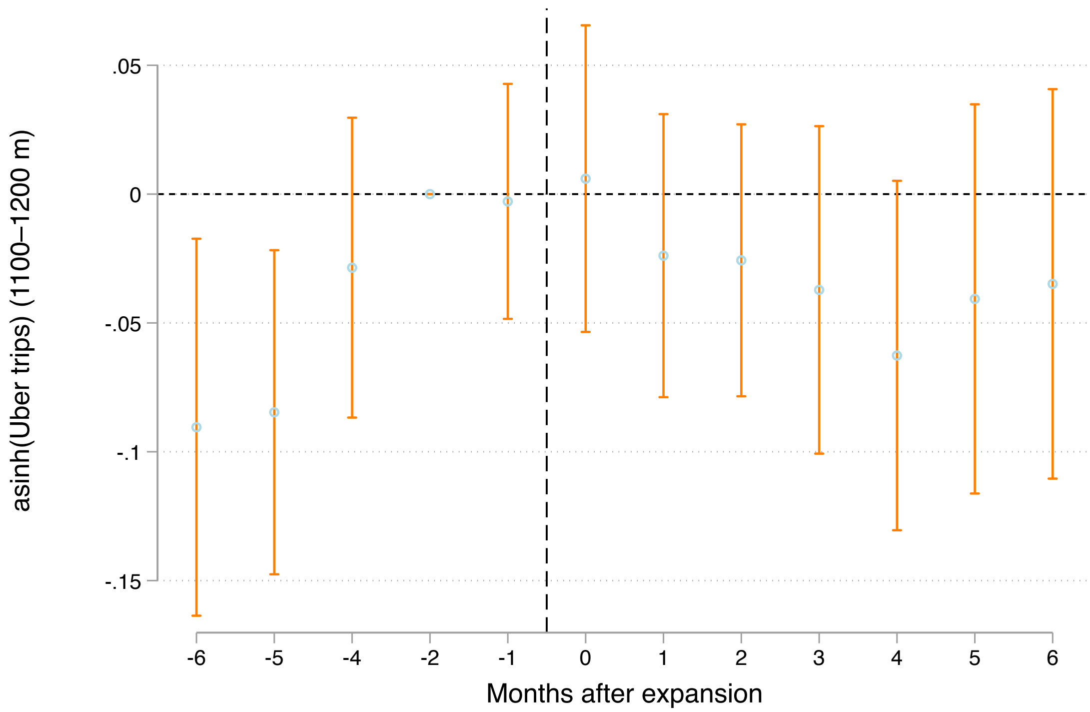

Figure 5: Event study estimates for 1100–1200 m

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) for the percentage effect of a new train station opening on Uber trips 1100–1200 m away. The coefficients are plotted for each month between six months prior and six months after an expansion. The regression is an event study model of station openings on the inverse hyperbolic sine of Uber trips that either originated or terminated at a given distance band from the opening. The model contains station by distance band and city by time fixed effects; additionally, fixed effects for distance band by more than 6 months before or after an expansion are included. All errors are clustered at the station-level.

use a version of equation (2) that allows the coefficient α to be city specific. We plot the results of doing so for Uber trips within 0–100 m of a transit station using a funnel graph in Figure 7. In this figure, the x-axis shows estimated coefficients and the y-axis shows standard errors. The region in white contains estimates that are not statistically different from zero. The light, medium, and dark gray regions contain estimates that are statistically different from zero at 10%, 5%, and 1%, respectively. The figure shows that the estimated coefficients are clustered around our average estimate (marked with the large triangle), showing that the treatment effect at 0–100 m is consistent across cities.

There are a variety of mechanisms by which Uber and public transportation affect each other, and there exist trips for which they are substitutes and other trips for which they are complements. Thus, it is reasonable to expect that in some cities the net effect would

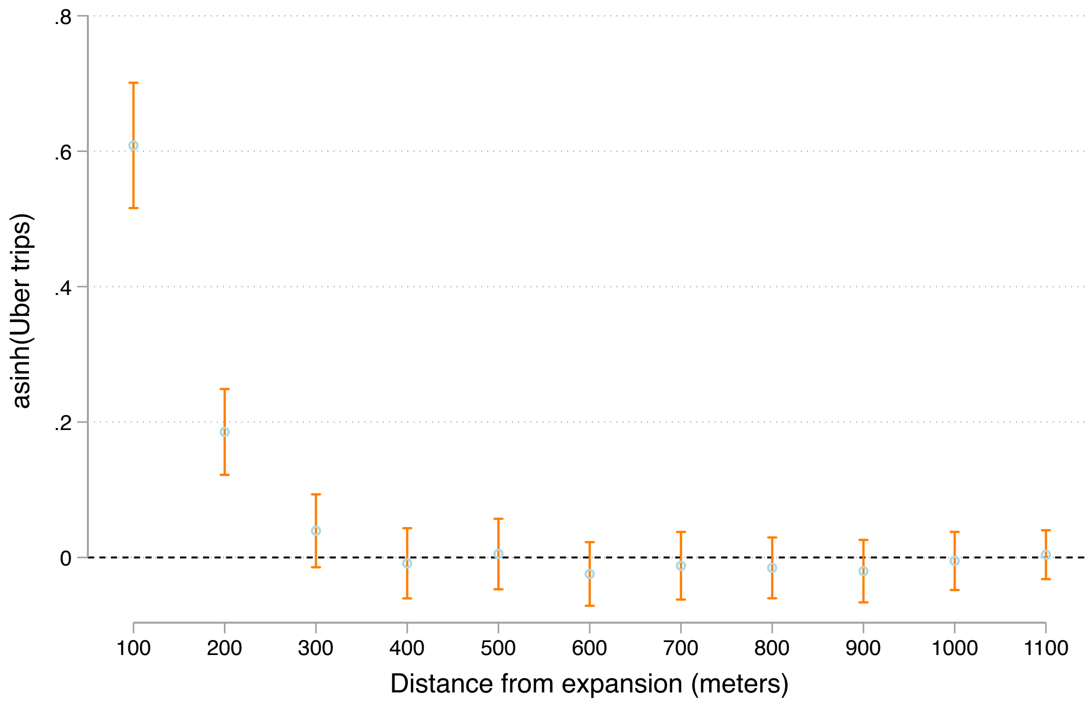

Figure 6: Effect by distance of new train stations on Uber trips

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) for the effect of a new train station opening on Uber trips at varying distances from the train station. Each distance is measured over a 100 m band ending at the given distance; for example, the coefficient at 400 m reports the change in trips between 300 and 400 m away from a train station.

be that of substitutes while in other cities the net effect would be that of complements. While we do not see this at 0–100 m, to give ourselves the best chance of detecting this, we also investigate heterogeneity at 300–400 m. This is where the average effect is first zero, and so seems the best place to look for cities with a negative effect. We plot the city-level heterogeneity in the relative treatment effect at 300–400 m in Figure 8. We find that most of the city-level effects are clustered around zero, with few of them statistically significant.

4.3 Mechanisms

We now turn to identifying the mechanisms by which public transit and Uber affect each other. If Uber is being used to help with the first- and last-mile of transit trips, then we expect to see that, close to transit stations, the average length of an Uber trip that starts or ends decreases after a transit station opens. To test this, we use the same specification as in

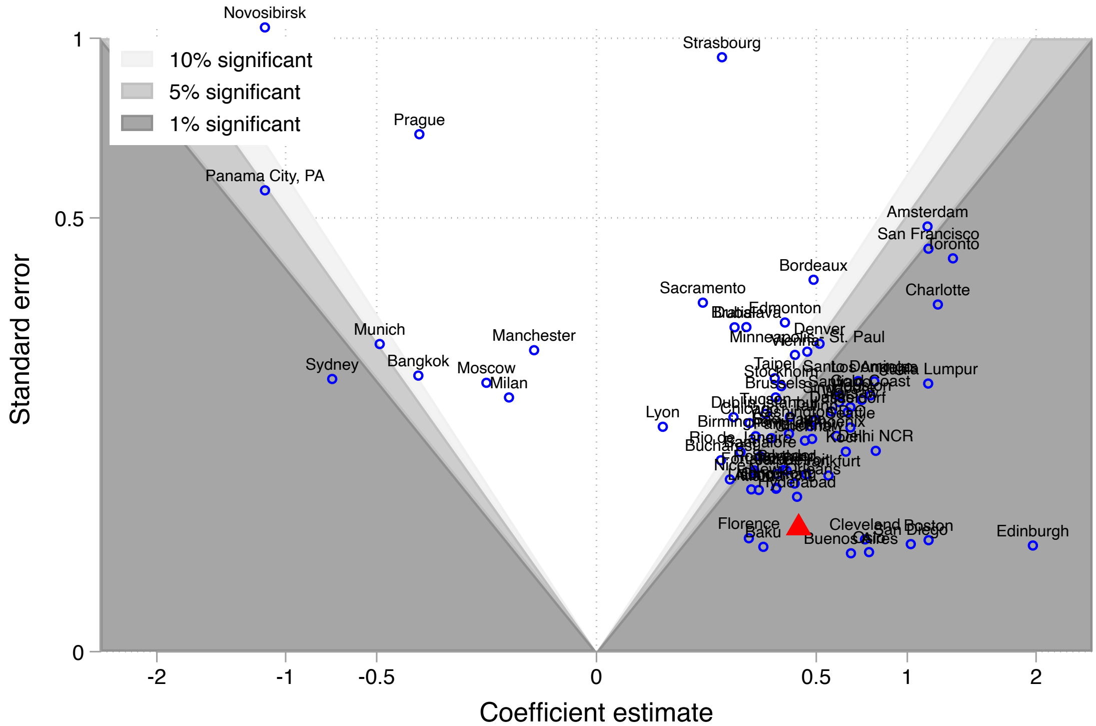

Figure 7: Heterogeneous effect by city, 0–100 m

Notes: This funnel graph shows city-specific treatment effects based on equation (2) where the coefficient α is allowed to vary by city. The x-axis shows coefficient estimates, the y-axis shows standard errors. The region in white contains estimates that are not statistically different from zero. The light, medium, and dark gray regions contain estimates that are statistically different from zero at 10%, 5%, and 1%, respectively. The large red triangle indicates the average effect reported in Figure 4.

(2), except that we use the average trip length as our outcome. The results from doing so are plotted in Figure 9. We find that the a new transit station opening reduces the average trip length by 1.26 km within 0–100 m and 0.43 km within 100–200 m of a transit station. We find no change at 200–300 m, which suggests that the increase in trips we observe at that distance is not due to first- and last-mile usage.

4.4 Robustness

This section contains four robustness tests. We start by conducting placebo tests and showing our results are robust to alternative specifications. We start with placebo tests of our estimates of the effect of a transit expansion on Uber trips at different distance bands. For each placebo test, we assign each transit station a fake opening date randomly drawn

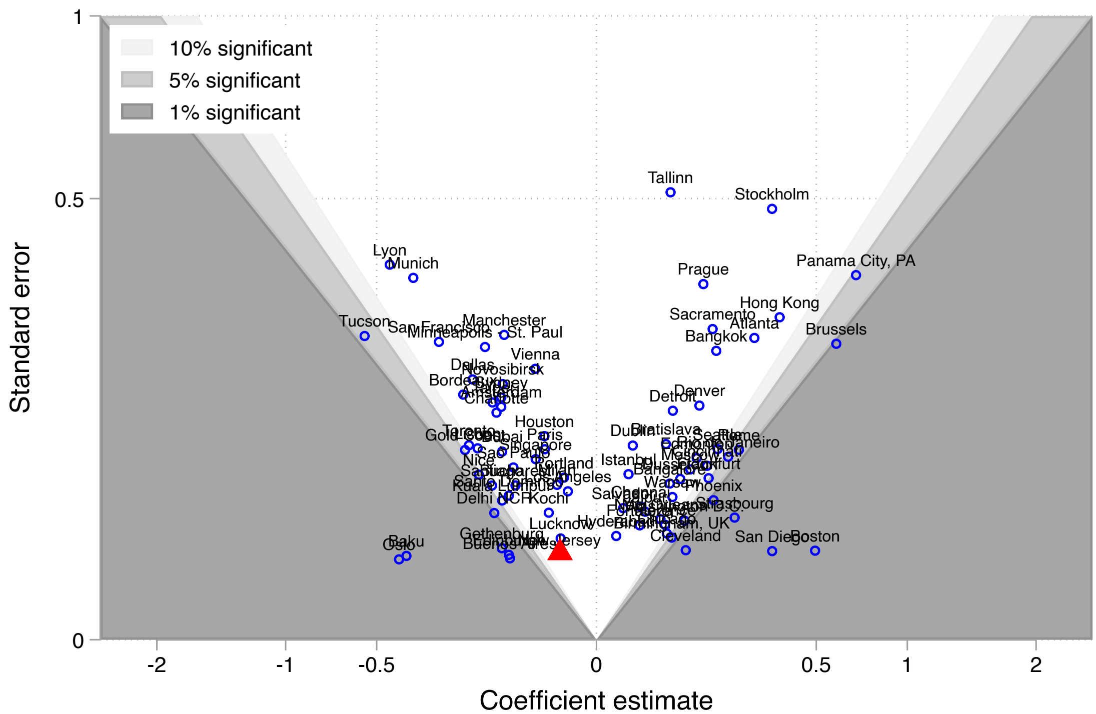

Figure 8: Heterogeneous effect by city, 300–400 m

Notes: This funnel graph shows city-specific treatment effects based on equation (2) where the coefficient α is allowed to vary by city. The x-axis shows coefficient estimates, the y-axis shows standard errors. The region in white contains estimates that are not statistically different from zero. The light, medium, and dark gray regions contain estimates that are statistically different from zero at 10%, 5%, and 1%, respectively. The large red triangle indicates the average effect reported in Figure 4.

from the list of dates for which that station has sufficient data to run our specification.[9] We repeat this process 100 times. Figure 10 uses a box and whisker plot to compare the 5th, 25th, 50th, 75th, and 95th percentiles of distribution of placebo results to our actual estimates, denoted by blue circles. It confirms the results from Figure A1 that a transit expansion increases Uber ridership close to the station while reducing it further way.

Appendix Figure A2 reports results using pseudo-Poisson maximum likelihood. Appendix Figure A3 reports results from a simpler "donut" design, using 1100–1200 m as a comparison group for all closer distances. Appendix Figure A4 reports results from the first-difference specification, which uses city by month fixed effects to control for time trends.

[9] We require six months of data before and after the station opens.

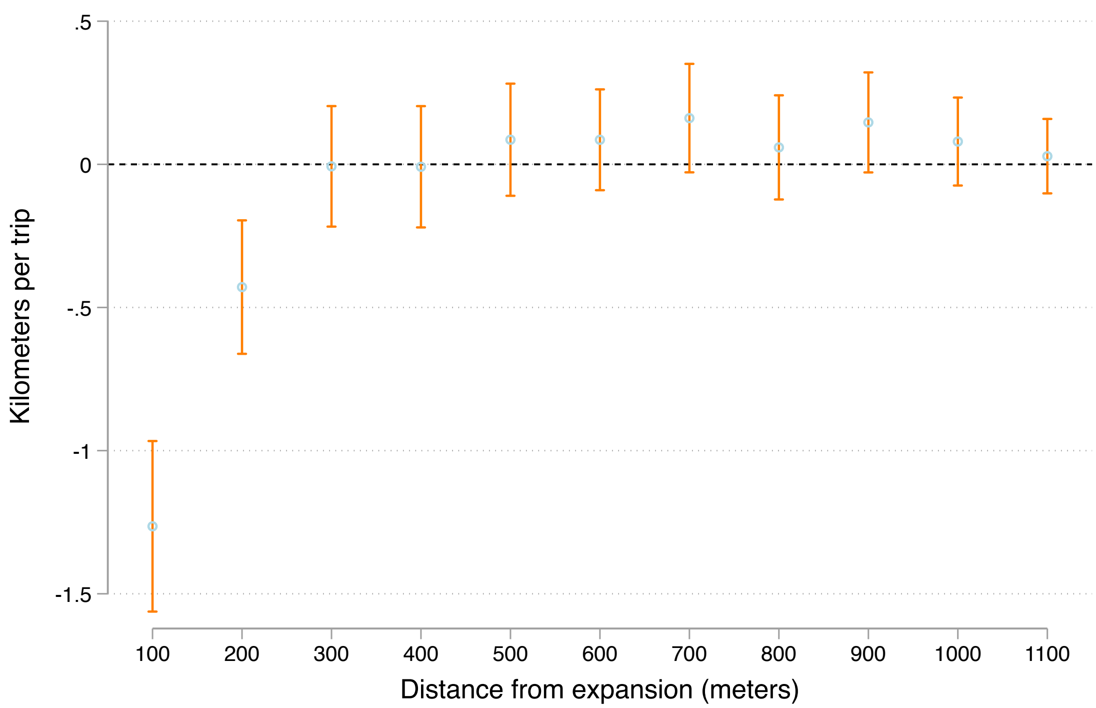

Figure 9: Effect by distance of new train stations on kilometers per Uber trip

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) for the effect of a new train station opening on kilometers travelled per Uber trip at varying distances from the train station. Each distance is measured over a 100 m band ending at the given distance; for example, the coefficient at 400 m reports the change in trips between 300 and 400 m away from a train station.

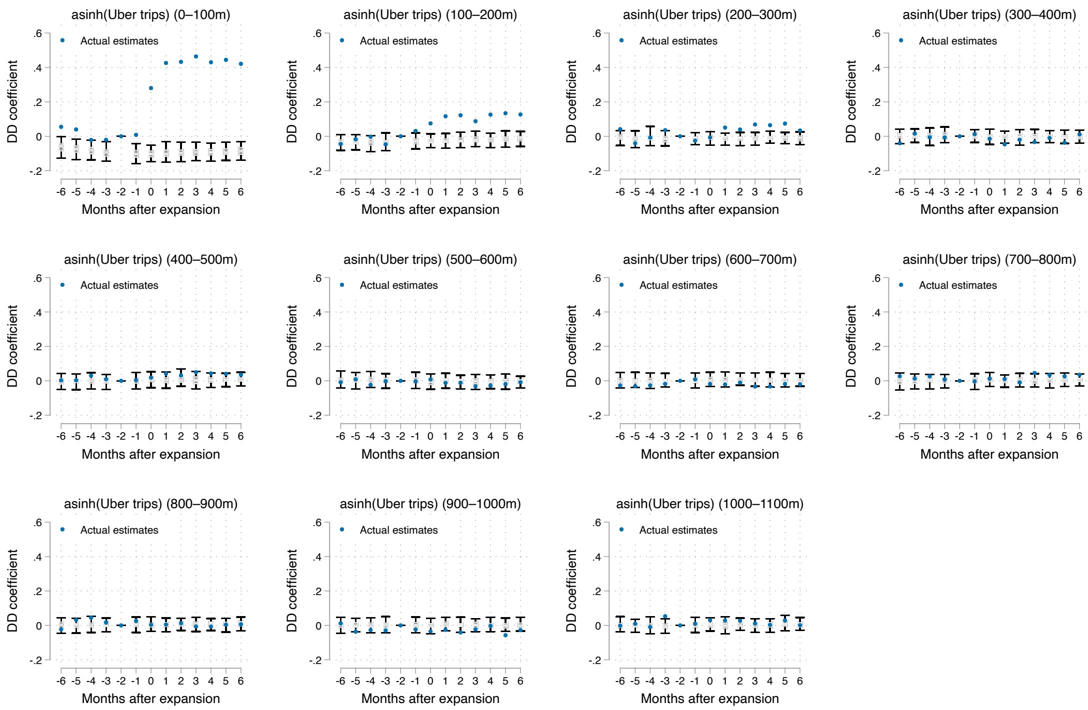

Figure 10: Placebo tests for all 100 m distance bands, up through 1000–1100 m

Notes: This figure plots the results of a placebo test where we randomly assign placebo opening dates to each transit station.

5 Conclusion

There is an on-going debate on whether ride-hailing complements or substitutes public transportation. The answer to this question has important public policy implications regarding how ride-hailing is regulated and taxed, for transit service and infrastructure planning, and for transit-ride-hailing partnerships. However, there remains great uncertainty as existing estimates vary in sign and magnitude.

We address this question using novel data and an innovative identification strategy. Our identification strategy relies on exogenous variation in transit availability caused by rail expansions. Using proprietary trip data from Uber, we use a dynamic difference-in-differences strategy to estimate how transit expansions affect local Uber ridership. Our method controls for hyper-local trends in Uber ridership. We find that a new rail station opening increases Uber ridership within 300 m, and has no impact between 300–1200 m. This implies that Uber and rail transit complement each other.

References

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

Anderson, Michael L. and Lucas W. Davis (2021) "Uber and Alcohol-Related Traffic Fatalities," *Working Paper*.

Babar, Yash and Gordon Burtch (2020) "Examining the Heterogeneous Impact of Ride-Hailing Services on Public Transit Use," *Information Systems Research*, 31 (3), 820–834, 10.1287/isre.2019.0917.

Barreto, Yuri, Raul Silveira Neto, and Luís Carazza (2020) "Uber and Traffic Safety: Evidence from Brazilian Cities," *Working Paper*.

Barrios, John Manuel, Yael V. Hochberg, and Hanyi Yi (2020) "The Cost of Convenience: Ridehailing and Traffic Fatalities," SSRN Scholarly Paper ID 3288802, Rochester, NY, 10.2139/ssrn.3288802.

Basso, Leonardo J. and Hugo E. Silva (2014) "Efficiency and Substitutability of Transit Subsidies and Other Urban Transport Policies," *American Economic Journal: Economic Policy*, 6 (4), 1–33, 10.1257/pol.6.4.1.

Baum-Snow, Nathaniel (2007) "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics*, 122 (2), 775–805, 10.1162/qjec.122.2.775.

Burgdorf, Jacob, Conor Lennon, and Keith Teltser (2019) "Do Ridesharing Services Increase Alcohol Consumption?" *SSRN Electronic Journal*, 10.2139/ssrn.3484845.

Cairncross, John, Jonathan D. Hall, and Craig Palsson (2021) "VancUber: The effect of ride-hailing on public transportation, congestion, and traffic fatalities," *Working Paper*.

Castillo, Juan Camilo (2020) "Who Benefits from Surge Pricing?" *Working Paper*.

Castillo, Juan, Daniel T. Knoepfle, and E. Glen Weyl (2021) "Surge Pricing Solves the Wild Goose Chase," *SSRN Electronic Journal*, 10.2139/ssrn.2890666.

Diamond, Rebecca and Tim McQuade (2019) "Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development," *Journal of Political Economy*, 127 (3), 1063–1117.

Diao, Mi, Hui Kong, and Jinhua Zhao (2021) "Impacts of transportation network companies on urban mobility," *Nature Sustainability*, 10.1038/s41893-020-00678-z.

Erhardt, Gregory D., Richard Alexander Mucci, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione (2021) "Do Transportation Network Companies Increase or Decrease Transit Ridership? Empirical Evidence from San Francisco," *Transportation*, 10.1007/s11116-021-10178-4.

Gendron-Carrier, Nicolas, Marco Gonzalez-Navarro, Stefano Polloni, and Matthew A. Turner (forthcoming) "Subways and Urban Air Pollution," *American Economic Journal: Applied Economics*.

Gorback, Caitlin (2019) "Ridesharing and the Redistribution of Economic Activity," *Working Paper*.

Graehler, Michael, Richard Alexander Mucci, and Gregory D. Erhardt (2019) "Understanding the Recent Transit Ridership Decline in Major US Cities: Service Cuts or Emerging Modes?" in *Transportation Research Board 98th Annual Meeting*.

Greenwood, Brad N., , and Sunil Wattal (2017) "Show Me the Way to Go Home: An Empirical Investigation of Ride-Sharing and Alcohol Related Motor Vehicle Fatalities," *MIS Quarterly*, 41 (1), 163–187, 10.25300/MISQ/2017/41.1.08.

Hall, Jonathan D., Craig Palson, and Joseph Price (2018) "Is Uber a Substitute or Complement for Public Transit?" *Journal of Urban Economics*, 108 (1), 36–50, 10.1016/j.jue.2018.09.003.

Heblich, Stephan, Stephen J Redding, and Daniel M Sturm (2020) "The Making of the Modern Metropolis: Evidence from London," *Quarterly Journal of Economics*, 135 (4), 2059–2133, 10.1093/qje/qjaa014.

Nelson, Erik and Nicole Sadowsky (2018) "Estimating the Impact of Ride-Hailing App Company Entry on Public Transportation Use in Major US Urban Areas," *The B.E. Journal of Economic Analysis & Policy*, 19 (1), 10.1515/bejeap-2018-0151.

Parry, Ian W. H and Kenneth A Small (2009) "Should Urban Transit Subsidies Be Reduced?" *American Economic Review*, 99 (3), 700–724, 10.1257/aer.99.3.700.

A Additional figures and tables

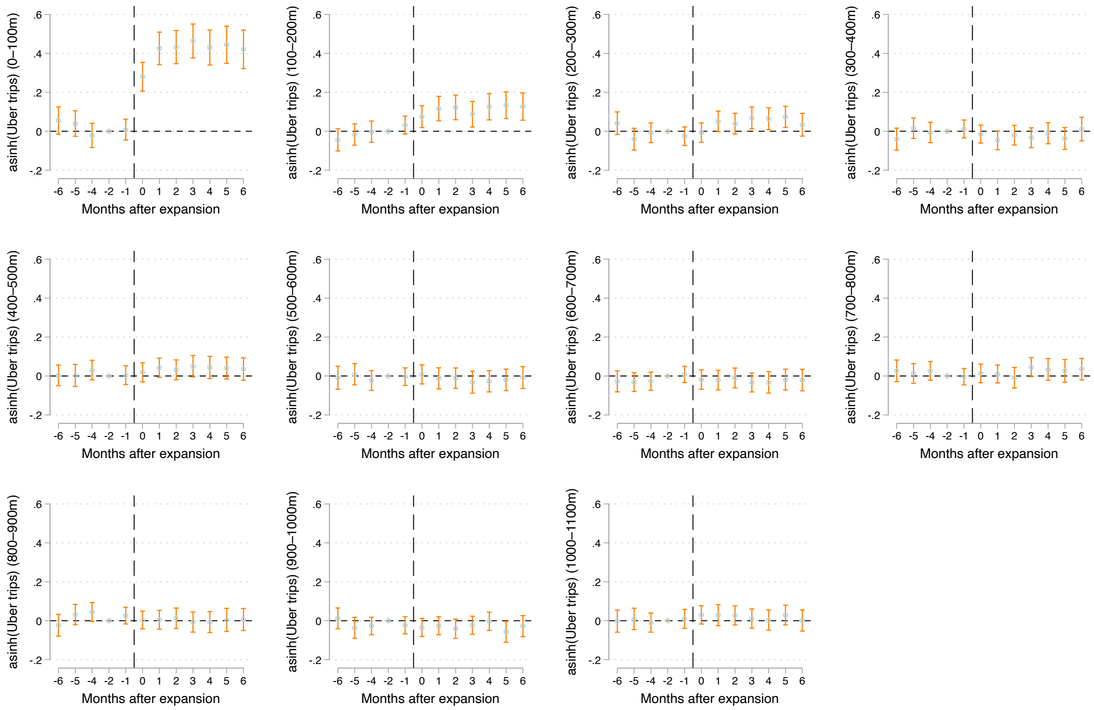

Figure A1: Difference-in-differences estimates, using the adjacent distance band as the control group, from 0–100 m to 1000–1100 m

Notes: These subfigures plot the coefficients (circles) and 95% confidence intervals (bars) for the percentage effect of a new train station opening on Uber trips for the labeled distance bands. The coefficients are plotted for each month between six months prior and six months after an expansion. The regression is a dynamic difference-in-differences model of station openings on the inverse hyperbolic sine of Uber trips that either originated or terminated at a given distance band from the opening, and the regression equation is given by equation (1). The model contains station by distance band and station by time fixed effects; additionally, fixed effects for distance band by more than 6 months before or after an expansion are included. All errors are clustered at the station-level.

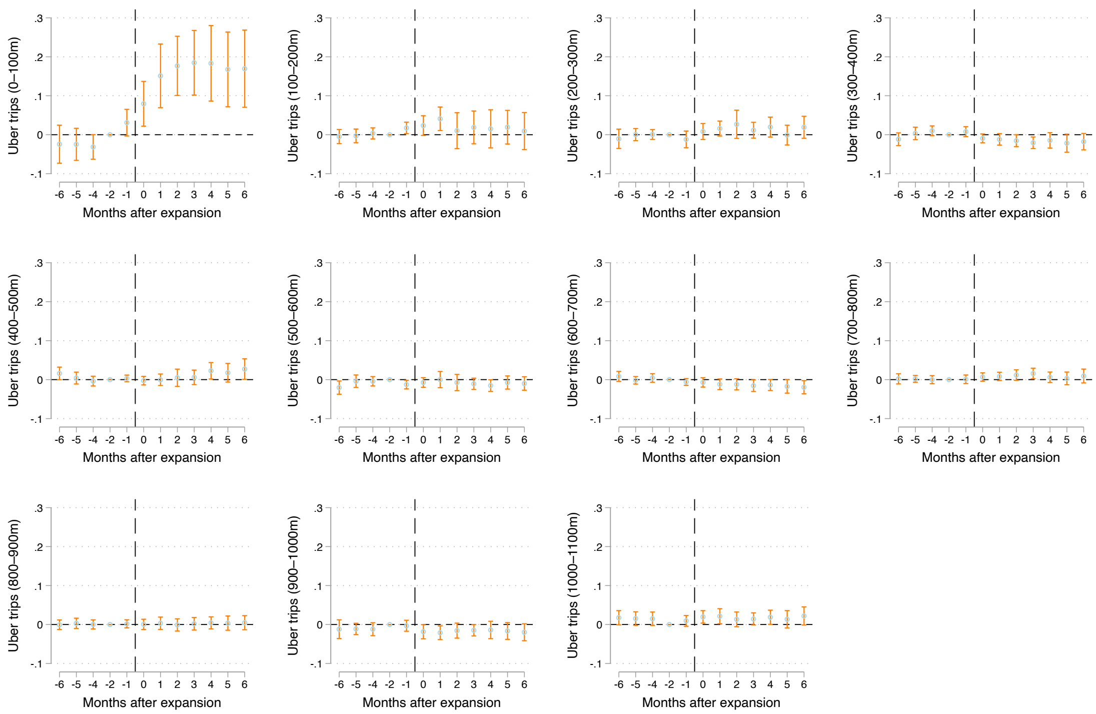

Figure A2: Difference-in-differences estimates, using adjacent distances as the control group, from 0–100 m to 1000–1100 m, using PPMLE

Notes: These subfigures report the results of re-estimating the analysis reported in Figure A1, except that we now use Poisson pseudo-likelihood regression rather than ordinary least squares.

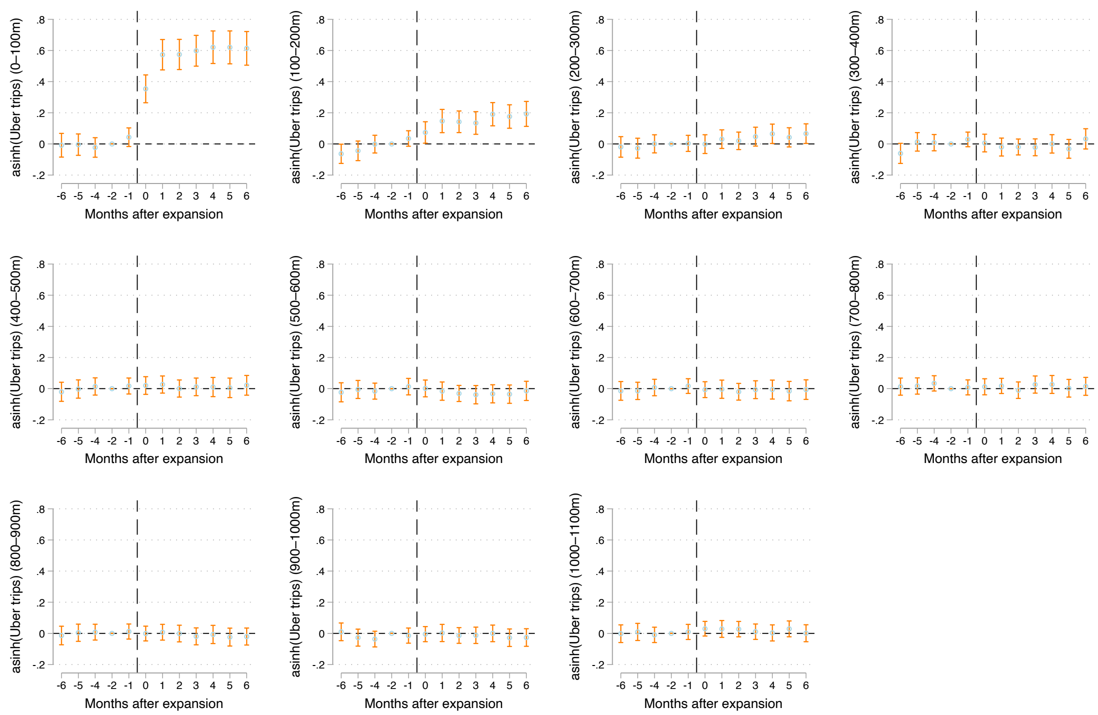

Figure A3: Difference-in-differences estimates, using 1100–1200 m as the control group, from 0–100 m to 1000–1100 m

Notes: These subfigures plot the coefficients (circles) and 95% confidence intervals (bars) for the percentage effect of a new train station opening on Uber trips for the labeled distance bands. The coefficients are plotted for each month between six months prior and six months after an expansion. The regression is a dynamic difference-in-differences model of station openings on the inverse hyperbolic sine of Uber trips that either originated or terminated at a given distance band from the opening, and the regression equation is given by equation (1). The model contains station by distance band and station by time fixed effects; additionally, fixed effects for distance band by more than 6 months before or after an expansion are included. All errors are clustered at the station-level.

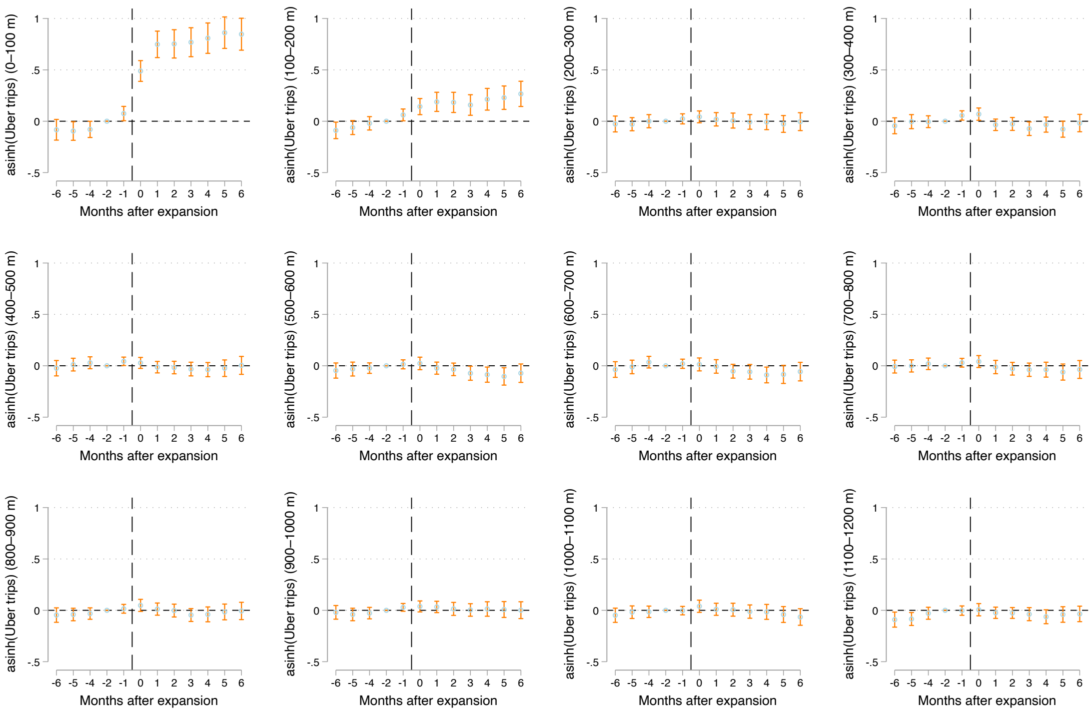

Figure A4: Event study estimates for 100 m distance bands, using city-month fixed effects

Notes: These subfigures report the results of re-estimating the analysis reported in Figure A1, except that we now use city-month fixed effects to adjust for time trends, rather than differencing out the effect at the next further distance band.

B Appendix: Methodology

B.1 Total effects by distance calculation

Our main difference-in-differences estimates come from (2), given by

$$
\begin{gathered}
y_{dit} = \gamma_{it} + \delta_{di} + \alpha_{\tilde{d}} \times \mathbb{1}_{\tau_{it} \in \{1,\ldots,6\}} \times \mathbb{1}_{d=\tilde{d}} \\
+ \left(\beta_{1,\tilde{d}} \times \mathbb{1}_{\tau_{it} < -6} + \beta_{2,\tilde{d}} \times \mathbb{1}_{\tau_{it} > 6} + \beta_{3,\tilde{d}} \times \mathbb{1}_{\tau_{it} \in \{-2,-1,0\}} \right) \mathbb{1}_{d=\tilde{d}} + \epsilon_{dit}, \\
\forall \ d \in \{\tilde{d}, \tilde{d}+1\}.
\end{gathered}
$$

Each regression is estimated separately by $\tilde{d}$, producing an array of expansion effects at varying distances relative to the next distance band $\tilde{d}+1$. The α_d can be estimated jointly by stacking all such distance pairs p, where $p \in \{100, ..., 1100\}$ refers to the "treated" distance in each pair. In the new data set, distance bands 0–100 m and 1100–1200 m will appear once, while all others will appear twice (once in the control group and once in the treatment group). The specification can then be written as follows.

$$
\begin{gathered}
y_{dipt} = \gamma_{ipt} + \delta_{dip} + \sum_{\tilde{d}} \alpha_{\tilde{d}} \times \mathbb{1}_{\tau_{it} \in \{1,\ldots,6\}} \times \mathbb{1}_{d=\tilde{d}} \times \mathbb{1}_{p=\tilde{d}} \\
+ \sum_{\tilde{d}} \left(\beta_{1,\tilde{d}p} \times \mathbb{1}_{\tau_{it} < -6} + \beta_{2,\tilde{d}p} \times \mathbb{1}_{\tau_{it} > 6} + \beta_{3,\tilde{d}p} \times \mathbb{1}_{\tau_{it} \in \{-2,-1,0\}} \right) \mathbb{1}_{d=\tilde{d}} \times \mathbb{1}_{p=\tilde{d}} + \epsilon_{dipt}
\end{gathered}
$$

The $\alpha_{\tilde{d}}$ will be numerically equivalent to those from (2) when errors are clustered at the station-level. As before, the α estimate the effect at one distance relative to another. In order to estimate the total effect of an expansion on Uber trips at a given distance, we sum all estimates from farther distances. To implement this, we simplify the above regression as follows.

$$
\begin{gathered}
y_{dipt} = \gamma_{ipt} + \delta_{dip} + \sum_{\tilde{d}} \alpha^*_{\tilde{d}} \times \mathbb{1}_{\tau_{it} \in \{1,\ldots,6\}} \times \mathbb{1}_{d=\tilde{d}} \\
+ \sum_{\tilde{d}} \left(\beta_{1,\tilde{d}} \times \mathbb{1}_{\tau_{it} < -6} + \beta_{2,\tilde{d}} \times \mathbb{1}_{\tau_{it} > 6} + \beta_{3,\tilde{d}} \times \mathbb{1}_{\tau_{it} \in \{-2,-1,0\}} \right) \mathbb{1}_{d=\tilde{d}} + \epsilon_{dipt}
\end{gathered}
$$

We no longer estimate the α's by comparisons within distance-pairs. By not fully-saturating on pairs p, we allow the treatment effect from one distance to pass through into the next, closer distance. Numerically, $\alpha^*_{\tilde{d}} = \sum_{d \geq \tilde{d}} \alpha_d$. Additionally, the regression implementation of summing the coefficients correctly calculates standard errors. These estimates are reported in Figure 6.