Uber versus Trains?
Worldwide Evidence from Transit Expansions*

Marco Gonzalez-Navarro†

Jonathan D. Hall§

Harrison Wheeler‡

Rik Williams¶

December 2025

Abstract

We study how ride-hailing and public transit interact – an open question with important implications for transportation policy and urban form. We develop a dynamic model of mode choice that features three key mechanisms: (i) *substitution*, whereby riders switch from Uber to transit near stations; (ii) *last-mile complementarity*, whereby Uber is used to access transit; and (iii) *mode lock-in*, where outbound mode choice constrains return options, giving Uber option value. We test the model's implications using a dynamic difference-in-differences design combined with new data on 650 global rail transit expansions and the universe of nearby Uber trips. We document three patterns consistent with each of the above mechanisms: Uber usage rises sharply within 100 meters of new stations (last-mile complementarity), declines locally among nearby residents and workers (substitution), and falls modestly overall for these users as local reductions are partially offset by increased Uber usage elsewhere (mode lock-in). We find that while transit and Uber are substitutes in the aggregate, their local interaction depends on competing forces often overlooked in prior work.

*We are grateful for helpful feedback from Enrico Moretti and Chandra Krishnamurthy, as well as seminar participants at the 2023 UEA North America Conference and the 2024 LA Urban Conference. 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].

†University of California Berkeley, marcog@berkeley.edu

‡University of Toronto, harry.wheeler@rotman.utoronto.ca

§University of Alabama, jonathan.hall@ua.edu

¶rjw297@gmail.com

1 Introduction

The rapid global expansion of ride-hailing services has sparked a critical debate over their impact on cities. Policymakers and urban planners are grappling with how best to regulate these services, with some countries – including Denmark, Hungary, and Bulgaria – even banning them outright. A central question in this debate is whether ride-hailing substitutes for or complements public transit. The answer has broad implications. First, declines in transit ridership can lead to budgetary shortfalls for transit agencies. Second, they may reduce social welfare, as transit fares typically exceed marginal social cost due to economies of scale and pricing constraints.[1] Third, if travelers shift away from transit, congestion and pollution may worsen.[2] Finally, changes in transportation technology have historically reshaped urban form: steam railways enabled London's expansion, highways hollowed out U.S. city centers, and ride-hailing may now be altering spatial patterns in real time.[3]

These economic, environmental, and spatial factors underscore the importance of understanding how ride-hailing and public transit interact. Yet the evidence remains unsettled. Estimates of ride-hailing's effect on transit ridership vary widely: some studies find increases of around 5% within two years of Uber's entry (Hall et al., 2018), while others report decreases of 16% after four years (Diao et al., 2021). Moreover, most existing work focuses on aggregate city-level effects, providing limited insight into when, where, or for whom substitution or complementarity occurs. The reverse relationship – how transit expansions affect ride-hailing demand – has been even less studied. Yet this question is central to understanding whether improvements to public transit reduce reliance on ride-hailing, or instead stimulate new demand through mechanisms like last-mile connectivity or the option value of Uber as a flexible backup when transit is delayed or unavailable.

This paper addresses both gaps by combining a dynamic model of transit choice with novel data and an innovative empirical strategy. To fix ideas, we begin by modeling the decision a traveler makes when choosing an outbound transit mode from her origin to a destination. She may drive, take Uber, or use public transit. Accessing the transit station requires either walking or using ride-hailing, with each mode incurring fixed and marginal costs based on first- or last-mile distance. At the destination, she may undertake an intermediate trip before returning home. Her preferences across modes and trip legs include both systematic and idiosyncratic components. Crucially, her outbound mode choice constrains her return options: if she drives, her car becomes available for intermediate travel, but she is committed to returning home by car. The traveler is forward-looking and anticipates the downstream consequences of outbound mode choice when making her initial decision.

The model yields several implications for how improvements to public transit may affect

[1] Transit fares are usually above marginal cost—though below average cost—due to scale economies. See Parry and Small (2009), Basso and Silva (2014).

[2] See Anderson (2014); Gendron-Carrier et al. (2022).

[3] See Heblich et al. (2020); Baum-Snow (2007); Gorback (2024).

demand for ride-hailing services. Close to the station, better public transit attracts individuals who would otherwise drive or take Uber. However, improved public transit can also complement Uber through two channels. First, it may increase the use of Uber for first- and last-mile connections to the transit system. Second, Uber increases traveler welfare by providing additional flexibility and choice, especially for non-drivers who would otherwise have limited transportation options. For example, a traveler who planned to take the subway home may switch to Uber if she encounters severe delays. At greater distances from the station, improved transit similarly draws travelers away from car and Uber use. However, those who did not drive for the outbound leg are then effectively "locked in" to using Uber or transit for intermediate travel. This complementary effect can lead to increased Uber use in other parts of the city. These competing forces – both near to and far from stations – can rationalize a wide range of responses in Uber ridership to improvements in public transit.

We test these predictions using proprietary data on Uber trips across 35 countries and exploit the sharp timing of 650 rail transit station openings. Our identification strategy compares changes in Uber usage within concentric distance bands around newly opened stations, using trips 1100–1200 meters away as a local control group. This approach offers several advantages. First, the timing of rail expansions is plausibly exogenous to trends in Uber ridership, as they are planned years in advance – often before Uber even existed. Second, our detailed, geographically precise data enable us to test the mechanisms through which transit and ride-hailing interact locally. Third, our research design allows us to flexibly control for hyper-local, highly variable time trends in Uber ridership. Fourth, we assess the general applicability of our results by examining 650 rail transit openings across 55 cities in 35 countries.

Our research design relies on two key identifying assumptions. First, that Uber trips occurring 1100-1200 meters from the station form an appropriate control group for those occurring within 100 meters. Second, that this control group is sufficiently distant to remain unaffected by the station opening.

To interpret the spatial and behavioral heterogeneity in responses to new transit, we organize our empirical analysis around three quadrants – each corresponding to a set of mechanisms in the model:

- **Quadrant I** examines how Uber usage responds in space. We estimate large and immediate increases in trip volumes within 100 meters of new stations (48% on average), consistent with first- and last-mile complementarity. However, ridership declines by about 2% at intermediate distances (300–700 meters), consistent with substitution from Uber to transit among travelers within walking distance. Aggregated across the catchment area, these opposing effects balance.

- **Quadrant II** removes last-mile behavior by focusing on travelers with a regular connection to the station area, as in those who might live or work nearby. Holding composition constant, we find substantial declines in their Uber usage near the station post-opening, confirming that substitution away from ride-hailing is a central behavioral response among affected users.

- **Quadrant III** examines broader behavioral adjustments among the same local travelers as in Quadrant II by tracking their citywide Uber use. We find that while their trips near the station decline, non-local Uber usage increases modestly. This pattern aligns with our model's prediction that travelers who switch to transit in the outbound leg may be "locked in" to Uber or transit for subsequent trips – an option value or composition effect that operates beyond the immediate vicinity of the station.

Each quadrant maps to a different structural prediction: Quadrant I documents last-mile complementarity; Quadrant II cleanly identifies substitution; and Quadrant III is informative about the role of downstream compositional shifts on transit mode choice. This structure also clarifies why the same transit expansion can both raise and lower Uber demand depending on user location and trip type.

We provide empirical validation of the assumptions underlying this design. First, prior to the station opening, Uber ridership trends within 100 meters are statistically indistinguishable from those within 1100-1200 meters. Second, the effect on Uber ridership decays towards zero with increasing distance from the station. Third, when leveraging only the variation in timing of rail expansions in an event study framework, we find no effect on Uber trips at 1100-1200 meters. This set of evidence is reinforced by our tight study window (six months before and after station openings) and the quasi-exogenous nature of transit expansions.

Our results are highly robust. We find similar patterns across time of the day/week and geography. Treatment effects are especially large near terminal stations. Uber trips near new stations also become shorter on average, consistent with a rise in first- and last-mile usage. Moreover, there is no increase in station-adjacent trips for individuals whose inferred home or work location lies within 500 meters of the station – further confirming that last-mile usage is concentrated among non-local riders accessing transit.

Finally, we show that these effects are specific to improvements in transit quality. Transit openings that do not offer meaningful speed gains – such as trams or streetcars – generate no detectable changes in Uber use, reinforcing the interpretation that our findings are driven by improved access rather than neighborhood change.

This paper builds on a quickly growing literature that seeks to determine the conditions under which ride-hailing and public transportation act as complements or substitutes. Most existing studies exploit variation in the timing of Uber's entry across U.S. metropolitan areas to estimate its impact on public transit. Hall et al. (2018) finds that ride-hailing complements public transit on average, whereas Graehler et al. (2019), Erhardt et al. (2021), and Diao et al. (2021) suggest that it serves as a substitute. Other studies, including Nelson and Sadowsky (2018), Babar and Burtch (2020), and Cairncross et al. (2021), report mixed or statistically insignificant results.[4] In contrast

[4] Complementary to this work is (Christensen et al., 2023), who study how travel patterns change with discounts to

to this work, we leverage *within-city* variation in transit access and granular trip data to identify responses with high spatial and temporal precision. This approach avoids confounding aggregate trends that can affect metro-level comparisons and offers an approach that can be applied to studying other transit expansions. We contribute by combining a new model of transit mode choice, a novel identification strategy and detailed administrative ride-hailing data across many cities globally to consider the local impacts of transit access on ride-hailing trips and users.

First, we introduce a new dynamic model of transit mode choice. To our knowledge, it is the first model to formalize the interaction between ride-hailing and public transit across multiple trip legs and access modes (see Agrawal and Zhao (2023) for another recent model that includes last-mile use of Uber). The model captures three distinct mechanisms: substitution, last-mile complementarity, and option value (or mode lock-in). Our model is tractable enough to yield clear insights on mechanisms, but rich enough to rationalize an array of responses in Uber ridership to a new train station opening. An appealing feature of our approach is that each mechanism maps directly to a reduced-form empirical margin, transparently linking the model to the data. This structure also allows us to estimate the model parameters from the reduced-form evidence and simulate counterfactuals, including the effect of Uber entry or exit on public transit usage and traveler welfare.

Second, we use the model to interpret a series of empirical tests that isolate each mechanism. The granularity of our Uber data – capturing origin and destination coordinates for all trips over time – enables us to identify last-mile behavior near stations, substitution by local travelers, and broader lock-in or composition effects. Each mechanism corresponds to a distinct empirical margin in our analysis: (i) the response of trips that originate/terminate near the station, (ii) the response of local Uber trips for users that work or reside near the station, and (iii) the response of citywide Uber trips for users that work or reside near the station. In particular, this paper is the first to empirically document the importance of last-mile trips in the interaction between ride-hailing and transit.[5]

Third, our empirical strategy exploits a new dataset on 650 rail station openings. This setting offers a clean source of variation: new stations provide a sharp improvement in transit access and are independently timed to trends in Uber ridership. The setting is also particularly policy-relevant to cities considering their own investments in transit infrastructure. Our identification strategy compares outcomes across concentric distance bands, which allows us to flexibly control for local trends while targeting the catchment area where responses in Uber ridership are expected. Moreover, while our reduced-form estimates measure the effect of transit expansions on Uber usage, our model allows us to invert the relationship and speak to the reverse question – how Uber

ride-hailing services in Cairo, Egypt.

[5]The potential for Uber to complement public transit in this way has been theorized (King et al., 2020, Siddiq et al., 2023). Li et al. (2022) find evidence of last-mile complementarity in Austin, while Yan et al. (2019) use survey data to show travelers' willingness to combine ride-hailing with transit. Uber itself introduced a dedicated "last mile" service in Berlin and Munich in 2021 (Uber, 2021, 2023). This service allowed riders to book fixed-price trips home from U-Bahn and S-Bahn stations. As discussed before, Agrawal and Zhao (2023) present a theoretical model that includes last-mile complementarities to study the optimal taxation of Uber.

availability affects transit ridership – thereby unifying two strands of the literature. Additionally, we contribute to the growing literature on how public transit shapes urban economic geography more broadly (Billings, 2011, Gonzalez-Navarro and Turner, 2018, Gendron-Carrier et al., 2022, Gupta et al., 2022, Zárate, 2024, Tsivanidis, 2024).

Finally, we show that the patterns we document are consistent across a wide range of urban contexts. While much of the literature has focused on single-city studies – often in the U.S. – our analysis covers 55 cities across 35 countries, including major cities in Asia, Europe, and Latin America. This coverage allows us to examine how the strength of last-mile and substitution effects varies with baseline transit quality and urban form. It also provides the first comparative international evidence on these mechanisms, showing that the forces we identify operate similarly across highly diverse settings.

There is also a broader literature working to understand the effect of ride-hailing on cities. This includes understanding the effects of surge pricing (Castillo, 2020, Castillo et al., 2021) and taxes on ride-hailing (Tarduno, 2025), as well as the impact of ride-hailing on traffic safety (Greenwood et al., 2017, Burgdorf et al., 2019, Barrios et al., 2020, Barreto et al., 2020, Anderson and Davis, 2021, Harmon and Wozniak, 2023) and congestion (Tarduno, 2021). The role of Uber in local labor markets (Angrist et al., 2021, Hall and Krueger, 2018) and housing costs (Gorback, 2024) have also been studied.

The remainder of this paper is structured as follows. Section 2 presents a simple model illustrating potential complementarities and substitution effects between Uber and transit following a new station opening. Section 3 details the data sources used in this study. Section 4 discusses our methodology and Section 5 presents our main empirical findings on the effect of transit expansions on Uber ridership. Section 6 estimates the parameters of our model based on our reduced-form evidence and considers counterfactuals. Section 7 concludes.

2 Model

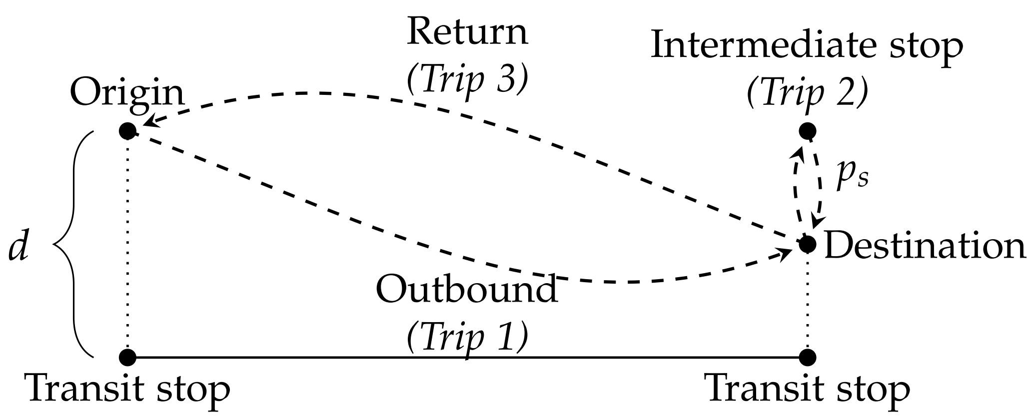
We develop a dynamic discrete-choice model of mode selection among public transit, Uber, and private automobile. Our goals are fourfold. First, provide insight on how choices vary with distance from the transit station. Second, illustrate four mechanisms by which improved transit can affect Uber ridership: substitution, last-mile, insurance, and lock-in. Third, be analytically tractable enough to provide qualitative insights and guide our reduced-form empirical analysis. Fourth, remain flexible enough to rationalize many different responses of Uber ridership to improvements in public transit, allowing us to combine the reduced-form effects with the theoretical framework to estimate different counterfactuals.

2.1 Model setup

Consider a single public transit route with two stops and a risk-neutral representative traveler who needs to go from an origin that is d meters from the first stop to a destination near the second.

Later in the day, she will make the return trip, and, with probability p_s, she will need to make another trip during the day. Figure 1 depicts how trips are linked between origin and destination in our model.

Figure 1: Model geography of trips



Notes: This figure depicts how trips are taken by the representative traveler in our model.

The traveler chooses between traveling using personal automobile, public transit, or Uber. If she uses transit, she must either walk or take Uber to and from the transit stop. If she drives a personal automobile for the outbound trip, she must use the automobile for the return journey but can use any mode for the additional trip. If she does not drive a personal automobile for the outbound trip, she cannot drive for the remainder of the day's trips.

We define the following parameters. Let the generalized in-vehicle base utility of mode m be denoted by α_m for the outbound and return journeys, and β_m for the additional trip during the day. Using access/egress mode a has fixed cost f_a and marginal cost $c_a \cdot d$. We assume walking has no fixed cost, but has a higher marginal cost than taking Uber to the transit stop ($f_{\text{walk}} = 0$; $c_{\text{walk}} > c_{\text{Uber}}$).

The utility of a one-way trip from the origin to the destination depends on the distance to the transit stop (d), the primary transportation mode chosen ($m_i \in \{\text{car}, \text{transit}, \text{Uber}\}$ for period $i \in \{1, 2, 3\}$), the access mode chosen if using transit ($a_i \in \{\text{walk}, \text{Uber}\}$ for $i \in \{1, 3\}$), and a random shock over transit mode preferences.

We model the shock as additive in a persistent component and a transitory component. The persistent component, η_m, is realized at the start of the day and captures shocks to a traveler's preferences across travel modes that persist across both outbound and return trips. Maybe one traveler has a unique preference for driving, or today's destination is unusually well suited to traveling by public transit. The η_m are independent draws from a normal distribution and η_{auto} is normalized to zero.[6]

$$\eta_{\text{Uber}} \sim \mathcal{N}(0, \sigma^2_{\text{Uber}}),$$
$$\eta_{\text{transit}} \sim \mathcal{N}(0, \sigma^2_{\text{transit}}), \text{ and}$$
$$\eta_{\text{car}} \equiv 0.$$

The σ^2_{Uber} and $\sigma^2_{\text{transit}}$ govern the dispersion of the persistent shock.

The transitory components, ϵ_{mai} for $i \in \{1, 3\}$ and ϵ_{m2}, are realized at the start of each period and captures shocks to a traveler's preferences that only affect the current trip. Perhaps the bus is about to arrive, or it is raining, which makes transit unattractive. The transitory shocks are independent and identically distributed random shocks with a type I extreme value distribution.

Thus, the utility for the first period trip is given by

$$U_1(m_1, a_1; d) = \alpha_{m_1} - \mathbb{1}_{m_1=\text{transit}}\left(f_{a_1} + d \cdot c_{a_1}\right) + \eta_m + \epsilon_{ma1}, \tag{1}$$

and the utility of the return trip is similarly given by

$$U_3(m_3, a_3; d) = \alpha_{m_3} - \mathbb{1}_{m_3=\text{transit}}\left(f_{a_3} + d \cdot c_{a_3}\right) + \eta_m + \epsilon_{ma3}. \tag{2}$$

The traveler needs to make an additional trip during the day with probability p_s. The utility of the additional trip during the day only depends on the mode chosen and a transitory shock, ϵ_{m2}, which is drawn from the same distribution as the ϵ_{mai}.

$$U_2(m_2) = \beta_{m_2} + \epsilon_{m2} \tag{3}$$

Two features of the model are worth reiterating. First, the traveler has all transit modes available for the additional trip in period 2 only if they chose car in period 1. Otherwise, they can only choose between public transit and Uber for the additional trip. Second, the traveler *must* return via car in period 3 if they drive their car in period 1.

The traveler chooses a sequence of transit options to maximize expected utility. We can derive the expected utility from each period's mode choice using backward induction, and then find conditional choice probabilities using the familiar logit formulation. Unconditional probabilities are calculated by integrating over the persistent shocks η_m. The details of these calculations are in Appendix C.

2.2 Theoretical implications from improving public transit

We use the model to assess how an improvement in public transit affects Uber ridership for the outbound trip, the return trip, and a potential additional trip later in the day. In the model, this improvement corresponds to an increase in the parameter α_{transit}, which captures the utility of

[6]When we try to fit the parameters of our model to the data, a more relevant interpretation of the η_m are persistent preferences across the larger population of travelers. Also, we do not explicitly model the fact that travelers' true destinations often lie beyond the public transit terminus. This discrepancy is absorbed by the model's error term.

using transit. While our empirical analysis focuses on the opening of new train stations, many cities already operate bus networks that serve these areas. We therefore view the opening of a new train station as a natural and tractable way to model an improvement in overall public transit quality.

As we show below, the model generates competing forces – substitution, complementarity, and composition / option value – that determine the net effect of transit improvements on Uber use. Depending on the strength of these forces, Uber ridership may rise or fall. These distinctions will guide our interpretation of empirical results later in the paper.

2.2.1 Uber ridership for outbound trips

Proposition 1 *The derivative of Uber use in period 1 with respect to the transit utility parameter $\alpha_{transit}$ (conditional on distance d and transit preference η) is:*

$$\frac{\partial}{\partial \alpha_{transit}}\left(P_1(Uber|d,\eta) + P_1(transit{+}Uber|d,\eta)\right)$$
$$= -\underbrace{P_1(Uber|d,\eta)\cdot P_1(transit|d,\eta)}_{\text{Direct substitution effect } (\geq 0)} +$$
$$\underbrace{P_1(transit + Uber|d,\eta)\cdot P_1(transit|d,\eta)\cdot\left[P_1(Uber|d,\eta) + P_1(car|d,\eta)(1+P_3(transit|m_1\neq car,d,\eta))\right]}_{\text{Last-mile complementarity } (\geq 0)} +$$
$$\underbrace{P_1(Uber|d,\eta)\cdot P_1(car|d,\eta)\cdot P_3(transit|m_1\neq car,d,\eta)}_{\text{Option value effect } (\geq 0)}$$

The proof can be found in Appendix C.5. The proposition says that the effect on outbound Uber use is indeterminate and can be decomposed into the effect of three competing factors:

1. **Direct substitution effect**: Improved transit quality makes transit more attractive relative to Uber as the primary mode in period 1. Because both Uber and public transit have the same choice set in period 3, the substitution effect does not depend on forward-looking terms.

2. **Last-mile complementarity**: Some of the shift towards transit creates additional demand for Uber as an access mode, partially offsetting the substitution effect. This is captured by the $P_1(\text{Uber}|d,\eta)$ term. The second term $P_1(\text{car}|d,\eta)(1+P_3(\text{transit}|d,\eta))$ captures substitution from car to last-mile trips. Those users benefit from improved transit in the outbound trip as well as access to improved transit in the return trip, hence their dependence on $P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)$.

3. **Option value effect**: This reflects the traveler's forward-looking behavior where better public transit options for the return trip increase the value of maintaining modal flexibility. Thus, some car users will switch to Uber to access improved public transit in the return trip.

The aggregate market share of Uber in period 1 integrates the terms in Proposition 1 over the distribution of η and across distances d, and is the function of a population-level substitution effect, a population-level last-mile complementarity, and a population-level option value effect.

2.2.2 Uber ridership for return trips

We now use the model to assess the impact of an improvement in public transit on Uber ridership for the return leg.

Proposition 2 *The derivative of Uber use in the return trip (period 3), with respect to the transit utility parameter αtransit (conditional on d and η) is:*

$$\frac{\partial}{\partial \alpha_{transit}}\left(P_3(Uber|d,\eta) + P_3(transit{+}Uber|d,\eta)\right) =$$

$$\underbrace{P_1(car|d,\eta)\cdot\Gamma\cdot[P_3(Uber|m_1 \neq car,d,\eta) + P_3(transit + Uber|m_1 \neq car,d,\eta)]}_{\textit{Composition effect } (\geq 0)} -$$

$$\underbrace{(1-P_1(car|d,\eta))\cdot P_3(Uber|m_1 \neq car,d,\eta)\cdot P_3(transit|m_1 \neq car,d,\eta)}_{\textit{Direct substitution effect } (\geq 0)} +$$

$$\underbrace{(1-P_1(car|d,\eta))\cdot P_3(Uber|m_1 \neq car,d,\eta)\cdot P_3(transit|m_1 \neq car,d,\eta)\cdot P_3(a_3 = Uber|m_3 = transit,d)}_{\textit{Last-mile complementarity } (\geq 0)}$$

where:

- $\Gamma = P_1(transit|d,\eta) + (1-P_1(car|d,\eta))\cdot P_3(transit|m_1 \neq car,d,\eta) \geq 0$

The proof can be found in Appendix C.5. As before, the aggregate market share of Uber in period 3 can be derived by integrating across preferences types and distances, and is the function of a population-level composition effect, population-level direct substitution effect, and population-level last-mile complementarity. The proposition says that the effect on return trip Uber use is indeterminate and can be decomposed into the effect of three competing factors:

1. **Composition effect**: Higher α_{transit} shifts travelers from car to transit/Uber in period 1, increasing the fraction with access to Uber in period 3.

2. **Direct substitution effect**: Improved transit quality makes transit more attractive relative to Uber as the primary mode in period 3.

3. **Last-mile complementarity**: Some of the shift toward transit creates additional demand for Uber as an egress mode, partially offsetting the direct substitution effect.

2.2.3 Uber ridership far from the station

The additional trip during the day allows us to consider how a local transit improvement affects Uber use elsewhere in the city. We assume that the utility of taking transit for this intermediate trip depends partly on the value of using transit near home, so that an improvement in local access also raises the attractiveness of transit farther away. Formally, this is captured by the condition $\beta'_{\text{transit}}(\alpha_{\text{transit}}) \geq 0$. Intuitively, if a new train station opens near your home, then being on transit downtown is more appealing as well, because you know the network is now better connected for your return trip.

Proposition 3 *Consider the three-period mode choice model described above, where $\beta_{transit} = \beta_{transit}(\alpha_{transit})$ is a differentiable function with $\beta'_{transit}(\alpha_{transit}) \geq 0$. The derivative of the traveler's Uber mode share in period 2 (conditional on period 2 occurring) with respect to the transit utility parameter $\alpha_{transit}$ is:*

$$\frac{\partial P_2(Uber|\eta)}{\partial \alpha_{transit}} = \underbrace{P_1(car|\eta) \cdot \Gamma \cdot [P_2(Uber|m_1 \in \{Uber,\ transit\}) - P_2(Uber|car)]}_{\text{Composition effect } (\geq 0)} - \underbrace{\beta'_{transit}(\alpha_{transit}) \cdot \Omega}_{\text{Substitution effect } (\geq 0)}$$

where:

- $\Gamma = P_1(transit|\eta) + (1 - P_1(car|\eta)) \cdot P_3(transit|m_1 \in \{Uber,\ transit\}, d, \eta) \geq 0$
- $\Omega = P_1(car|\eta) \cdot P_2(Uber|car) \cdot P_2(transit|car) + [1 - P_1(car|\eta)] \cdot P_2(Uber|m_1 \in \{Uber,\ transit\}) \cdot P_2(transit|m_1 \in \{Uber,\ transit\}) \geq 0$

The proof is described in Appendix C.5. As for the outbound and return trips, aggregate market shares of Uber ridership in the city are the result of aggregated compositional and substitution effects across distances and riders. The proposition says that the effect on Uber ridership elsewhere in the city is directionally indeterminate. It can be decomposed into the effect of two competing and opposite forces:

1. **Composition effect**: Higher α_{transit} shifts travelers from car to transit/Uber in period 1, increasing the fraction with restricted choice sets in period 2, which increases Uber's mode share.

2. **Substitution effect**: When β_{transit} increases with α_{transit}, transit becomes more attractive relative to Uber in period 2, reducing Uber's mode share.

The proposition has a natural corollary. If the improvement to public transit in period 1 does not improve the value of transit elsewhere in the city (through β_{transit}), then there is no substitution effect. Consequently, the additional trip Uber mode share will unambiguously increase.[7]

[7] Technically, there would be no change in the Uber mode share if $P_1(\text{car}|\eta) \in \{0, 1\}$. So long as the outbound trip does not always or never rely on car, then the Uber mode share will increase for intermediate trips.

2.2.4 Discussion

There are four possible transportation modes in our model: car, Uber, public transit with walking access, and public transit with Uber access. A change in the desirability of public transit yields 12 possible (off-diagonal) substitution patterns between these modes. Of these, only 10 involve either Uber directly or as a complement to transit. Importantly, an improvement in transit cannot lead to more car use, nor does it induce substitution from public transit to Uber. Moreover, we assume that the improvement does not directly affect the access mode decision. After eliminating these cases, we are left with four relevant substitution patterns – all of which correspond to the three effects described above:

(1) From car to Uber *(option value / composition effects)*

(2) From Uber to transit-by-walk *(substitution effect)*

(3) From Uber to transit-by-Uber *(substitution + last-mile complementarity)*

(4) From car to transit-by-Uber *(substitution + option value / composition + last-mile)*

The theoretical results establish that Uber ridership responds in ambiguous ways to transit improvements, depending on the balance of substitution, complementarity, and composition / option value effects. To quantify which effects dominate under plausible parameters, we simulate the model in the next section.

2.3 Model simulations

To illustrate the mechanisms embedded in the model and their empirical relevance, we simulate mode choice responses to improvements in public transit. These simulations trace how Uber usage changes across space and trip types. While stylized, the simulations provide intuition for the various forces at play and guide interpretation of the reduced-form estimates presented in Section 5.

One of our goals is to map the model's predictions onto observable changes in Uber trips at varying distances from a transit station. To do this, we must assign first- and last-mile trips to the station itself (located at $d = 0$). In practice, these access trips are measured with error in the data. To account for this, we assume that the origins and destinations of first- and last-mile trips follow an exponential distribution centered at the station. This assumption implies that the density of such trips decays with distance, consistent with how we expect access trips to be distributed in real settings.

Figure 2 presents three versions of the model, chosen to illustrate the range of empirical patterns it can rationalize. Each column corresponds to a different parameterization. The first row reports average outbound and return trip mode shares under baseline preferences. The second row shows how those mode shares respond to a 0.5-unit increase in α_{transit}, which captures an improvement in public transit. The third row plots the log change in Uber trips – those originating or terminating near the station – induced by the same transit improvement. All figures report values for locations within 1200 meters of the station. Parameter values for each simulation are described in Section C.7.

In the first model, we set low fixed and marginal costs for using Uber as an access mode. This leads to a substantial increase in last-mile Uber use following the transit improvement. Aggregating across trips, the model predicts an increase in Uber usage at all distances from the station. The second model shuts down the last-mile channel by imposing high fixed access costs. It also introduces a higher dispersion of persistent preferences over Uber (σ_{Uber}), increasing the mass of marginal Uber users who switch to transit. In this case, we observe a decline in Uber usage across all distances.

The third model combines low access costs with high preference dispersion. It predicts a sharp increase in Uber use near the station—driven by first- and last-mile behavior—and a decline at intermediate distances, where former Uber users now walk to improved transit. This pattern closely mirrors what we observe empirically in Section 5. We conclude this section by revisiting how the model informs our interpretation of the causal effects estimated in the data.

2.4 Relating the model to data on Uber trips

The model yields clear, distance-indexed predictions for how Uber ridership should respond to the opening of a new transit station. To bring these predictions into our empirical setting, we must determine which *trips* and which *travelers* to include in our research design.[8]

The model provides two key insights that inform this choice: (i) substitution and complementarity effects are strongest for trips originating near a station, and (ii) these effects are most pronounced among individuals who live or work within the station's catchment area. Accordingly, we classify our empirical margins along two dimensions:

Table 1: Classification of empirical margins by trip and traveler proximity

Trips/People	Near station	Anyone
Near station	II	I
Everywhere	III	IV

I: *All trips near station, all travelers.* This margin captures changes in Uber trip volumes geolocated near the new station, irrespective of the traveler's home or work location. It is particularly informative for assessing the extent of last-mile complementarity, as it includes both local and non-local users accessing transit.

II: *Trips near station by travelers near station.* By restricting attention to travelers residing or working within the station's catchment area, this specification isolates behavioral responses

[8] To map model-generated probabilities to our Uber trips data, we aggregate the model's choice probabilities across origins at varying distances from each station, weighting by the geographic area at each distance band. Appendix C.6 describes this mapping in detail.

among local users. It helps identify substitution from Uber to public transit while minimizing the influence of last-mile effects.

III: *All trips by travelers near station.* This margin examines changes in Uber usage across the entire city for the subset of individuals affected by the new station. It provides insight into how overall Uber reliance adjusts for local residents and workers, reflecting both substitution and broader composition effects (as highlighted in Proposition 2).

IV: *All trips by all travelers.* This specification corresponds to aggregate, citywide effects studied in prior literature. As our identification strategy is station-level, we do not focus on this margin directly.

Additional sources of variation allow us to isolate other model mechanisms. For instance, differences in the Uber response during morning versus evening hours – likely reflecting outbound and return trips, respectively – shed light on the relative magnitudes of the option value (Proposition 1) and composition effects (Proposition 2). In Section 6, we estimate the model by fitting it to the causal evidence across these margins.

3 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 and riders 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.

3.1 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. We use this dataset for the analysis in Quadrant I of Table 1.

We also make use of a fixed area hexagon in robustness exercises as well as to identify individuals with a point-of-interest (POI) near the station. The hexagons have an average edge length of 174 m and an average area of 0.105 km^2. Using these hexagons, we construct what we refer to as our POI sample. These are individuals that had 3 or more trips within a hexagon 12 to 7 months prior to the station opening, where the average location of their trips is within 1200 m of the station. We use this dataset for the analyses in Quadrants II and III of Table 1. Both hexagons and concentric ring spatial units are depicted for train openings in Toronto in Figure

A1.[9] We emphasize that throughout our analysis, as per our data agreement, we have access to data aggregated from the individual trip and rider databases, but not the underlying microdata itself.

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. Geography-month observations with less than five trips have been set to zero due to privacy concerns.

3.2 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. (2022), 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. We restrict our sample to transit that does not interact with traffic. This excludes tram and street car station openings. We are left with 650 station openings across 55 cities and 35 countries.

Figure 3 plots the geographic distribution of cities used in our analysis. The size of each circle is proportional to the number of Uber trips we have in our data from that city. The map reveals that transit openings in our sample are heavily concentrated in India and Southeast Asia, Europe, and North America. Table A1 reports the cities in our sample that expanded their rail transit between the date Uber entered and 2018. Figure A2 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.[10]

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 these data.

Table 2 reports summary statistics for a trip data over the analysis window. The first four

[9] These hexagons are defined using the H3 hexagonal hierarchical geospatial indexing system, designed by Uber. The advantage of hexagons is that they have a uniform area, unlike the concentric rings we use in our main analysis. For more information, see https://h3geo.org/. We have data on all hexagons whose centroids are within 1,200 m of a transit station.

[10] Figure A3 plots the distribution of days and months in our sample, and shows no discernible patterns outside of the drop in transit openings during the winter.

rows report mean and median values for the number of pick-ups, the number of drop-offs, the total kilometers traveled, and average trip length. The fifth row reports the share of distance bin - month observations with zero Uber trips. Rows 6 through 10 report the share of trips that occurred during a weekday morning, a weekday afternoon, Friday and Saturday night, and the weekend during the day, respectively. We report these summary statistics across our entire dataset in Column (1), and across four distance bins in Columns (2) through (5).

On average, our data contain 1,659 Uber pick-ups and 1,652 drop-offs in each 100 m distance band by month cell. Uber trips are on average 11.20 kilometers long. Within 100 m of the station opening, 25% of observations are zeros. That same number is 15% for our 1100–1200 m. The median values show that our data is severely right-skewed towards some larger cities and dense locations. The existence of zeros and outliers motivates our use of the Pseudo-Poisson Maximum Likelihood estimator, which we discuss in further detail in Section 4.

3.3 Additional variables

We use data on city centers to explore heterogeneity in our treatment effects. We define the center of each city using United Nations and Social Affairs (2018). To find the footprint of each city, we use the Global Human Settlement Layer (Schiavina et al., 2023) which defines urban clusters "via a logic of cell clusters population size, population and built-up area densities as defined by the stage I of the Degree of Urbanisation." This is measured for 1 km x 1 km cells. For the purpose of finding average values, we define the footprint of a city as the urban cells within 25 km of the city center. We also use the city center and footprints to determine how close a new train station is to the city center. We measure "closeness" in two ways. First, we measure the straight-line distance, and second, we measure the travel time using public transit.[11]

4 Methodology

Section 2.4 introduced three empirical margins in Table 1 – Quadrants I, II, and III – each corresponding to a distinct set of testable implications from the model. For clarity, we begin by outlining our empirical strategy for Quadrant I, which analyzes the effect of new station openings on *station-adjacent* Uber trips – those that start or end near the station – regardless of the rider's home or work location. In contrast, Quadrants II and III focus on riders with a fixed location near the station, allowing us to study substitution effects and broader behavioral adjustments among the affected population. We then describe how the empirical strategies for Quadrants II and III build on the core design to examine traveler-specific outcomes.

Our identification strategy uses a dynamic difference-in-differences approach, leveraging high-frequency Uber trip data and the precise timing of station openings. The first difference compares Uber ridership before and after a station opens, within a given distance band around the

[11] In calculating this, we remove walking time from the last transit stop and the city center

station. The second difference controls for broader trends in Uber usage over time – an important adjustment, since Uber ridership was growing rapidly throughout our sample period. To flexibly account for these local time trends, we use trips occurring at distances between 1100–1200 meters from the station as a reference group.

Quadrant II applies a similar design but restricts attention to Uber trips taken by individuals with a point of interest (POI) – such as home or work – near the station. This removes last-mile behavior and sharpens our focus on substitution effects. Quadrant III instead tracks citywide Uber use by individuals with a POI near the station, capturing how a new station affects broader Uber usage for the affected population. In both cases, we use Uber trips from individuals with POIs farther from the station as the reference group.

We define the research design for Quadrant I as follows. Let y_{dit} denote the number of station-adjacent trips beginngin or ending in distance band d around transit station i during month t. 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.

We model the effect of a new transit station opening on Uber ridership in distance band d^* using observations from both distance band d^* and the farthest distance band $\bar{d} = 1200$ m. The dynamic difference-in-differences specification is as follows:

$$\begin{aligned} y_{dit} = \gamma_{it} + \delta_{di} + \sum_{j \in \{-6,-5,\ldots,6\} \setminus \{-2\}} \alpha_{d^*j} \times \mathbb{1}_{\tau_{it}=j} \times \mathbb{1}_{d=d^*} \\ + \left(\beta_{1,d^*} \times \mathbb{1}_{\tau_{it}<-6} + \beta_{2,d^*} \times \mathbb{1}_{\tau_{it}>6}\right) \mathbb{1}_{d=d^*} + \epsilon_{dit}, \\ \forall\, d \in \{d^*, \bar{d}\}, \quad (4) \end{aligned}$$

where $\mathbb{1}_{\tau_{it}=j}$ is an indicator function for whether the given observation occurs j months after the station opens, $\mathbb{1}_{d=d^*}$ is an indicator function for whether the given observation is at distance d^*, γ_{it} is a station-time fixed effect, and δ_{di} is a station-distance fixed effect. The index j runs from -6 to 6, excluding -2, so the second month prior to the transit station openings is the reference category.[12]

The coefficients of interest are the α_{d^*j}, which in our estimation can be interpreted as the percentage change in Uber trips j months after a new transit station opens. For $j \geq 0$, the α_{d^*j} are the treatment effects we seek to measure at distance d^* in each month after the opening. For $j < 0$, the α_{d^*j} allow us to test whether there are pre-trends in Uber trips before stations open. The coefficients β_{1,d^*} and β_{2,d^*} are event study coefficients for before and after our treatment window. This specification leverages up to 79 months of data per station, while isolating variation around the opening date.

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

We estimate the following model via Poisson Pseudo-Maximum Likelihood (PPML) (Silva and Tenreyro, 2006, 2011). The PPML estimator has several advantages for our setting. First, it allows for a natural interpretation of our treatment effects as percentage changes. This is an attractive feature for our setting since farther away distance bands are mechanically larger in area, and thus contain more trips. Second, it handles zeros and outliers in the dependent variable well. Many 100 m distance band by month observations will contain zero trips. As evidenced in Table 2, 25% and 15% of observations have zero trips in the 0–100 m and 1100–1200 m distance bands, respectively. Lastly, it handles high-dimensional fixed effects well relative to other non-linear models. Our main specification is fully saturated with station by distance and station by calendar time fixed effects. Throughout, 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 in a single table or figure. We will want to report these coefficients for each distance band, and so we rely on the following difference-in-difference specification:

$$\begin{aligned} y_{dit} = \gamma_{it} + \delta_{di} &+ \sum_{d^* \neq 1200 \text{ m}} \alpha_{d^*} \times \mathbb{1}_{\tau_{it} \in \{1,...,6\}} \times \mathbb{1}_{d=d^*} \\ &+ \sum_{d^* \neq 1200 \text{ m}} \left(\beta_{1,d^*} \times \mathbb{1}_{\tau_{it} < -6} + \beta_{2,d^*} \times \mathbb{1}_{\tau_{it} > 6} + \beta_{3,d^*} \times \mathbb{1}_{\tau_{it} \in \{-2,-1,0\}} \right) \mathbb{1}_{d=d^*} + \epsilon_{dit}, \\ & \qquad \forall d, \quad (5) \end{aligned}$$

In this regression, the omitted category is the third through sixth months before the station opened and the coefficients of interest are the α_{d^*}. The α_{d^*} capture the average percentage effect of a station opening on Uber trips at distance d^* relative to $\bar{d} = 1200$ m, for months $\tau_{it} \in \{1, ..., 6\}$ relative to months $\tau_{it} \in \{-6, -5, -4, -3\}$. We separately control for the effect two months prior through 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. (2022) 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. Moreover, we have chosen a narrow study window - six months before and after the station opening - for our analysis. Any effects we observe are most likely driven by the sudden opening of public transit, rather than other changes to public policy. Second, we make the standard difference-in-differences assumptions that Uber ridership at all distances would have moved in parallel in the absence of a new transit station opening. A strength of our approach is the granularity of our data, which uses nearby distances to provide a more plausible counterfactual. We will present evidence that

Uber ridership in the 1100-1200 m band moved in parallel to closer distances prior to the station opening.

Given our station openings occur over the period 2013–2018, a concern might be whether our setting is subject to the usual issues of difference-in-differences estimators in the presence of staggered treatment timing and treatment effect heterogeneity (Goodman-Bacon, 2021, Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Borusyak et al., 2021). Our main specification is fully saturated with station by distance and station by month fixed effects. Consequently, the α_{dj} will be a variance weighted-average of each station's individual difference-in-differences estimate at event time j and will not suffer from contaminated control units. An alternative way to say this is for each station, each distance is subject to the same, contemporaneous treatment timing - including our control group 1100–1200 m away (Baker et al., 2022). This is equivalent to a "stacked" staggered adoption specification (Cengiz et al., 2019).

The specification discussed above uses Uber trips far from a station as a control group for trips near to a station that are impacted by its opening. This design is relevant to the analysis in Quadrant I, which considers the effect of a station opening on Uber trip usage for everyone in the city near the station. Quadrants II and III examine how new station openings affect Uber usage among riders with a fixed point of interest (POI) near the station, both locally (II) and across the city (III). To study these effects, we first bin riders into three groups according to the location of their POI relative to the station: 0–500 m, 500–800 m, and 800–1200 m. We then use riders with a POI 800–1200 m away as a control group for riders in our 0–500 m and 500–800 m groups. We pool bins because of the smaller amount of trips in this sample. The specification is the same as above, except observations y_{dit} refer to the number of trips for a rider with a POI in distance group d from station i during month t. Our identifying assumption is now that riders who live or work 800–1200 m away from the station would have behaved similarly to riders who live or work closer to the station in the absence of the station opening.

While our main outcome of interest is Uber trips, we also consider the effects on pick-ups and drop-offs separately. We also study trips that originate or end at certain times of day - for example, weekday morning, weekday afternoon, or the weekend. We also consider other margins of Uber usage, like total distance traveled and total fares. We use average trip distance as an outcome to consider if trips get longer or shorter after the station opening. These alternative outcomes are considered in Section 5.1.

5 The Effect of a New Station on Uber Ridership

This section presents our core empirical results and interprets them through the lens of the model. We organize the analysis around the three empirical quadrants introduced in Table 1, each corresponding to a distinct behavioral margin. Quadrant I analyzes changes in Uber usage near the station for all riders and serves as the empirical counterpart to the model's predicted shift in aggregate mode shares. Quadrant II focuses on station-adjacent Uber trips taken by individuals with a fixed POI near the station, allowing us to identify substitution behavior between Uber and Transit by abstracting from trips by transient individuals. Finally, Quadrant III examines citywide Uber usage among the same POI sample, capturing broader behavioral responses, including the importance of composition effects. We begin with a detailed analysis of Quadrant I, which both motivates the structural modeling and forms the basis for several key robustness exercises. We then turn to Quadrants II and III, where we highlight how heterogeneity in the response by location and time of day informs the underlying mechanisms in our model.

5.1 Last-mile and substitution responses to transit openings (Quadrant I)

We begin by estimating how overall Uber trips respond to the opening of a new transit station, focusing on variation in distance from the station. Panel A of Figure 4 plots the estimated effect on trips within 0–100 meters of a station, relative to trips within 1100–1200 meters. These estimates correspond to the $\alpha_{0\text{–}100\text{ m},j}$ in Equation 4. We find a large and statistically significant increase in ridership of 42% within one month of the opening (0.35 log points). In the two months prior to opening, there were an average of 650 trips within 100 m of a station, so the estimated effect corresponds to roughly 273 additional trips. Six months after opening, trips had increased by 0.45 log points, and we can rule out increases smaller than 0.33 log points at the 95% level. This effect is persistent and shows no signs of decay over the six-month window.

The smaller increase in month 0 likely reflects partial exposure: stations do not always open on the first day of the month. In addition, travelers may take time to adjust their routines and re-optimize travel behavior. We also observe a modest effect one month prior to opening, consistent with "soft" openings or early access at some stations. Overall, we find a swift and large rise in Uber trips within 100 m of a new transit station.

The immediacy of the ridership response also helps rule out alternative explanations. For example, it is unlikely that new restaurants or retail developments – whose impacts typically unfold gradually – could account for such a sharp jump. While businesses may attempt to coordinate openings with new stations, doing so precisely is difficult. The high frequency of our data allows us to isolate the effects of sudden infrastructure shocks from slower-moving neighborhood change.

The figure also provides evidence in support of our identification strategy. There are no signs of pre-trends in Uber trips from six to three months before the opening, and the large increase in ridership only occurs once the station opens. Our design is well-powered: we can reject pre-treatment differences between the 0–100 m and 1100–1200 m zones greater than 5.8% at the 95% level.

To assess substitution responses, we next examine effects slightly farther from the station, where the model suggests substitution responses may dominate. Panel B of Figure 4 presents the event study for trips 300—400 m from the station, where we do not expect last-mile trip effects to take place. Unlike the sharp increases observed in the immediate vicinity, we observe a moderate but statistically significant decline in Uber usage beginning shortly after the station opens. This pattern is consistent with substitution away from Uber toward improved public transit access, particularly for travelers within comfortable walking distance of the station.

Figure 5 summarizes the average effect during months 1–6 of a new transit station on Uber ridership changes with distance. The reported coefficients reported are those from the model in Equation 5. On average, a new station opening increases Uber trips by 48% or 0.39 log points (s.e. 0.045) within 100 m. Farther out, we find that a new transit station opening increases Uber ridership at 100–200 m by 0.05 log points (a 5% effect, significant at the 95%-level). For distances 200–700 m, we find evidence that new rail openings *decrease* Uber trips (though insignificantly at 400–500 m and 600–700 m). We find significant effects on the order of a 2% decline in trips at these distances. Thus, while Uber users show strong evidence of complementarities at the station, we do find evidence of substitution away from Uber at intermediate distances. This is a point we take up in more detail in the next section.

Beyond these distances there is no detectable effect. Moreover, the magnitude of the estimates converges towards zero. For example, we can rule out an effect of the station opening on Uber trips at 1000–1100 m of greater than 2.1%. We take this as additional evidence in favor of our identification assumption: that 1100–1200 m is sufficiently far from the station as to be unaffected by its opening. Appendix Figure A4 plots the coefficients for the full dynamic difference-in-differences specification for each distance. Importantly, the plots consistently show no pre-trends in the months leading up to the station opening across each distance band.

Together, these findings indicate that new station openings trigger a discrete shift in Uber demand – consistent with strong last-mile complementarities near the station and moderate substitution at intermediate distances.

Stability of the control group 1100–1200 m

Our finding of no effect beyond 600 m, relative to the 1100–1200 m band, suggests that transit–Uber interactions at these distances are minimal. An alternative possibility is that effects are constant across 600–1200 m and thus net out in our design. To test this, we estimate an event study at 1100–1200 m as a placebo check on our identifying assumptions.

This test is motivated by our model, which highlights the importance of the catchment area. If riders are willing to walk long distances to access public transit, then even those far from a station may substitute away from Uber. In North American transportation planning, a 1/2-mile (approximately 800 m) radius is widely used as the default catchment area for rail transit stations (Guerra, 2012). Empirical studies show that ridership tends to decline sharply beyond this range (Agrawal et al., 2008, Daniels and Mulley, 2013). Our control group lies well outside this threshold, but we formally assess its validity below.

We estimate an event study using only observations from the 1100–1200 m band, treating not-yet-treated locations as controls. Following de Chaisemartin and D'Haultfœuille (2024), we include city-by-month fixed effects to absorb secular trends. We use a linear model with the inverse hyperbolic sine of trips (in hundreds) as the outcome.[13] Full specification details are provided in Appendix Section D.1. This design isolates variation at a single distance band and serves as a placebo test: we expect no treatment effect at 1100–1200 m if our identifying assumptions hold.

Results are shown in Figure 6. Consistent with our identifying assumptions, we find no significant changes in Uber trips before or after the opening. The absence of pre-trends is expected, given the exogenous timing of station openings. While this specification lacks the precision and localized trend control of our main difference-in-differences design, it reinforces the validity of the 1100–1200 m band as a counterfactual. Appendix Figures A5 and A6 confirm similar results using the PPML estimator and an IHS-transformed linear model, respectively.

Aggregate effect within 700 m

While we find large increases in Uber ridership very close to new transit stations and small declines at intermediate distances, it is not immediately clear which effect dominates in aggregate. Although the impact is strongest within 0–100 m, that zone covers a much smaller area—for example, the 300–400 m band is seven times larger. Table 2 shows that, on average, the 300–400 m band sees 3.7 times more Uber trips than the 0–100 m zone.

To assess the net effect, we aggregate trips within 700 m and estimate a difference-in-differences model using 1100–1200 m as the control group (as in Equation 4). Appendix Figure 8 plots the estimated treatment effects. We find no statistically significant differences between the treated and control zones, either before or after the station opening. This implies that the positive effects near the station are offset by substitution at intermediate distances. Appendix Figure A16 shows this balance holds across most cities in our sample.

Effects across other ridership margins

Our main outcome thus far has been the total number of Uber trips that either originated or terminated near the station. We now explore whether our results hold across alternative ride metrics. Appendix Figure A13 shows that the effects are nearly identical when estimating impacts on pickups and dropoffs separately. We also find consistent results when examining total fares paid and kilometers traveled, as shown in Appendix Figure A14.

Given our broad sample of 55 cities across four continents, we next test whether the average effects mask regional heterogeneity. Appendix Figure A15 shows that the pattern of last-mile increases and intermediate-distance substitution holds across continents, with the strongest effects observed in Asian cities and attenuated patterns in Europe. To explore this further, we estimate city-specific treatment effects within 0–100 m and 300–400 m of a station and plot the results in funnel graphs (Figure 7). Estimates are centered around the aggregate treatment effect, with most cities – especially those with larger Uber markets – falling within standard confidence

[13]We normalize by 100 so that estimated treatment effects are comparable to those from our PPML specifications. For discussion of scale invariance in IHS transformations, see Chen and Roth (2023).

bounds. While a handful of smaller cities show outlier responses,[14] the core pattern is remarkably consistent across the global sample.

These findings reinforce the external validity of our main results and highlight their applicability across the variety of contexts in our sample.

Effects depend on quality of transit improvements

We test whether the magnitude of our estimated effects depends on the quality of the transit improvement. In particular, we expect that openings which meaningfully improve transit access – by offering faster or more reliable service – are more likely to induce last-mile complementarities near the station and substitution from Uber use farther away.

As a test, we analyze a set of 378 station openings for transit modes that interact with traffic, such as trams and streetcars. While these systems expand public transit coverage, they generally do not offer speed advantages over cars or Uber due to shared road use. These transit openings are not included in our main analysis. Appendix Figure A17 presents the results. Panel A plots dynamic estimates from Equation 4, showing no pre-trends in Uber trips. Panel B reports average effects by distance from Equation 5. We find a modest increase in Uber trips at 200–300 m, but no significant effects at other distances. These findings suggest that the observed responses in our main sample are indeed driven by substantive improvements in transit accessibility, rather than confounding neighborhood changes or measurement artifacts.

Last mile behavior

Uber is a natural complement to public transit for last-mile travel. If the increase in ridership near stations reflects last-mile use, we would expect Uber trips in those areas to be shorter on average. To test this, we compute the average kilometers per trip across all station-month observations (excluding those with zero trips). Estimates from our main specification are shown in Figure 9.

Consistent with last-mile behavior, we find that average trip length declines sharply near the station. Within 100 meters, trips are 0.15 log points (16%) shorter, corresponding to a reduction of 1.8 kilometers. At 100–200 m, trips are 5% shorter. In contrast, trip length increases slightly at intermediate distances – for example, by 0.02 log points (2%) at 400–500 m. This pattern is consistent with substitution away from previous last-mile trips: if travelers previously took Uber for longer trips to access a farther station, but now switch to the new one nearby, the remaining trips at those distances will appear longer on average.

We next examine whether these patterns are stronger for stations where last-mile behavior is likely more important. Terminal stations – those at the end of a transit line – are natural candidates. Appendix Figure A18 reports treatment effects by distance band for terminal and non-terminal stations. Though noisier, estimates for terminal stations are substantially larger. We find an 82% increase in Uber ridership within 100 m (0.60 log points), more than double the 38%

[14] Negative effects are concentrated in a few cities, including Panama City, Prague, and Lisbon.

increase (0.32 log points) observed near non-terminal stations. At 100–200 m, terminal stations induce an additional 0.30 log point increase in trips relative to non-terminal ones.

We further examine whether ridership gains are larger for stations on the urban periphery. Appendix Figure A19 shows that effects within 100 m are significantly larger for more peripheral stations, both when measured by travel time and by kilometers from the central business district (significant at the 5% and 10% levels, respectively).

Finally, we assess whether these increases are concentrated among riders less likely to live or work near the station—those more likely to require a last-mile connection. We explore this directly in the next section by examining Uber usage for individuals with persistent activity near stations using our POI data.

5.2 Substitution for travelers with nearby POIs (Quadrant II)

Quadrant I showed that Uber usage increases sharply within 100 meters of a new transit station and declines modestly at intermediate distances. While these patterns are consistent with a mix of complementarity and substitution, the evidence is inherently spatial and may conflate shifts in who uses Uber with changes in individual behavior. To better isolate substitution responses, we now turn to a sample of riders with a persistent connection to a specific location near the new station.

This section focuses on a subset of Uber trips taken by users whose regular activity (either home, work, or habitual travel) is near the station – what we refer to as the POI sample. These riders have a fixed relationship to the treated area, allowing us to hold composition constant and directly observe how their mode choice responds to improved transit access.

This setting corresponds closely to the substitution mechanism in our model. In the presence of a new transit option, users who previously relied on Uber for travel to or from their POI may now substitute toward transit. Moreover, the location-specific nature of the treatment and fixed POI ensures that observed changes are not driven by shifting demand patterns across neighborhoods.

We begin by examining how Uber usage changes across distance bands for users with a fixed point of interest (POI) near the station, following the structure of our main difference-in-differences analysis. Specifically, we compare trips taken 0–500 meters from the station by users with a POI in that zone to trips taken 800–1200 meters from the station by users with a POI in that corresponding outer zone.[15] Because the sample size is smaller, we pool distance band for the control group. We also pool the 0–100 m and 100–200 m groups due to the resolution of our POI hexes, which are approximately 174 meters in width.[16]

Dynamic estimates from Equation 4 for each distance bin through 700–800 m are presented in

[15]Due to data constraints, we observe Uber trips for POI samples only in two distance bands: 0–500 m and 500–1200 m. However, trips taken 800–1200 m from the station by users with a POI 500–1200 m away will primarily come from those located closer to the 800–1200 m range. While this approximation may introduce some attenuation bias, we do not expect it to materially affect the results.

[16]Even with these adjustments, 110–115 stations (depending on the specification) are ommitted from the regression due to sparsity of the data and our saturated fixed effects.

Table 5. We generally find balanced trends in Uber usage prior to the station opening. While Uber trips are 2% higher within 0–300 m four months before opening, the magnitude is small relative to the post-treatment effects. We also observe some evidence of anticipation effects 200–400 m from the station, consistent with soft openings.

Following the opening, we find a significant and sustained decline in Uber usage. Uber trips fall by 9% at 200–300 m, 8% at 300–400 m, and 6% at 400–500 m. The effects taper off beyond this range, becoming smaller at 500–600 m and nearly disappearing by 600–700 m. To confirm that the treatment effect dissipates beyond the immediate catchment area, we estimate the same model using 1100–1200 m as the control group for 800–900 m, 900–1000 m, and 1000–1100 m bands. We find no significant effects at these distances, as shown in Figure A20.

These results provide strong evidence of substitution away from Uber following transit improvements among travelers with persistent ties to the treated area. Because the POI-based design holds rider composition constant, the observed declines in usage are best interpreted as true behavioral responses to improved access to public transit.

Outbound vs. return trips: heterogeneity by time/day of week

The model highlights two reasons why Uber use may persist in the presence of better transit access: an *option value* channel, in which travelers take Uber for outbound trips to preserve flexibility for the return leg, and a *composition* channel, in which fewer outbound car trips increase subsequent reliance on Uber for return travel. If substitution away from Uber is symmetric across outbound and return trips, then differences across time-of-day can help reveal the relative strength of these two forces.

To investigate this, we explore heterogeneity in our estimated effects by time of day and day of week. We focus on the 300–400 meter band, which exhibited the strongest substitution effects in the previous analysis, and estimate Equation 4 separately for different types of trips. Table 6 reports these results. Column (1) replicates our baseline estimate. Columns (2) and (3) isolate weekday morning (AM) and weekday evening (PM) trips, respectively, which are likely to reflect work-related outbound and return travel. Columns (4) and (5) focus on leisure travel: Friday and Saturday nights, and weekend daytime trips. Column (6) covers other remaining periods.

Across all groups, we find consistent evidence of substitution: Uber usage declines by 6–10% for users with POIs near the station, with limited variation across time blocks. Weekday afternoon trips fall by 12%, compared to a 7.6% decline in weekday morning trips. However, these differences – averaged over the 1–6 months following the station opening – are not statistically significant at conventional levels. To test for differential responses between morning and afternoon trips across all distances within 0–500 meters, we estimate an average difference of 0.2 percentage points (s.e. 0.9). This confirms that substitution patterns are statistically indistinguishable across the two periods.

These results suggest that substitution away from Uber occurs across a broad range of travel contexts. Importantly, we find no meaningful difference between periods likely to capture outbound versus return trips. This pattern is consistent with similar magnitudes of the option value effect (on outbound trips) and the composition effect (on return trips) predicted by the model. We next turn to Quadrant III, which expands the analysis beyond the local area to examine how Uber usage changes citywide for the same set of individuals – allowing us to evaluate broader behavioral shifts, including changes in trip origins and destinations, as well as investigate the empirical relevance of the composition effect.

5.3 Citywide Uber ridership (Quadrant III)

We now turn to Quadrant III, focusing on how the opening of a new transit station affects Uber usage among individuals with a routine destination. Using the same POI sample from the previous section, we examine the total number of Uber trips taken by each user across the city – what we refer to as *citywide* Uber ridership. To better understand the nature of these changes, we disaggregate trips into those that are *local* (originating or ending near the new station) and *non-local* (all other trips).

We assign each user to one of three groups based on the distance between their primary POI and the new station: 0–500 meters, 500–800 meters, and 800–1200 meters. Following the strategy used in the previous section, we treat users whose POI is 800–1200 meters from the station as a control group, under the identifying assumption that – absent the transit expansion – they would have experienced similar trends in Uber use as the closer groups.

Table 7 presents estimates from Equation 4. The first three columns report results for users with POIs within 0–500 meters of the station. Column (1) shows that, prior to the station opening, citywide Uber usage trends for this group track closely with the control group. After the station opens, however, citywide trips decline significantly by 3.3%. Disaggregating this effect, Column (2) shows a 7.5% decline in local trips—confirming the drop in station-adjacent travel observed in the previous section. Column (3) shows that non-local trips increase modestly, with an average post-opening effect of 3.8% (significant at the 10% level), although this estimate should be interpreted cautiously due to pre-treatment imbalances.

The next three columns focus on users with POIs located 500–800 meters from the station. Here, we find little evidence of change in citywide or local Uber usage following the station opening. However, non-local trips rise steadily over time, increasing from 1.7% one month after the opening to 4.4% after six months. The average effect is a 3.4% increase in non-local trips for this group.

These findings help disentangle substitution from composition effects, as outlined in our model. The increase in non-local trips – especially among users located 500–800 meters from the new station – mirrors the dynamics described in our Period 2 analysis and points to the importance of composition effects. In particular, some users may choose not to drive in the morning so they can take advantage of improved transit options for the return leg later in the day. As a result, they rely on both transit and Uber more for midday or intermediate trips that arise. Importantly, while the value of transit access declines with distance, the relevance of the composition margin increases:

users farther from the station are more likely to drive by default due to higher access costs, and thus more likely to change their travel mode when transit is improved. This helps explain why the composition effect – reflected in increased non-local Uber trips – is more stable across distance.

In contrast, the decline in local Uber trips is sharper for those located 0–500 meters from the new station, producing a net decrease in citywide Uber usage for that group. This pattern is consistent with substitution effects being strongest where access costs are lowest and the gains from transit improvements are greatest. Taken together with the previous section's time-of-day results, which showed no significant differences between outbound and return trip periods, the findings suggest that the option value channel – where Uber provides access to desirable transit in the afternoon, as well as a fallback option if it is unreliable – may be of similar magnitude to the composition effect. While our reduced-form evidence cannot fully isolate the two, the model estimated in Section 6 allows us to formally disentangle their relative contributions.

5.4 Robustness

We conduct a series of robustness checks focused on Quadrant I, where statistical power is greatest and behavioral responses aggregate across all margins. Importantly, Quadrant I is the only setting that captures last-mile behavior, making these estimates especially relevant for the model-based analysis in Section 6. We show that the results are consistent across alternative specifications, placebo tests, and variations in geography and event timing.

Placebo tests. We assign each station a fake opening date randomly drawn from the set of dates with sufficient pre/post data (6 months on each side) and re-estimate our main specification 100 times. Appendix Figure A7 compares the distribution of placebo coefficients to the true estimates (blue circles), confirming that transit expansions increase Uber trips near the station and reduce them farther away.

Alternative fixed effects. Tables 3 and 4 show results at 0–100 m and 300–400 m using different sets of fixed effects. Column (3) is our main design, which includes station-by-distance and station-by-month fixed effects to flexibly control for unobserved heterogeneity and hyper-local trends. Column (1) omits these controls, and Column (2) adds city-by-month fixed effects. Despite these variations, treatment effects remain stable and pre-trends are absent, reinforcing the importance of exogenous timing variation for identification. Column (4) adds distance-bin-by-month fixed effects to address concerns about broader neighborhood change. Results are consistent across all specifications.

Structural break test. We follow Andrews (1993, 2003) in estimating the treatment effect under shifted "placebo" opening dates. Appendix Figure A9 shows that the estimated effect is most significant at the true opening date, with a Wald statistic of 66.59, rejecting the null of no structural break at conventional levels.

Adjacent distance comparisons. We re-estimate effects using adjacent 100-meter bands as controls, following a discrete version of the gradient estimator in Diamond and McQuade (2019). Appendix Figure A10 shows both the marginal and aggregate effects of station openings. The results closely track our main estimates in shape, magnitude, and significance.

Geographic robustness. To address concerns about ring geometry, we replicate our analysis using fixed-area hexagonal bins (Appendix Figure A11). Results are muted, as expected, due to spatial averaging within bins, but show similar qualitative patterns.

Time window. Expanding the analysis to an 18-month window (Appendix Figure A12) reveals no pre-trends and, if anything, stronger post-opening effects. Uber trips within 0—100 m increase by 0.60 log points (82%) sixteen months after opening.

6 Model Estimation and Counterfactuals

While our reduced-form findings provide credible evidence of substitution and complementarity across trips and user types, they cannot isolate the relative contributions of specific mechanisms – such as the role of option value versus composition effects – or predict behavior under alternative policy scenarios. We therefore estimate the parameters of the model from Section 2 to match key reduced-form moments from Section 5. This approach allows us to decompose the drivers of Uber–transit interaction and simulate counterfactual policies, such as the effect of Uber exiting the market.

[TBD] ...

7 Conclusion

There is an ongoing debate over whether ride-hailing complements or substitutes for public transportation. The answer has important policy implications for the regulation and taxation of ride-hailing platforms, for transit service and infrastructure planning, and for the design of transit–ride-hailing partnerships. Despite a growing literature, there remains substantial uncertainty, as existing estimates vary widely in both sign and magnitude.

We develop a dynamic discrete-choice model of transit choice to clarify the mechanisms underlying these mixed findings. The model shows that ride-hailing can both complement and substitute for public transportation, with the direction and magnitude of the response depending on where travelers are located and how they use the transportation network. Improvements in public transit generate multiple, offsetting responses in ride-hailing demand that vary systematically over space and across travelers.

We study these predictions empirically using data on the universe of Uber trips combined with 650 rail station openings across 55 cities in 35 countries. Our identification strategy compares changes in Uber usage within concentric distance bands around newly opened stations, using trips 1100–1200 meters away as a local control group. This dynamic difference-in-differences design leverages the quasi-random timing of station openings – planned years in advance, typically before Uber existed – while flexibly controlling for hyper-local time trends in ride-hailing activity.

We document three patterns in how Uber ridership responds to new train stations. First, Uber usage rises sharply very close to new stations: within 100 meters, trips increase by 48%. These trips are short and are disproportionately taken by travelers without persistent ties to the area, consistent with first- and last-mile use of Uber to access transit. Second, at intermediate distances of roughly 300–700 meters, Uber usage declines by about 2%, reflecting substitution toward improved transit among nearby residents and workers. Third, when aggregated across the station catchment area, these effects largely offset one another, resulting in no detectable change in total Uber trips within 1200 meters of the station.

By following individuals with a point-of-interest near station areas, we test for true behavioral changes in Uber ridership. Among travelers who live or work within 500 meters of a new station, Uber usage near the station falls by 6–10%. At the same time, their overall citywide Uber usage declines by only 3%, as reductions near the station are partially offset by increased ride-hailing elsewhere in the city. This pattern is consistent with the model's prediction that outbound mode choice constrains subsequent travel options over the day: travelers who switch from driving to transit for the outbound leg become more reliant on transit and ride-hailing for later trips.

These findings help reconcile mixed evidence in the literature. Complementarity arises in settings where last-mile frictions are binding, while substitution dominates among travelers with good direct access to transit. Both forces operate simultaneously, and the net effect depends on the spatial distribution of travelers relative to transit infrastructure.

From a policy perspective, the results imply that transit expansions need not uniformly reduce or increase ride-hailing demand. Some effects are highly localized to transit improvements, and their fiscal and welfare implications depend on the relative magnitudes of last-mile usage near stations and substitution by nearby residents. Transit improvements may also affect ride-hailing demand beyond the station area by increasing the value of flexible travel modes that allow travelers to rely on transit when it is available and revert to ride-hailing when it is not.

Several limitations remain. Our estimates capture short-run responses within six months of station openings and may not reflect longer-run adjustments in residential location, car ownership, or travel patterns. We focus on how transit expansions affect ride-hailing demand, rather than the reverse relationship. Extending the analysis to incorporate welfare effects, congestion, and environmental externalities remains an important direction for future research.

More broadly, the results highlight the complicated interactions between new transportation technologies and existing infrastructure. Understanding these interactions is essential for predicting how continued transit investment and emerging technologies (for example, autonomous vehicles) will shape future transportation networks.

Bibliography

Agrawal, Atin, Norbert Schüssler, and Alexander Erath (2008) "Exploring the Distances People Walk to Access Public Transport," *Journal of Transport and Land Use*, 1 (2), 43–63, 10.5198/jtlu.v1i2.45, Mean walking distance to LRT stations: 756 m.

Agrawal, David R. and Weihua Zhao (2023) "Taxing Uber," *Journal of Public Economics*, 221 (C), 104862, 10.1016/j.jpubeco.2023.104862.

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*.

Andrews, Donald W. K. (1993) "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61 (4), 821–856, 10.2307/2951764.

——— (2003) "Tests for Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum," *Econometrica*, 71 (1), 395–397, 10.1111/1468-0262.00405.

Angrist, Joshua D., Sydnee Caldwell, and Jonathan V. Hall (2021) "Uber versus Taxi: A Driver's Eye View," *American Economic Journal: Applied Economics*, 13 (3), 272–308, 10.1257/app.20190655.

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.

Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang (2022) "How Much Should We Trust Staggered Difference-In-Differences Estimates?" *Journal of Financial Economics*, 144 (2), 370–395, 10.1016/j.jfineco.2022.01.004.

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.

Billings, Stephen B. (2011) "Estimating the value of a new transit option," *Regional Science and Urban Economics*, 41 (6), 525–536, https://doi.org/10.1016/j.regsciurbeco.2011.03.013.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2021) "Revisiting Event Study Designs: Robust and Efficient Estimation," *arXiv preprint arXiv:2108.12419*, https://arxiv.org/abs/2108.12419.

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*.

Callaway, Brantly and Pedro H. C. Sant'Anna (2021) "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225 (2), 200–230, 10.1016/j.jeconom.2020.12.001.

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.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer (2019) "The Effect of Minimum Wages on Low-Wage Jobs," *The Quarterly Journal of Economics*, 134 (3), 1405–1454, 10.1093/qje/qjz014.

de Chaisemartin, Clément and Xavier D'Haultfœuille (2024) "Difference-in-Differences Estimators of Intertemporal Treatment Effects," *The Review of Economics and Statistics*, 10.1162/rest_a_01414.

Chen, Jiafeng and Jonathan Roth (2023) "Logs with zeros? Some problems and solutions," *arXiv preprint arXiv:2212.06080*, https://arxiv.org/abs/2212.06080.

Christensen, Peter, Gustavo Nino, and Adam Osman (2023) "The Demand for Mobility: Evidence from an Experiment with Uber Riders," Working Paper 31330, National Bureau of Economic Research, 10.3386/w31330.

Daniels, Rhonda and Corinne Mulley (2013) "Explaining Walking Distance to Public Transport: The Dominance of Public Transport Supply," *Journal of Transport and Land Use*, 6 (2), 59–72, 10.5198/jtlu.v6i2.308, Mean walking distance to train stations: 805 m (Sydney).

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 (2022) "Subways and Urban Air Pollution," *American Economic Journal: Applied Economics*, 14 (1), 164–196.

Gonzalez-Navarro, Marco and Matthew A. Turner (2018) "Subways and Urban Growth: Evidence from Earth," *Journal of Urban Economics*, 108, 85–106, 10.1016/j.jue.2018.09.002.

Goodman-Bacon, Andrew (2021) "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225 (2), 254–277, 10.1016/j.jeconom.2021.03.014.

Gorback, Caitlin (2024) "Ridesharing and the Redistribution of Economic Activity," *Journal of Urban Economics*, Forthcoming.

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.

Guerra, Erick (2012) "Planning for cars that drive themselves: Metropolitan planning organizations, regional transportation plans, and autonomous vehicles," *Journal of Planning Education and Research*, 36 (2), 210–224, 10.1177/0739456X15613591.

Gupta, Arpit, Stijn Van Nieuwerburgh, and Constantine Kontokosta (2022) "Take the Q Train: Value Capture of Public Infrastructure Projects," *Journal of Urban Economics*, 129, 103422, 10.1016/j.jue.2022.103422.

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.

Hall, Jonathan V. and Alan B. Krueger (2018) "An Analysis of the Labor Market for Uber's Driver-Partners in the United States," *ILR Review*, 71 (3), 705–732, 10.1177/0019793917717222.

Harmon, Nikolaj A. and Przemysław Marc Wozniak (2023) "Uber and Traffic Fatalities," *The Review of Economics and Statistics*, 10.1162/rest_a_01385.

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.

King, David, Matthew Wigginton, and Deborah Salon (2020) "Do For-Hire Vehicles Provide First Mile/Last Mile Access to Transit?" *Findings Press*, https://findingspress.org/article/12872-do-for-hire-vehicles-provide-first-mile-last-mile-access-to-transit, Accessed: 2025-01-28.

Li, Shengxiao(Alex), Wei Zhai, Junfeng Jiao, and Chao (Kenneth) Wang (2022) "Who loses and who wins in the ride-hailing era? A case study of Austin, Texas," *Transport Policy*, 120, 130–138, https://doi.org/10.1016/j.tranpol.2022.03.009.

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.

Schiavina, Marcello, Michele Melchiorri, and Martino Pesaresi (2023) "GHS-SMOD R2023A - GHS settlement layers, application of the Degree of Urbanisation methodology (stage I) to GHS-POP R2023A and GHS-BUILT-S R2023A, multitemporal (1975-2030)," 10.2905/A0DF7A6F-49DE-46EA-9BDE-563437A6E2BA.

Siddiq, Auyon, Christopher Tang, and Jingwei Zhang (2023) "Is Uber the Answer to Public Transit's Last-Mile Problem?," https://anderson-review.ucla.edu/last-mile/, Accessed: 2025-01-28.

Silva, Joao M. C. Santos and Silvana Tenreyro (2006) "The Log of Gravity," *The Review of Economics and Statistics*, 88 (4), 641–658, 10.1162/rest.88.4.641.

——— (2011) "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator," *Economics Letters*, 112 (2), 220–222, 10.1016/j.econlet.2011.05.008.

Sun, Liyang and Sarah Abraham (2021) "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Journal of Econometrics*, 225 (2), 175–199, 10.1016/j.jeconom.2020.09.006.

Tarduno, Matthew (2021) "The congestion costs of Uber and Lyft," *Journal of Urban Economics*, 122, 103318, https://doi.org/10.1016/j.jue.2020.103318.

——— (2025) "Elasticities and tax incidence in urban ridesharing markets: Evidence from Chicago," *Journal of Urban Economics*, 150, 103809, https://doi.org/10.1016/j.jue.2025.103809.

Tsivanidis, Nick (2024) "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," https://www.nicktsivanidis.com/research.

Uber (2021) "Uber Last Mile: New Stations," https://www.uber.com/en-DE/blog/uber-last-mile-new-stations/, Accessed: 2025-01-28.

——— (2023) "Uber Last Mile: Phase Two," https://www.uber.com/en-DE/blog/uber-last-mile-phase-two/, Accessed: 2025-01-28.

United Nations, Department of Economic and Population Division Social Affairs (2018) "World Urbanization Prospects: The 2018 Revision, Online Edition. File 22: Annual Population of Urban Agglomerations with 300,000 Inhabitants or More in 2018, by Country, 1950-2035 (thousands)," https://population.un.org/wup/.

Yan, Xiang, Jonathan Levine, and Xilei Zhao (2019) "Integrating ridesourcing services with public transit: An evaluation of traveler responses combining revealed and stated preference data," *Transportation Research Part C: Emerging Technologies*, 105, 683–696, https://doi.org/10.1016/j.trc.2018.07.029.

Zárate, Román David (2024) "Spatial Misallocation, Informality and Transit Improvements: Evidence from Mexico City," https://www.romandavidzarate.com/research.

Tables

Table 2: Uber trips descriptive statistics

Summary Statistics					
	Full: $t \in [-6, 6]$	*0-100m*	*3-400m*	*7-800m*	*11-1200m*
Uber trips	mean [median]	mean [med.]	mean [med.]	mean [med.]	mean [med.]
Pickups	1,659 [188]	417 [47]	1,527 [181]	1,858 [255]	2,106 [227]
Dropoffs	1,652 [187]	423 [47]	1,580 [188]	1,824 [247]	2,088 [227]
Kilometers	29,113 [3,979]	7,698 [940]	28,592 [3,882]	31,616 [5,275]	35,722 [5,003]
Km/trip	11.20 [10.06]	10.24 [9.10]	11.31 [10.12]	11.33 [10.23]	11.32 [10.12]
Zero Trips Shr	0.16	0.25	0.15	0.14	0.15
Weekday AM Shr	0.09 [0.09]	0.08 [0.08]	0.09 [0.09]	0.09 [0.09]	0.09 [0.09]
Weekday PM Shr	0.11 [0.13]	0.11 [0.12]	0.12 [0.13]	0.12 [0.13]	0.11 [0.13]
Fri-Sat Night Shr	0.08 [0.07]	0.06 [0.06]	0.08 [0.07]	0.09 [0.07]	0.08 [0.08]
Weekend Day Shr	0.11 [0.11]	0.10 [0.11]	0.11 [0.11]	0.11 [0.12]	0.11 [0.12]
N	100,032	8336	8336	8336	8336

Notes: This table contains summary statistics on Uber trips in our sample. All data is from the study window of 6 months prior and after the station opening. Each column contains means and medians (in brackets) for some relevant Uber trip outcomes. Column (1) is for our full sample through 1200 meters away from the station. Columns (2) through (5) disaggregate those means by four 100 meter distance bands. Rows (1) through (4) contains average counts for the number of pick-ups, drop-offs, kilometers travelled, and average trip length. Row (5) contains the share of month-distance band cells that had zero trips in them. This includes true zeros as well as those that had fewer than 5 trips, and so were censored. Rows (6) through (9) contain the share of trips during different parts of the week. Row (10) contains observation counts.

Table 3: Difference-in-differences estimates for 0–100 m with varied fixed effects

	(1) Uber trips	(2) Uber trips	(3) Uber trips	(4) Uber trips
-6 months after	-0.00966	0.00671	0.00676	0.0399
	(0.0262)	(0.0260)	(0.0259)	(0.0304)
-5 months after	-0.00417	0.00506	-0.00327	0.0239
	(0.0223)	(0.0212)	(0.0234)	(0.0263)
-4 months after	0.0177	0.00980	0.00756	0.0229
	(0.0211)	(0.0181)	(0.0194)	(0.0205)
-3 months after	0.0125	0.00807	0.0124	0.0201
	(0.0157)	(0.0141)	(0.0160)	(0.0169)
-1 months after	0.0324	0.0365*	0.0482**	0.0478**
	(0.0233)	(0.0188)	(0.0225)	(0.0211)
Month of expansion	0.141***	0.161***	0.180***	0.177***
	(0.0285)	(0.0266)	(0.0292)	(0.0296)
1 months after	0.328***	0.343***	0.350***	0.342***
	(0.0415)	(0.0417)	(0.0428)	(0.0429)
2 months after	0.336***	0.366***	0.366***	0.350***
	(0.0437)	(0.0440)	(0.0432)	(0.0432)
3 months after	0.353***	0.379***	0.372***	0.351***
	(0.0459)	(0.0470)	(0.0456)	(0.0460)
4 months after	0.391***	0.410***	0.396***	0.375***
	(0.0494)	(0.0488)	(0.0454)	(0.0459)
5 months after	0.407***	0.427***	0.406***	0.386***
	(0.0605)	(0.0584)	(0.0556)	(0.0560)
6 months after	0.436***	0.469***	0.451***	0.423***
	(0.0629)	(0.0624)	(0.0603)	(0.0610)
Observations	72,618	70,998	58,298	58,298
Station Openings	635	634	608	608
Station x Dist FE	✓	✓	✓	✓
Cal. and Evt. Time FE	✓	✓	✓	✓
City x Date FE		✓		
Station x Date FE			✓	✓
Dist x Date FE				✓

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table contains coefficient estimates and standard errors from equation (4) for trips 0–100 meters from the station under varying fixed effects. These coefficients estimate the effect of a new train station opening on Uber trips 0–100 meters away in every month leading up to and after the station opening relative to Uber trips 1100–1200 meters away. Coefficients are relative to two months prior to the station opening. All specifications include station and calendar and event time fixed effects. Column (2) adds city by calendar time fixed effects. Column (3) adds station by date fixed effects, and is our main specification. Column (4) includes distance by date fixed effects. All errors are clustered at the station-level.

Table 4: Difference-in-differences estimates for 300–400 m with varied fixed effects

	(1) Uber trips	(2) Uber trips	(3) Uber trips	(4) Uber trips
-6 months after	-0.0117	0.000533	0.00584	-7.43e-06
	(0.0137)	(0.0107)	(0.0113)	(0.0114)
-5 months after	0.00198	0.00527	0.00755	0.00394
	(0.0118)	(0.00843)	(0.00873)	(0.00892)
-4 months after	0.00400	0.00563	0.00864	0.00817
	(0.0119)	(0.00935)	(0.00979)	(0.00998)
-3 months after	0.00527	0.00476	0.00603	0.00609
	(0.00813)	(0.00627)	(0.00658)	(0.00650)
-1 months after	-0.00716	-0.00964	-0.00998	-0.00825
	(0.0120)	(0.00854)	(0.00674)	(0.00690)
Month of expansion	-0.0297***	-0.0251**	-0.0163*	-0.0134
	(0.0110)	(0.0104)	(0.00918)	(0.00927)
1 months after	-0.0472***	-0.0431***	-0.0255**	-0.0208**
	(0.0130)	(0.0120)	(0.0110)	(0.0106)
2 months after	-0.0562***	-0.0447***	-0.0299**	-0.0235*
	(0.0152)	(0.0139)	(0.0130)	(0.0125)
3 months after	-0.0533***	-0.0437***	-0.0340**	-0.0267**
	(0.0148)	(0.0140)	(0.0135)	(0.0132)
4 months after	-0.0444**	-0.0385**	-0.0302**	-0.0220
	(0.0180)	(0.0157)	(0.0140)	(0.0152)
5 months after	-0.0654***	-0.0627***	-0.0632***	-0.0546***
	(0.0229)	(0.0167)	(0.0154)	(0.0159)
6 months after	-0.0732***	-0.0556***	-0.0482***	-0.0381**
	(0.0234)	(0.0150)	(0.0135)	(0.0151)
Observations	73,407	71,706	60,624	60,624
Station Openings	637	637	620	620
Station x Dist FE	✓	✓	✓	✓
Cal. and Evt. Time FE	✓	✓	✓	✓
City x Date FE		✓		
Station x Date FE			✓	✓
Dist x Date FE				✓

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: This table contains coefficient estimates and standard errors from equation (4) for trips 300–400 meters from the station under varying fixed effects. These coefficients estimate the effect of a new train station opening on Uber trips 300–400 meters away in every month leading up to and after the station opening relative to Uber trips 1100–1200 meters away. Coefficients are relative to two months prior to the station opening. All specifications include station and calendar and event time fixed effects. Column (2) adds city by calendar time fixed effects. Column (3) adds station by date fixed effects, and is our main specification. Column (4) includes distance by date fixed effects. All errors are clustered at the station-level.

Table 5: Difference-in-differences estimates for station-adjacent trips in POI sample

	(1) Log Uber trips	(2) Log Uber trips	(3) Log Uber trips	(4) Log Uber trips	(5) Log Uber trips	(6) Log Uber trips	(7) Log Uber trips
-6 months after	0.0199*	0.0109	0.00415	0.00766	-0.0151***	-0.000494	-0.00732
	(0.0118)	(0.0149)	(0.01000)	(0.00727)	(0.00551)	(0.00542)	(0.00515)
-5 months after	0.0135	0.0159**	0.00385	0.00343	-0.00946*	-0.00558	-0.00423
	(0.00917)	(0.00736)	(0.00634)	(0.00589)	(0.00555)	(0.00552)	(0.00431)
-4 months after	0.0204**	0.0205**	-0.00208	-0.00257	-0.00648	-0.00409	-0.00709*
	(0.00952)	(0.0102)	(0.00679)	(0.00633)	(0.00535)	(0.00488)	(0.00418)
-3 months after	0.00923	0.0165	0.00422	0.00208	-0.00244	-0.00190	0.00456
	(0.00764)	(0.0130)	(0.00545)	(0.00457)	(0.00403)	(0.00366)	(0.00294)
-1 months after	-0.00651	-0.0287**	-0.0225***	-0.00672	-0.00201	-0.00183	0.00388
	(0.00889)	(0.0119)	(0.00575)	(0.00471)	(0.00465)	(0.00393)	(0.00372)
Month of expansion	-0.0281***	-0.0526***	-0.0384***	-0.0228***	-0.00790*	-0.00706	-0.00582
	(0.00839)	(0.0133)	(0.00814)	(0.00515)	(0.00476)	(0.00502)	(0.00423)
1 months after	-0.0457***	-0.0848***	-0.0688***	-0.0520***	-0.0158***	-0.0120**	-0.00556
	(0.0110)	(0.00821)	(0.00688)	(0.00571)	(0.00523)	(0.00531)	(0.00431)
2 months after	-0.0334***	-0.0829***	-0.0641***	-0.0523***	-0.0184***	-0.00873	0.00101
	(0.0119)	(0.00939)	(0.00741)	(0.00679)	(0.00648)	(0.00563)	(0.00492)
3 months after	-0.0512***	-0.0832***	-0.0685***	-0.0530***	-0.00895	-0.00140	0.00955**
	(0.0129)	(0.0106)	(0.00869)	(0.00749)	(0.00654)	(0.00590)	(0.00470)
4 months after	-0.0511***	-0.0973***	-0.0827***	-0.0635***	-0.0121	-0.00771	-0.00208
	(0.0159)	(0.0133)	(0.00899)	(0.00901)	(0.00754)	(0.00598)	(0.00587)
5 months after	-0.0460**	-0.0890***	-0.0784***	-0.0544***	-0.0141*	-0.00339	-0.00533
	(0.0181)	(0.0112)	(0.0106)	(0.0105)	(0.00781)	(0.00735)	(0.00677)
6 months after	-0.0503**	-0.0941***	-0.0902***	-0.0645***	-0.0126	-0.00368	-0.00868
	(0.0199)	(0.0143)	(0.0121)	(0.0105)	(0.00801)	(0.00655)	(0.00688)
Observations	135,470	116,321	117,025	117,361	118,835	118,267	117,787
Distance	0–200 m	200–300 m	300–400 m	400–500 m	500–600 m	600–700 m	700–800 m
Station Openings	540	539	542	538	539	536	535
Station x Dist FE	✓	✓	✓	✓	✓	✓	✓
Station x Date FE	✓	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: This table contains coefficient estimates and standard errors from equation (4) for station-adjacent trips in 100 meter bins from the station in the POI sample. These coefficients estimate the effect of a new train station opening on POI sample Uber trips in 100 meter bins from the station in every month leading up to and after the station opening relative to POI sample Uber trips 800–1200 meters away. Coefficients are relative to two months prior to the station opening. All specifications include station by distance, station by date and event time fixed effect. All errors are clustered at the station-level.

Table 6: DD estimates for station-adjacent trips in POI sample by time/day of week

	(1) Log Uber trips	(2) Log Uber trips	(3) Log Uber trips	(4) Log Uber trips	(5) Log Uber trips	(6) Log Uber trips
-6 months after	0.00415	0.0122	0.00346	0.0233	-0.00808	0.00314
	(0.01000)	(0.0142)	(0.0157)	(0.0162)	(0.0124)	(0.0106)
-5 months after	0.00385	0.00741	0.00304	0.00688	0.00974	0.00329
	(0.00634)	(0.0139)	(0.0108)	(0.0129)	(0.0121)	(0.00673)
-4 months after	-0.00208	-0.00648	0.000129	0.00221	-0.00328	-0.00208
	(0.00679)	(0.0126)	(0.0121)	(0.0129)	(0.0107)	(0.00683)
-3 months after	0.00422	0.00739	0.00414	0.00857	-0.00734	0.00601
	(0.00545)	(0.0101)	(0.0101)	(0.0119)	(0.0114)	(0.00586)
-1 months after	-0.0225***	-0.0106	-0.0287***	-0.0288**	-0.0287**	-0.0210***
	(0.00575)	(0.0101)	(0.00986)	(0.0128)	(0.0117)	(0.00636)
Month of expansion	-0.0384***	-0.0390***	-0.0384***	-0.0354**	-0.0418***	-0.0383***
	(0.00814)	(0.0112)	(0.0126)	(0.0146)	(0.0119)	(0.00931)
1 months after	-0.0688***	-0.0696***	-0.0708***	-0.0526***	-0.0887***	-0.0671***
	(0.00688)	(0.0136)	(0.0130)	(0.0132)	(0.0119)	(0.00759)
2 months after	-0.0641***	-0.0607***	-0.0725***	-0.0602***	-0.0791***	-0.0612***
	(0.00741)	(0.0142)	(0.0133)	(0.0134)	(0.0117)	(0.00845)
3 months after	-0.0685***	-0.0577***	-0.0853***	-0.0564***	-0.0884***	-0.0662***
	(0.00869)	(0.0155)	(0.0117)	(0.0156)	(0.0142)	(0.00983)
4 months after	-0.0827***	-0.0778***	-0.0962***	-0.0755***	-0.0779***	-0.0831***
	(0.00899)	(0.0179)	(0.0128)	(0.0136)	(0.0136)	(0.00982)
5 months after	-0.0784***	-0.0768***	-0.0962***	-0.0660***	-0.0810***	-0.0772***
	(0.0106)	(0.0177)	(0.0154)	(0.0184)	(0.0152)	(0.0117)
6 months after	-0.0902***	-0.0756***	-0.116***	-0.0845***	-0.0888***	-0.0892***
	(0.0121)	(0.0200)	(0.0157)	(0.0200)	(0.0181)	(0.0133)
Observations	117,025	80,816	84,739	81,045	81,108	108,821
Distance	300–400 m	300–400 m	300–400 m	300–400 m	300–400 m	300–400 m
Time Block	All	Weekday AM	Weekday PM	Fri/Sat night	Weekend day	Rest of week
Station Openings	542	451	449	446	450	516
Station x Dist FE	✓	✓	✓	✓	✓	✓
Station x Date FE	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: This table contains coefficient estimates and standard errors from equation (4) for trips 300–400 meters from the station across various times/days of the week in the POI sample. These coefficients estimate the effect of a new train station opening on POI sample Uber trips in 300–400 meters from the station in every month leading up to and after the station opening relative to POI sample Uber trips 800–1200 meters away. Coefficients are relative to two months prior to the station opening. All specifications include station by distance, station by date and event time fixed effect. All errors are clustered at the station-level.

Table 7: Difference-in-differences estimates for citywide, local and non-local trips in POI sample

	(1) Log Uber trips Citywide	(2) Log Uber trips Local	(3) Log Uber trips Non-Local	(4) Log Uber trips Citywide	(5) Log Uber trips Local	(6) Log Uber trips Non-Local
-6 months after	-0.00183	0.0102	-0.0355**	-0.00466	-0.00739**	-0.00312
	(0.00608)	(0.00877)	(0.0171)	(0.00311)	(0.00373)	(0.00895)
-5 months after	-0.00435	0.00913*	-0.0346**	-0.00521*	-0.00621*	-0.0136*
	(0.00393)	(0.00496)	(0.0171)	(0.00294)	(0.00369)	(0.00734)
-4 months after	-8.02e-06	0.00718	-0.0162	-0.00365	-0.00571	-0.00713
	(0.00409)	(0.00616)	(0.0125)	(0.00294)	(0.00374)	(0.00707)
-3 months after	0.00347	0.00832	-0.00873	-0.000744	-0.000188	-0.00524
	(0.00414)	(0.00604)	(0.00968)	(0.00223)	(0.00249)	(0.00565)
-1 months after	-0.00979***	-0.0165***	0.00471	-0.00227	1.44e-05	0.00221
	(0.00358)	(0.00533)	(0.00930)	(0.00266)	(0.00322)	(0.00600)
Month of expansion	-0.0187***	-0.0359***	0.00740	-0.00618**	-0.00725**	0.00484
	(0.00418)	(0.00633)	(0.0105)	(0.00294)	(0.00328)	(0.00722)
1 months after	-0.0348***	-0.0629***	0.00529	-0.00609*	-0.0112***	0.0168**
	(0.00421)	(0.00512)	(0.0132)	(0.00318)	(0.00355)	(0.00782)
2 months after	-0.0310***	-0.0587***	0.0112	-0.00367	-0.00891**	0.0262***
	(0.00519)	(0.00639)	(0.0140)	(0.00335)	(0.00404)	(0.00943)
3 months after	-0.0332***	-0.0634***	0.00428	0.000762	0.000107	0.0170*
	(0.00527)	(0.00653)	(0.0157)	(0.00311)	(0.00412)	(0.00912)
4 months after	-0.0338***	-0.0745***	0.0207	-0.00271	-0.00754	0.0303***
	(0.00482)	(0.00721)	(0.0196)	(0.00433)	(0.00494)	(0.0111)
5 months after	-0.0328***	-0.0680***	0.0211	-0.00234	-0.00795	0.0392***
	(0.00553)	(0.00815)	(0.0261)	(0.00509)	(0.00568)	(0.0140)
6 months after	-0.0349***	-0.0778***	0.0263	-0.00153	-0.00852*	0.0437***
	(0.00619)	(0.00922)	(0.0223)	(0.00526)	(0.00516)	(0.0138)
Observations	49,830	48,568	49,434	51,346	49,938	50,326
Distance	0–500 m	0–500 m	0–500 m	500–800 m	500–800 m	500–800 m
Avg. Effect (%)	-3.26***	-7.54***	3.78*	0.09	-0.23	3.43***
Station Openings	508	522	508	533	539	530
Station x Dist FE	✓	✓	✓	✓	✓	✓
Station x Date FE	✓	✓	✓	✓	✓	✓

Robust standard errors in parentheses

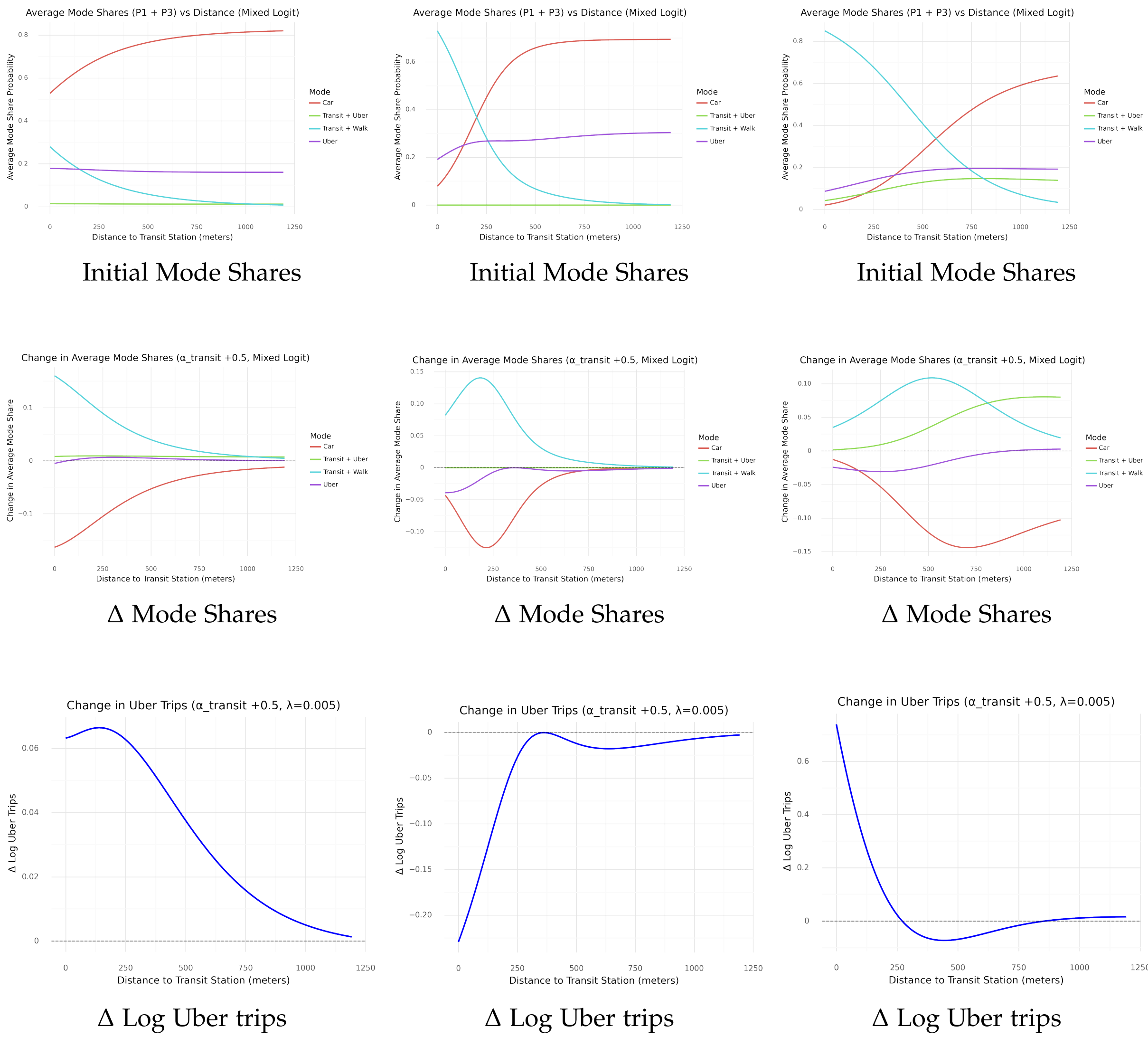
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table contains coefficient estimates and standard errors from equation (4) for citywide, local, and non-local trips, for those with a POI 0–500 meters and 500–800 meters from the station. These coefficients estimate the effect of a new train station opening on citywide, local, and non-local POI sample Uber trips in every month leading up to and after the station opening relative to POI sample Uber trips 800–1200 meters away. Coefficients are relative to two months prior to the station opening. All specifications include station by distance, station by date and event time fixed effect. All errors are clustered at the station-level.

Figures

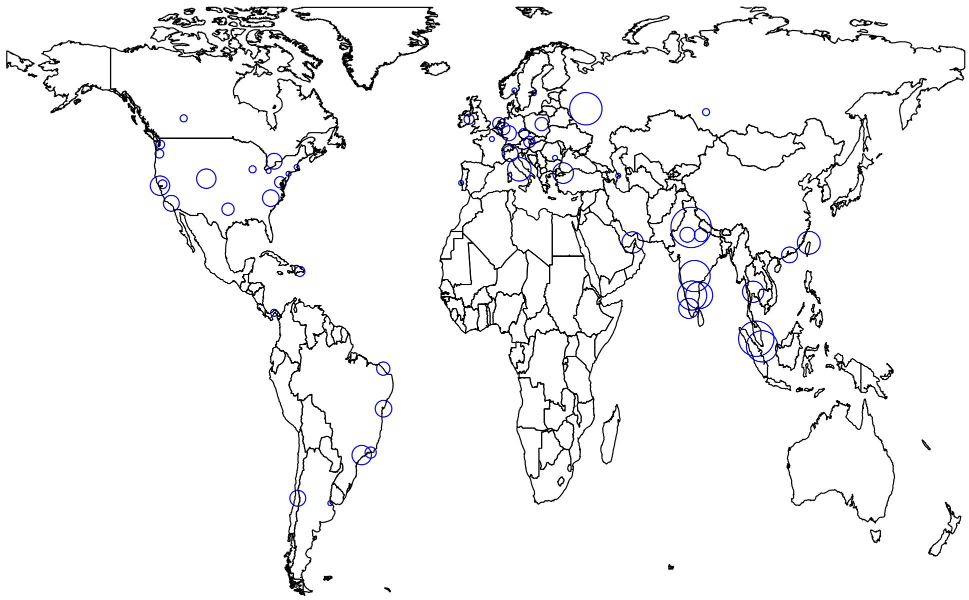
Figure 2: Model simulations of the response to an improvement in public transit

Positive impact on Uber trips $\forall d$	**Negative impact on Uber trips $\forall d$**	**Mixed impact on Uber trips $\forall d$**



Notes: This panel contains figures from three separate model simulations. Each model simulation corresponds to a column of figures. Row 1 contains average initial mode shares for outbound and return trips plotted by distance from the transit station. Row 2 contains changes in average outbound and return trip mode shares by distance from the transit station. Row 3 contains log changes in the number of trips at each distance from the transit station. First- and last-mile trips are assumed to be located at $d = 0$ but have some measurement error, following an exponential distribution with scale parameter $\lambda = 0.005$. Details of the model and calculations are included in the text and Appendix C.

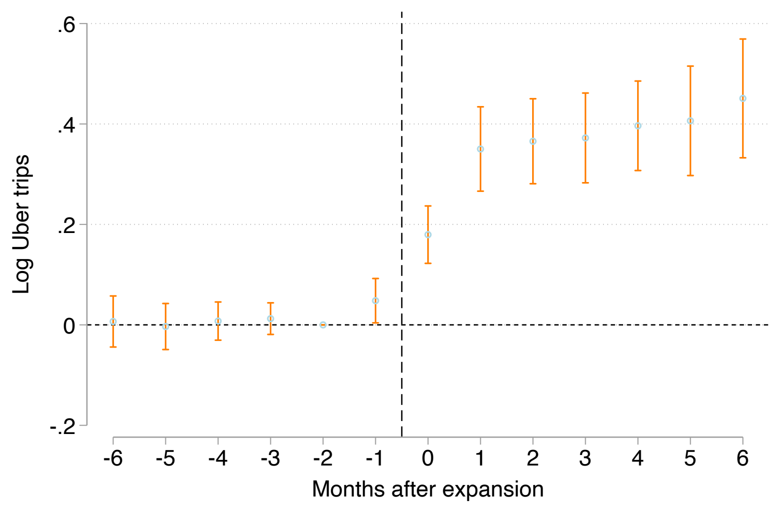
Figure 3: Locations of transit expansions in our data



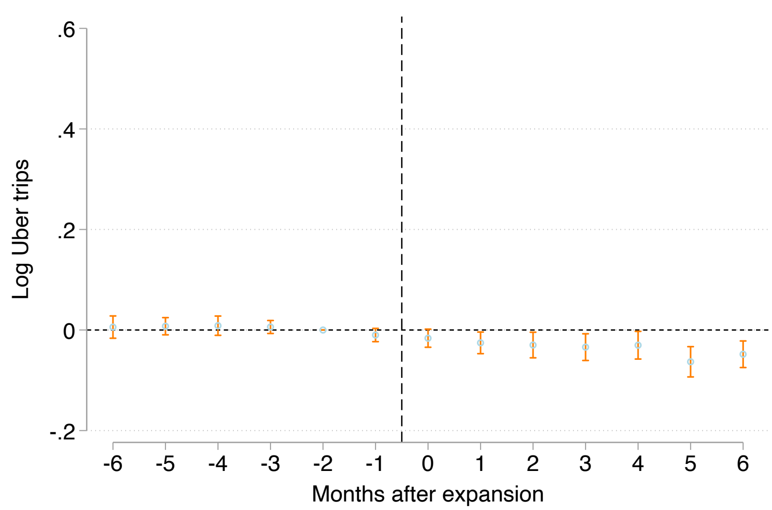
Notes: This map plots the locations of new transit openings in our data. The size of each circle is proportional to the number of new stations that opened between Uber's entry date and 2018.

Figure 4: Dynamic difference-in-difference estimates relative to 1100–1200 m

Effects 0–100 m from the station

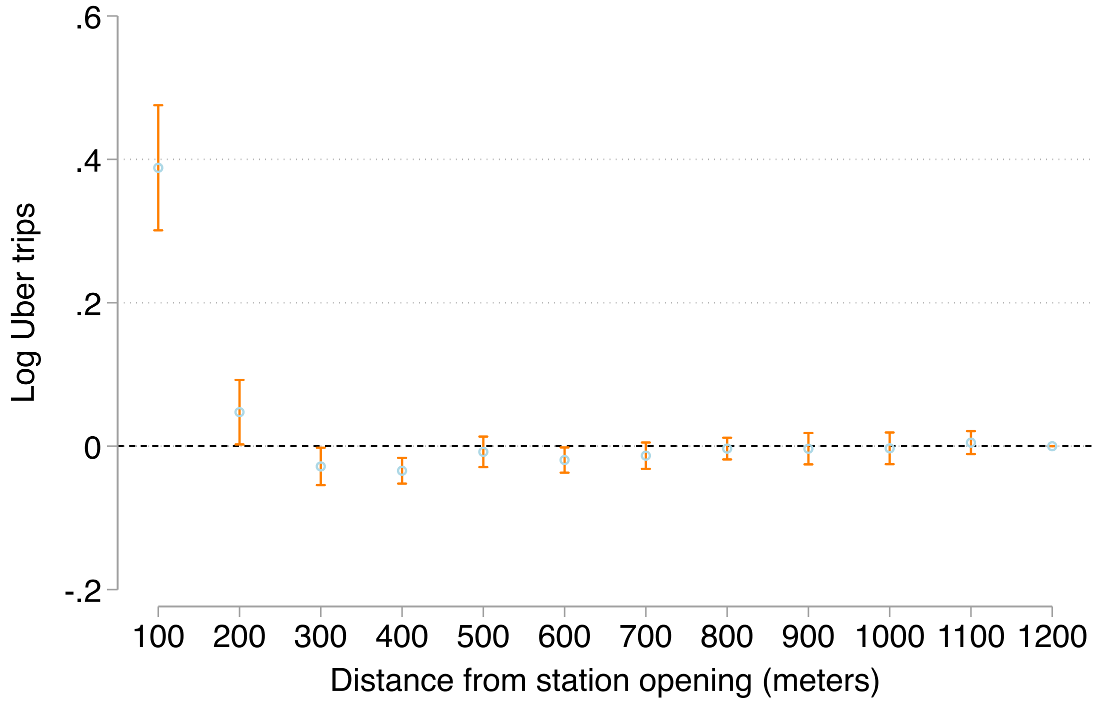


Effects 300–400 m from the station



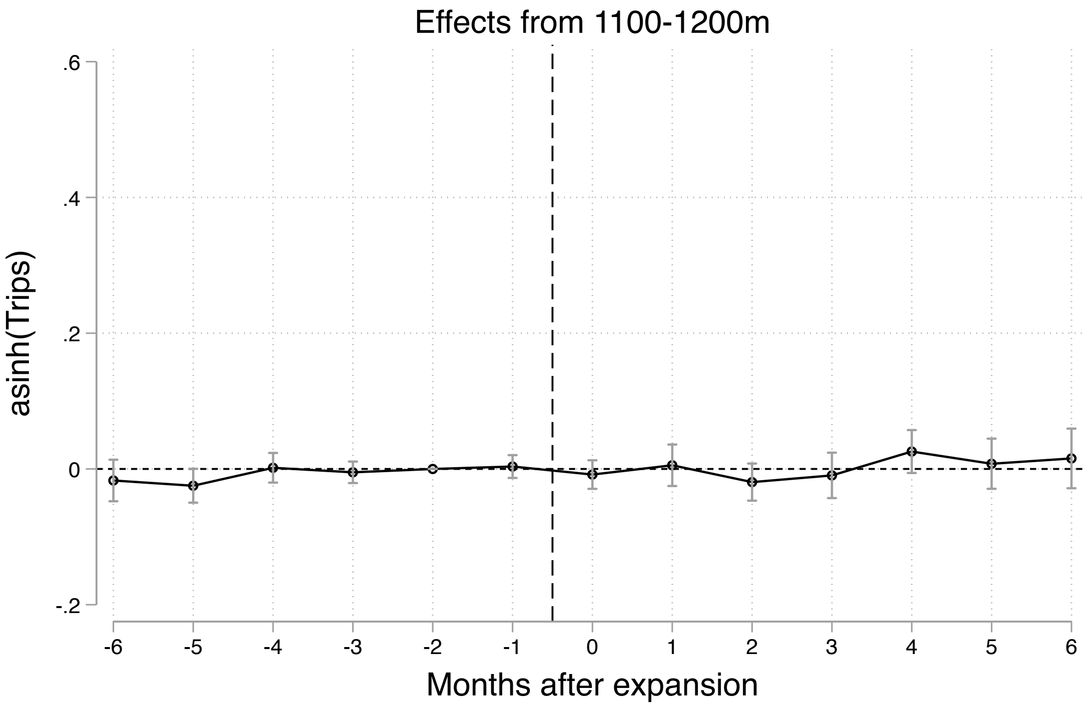
Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) for the percentage effect of a new train station opening on Uber trips 0–100 m away (Panel A) and 300–400 m (Panel B), relative to 1100–1200 m away. The coefficients are plotted for each month between six months prior and six months after an expansion. The PPML estimation follows a dynamic difference-in-differences model of station openings on 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.

Figure 5: Relative effect by distance of new train stations on Uber trips



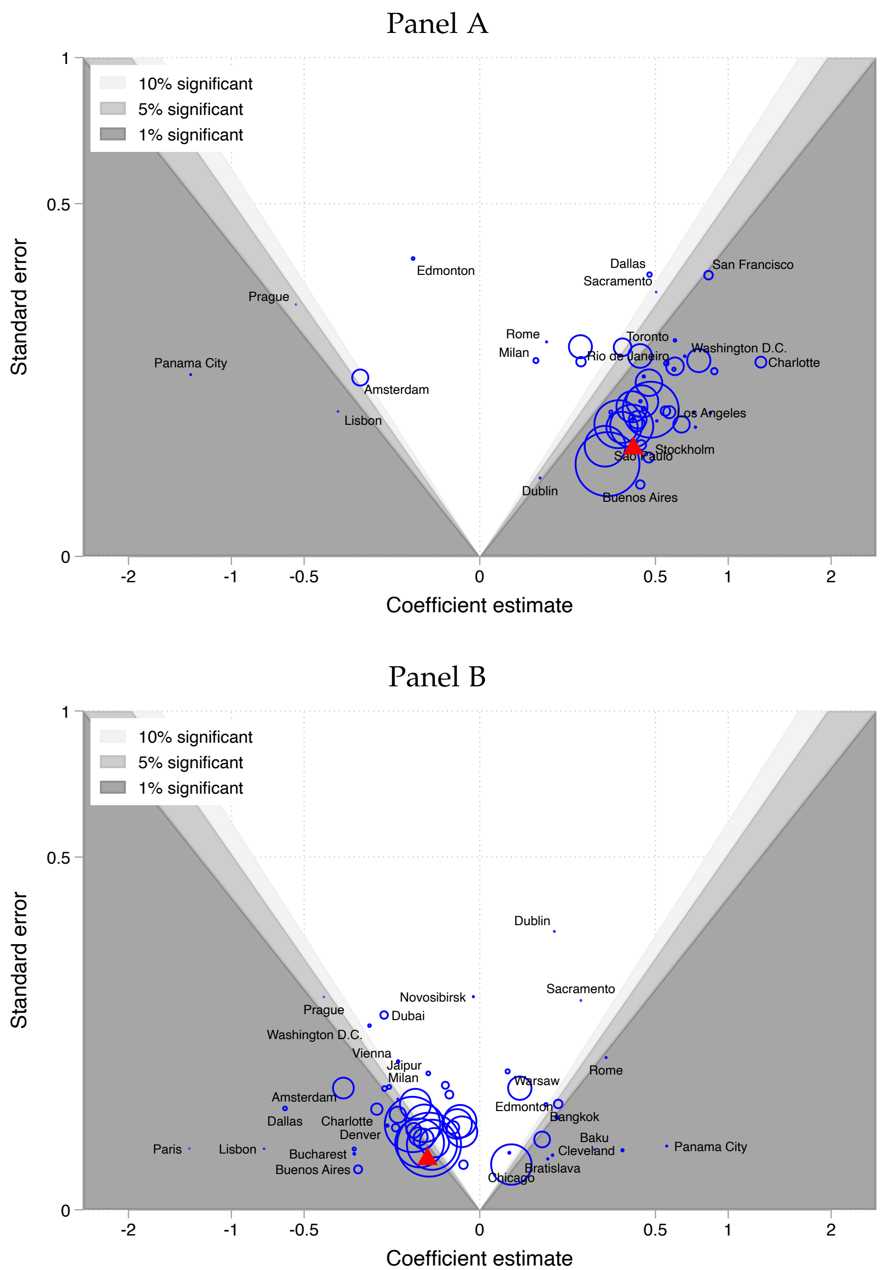
Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equation (5). These coefficients estimate the effect of a new train station opening on Uber trips at varying distances from the train station during months 1 through 6, relative to the farthest distance band 1100–1200 m, 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.

Figure 6: Event study estimates for 1100–1200 m



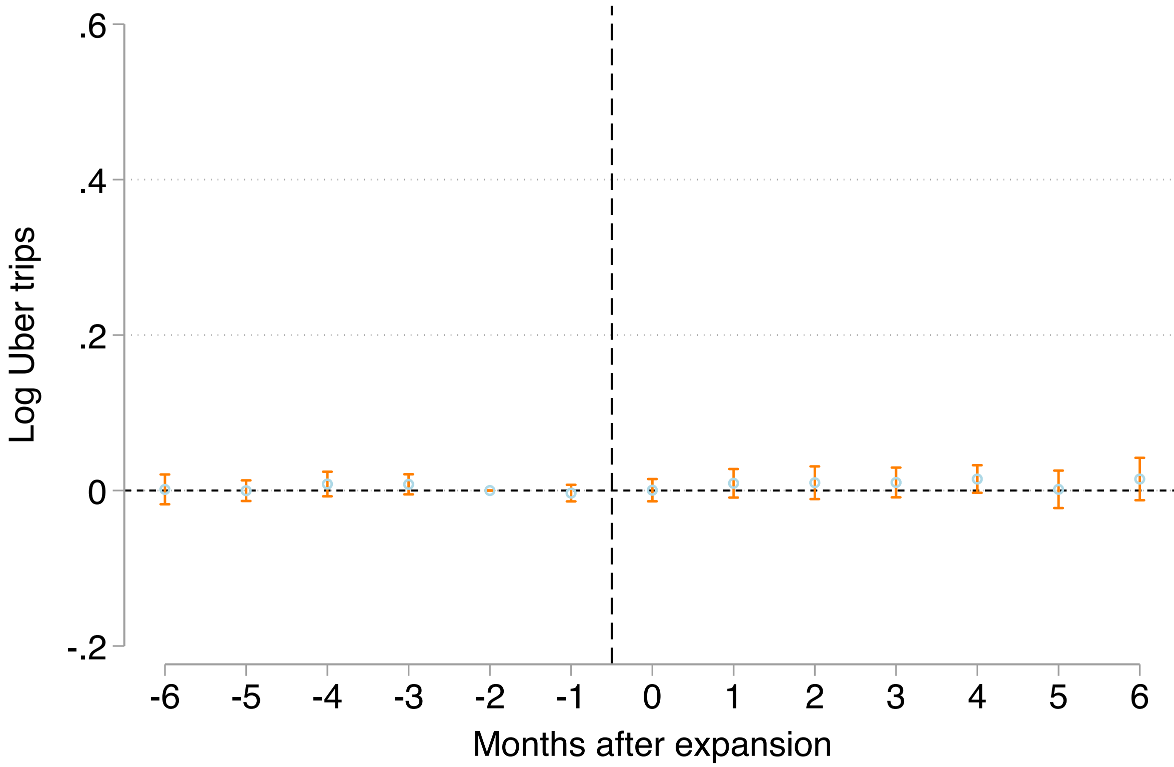
Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) for the percentage effect of a new train station opening on Uber trips 1100–1200 m away. The coefficients are plotted for each month between six months prior and six months after an expansion. The regression is an event study model of station openings on the inverse hyperbolic sine of Uber trips (in 100s) that either originated or terminated at a given distance band from the opening. The model contains station and city by time fixed effects; additionally, fixed effects for more than 6 months before or after an expansion are included. The model is estimated using (de Chaisemartin and D'Haultfœuille, 2024). All errors are clustered at the station-level.

Figure 7: Heterogeneous effect by city, 0–100 m and 300–400 m



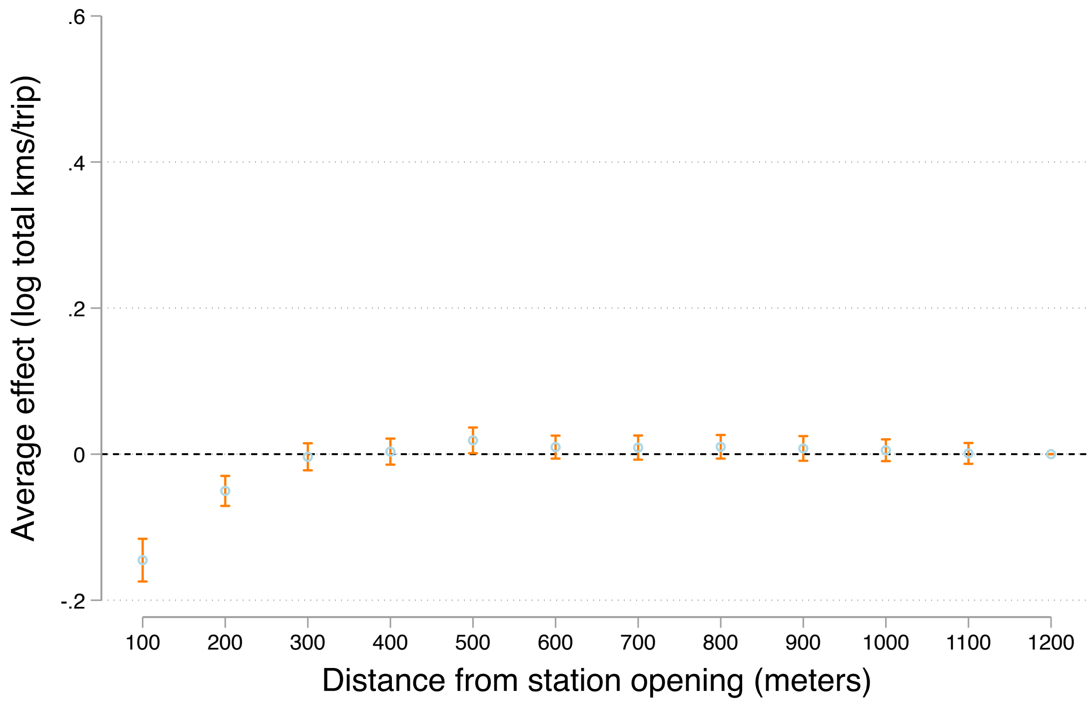
Notes: This funnel graph shows city-specific treatment effects based on equation (5) where the coefficient α is allowed to vary by city. Panel A contains estimates for 0–100 m whereas Panel B contains estimates for 300–400 m. The x-axis shows coefficient estimates, the y-axis shows standard errors. The 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 size of each circle is proportional to baseline Uber ridership. The large red triangle indicates the average effect reported in Figure 5.

Figure 8: Aggregate 0 – 700 m effect



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 0–700 m away, relative to 1100–1200 m away. The coefficients are plotted for each month between six months prior and six months after an expansion. The PPML estimation follows a dynamic difference-in-differences model of station openings on 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.

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) from estimating equation (5) 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.

Online Appendix

Contents

A Supplemental tables **2**

A.1 Descriptives . . . 2

A.2 Robustness of POI sample findings . . . 3

B Supplemental figures **4**

B.1 Descriptives . . . 4

B.2 Robustness . . . 7

B.3 Additional outcomes . . . 16

B.4 Heterogeneity in transit quality / location . . . 20

B.5 POI sample . . . 23

C Model appendix **24**

C.1 Solving the model . . . 24

C.2 Conditional choice probabilities . . . 25

C.3 Solving the model without ride-hailing services . . . 26

C.4 Choice probabilities Without Uber . . . 27

C.5 Theoretical results . . . 29

C.6 Mapping the model to the data . . . 36

C.7 Model simulations . . . 37

D Empirics appendix **38**

D.1 Event study for far distances . . . 38

D.2 Adjacent distance bins methodology . . . 38

A Supplemental tables

A.1 Descriptives

Table A1: City-level descriptive statistics

City	# Exp.	# Stn.	Light	Heavy	Subway	First Yr	Last Yr	Uber Yr.
Amsterdam	1	6	0	0	6	2018	2018	2012
Baku	1	1	0	0	1	2016	2016	2015
Bangalore	5	34	0	0	34	2014	2017	2013
Bangkok	4	18	0	0	18	2016	2017	2014
Boston	1	1	0	0	1	2014	2014	2012
Bratislava	1	1	1	0	0	2016	2016	2015
Bucharest	1	1	0	0	1	2017	2017	2015
Buenos Aires	1	1	0	0	1	2018	2018	2016
Charlotte	1	11	11	0	0	2018	2018	2013
Chennai	5	29	0	0	29	2015	2018	2014
Chicago	2	2	0	0	2	2015	2017	2012
Cleveland	1	1	0	0	1	2015	2015	2014
Dallas	3	6	6	0	0	2014	2016	2012
Delhi	13	65	0	0	65	2014	2018	2013
Denver	2	15	15	0	0	2013	2017	2012
Dubai	5	18	10	0	8	2014	2017	2013
Dublin	1	4	4	0	0	2017	2017	2014
Dusseldorf	1	6	6	0	0	2016	2016	2014
Edmonton	1	2	2	0	0	2015	2015	2014
Fortaleza	3	7	6	0	1	2017	2018	2016
Frankfurt	1	9	0	0	9	2016	2016	2014
Hong Kong	4	10	0	0	10	2014	2016	2014
Hyderabad	2	39	0	0	39	2017	2018	2014
Istanbul	5	17	0	0	17	2015	2017	2014
Jaipur	1	9	0	0	9	2015	2015	2014
Kochi	2	16	0	0	16	2017	2017	2014
Kuala Lumpur	5	50	22	0	28	2015	2017	2013
Lisbon	1	1	0	0	1	2016	2016	2014
Los Angeles	2	10	10	0	0	2016	2016	2012
Lucknow	1	7	0	0	7	2017	2017	2016
Milan	7	10	0	0	10	2014	2015	2013
Moscow	13	42	0	0	42	2014	2017	2013
New Jersey	1	1	0	1	0	2016	2016	2013
Novosibirsk	1	2	2	0	0	2016	2016	2015
Oslo	1	1	0	0	1	2016	2016	2014
Panama City	2	2	0	0	2	2015	2015	2014
Paris	1	1	0	0	1	2013	2013	2012
Portland	1	3	3	0	0	2015	2015	2014
Prague	1	4	0	0	4	2015	2015	2014
Rio de Janeiro	1	5	0	0	5	2016	2016	2014
Rome	3	22	0	0	22	2014	2015	2013
Sacramento	1	3	3	0	0	2015	2015	2013
Salvador	4	11	0	0	11	2016	2018	2016
San Francisco	5	15	0	11	4	2014	2018	2012
Santiago	1	10	0	0	10	2017	2017	2013
Santo Domingo	1	4	0	0	4	2018	2018	2015
Sao Paulo	9	15	0	0	11	2017	2018	2014
Seattle	2	3	3	0	0	2016	2016	2012
Singapore	9	39	6	0	33	2013	2017	2013
Stockholm	1	1	0	1	0	2017	2017	2013
Taipei	3	21	0	0	21	2014	2017	2013
Toronto	2	10	0	4	6	2015	2017	2012
Vienna	1	5	0	0	5	2017	2017	2014
Warsaw	1	7	0	0	7	2015	2015	2014
Washington D.C.	1	5	0	0	5	2014	2014	2012

Notes: This table contains summary statistics on the 650 stations in our sample. Column (1) contains the 55 cities in our sample. Columns (2) through (6) count the number of expansions, the number of stations associated with those expansions, and whether they were for light rail, heavy rail, or subways, respectively. Column (7) and (8) denote the first and last year we have Uber data for that city. Column (9) contains the year that Uber entered the market.

A.2 Robustness of POI sample findings

Table A2: Difference-in-differences estimates for 0–500 m POI sample

	(1) Uber trips	(2) Uber trips	(3) Uber trips	(4) Uber trips
-6 months after	-0.0109	-0.00332	-0.00183	0.00927
	(0.00852)	(0.00714)	(0.00608)	(0.00816)
-5 months after	-0.0110*	-0.00708	-0.00447	0.00470
	(0.00579)	(0.00472)	(0.00393)	(0.00524)
-4 months after	-0.00637	-0.00430	-8.52e-05	0.00720
	(0.00561)	(0.00474)	(0.00408)	(0.00509)
-3 months after	0.00149	0.00189	0.00345	0.00770
	(0.00485)	(0.00422)	(0.00413)	(0.00474)
-1 months after	-0.00751	-0.0106**	-0.00988***	-0.0144***
	(0.00527)	(0.00425)	(0.00358)	(0.00480)
Month of expansion	-0.0215***	-0.0205***	-0.0188***	-0.0264***
	(0.00666)	(0.00450)	(0.00418)	(0.00548)
1 months after	-0.0399***	-0.0363***	-0.0349***	-0.0438***
	(0.00793)	(0.00490)	(0.00421)	(0.00514)
2 months after	-0.0401***	-0.0340***	-0.0311***	-0.0420***
	(0.00919)	(0.00619)	(0.00519)	(0.00546)
3 months after	-0.0453***	-0.0360***	-0.0332***	-0.0464***
	(0.0101)	(0.00626)	(0.00527)	(0.00630)
4 months after	-0.0523***	-0.0372***	-0.0338***	-0.0508***
	(0.0115)	(0.00636)	(0.00482)	(0.00734)
5 months after	-0.0493***	-0.0349***	-0.0327***	-0.0532***
	(0.0119)	(0.00616)	(0.00552)	(0.00921)
6 months after	-0.0511***	-0.0379***	-0.0349***	-0.0591***
	(0.0129)	(0.00773)	(0.00619)	(0.00955)
Observations	62,202	59,897	49,946	49,946
Station Openings	566	563	510	510
Station FE	✓	✓	✓	✓
Cal. and Evt. Time FE	✓	✓	✓	✓
City x Date FE		✓		
Station x Date FE			✓	✓
Dist x Date FE				✓

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: This table contains coefficient estimates and standard errors from equation (4) for individuals with a POI near the station, and with varying fixed effects. These coefficients estimate the effect of a new train station opening on *city-wide* Uber trips in every month leading up to and after the station opening, for those with a POI within 500 m of the station opening relative to those with a POI within 800–1200 meters. Coefficients are relative to two months prior to the station opening. All specifications include station and calendar and event time fixed effects. Column (2) adds city by calendar time fixed effects. Column (3) adds station by date fixed effects, and is our main specification. Column (4) includes distance by date fixed effects. All errors are clustered at the station-level.

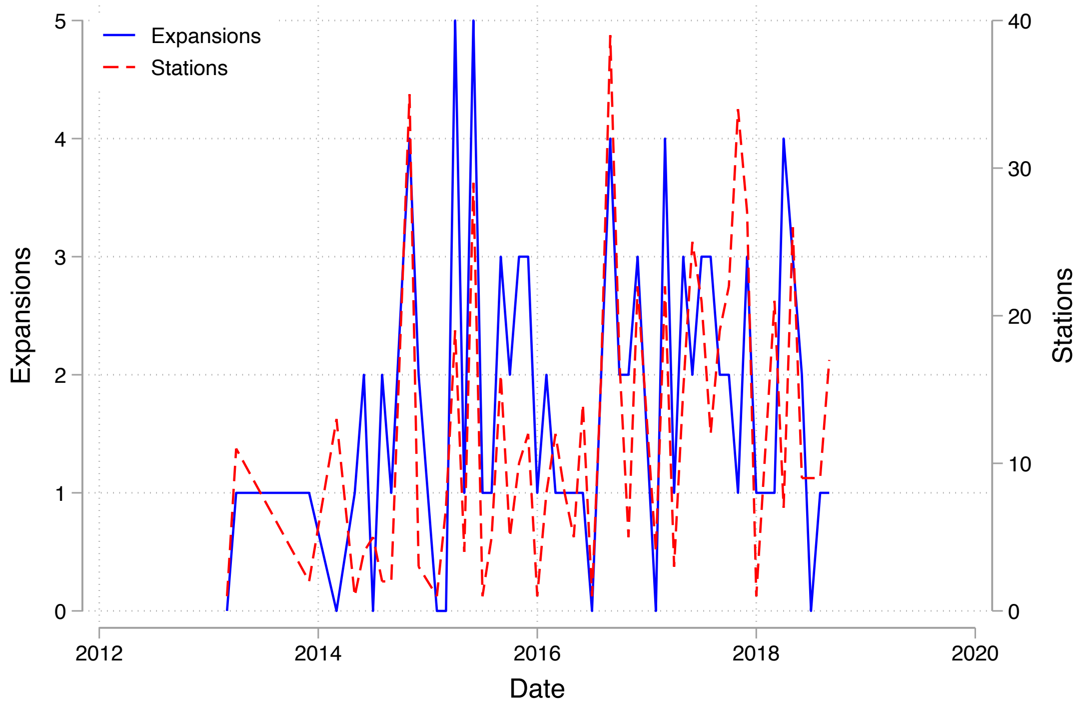
B Supplemental figures

B.1 Descriptives

Figure A1: Spatial sampling for Toronto-York Spadina subway extension in 2017

Notes: These maps depict two ways to sample Uber trips in our data. Blue dots indicate the locations of 6 subway openings along the 1-line in Toronto's subway system. The map on the left uses concentric rings of 100 meter width from the location of the station. Our main specifications rely on data sampled in this manner. The map on the right uses fixed width hexagons at varying distances from the new station. We use these to determine points-of-interest (POIs) for individual riders, as well as in some robustness exercises.

Figure A2: Transit expansions in our data over time



Notes: This figure plots the time series of expansions and station openings in our sample. The left-side y-axis contains counts of rail expansions, which are plotted in blue. The right-side y-axis contains counts of stations associated with those expansions, which are plotted in red.

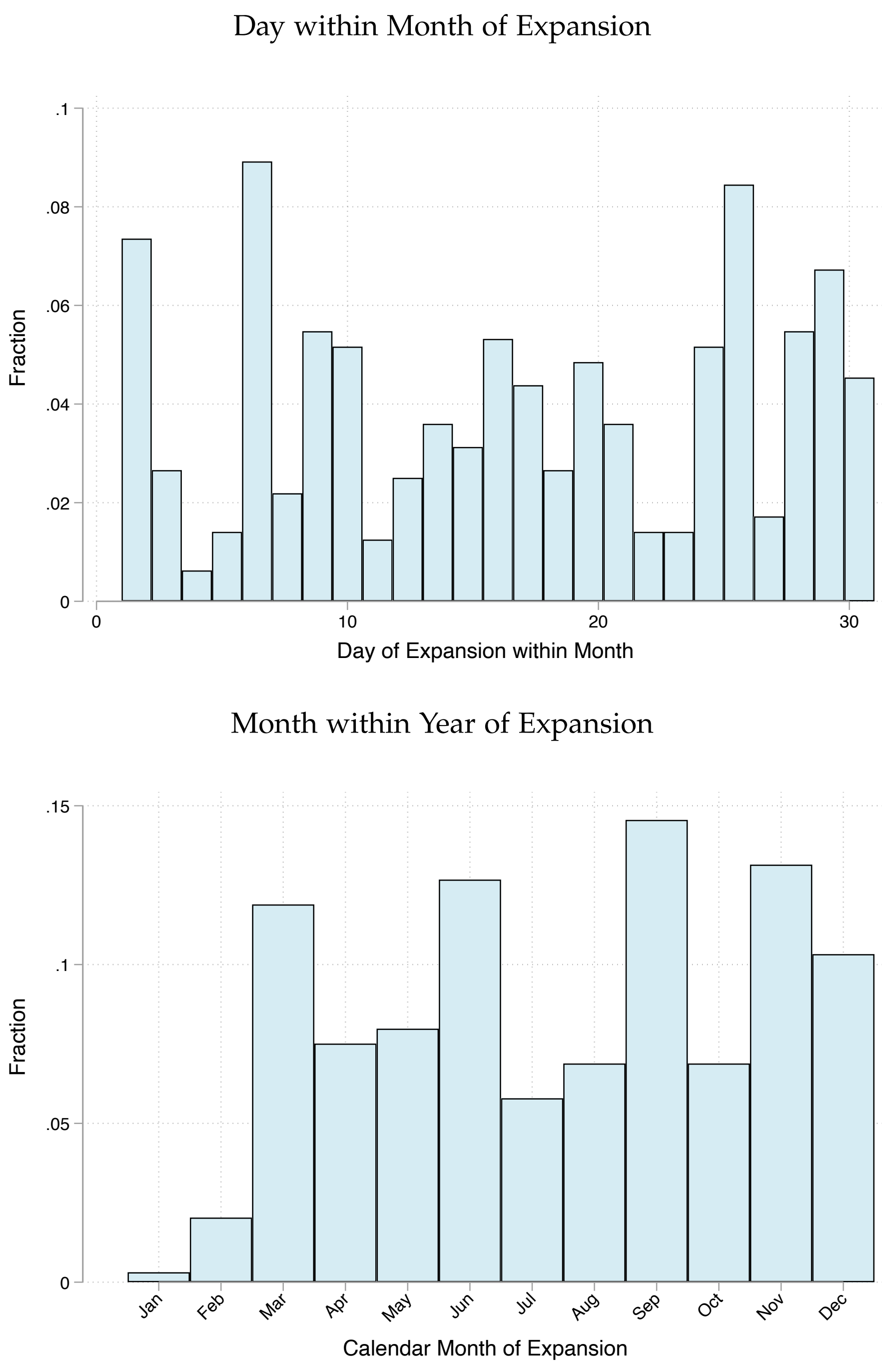


Figure A3: Distribution of days and months of station openings

Notes: These histograms plot the temporal distribution of station openings. The x-axis is the fraction of station openings in our sample that fall within a certain day/month. The top figure plots the distribution of numeric day within the month of expansion. The bottom figure plots the distribution of month within year of expansion.

B.2 Robustness

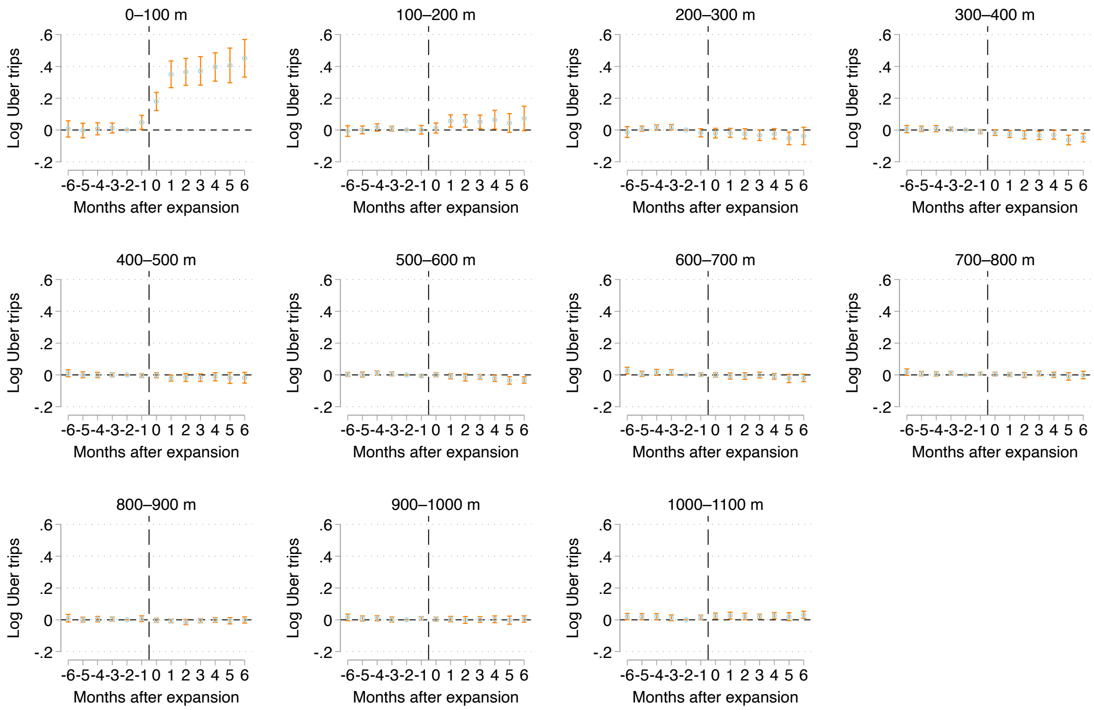


Figure A4: Difference-in-differences estimates, using the farthest distance band as the control group, from 0–100 m to 1000–1100 m

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 Uber trips that either originated or terminated at a given distance band from the opening, and the regression equation is given by equation (4). 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.

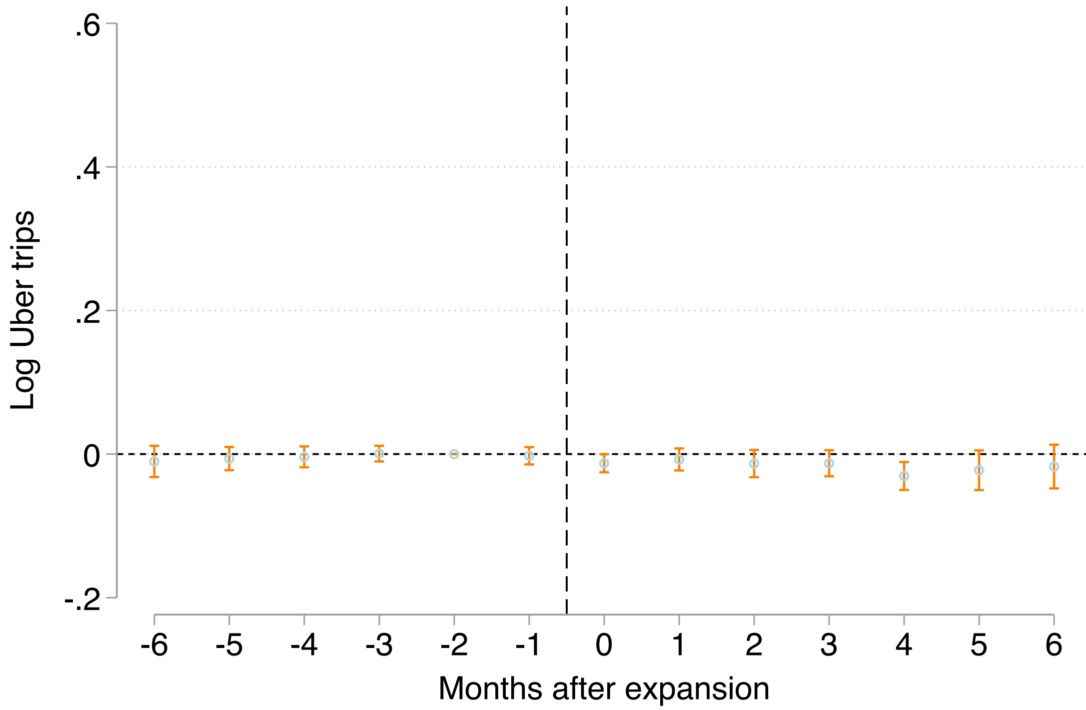


Figure A5: Event study estimates for 1100–1200 m using PPML estimator

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 at our farthest distance band: 1100–1200 meters. It is a complement to the event study analysis depicted in Figure 6. The coefficients are plotted for each month between six months prior and six months after an expansion. The PPML estimates follow an event study model of station openings on Uber trips that either originated or terminated at a given distance band from the opening. The model contains station and city by time fixed effects; additionally, fixed effects for more than 6 months before or after an expansion are included. All errors are clustered at the station-level.

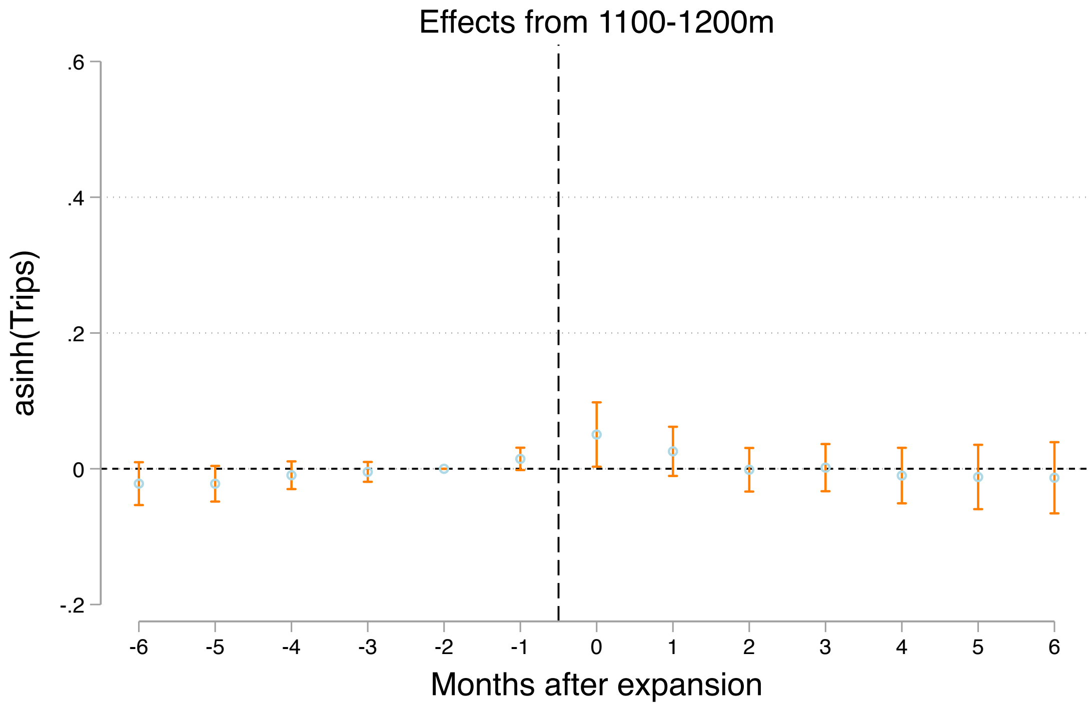


Figure A6: Event study estimates for 1100–1200 m using two-way fixed effects linear regression and inverse hyperbolic sine transformed Uber trips (in hundreds)

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 at our farthest distance band: 1100–1200 meters. It is a complement to the event study analysis depicted in Figure 6, and corresponds to the simpler two-way fixed effects linear regression model. 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 Uber trips that either originated or terminated at a given distance band from the opening. The model contains station and city by time fixed effects; additionally, fixed effects for more than 6 months before or after an expansion are included. All errors are clustered at the station-level.

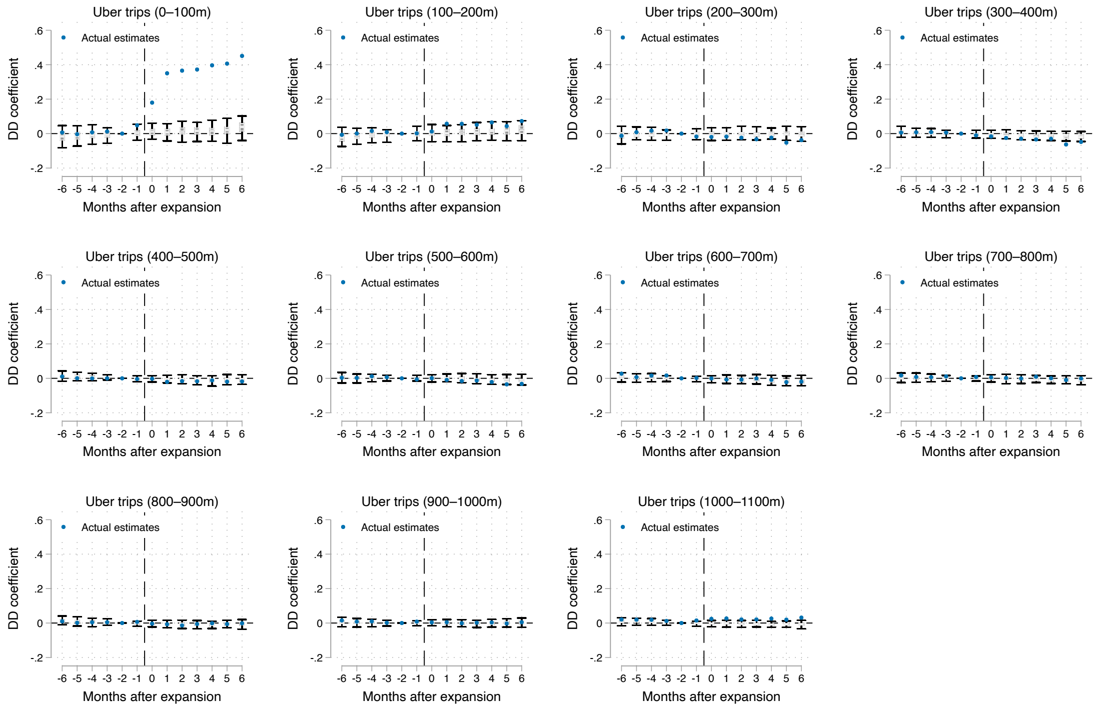


Figure A7: Placebo tests for all 100 m distance bands, up through 1000–1100 m

Notes: This figure plots the results of a placebo test where we randomly assign placebo opening dates to each transit station. To implement this test, we randomly assign each station a new opening date with uniform probabilities from 6 months after the station's first observation to 6 months prior to the station opening. This ensures that we have sufficient observations to estimate the event-time specific coefficients above. Also, it ensures that the new station opening is prior to any actual effects on Uber trips. We then re-estimate our baseline model using the "fake" opening dates. We perform this 100 times and use a box-whisker plot to depict the distribution of these estimates. The blue circle denotes estimates from the true station opening dates. We perform this placebo analysis for each distance band.

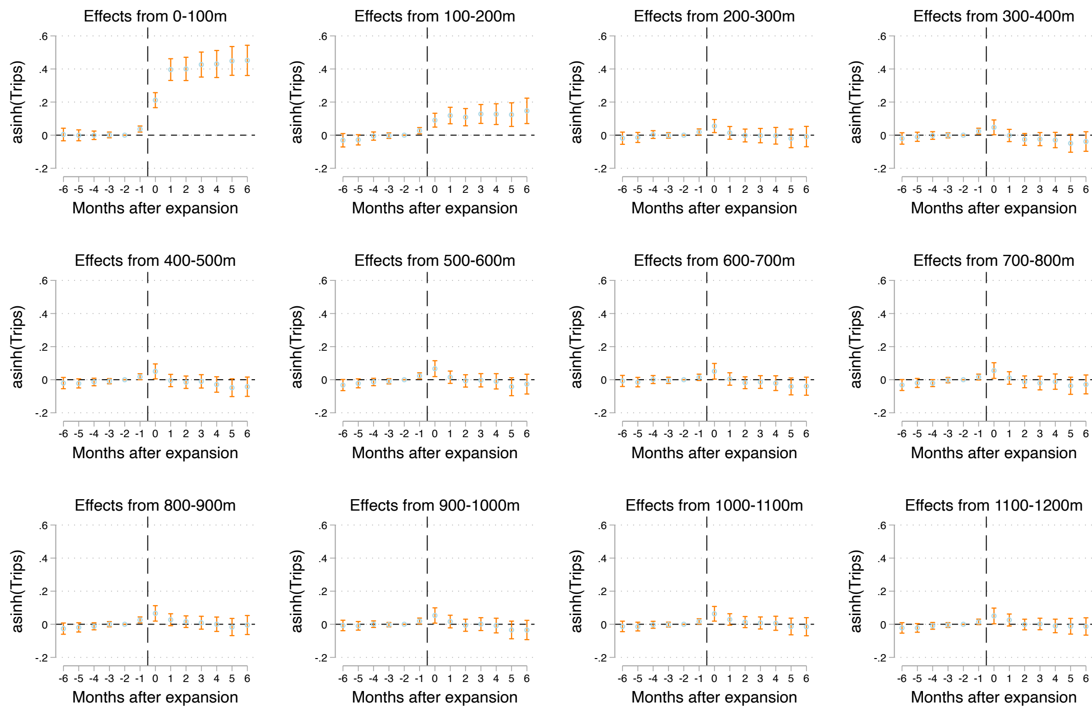


Figure A8: Event study estimates estimated via linear regression with the inverse hyperbolic sine transformed Uber trips (in hundreds) as the outcome

Notes: These figures plot the coefficients (circles) and 95% confidence intervals (bars) for the percentage effect of a new train station opening on Uber trips for each 100 meter distance band. It relies on the simpler two-way fixed effects linear regression model and the inverse hyperbolic sine transformation of the outcome Uber trips. 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 Uber trips that either originated or terminated at a given distance band from the opening. The model contains station and city by time fixed effects; additionally, fixed effects for more than 6 months before or after an expansion are included. All errors are clustered at the station-level.

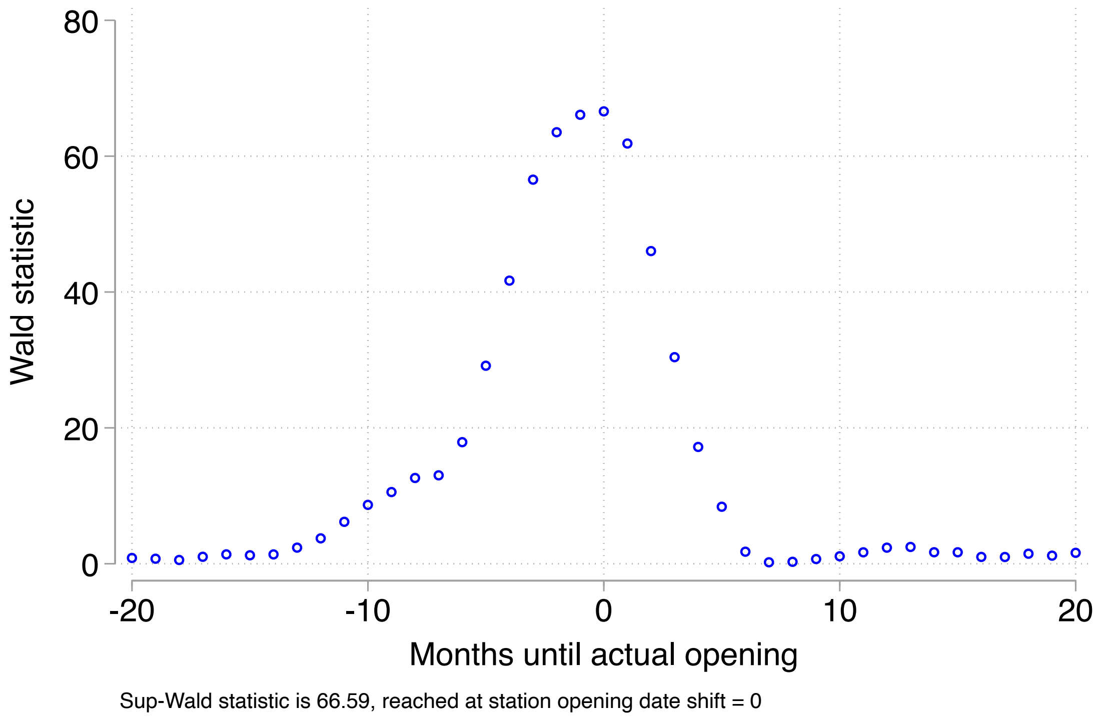
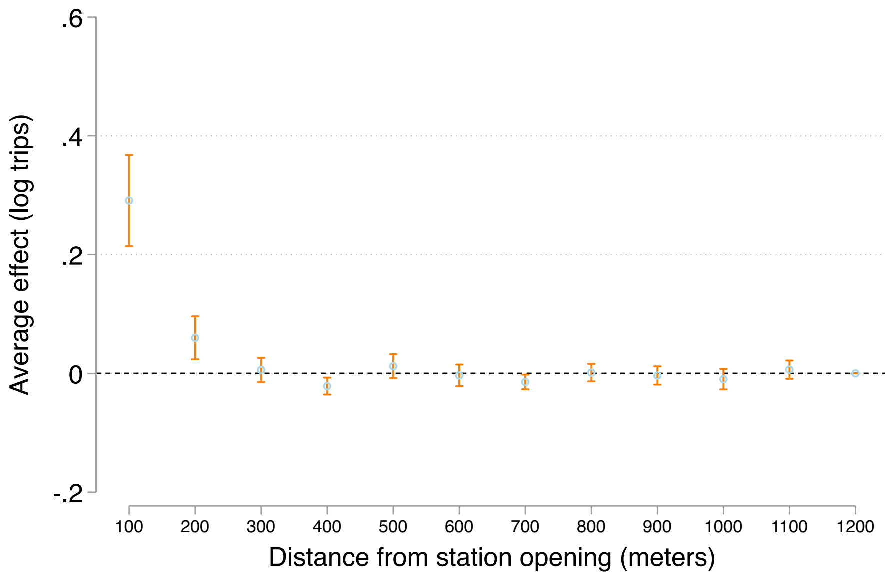


Figure A9: Wald test for 0–100 m difference-in-difference estimate

Notes: This figure plots the results of a structural break test on the effect of a station opening on Uber trips 0–100 meters away. The test follows Andrews (1993, 2003). To implement this test, we re-estimate the average effect for our main difference-in-differences specification as if the opening date were any number of months before or after the true station opening. We estimate this model for each event time, and plot the Wald-statistic of those estimates on the chart above. The sup-Wald statistic is achieved at the true station opening. Moreover, our model supports a structural break in Uber usage at any conventional level (Andrews, 2003). All errors are clustered at the station-level.

"Marginal effects"



"Aggregate effects"

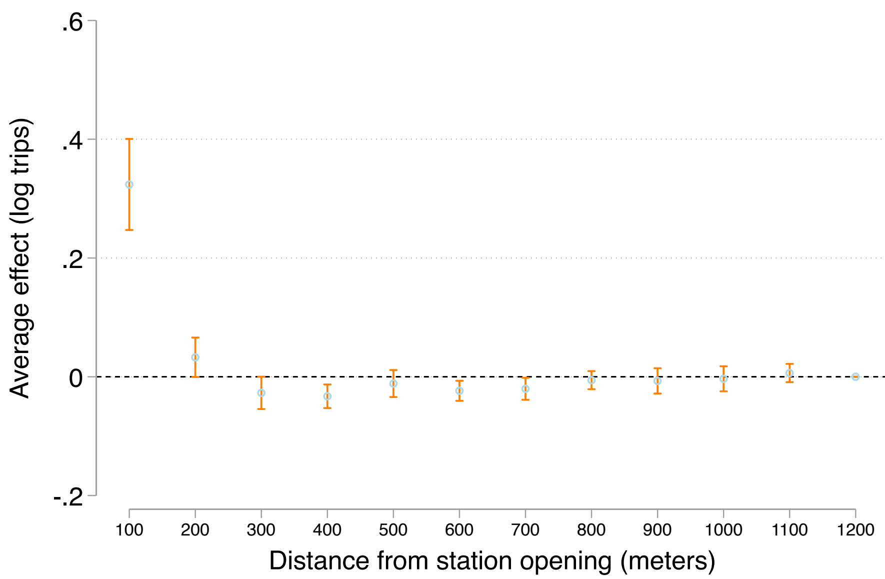


Figure A10: Difference-in-differences estimates using the adjacent bin as a control group

Notes: These figures plot the coefficients (circles) and 95% confidence intervals (bars) from estimating a version of equation (5) that relies on adjacent distance bins as the control group. In the top figure, 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* farthest 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 relative to 400 and 500 m away. In the bottom figure, those coefficients are summed over the entire range to account for effect the new station opening has on the "control" group at close enough distances. This procedure is described in Section 5 as well as Appendix D.2. All errors are clustered at the station-level.

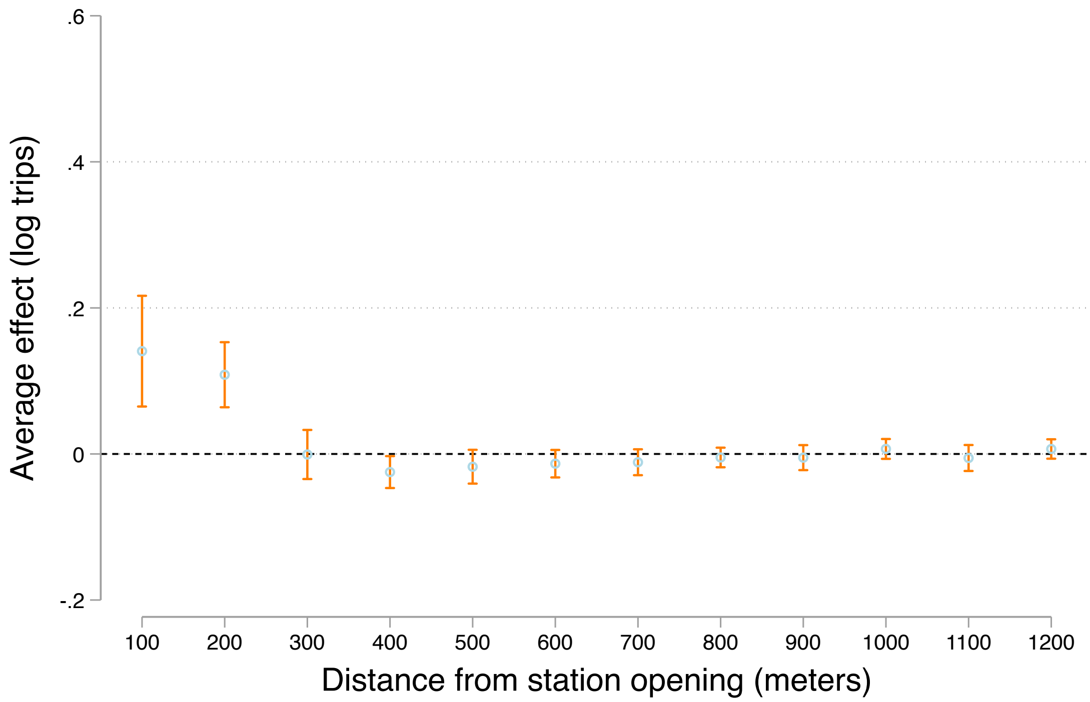


Figure A11: Difference-in-differences estimates across distances using hexagons

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating a version of equation (5) estimated on data aggregated to the hexagon level. Hexagons are assigned distances from the station according to their centroid. 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 farthest 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. All errors are clustered at the station-level.

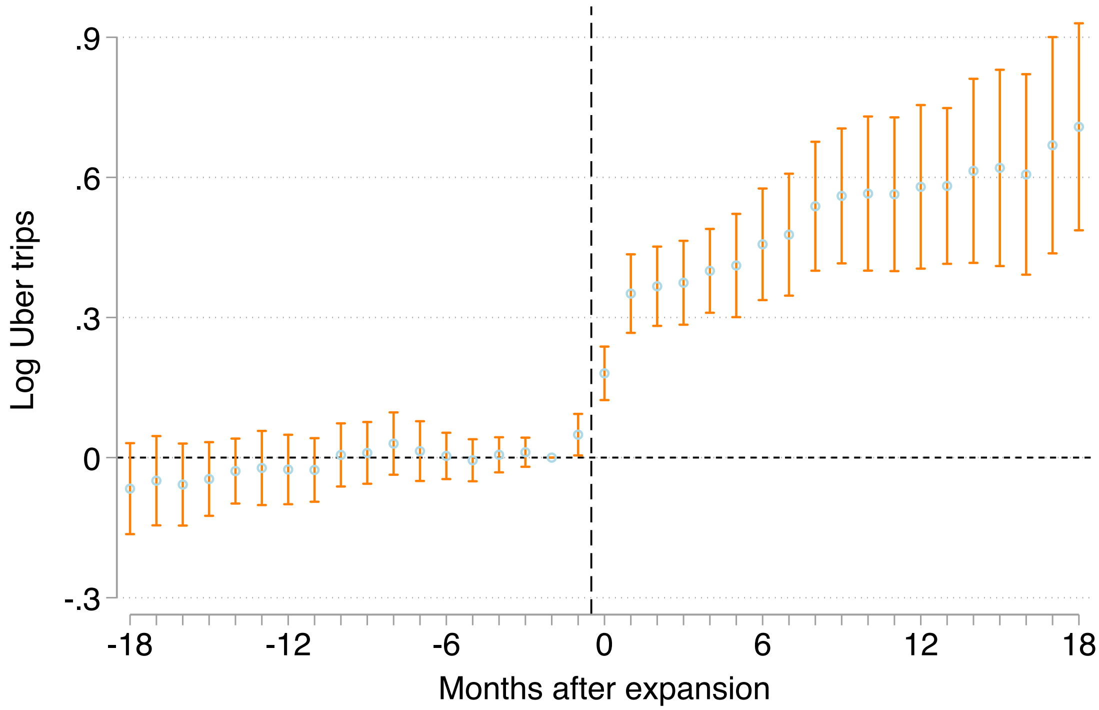


Figure A12: Difference-in-differences estimates using an 18-month window

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 from the station. The coefficients are plotted for each month between 18 months prior and 18 months after an expansion. The regression is a dynamic difference-in-differences model of station openings on Uber trips that either originated or terminated at a given distance band from the opening, and the regression equation is given by equation (4). The model contains station by distance band and station by time fixed effects; additionally, fixed effects for distance band by more than 18 months before or after an expansion are included. All errors are clustered at the station-level.

B.3 Additional outcomes

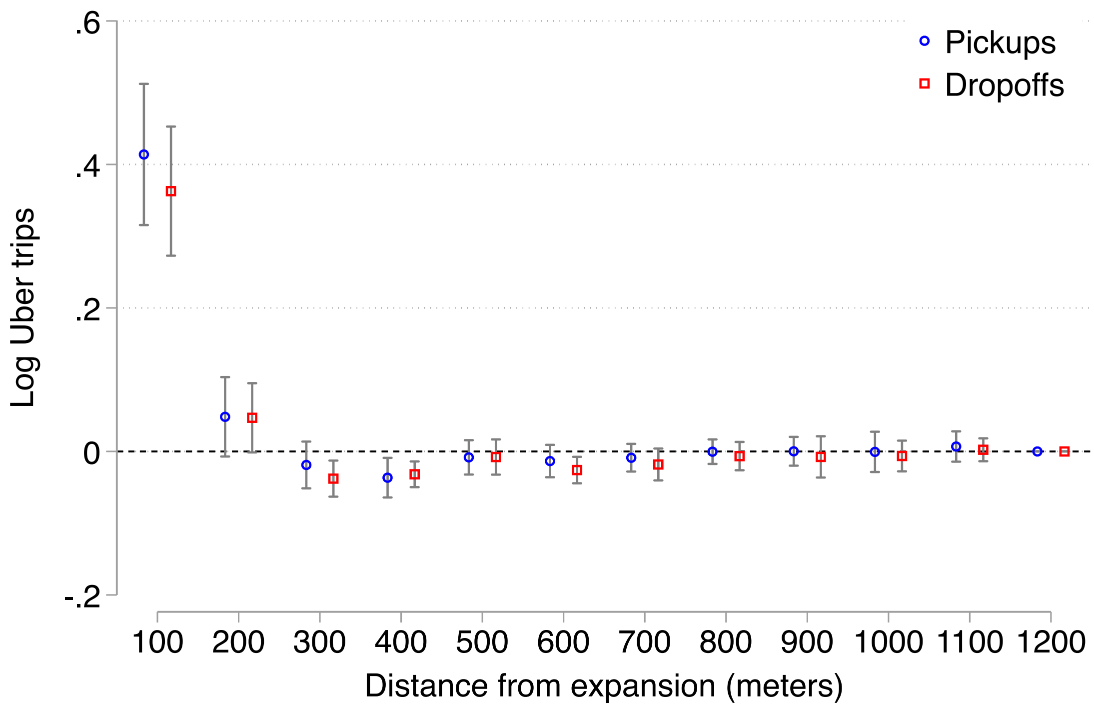
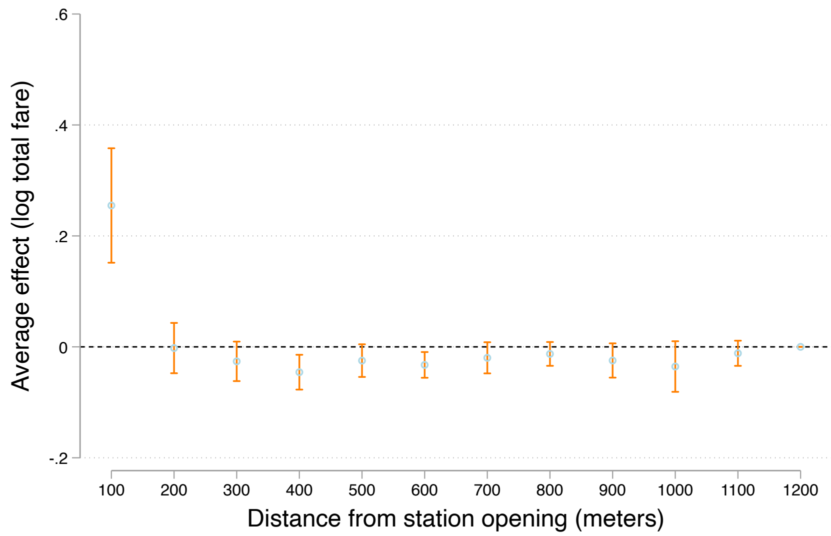


Figure A13: Estimates for pick-ups and drop-offs

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equation (5) separately for pick-ups and drop-offs. These coefficients estimate the effect of a new train station opening on Uber pick-ups and drop-offs at varying distances from the train station during months 1 through 6, relative to the farthest 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. All errors are clustered at the station-level.

Total Fares



Total Kilometers

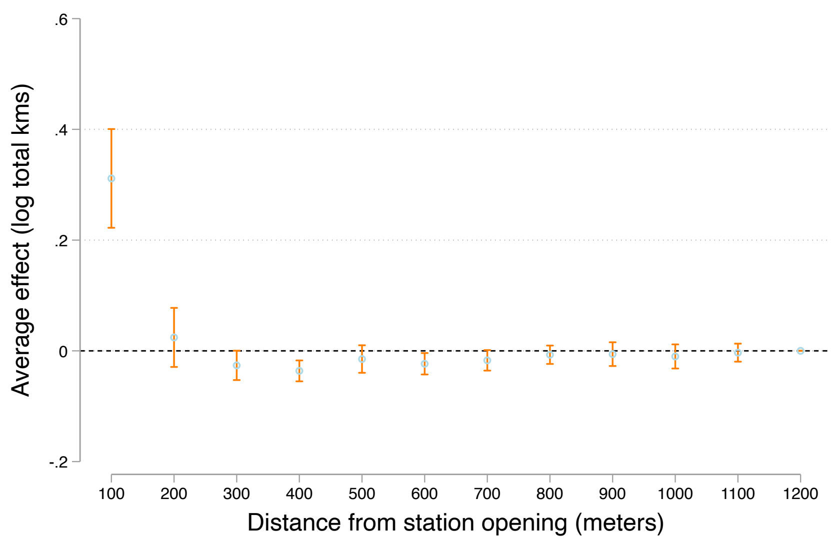


Figure A14: Estimates for total kilometers and fare of trips

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equation (5) for the total fare paid (top figure) and kilometers travelled (bottom figure) along Uber trips. These coefficients estimate the effect of a new train station opening on Uber fare and trip length at varying distances from the train station during months 1 through 6, relative to the farthest 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. All errors are clustered at the station-level.

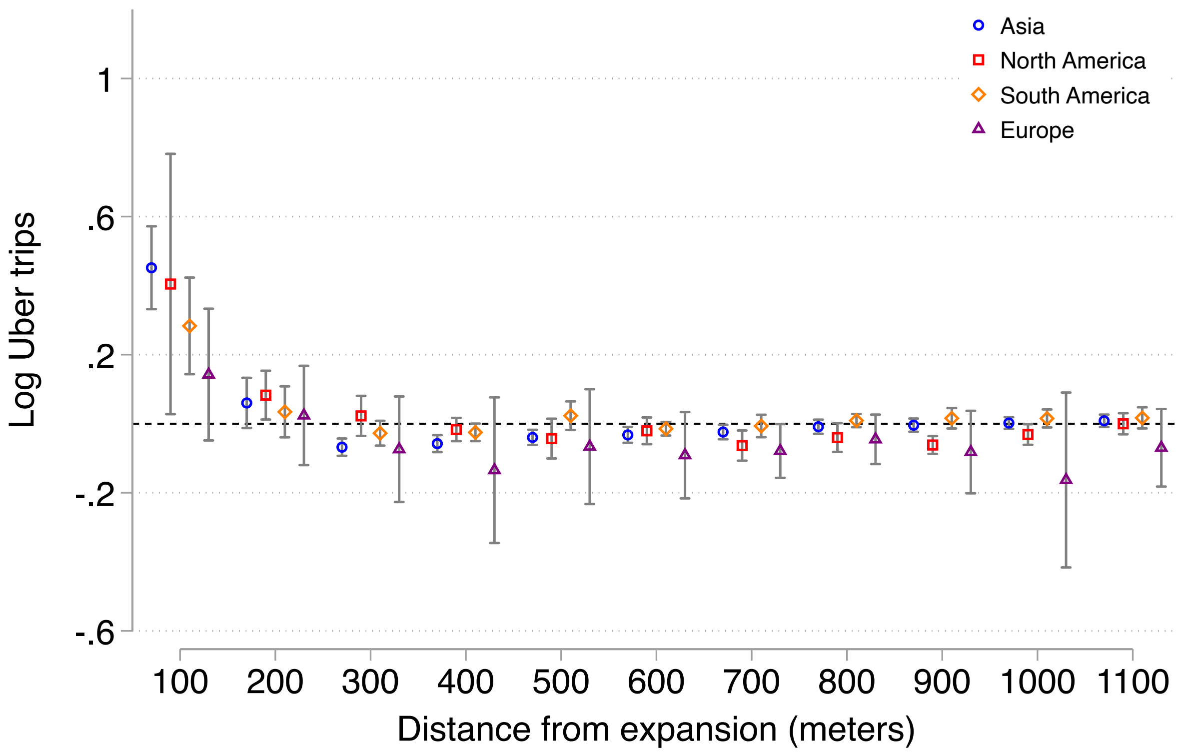


Figure A15: Estimates across continents

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equation (5) separately for each continent: Asia, North America, South America, and Europe. These coefficients estimate the effect of a new train station opening on Uber pick-ups and drop-offs at varying distances from the train station during months 1 through 6, relative to the farthest 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. All errors are clustered at the station-level.

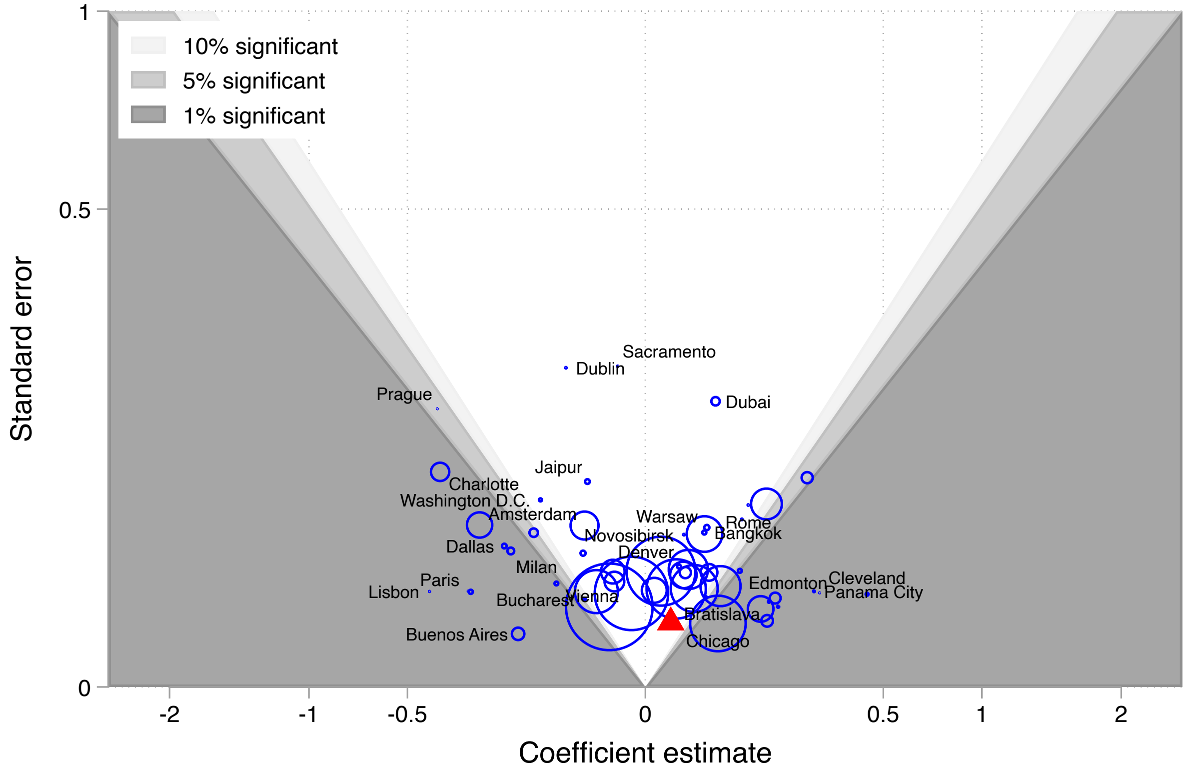
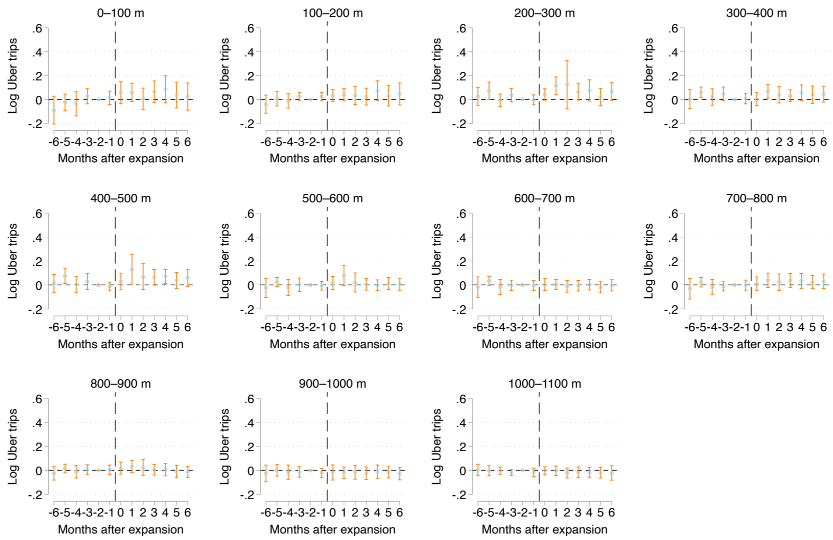


Figure A16: Aggregate 0 – 700 m effect by city

Notes: This funnel graph shows city-specific treatment effects based on equation (5) for Uber trips within 700 meters of the station opening, and 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 size of each circle is proportional to baseline Uber ridership. The large red triangle indicates the average effect reported in Figure 5. All error are clustered at the station-level.

B.4 Heterogeneity in transit quality / location

Panel A: Dynamic effects



Panel B: Average effect by distance

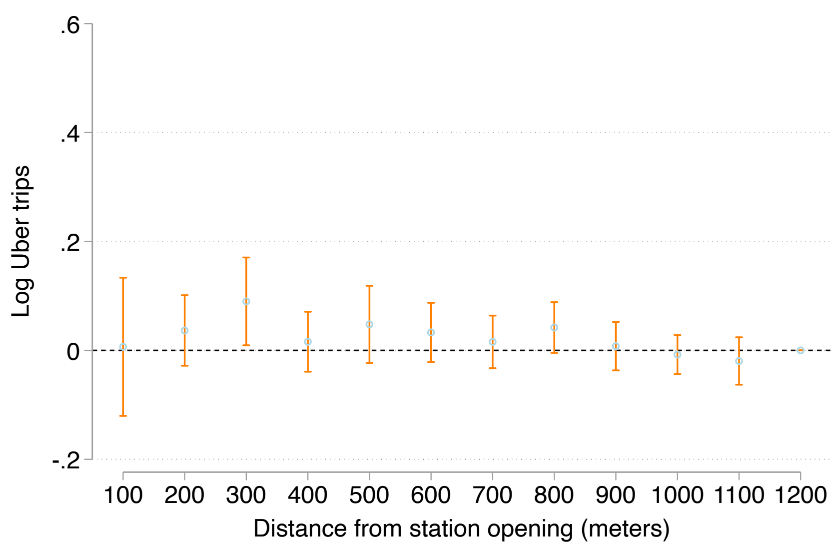


Figure A17: Dynamic and average effects for station openings that interact with traffic

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equations (4) and (5) for stations whose transit interacts with traffic. These estimates are contained in Panel A and Panel B, respectively. These coefficients estimate the effect of a new train station opening on Uber trips at varying times and distances from the train station opening. All errors are clustered at the station-level.

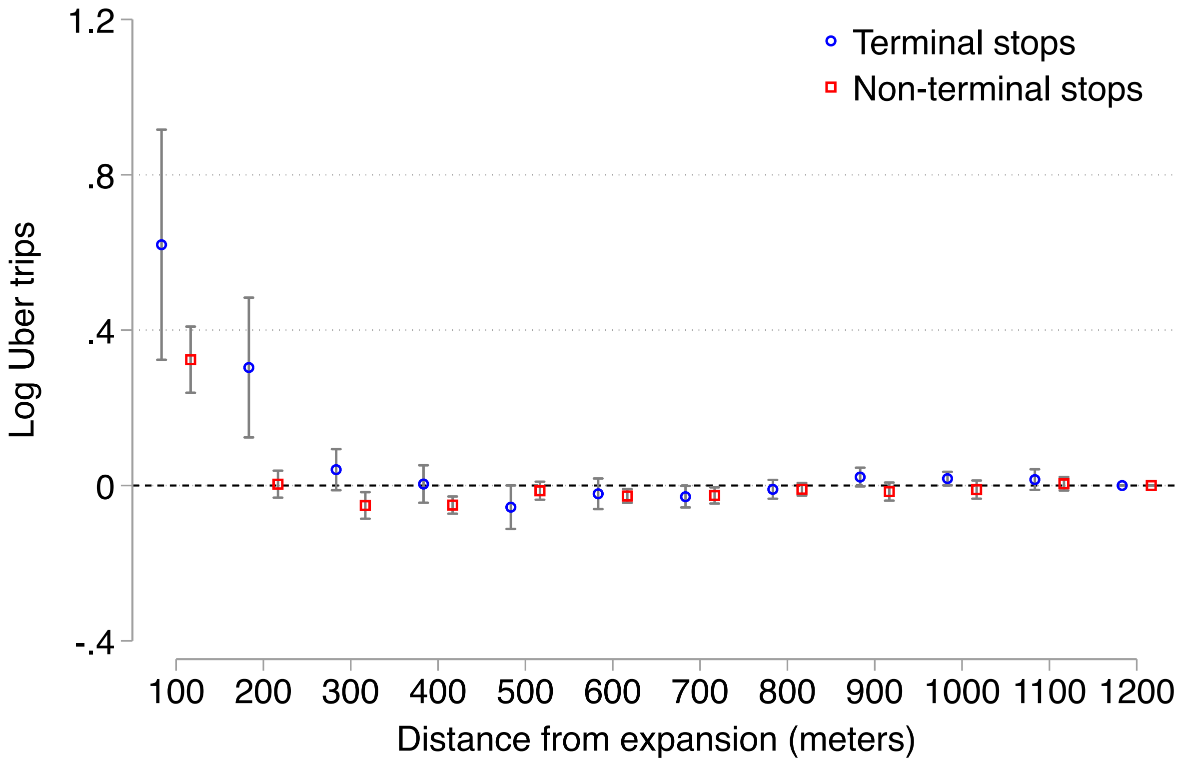
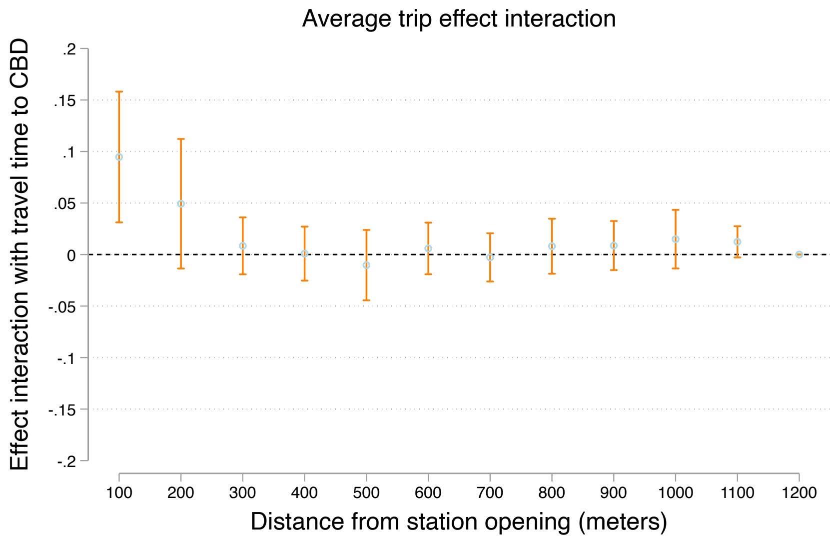


Figure A18: Differential effects for terminal versus non-terminal stations

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equation (5) separately for stations that were a terminal stop on their train line. The coefficients estimate the effect on Uber trips at varying distances from the train station opening, separately for each type of train station. All errors are clustered at the station-level.

Panel A: Interaction of effect by with normalized travel time from CBD



Panel B: Interaction of effect by distance with normalized travel time from CBD

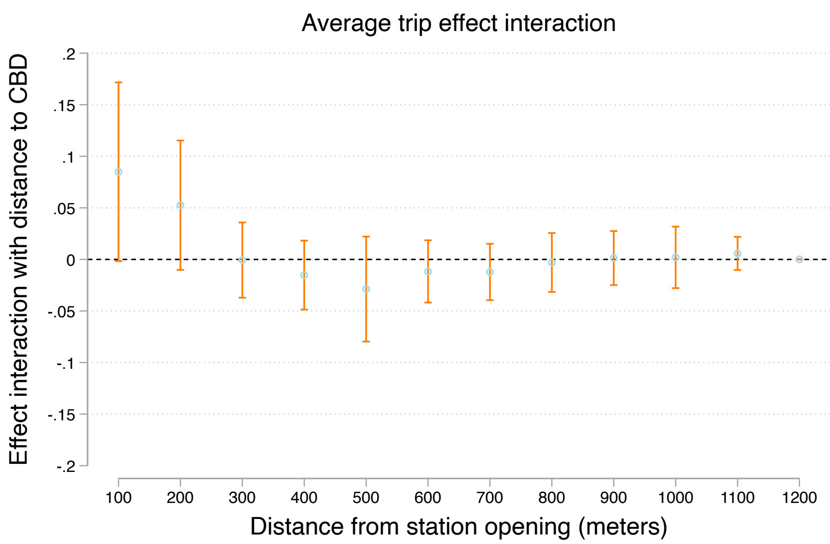


Figure A19: Differential effects for peripheral stations

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating a version of equation (5) that interacts each effect with a normalized measure of distance from the CBD. The measure uses travel time in Panel A and physical distance in Panel B. The coefficients estimate the different effect on Uber trips at varying distances from the train station opening. All errors are clustered at the station-level.

B.5 POI sample

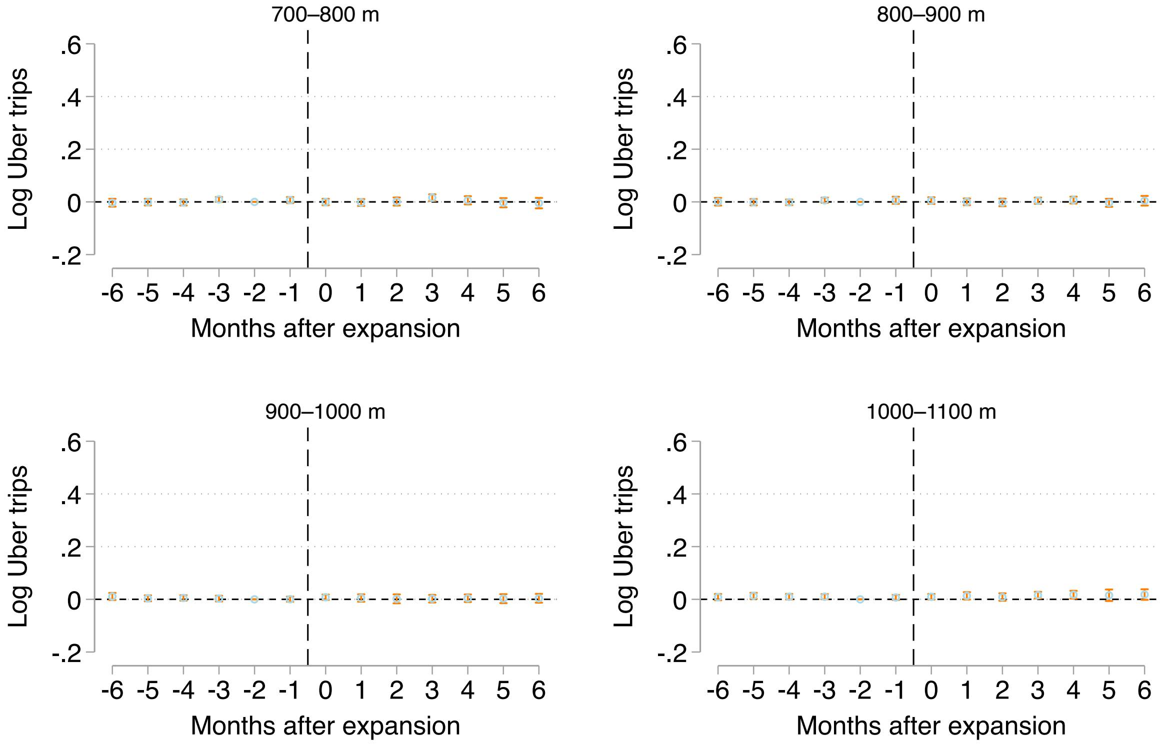


Figure A20: Dynamic effects for the POI sample 800–1100 m

Notes: This figure plots the coefficients (circles) and 95% confidence intervals (bars) from estimating equations (4) for individuals with a POI within 500–1200 m of the opening, calculated using trip data from 12 to 7 months prior to the station opening. These coefficients estimate the effect of a new train station opening on Uber trips at varying distances from the train station opening for this sample of Uber riders, relative to trips taken 1100–1200 m away. All errors are clustered at the station-level.

C Model appendix

Finding choice probabilities

In this section we first derive the expected utility from each period's mode choice using backward induction and then find choice probabilities using the familiar logit formulation. Both calculations are conditional on the permanent preference shock η_m, which is known to the traveler prior to deciding on transit. We then calculate unconditional probabilities by aggregating over the distribution of η_m.

C.1 Solving the model

Return Journey (Period 3)

When making the return journey, the traveler is constrained by whether or not they drove for the outbound journey: $m_3 = \text{car}$ if and only if $m_1 = \text{car}$. If she did not drive, she chooses between transit and Uber.

Transit involves choosing the egress mode. The inclusive value for transit is:

$$IV_3^{\text{transit}}(d, \eta) = \ln\big[\exp(\alpha_{\text{transit}} - d \cdot c_{\text{walk}} + \eta_{\text{transit}}) + \exp(\alpha_{\text{transit}} - f_{\text{Uber}} - d \cdot c_{\text{Uber}} + \eta_{\text{transit}})\big]. \tag{6}$$

The expected utility from the return journey, conditional on the outbound mode, is:

$$V_3(m_1, d, \eta) = \begin{cases} \alpha_{\text{car}} & \text{if } m_1 = \text{car} \\ \ln[\exp(IV_3^{\text{transit}}(d, \eta)) + \exp(\alpha_{\text{Uber}} + \eta_{\text{Uber}})] + \gamma & \text{if } m_1 \in \{\text{transit}, \text{Uber}\} \end{cases} \tag{7}$$

where $\gamma \approx 0.5772$ is Euler's constant, which arises from the expected maximum of Type I extreme value random variables.

Intermediate Trip (Period 2)

With probability p_s, the traveler needs to make an intermediate trip. If she didn't take her personal automobile already, she cannot do so now. This gives the following expected utility for the second period.

$$V_2(m_1) = p_s \cdot \begin{cases} \ln[\exp(\beta_{\text{car}}) + \exp(\beta_{\text{transit}}) + \exp(\beta_{\text{Uber}})] + \gamma & \text{if } m_1 = \text{car} \\ \ln[\exp(\beta_{\text{transit}}) + \exp(\beta_{\text{Uber}})] + \gamma & \text{if } m_1 \in \{\text{transit}, \text{Uber}\} \end{cases} \tag{8}$$

Outbound Journey (Period 1)

For the outbound journey, transit users choose an access mode. The inclusive value is:

$$IV_1^{\text{transit}}(d,\eta) = \ln\left[\exp(\alpha_{\text{transit}} - d \cdot c_{\text{walk}} + \eta_{\text{transit}}) + \exp(\alpha_{\text{transit}} - f_{\text{Uber}} - d \cdot c_{\text{Uber}} + \eta_{\text{transit}})\right] \quad (9)$$

The total expected utility for each outbound mode is:

$$V(m_1,d,\eta) = \begin{cases} \alpha_{\text{car}} + V_2(\text{car}) + V_3(\text{car},d,\eta) & \text{if } m_1 = \text{car} \\ IV_1^{\text{transit}}(d,\eta) + \gamma + V_2(\text{transit}) + V_3(\text{transit},d,\eta) & \text{if } m_1 = \text{transit} \\ \alpha_{\text{Uber}} + \eta_{\text{Uber}} + V_2(\text{Uber}) + V_3(\text{Uber},d,\eta) & \text{if } m_1 = \text{Uber} \end{cases} \quad (10)$$

Note that, per (8) and (7), $V_2(\text{transit}) = V_2(\text{Uber})$ and $V_3(\text{transit},d,\eta) = V_3(\text{Uber},d,\eta)$.

C.2 Conditional choice probabilities

We can now use the familiar logit formulation to derive choice probabilities conditional on the preference shock η_m.

Outbound Journey (Period 1)

The probability of choosing mode m_1 for the outbound journey is:

$$P_1(m_1 \mid d,\eta) = \frac{\exp(V(m_1,d,\eta))}{\sum_{m' \in \{\text{car,transit,Uber}\}} \exp(V(m',d,eta))} \quad (11)$$

For travelers who choose transit, the probability of using access mode a_1 is:

$$P_1(a_1 \mid m_1 = \text{transit},d,\eta) = \frac{\exp(\alpha_{\text{transit}} - f_{a_1} - d \cdot c_{a_1} + \eta_{\text{transit}})}{\sum_{a' \in \{\text{walk,Uber}\}} \exp(\alpha_{\text{transit}} - f_{a'} - d \cdot c_{a'} + \eta_{\text{transit}})}. \quad (12)$$

Intermediate Trip (Period 2)

Conditional on needing an intermediate trip (which occurs with probability p_s), the mode choice probabilities depend on the outbound mode:

For travelers who drove outbound:

$$P_2(m_2 \mid m_1 = \text{car}) = \frac{\exp(\beta_{m_2})}{\sum_{m' \in \{\text{car,transit,Uber}\}} \exp(\beta_{m'})} \quad \text{for } m_2 \in \{\text{car}, \text{transit}, \text{Uber}\} \quad (13)$$

For travelers who used transit or Uber outbound:

$$P_2(m_2 \mid m_1 \in \{\text{transit}, \text{Uber}\}) = \begin{cases} \frac{\exp(\beta_{m_2})}{\exp(\beta_{\text{transit}}) + \exp(\beta_{\text{Uber}})} & \text{if } m_2 \in \{\text{transit}, \text{Uber}\} \\ 0 & \text{if } m_2 = \text{car} \end{cases} \quad (14)$$

Return Journey (Period 3)

We start by noting that for travelers who drove outbound, the return mode is deterministic:

$$P_3(m_3 \mid m_1 = \text{car}) = \begin{cases} 1 & \text{if } m_3 = \text{car} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The conditional choice probabilities for the return journey for those who did not drive outbound are

$$P_3(m_3 \mid m_1 \in \{\text{transit}, \text{Uber}\}, d, \eta) = \begin{cases} \frac{\exp(IV_3^{\text{transit}}(d,\eta))}{\exp(IV_3^{\text{transit}}(d,\eta)) + \exp(\alpha_{\text{Uber}} + \eta_{\text{Uber}})} & \text{if } m_3 = \text{transit} \\ \frac{\exp(\alpha_{\text{Uber}} + \eta_{\text{Uber}})}{\exp(IV_3^{\text{transit}}(d,\eta)) + \exp(\alpha_{\text{Uber}} + \eta_{\text{Uber}})} & \text{if } m_3 = \text{Uber} \\ 0 & \text{if } m_3 = \text{car} \end{cases} \tag{16}$$

For travelers who choose transit for the return journey, the egress mode probabilities are

$$P_3(a_3 \mid m_3 = \text{transit}, d, \eta) = \frac{\exp(\alpha_{\text{transit}} - f_{a_3} - d \cdot c_{a_3} + \eta_{\text{transit}})}{\sum_{a' \in \{\text{walk,Uber}\}} \exp(\alpha_{\text{transit}} - f_{a'} - d \cdot c_{a'} + \eta_{\text{transit}})} \tag{17}$$

Unconditional choice probabilities

The unconditional choice probabilities can be found by integrating over possible realizations of η_m. Let $F(\eta)$ denote the joint distribution over all preference shocks η_m. Then the unconditional choice probabilities for the outbound trip and access modes are

$$P_1(m_1 \mid d) = \int P_1(m_1 \mid d, \eta) \mathrm{d}F(\eta)$$

$$P_1(a_1 \mid m_1 = \text{transit}, d) = \int P_1(a_1 \mid m_1 = \text{transit}, d, \eta) \mathrm{d}F(\eta)$$

Similarly, the unconditional choice probabilities for the return trip and access modes are

$$P_3(m_3 \mid m_1, d) = \int P_3(m_3 \mid m_1, d, \eta) \mathrm{d}F(\eta)$$

$$P_3(a_3 \mid m_3 = \text{transit}, d) = \int P_3(a_3 \mid m_3 = \text{transit}, d, \eta) \mathrm{d}F(\eta)$$

C.3 Solving the model without ride-hailing services

We also want to use our model to understand how ride-hailing services, such as Uber, have impacted public transportation. To do so, we derive the model predictions when travelers can only choose between a personal automobile and public transit.

Expected Utilities Without Uber

When Uber is not available, travelers face a more restricted choice set. Travelers can no longer use Uber as a primary or access mode. We still solve for expected utilities using backward induction.

Return Journey (Period 3)

Without Uber, travelers who took transit for the outbound journey must use transit for their return journey, and walking is the only egress mode. The expected utility from the return journey becomes:

$$\tilde{V}_3(m_1, d, \eta) = \begin{cases} \alpha_{\text{car}} & \text{if } m_1 = \text{car} \\ \alpha_{\text{transit}} - d \cdot c_{\text{walk}} + \eta_{\text{transit}} & \text{if } m_1 = \text{transit} \end{cases} \tag{18}$$

Note that there is no longer a choice in the return journey for those who took transit outbound—they must take transit and walk.

Intermediate Trip (Period 2)

The expected utility for the intermediate trip becomes:

$$\tilde{V}_2(m_1) = p_s \cdot \begin{cases} \ln[\exp(\beta_{\text{car}}) + \exp(\beta_{\text{transit}})] + \gamma & \text{if } m_1 = \text{car} \\ \beta_{\text{transit}} & \text{if } m_1 = \text{transit} \end{cases} \tag{19}$$

For those who didn't drive, transit is the only option for the intermediate trip.

Outbound Journey (Period 1)

For the outbound journey, the total expected utility for each mode is:

$$\tilde{V}(m_1, d, \eta) = \begin{cases} \alpha_{\text{car}} + \tilde{V}_2(\text{car}) + \tilde{V}_3(\text{car}, d) & \text{if } m_1 = \text{car} \\ \alpha_{\text{transit}} - d \cdot c_{\text{walk}} + \eta_{\text{transit}} + \tilde{V}_2(\text{transit}) + \tilde{V}_3(\text{transit}, d, \eta) & \text{if } m_1 = \text{transit} \end{cases} \tag{20}$$

C.4 Choice probabilities Without Uber

Outbound Journey (Period 1)

The probability of choosing each mode for the outbound journey is:

$$\tilde{P}_1(m_1 \mid d, \eta) = \frac{\exp(\tilde{V}(m_1, d, \eta))}{\exp(\tilde{V}(\text{car}, d, \eta)) + \exp(\tilde{V}(\text{transit}, d, \eta))} \tag{21}$$

Since walking is the only access mode for transit:

$$\tilde{P}_1(a_1 = \text{walk} \mid m_1 = \text{transit}, d, \eta) = 1 \tag{22}$$

Intermediate Trip (Period 2)

For travelers who drove outbound:

$$\tilde{P}_2(m_2 \mid m_1 = \text{car}) = \frac{\exp(\beta_{m_2})}{\exp(\beta_{\text{car}}) + \exp(\beta_{\text{transit}})} \tag{23}$$

For travelers who used transit outbound:

$$\tilde{P}_2(m_2 \mid m_1 = \text{transit}) = \begin{cases} 1 & \text{if } m_2 = \text{transit} \\ 0 & \text{if } m_2 = \text{car} \end{cases} \tag{24}$$

Return Journey (Period 3)

The return journey probabilities are deterministic based on the outbound mode:

$$\tilde{P}_3(m_3 \mid m_1) = \begin{cases} 1 & \text{if } m_3 = m_1 \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

And for transit users:

$$\tilde{P}_3(a_3 = \text{walk} \mid m_3 = \text{transit}, d, \eta) = 1 \tag{26}$$

Implications for Transit Ridership

The absence of Uber affects transit ridership through several channels:

1. **Eliminated competition**: Without Uber as a direct alternative to transit, some travelers who would have chosen Uber must now choose between car and transit.

2. **Last-mile problem**: Transit users must walk the entire distance to/from stops, making transit less attractive for those living farther from stops (higher d).

3. **Reduced flexibility**: The inability to use Uber for access/egress or as a backup option during transit disruptions reduces the overall attractiveness of not owning a car.

4. **Mode lock-in**: Without Uber, the decision to not drive for the outbound journey severely constrains options for the rest of the day, potentially pushing more people to drive.

The net effect on transit ridership depends on the relative magnitudes of these effects and the distribution of travelers across distances from transit stops and persistence preference shocks η.

C.5 Theoretical results

Proposition 1 *Consider the three-period mode choice model described in the text. The derivative with respect to the transit utility parameter $\alpha_{transit}$ of Uber use in period 1 conditional on the traveler's distance d and persistent transit preference η is:*

$$\frac{\partial}{\partial \alpha_{transit}}\Big(P_1(Uber|d,\eta) + P_1(transit{+}Uber|d,\eta)\Big)$$

$$= -\underbrace{P_1(Uber|d,\eta)\cdot P_1(transit|d,\eta)}_{\text{Direct substitution effect } (\geq 0)} +$$

$$\underbrace{P_1(transit + Uber|d,\eta)\cdot P_1(transit|d,\eta)\cdot \Big[P_1(Uber|d,\eta) + P_1(car|d,\eta)(1+P_3(transit|m_1 \neq car,d,\eta))\Big]}_{\text{Last-mile complementarity } (\geq 0)} +$$

$$\underbrace{P_1(Uber|d,\eta)\cdot P_1(car|d,\eta)\cdot P_3(transit|m_1 \neq car,d,\eta)}_{\text{Insurance effect } (\geq 0)}$$

Proof. The total Uber usage in period 1 consists of two components: direct Uber usage and Uber as an access mode to transit. We know that:

$$P_1(\text{transit+Uber}|d,\eta) = P_1(a_1 = \text{Uber}|m_1 = \text{transit}, d,\eta) \times P_1(m_1 = \text{transit}|d,\eta)$$

Taking the derivative of total Uber usage:

$$\frac{\partial}{\partial \alpha_{\text{transit}}}\Big(P_1(\text{Uber}|d,\eta) + P_1(\text{transit+Uber}|d,\eta)\Big)$$

$$= \frac{\partial P_1(\text{Uber}|d,\eta)}{\partial \alpha_{\text{transit}}} + \frac{\partial}{\partial \alpha_{\text{transit}}}\Big[P_1(a_1 = \text{Uber}|m_1 = \text{transit}, d,\eta) \times P_1(\text{transit}|d,\eta)\Big]$$

$$= \frac{\partial P_1(\text{Uber}|d,\eta)}{\partial \alpha_{\text{transit}}} + P_1(a_1 = \text{Uber}|m_1 = \text{transit}, d,\eta) \times \frac{\partial P_1(\text{transit}|d,\eta)}{\partial \alpha_{\text{transit}}}$$

The last line follows since the access mode choice depends only on the relative costs of walking versus Uber for accessing transit, not on α_{transit} directly:

$$\frac{\partial P_1(a_1 = \text{Uber}|m_1 = \text{transit}, d,\eta)}{\partial \alpha_{\text{transit}}} = 0$$

Step 1: Period 1 derivatives. From the multinomial logit model and backward induction:

$$\frac{\partial V(\text{car}, d, \eta)}{\partial \alpha_{\text{transit}}} = 0$$
$$\frac{\partial V(\text{transit}, d, \eta)}{\partial \alpha_{\text{transit}}} = 1 + P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)$$
$$\frac{\partial V(\text{Uber}, d, \eta)}{\partial \alpha_{\text{transit}}} = P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)$$

Using properties of multinomial logit derivatives:

$$\frac{\partial P_1(\text{transit}|d, \eta)}{\partial \alpha_{\text{transit}}} = P_1(\text{transit}|d, \eta) \times [1 + P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) - \bar{V}']$$

$$\frac{\partial P_1(\text{Uber}|d, \eta)}{\partial \alpha_{\text{transit}}} = P_1(\text{Uber}|d, \eta) \times [P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) - \bar{V}']$$

$$\frac{\partial P_1(\text{car}|d, \eta)}{\partial \alpha_{\text{transit}}} = P_1(\text{car}|d, \eta) \times [0 - \bar{V}'] = -P_1(\text{car}|d, \eta) \times \bar{V}'$$

where

$$\bar{V}' = P_1(\text{transit}|d, \eta) \times [1 + P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)] + P_1(\text{Uber}|d, \eta) \times P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)$$

Step 2: Substitute and combine terms. Substituting the expressions in, we get:

$$\begin{aligned}\frac{\partial P_1(\text{Uber}|d, \eta)}{\partial \alpha_{\text{transit}}} = {} & P_1(\text{Uber}|d, \eta) \times P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) - \\ & P_1(\text{Uber}|d, \eta) \times P_1(\text{transit}|d, \eta) \times [1 + P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)] - \\ & P_1(\text{Uber}|d, \eta)^2 \times P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)\end{aligned}$$

$$\begin{aligned}\frac{\partial P_1(\text{transit}|d, \eta)}{\partial \alpha_{\text{transit}}} = {} & P_1(\text{transit}|d, \eta) \times [1 + P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)] - \\ & P_1(\text{transit}|d, \eta) \times P_1(\text{transit}|d, \eta) \times [1 + P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)] \\ & - P_1(\text{transit}|d, \eta) \times P_1(\text{Uber}|d, \eta) \times P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)\end{aligned}$$

Take the first derivative. We can rewrite this expression as:

$$\frac{\partial P_1(\text{Uber}|d,\eta)}{\partial \alpha_{\text{transit}}} = P_1(\text{Uber}|d,\eta)\times$$
$$\left[P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \times (1 - P_1(\text{transit}|d,\eta) - P_1(\text{Uber}|d,\eta)) - P_1(\text{transit}|d,\eta)\right]$$
$$= P_1(\text{Uber}|d,\eta) \times \left[P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \times P_1(\text{car}|d,\eta) - P_1(\text{transit}|d,\eta)\right]$$

The second derivative can be rewritten as:

$$\frac{\partial P_1(\text{transit}|d,\eta)}{\partial \alpha_{\text{transit}}} = P_1(\text{transit}|d,\eta)\times$$
$$\left[P_1(\text{Uber}|d,\eta) + P_1(\text{car}|d,\eta)(1 + P_3(\text{transit}|m_1 \neq \text{car}, d, \eta))\right]$$

Combining the terms with the original product-rule derivative for total Uber use gives the expression in the proposition. ∎

Proposition 2 *Consider the three-period mode choice model described in the text. The derivative of the conditional Uber use in period 3 with respect to the transit utility parameter $\alpha_{transit}$ is:*

$$\frac{\partial}{\partial \alpha_{transit}}\left(P_3(Uber|d,\eta) + P_3(transit\text{+}Uber|d,\eta)\right)$$
$$= \underbrace{P_1(car|d,\eta) \cdot \Gamma \cdot [P_3(Uber|m_1 \neq car, d, \eta) + P_3(transit + Uber|m_1 \neq car, d, \eta)]}_{\text{Composition effect } (\geq 0)} -$$
$$\underbrace{(1 - P_1(car|d,\eta)) \cdot P_3(Uber|m_1 \neq car, d, \eta) \cdot P_3(transit|m_1 \neq car, d, \eta)}_{\text{Direct substitution effect } (\geq 0)} +$$
$$\underbrace{(1 - P_1(car|d,\eta)) \cdot P_3(Uber|m_1 \neq car, d, \eta) \cdot P_3(transit|m_1 \neq car, d, \eta) \cdot P_3(a_3 = Uber|m_3 = transit, d)}_{\text{Last-mile complementarity } (\geq 0)}$$

where:

- $\Gamma = P_1(transit|d,\eta) + (1 - P_1(car|d,\eta)) \cdot P_3(transit|m_1 \neq car, d, \eta) \geq 0$

Proof. The overall Uber mode share in period 3 consists of direct Uber trips and Uber used as the egress mode for transit.

$$P_3(\text{Uber}|d,\eta) + P_3(\text{transit+Uber}|d,\eta) =$$
$$(1 - P_1(\text{car}|d,\eta)) \cdot [P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) + P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta)]$$

Taking the derivative with respect to α_{transit}:

$$\frac{\partial}{\partial \alpha_{\text{transit}}}\Big(P_3(\text{Uber}|d,\eta) + P_3(\text{transit+Uber}|d,\eta)\Big) =$$
$$-\frac{\partial P_1(\text{car}|d,\eta)}{\partial \alpha_{\text{transit}}} \cdot [P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) + P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta)]$$
$$+ (1 - P_1(\text{car}|d,\eta)) \cdot \left[\frac{\partial P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta)}{\partial \alpha_{\text{transit}}} + \frac{\partial P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta)}{\partial \alpha_{\text{transit}}}\right]$$

Step 1: Period 1 derivatives. From the multinomial logit model and backward induction:

$$\begin{aligned}
\frac{\partial V(\text{car}, d, \eta)}{\partial \alpha_{\text{transit}}} &= 0 \\
\frac{\partial V(\text{transit}, d, \eta)}{\partial \alpha_{\text{transit}}} &= 1 + P_3(\text{transit}|m_1 \in \{\text{Uber, transit}\}, d, \eta) \\
\frac{\partial V(\text{Uber}, d, \eta)}{\partial \alpha_{\text{transit}}} &= P_3(\text{transit}|m_1 \in \{\text{Uber, transit}\}, d, \eta)
\end{aligned}$$

Using the multinomial logit derivative formula:

$$\begin{aligned}
\frac{\partial P_1(\text{car}|d,\eta)}{\partial \alpha_{\text{transit}}} &= -P_1(\text{car}|d,\eta) \cdot [P_1(\text{transit}|d,\eta) \cdot (1 + P_3(\text{transit}|m_1 \in \{\text{Uber, transit}\}, d, \eta)) \\
&\quad + P_1(\text{Uber}|d,\eta) \cdot P_3(\text{transit}|m_1 \in \{\text{Uber, transit}\}, d, \eta)] \\
&= -P_1(\text{car}|d,\eta) \cdot [P_1(\text{transit}|\eta) + (1 - P_1(\text{car}|d,\eta)) \cdot P_3(\text{transit}|m_1 \in \{\text{Uber, transit}\}, d, \eta)] \\
&= -P_1(\text{car}|d,\eta) \cdot \Gamma
\end{aligned}$$

Step 2: Period 3 conditional probability derivatives. For travelers who did not drive in period 1, from the multinomial logit model for period 3:

$$\frac{\partial P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta)}{\partial \alpha_{\text{transit}}} = -P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)$$

For the transit + Uber combination, we need to consider both the choice of transit as primary mode and Uber as egress mode.

$$P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta) = P_3(a_3 = \text{Uber}|m_1 = \text{transit}, d, \eta) \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)$$

Taking the derivative:

$$\begin{aligned}
\frac{\partial P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta)}{\partial \alpha_{\text{transit}}} &= \frac{\partial P_3(a_3 = \text{Uber}|m_3 = \text{transit}, d, \eta)}{\partial \alpha_{\text{transit}}} \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) + \\
&\quad P_3(a_3 = \text{Uber}|m_3 = \text{transit}, d, \eta) \cdot \frac{\partial P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)}{\partial \alpha_{\text{transit}}}
\end{aligned}$$

The egress mode choice depends only on the fixed and marginal costs. However, improving transit quality (α_{transit}) makes transit more attractive relative to Uber as the primary mode, so:

$$\frac{\partial P_3(a_3 = \text{Uber}|m_3 = \text{transit}, d, \eta)}{\partial \alpha_{\text{transit}}} = 0$$

$$\frac{\partial P_3(\text{transit}|m_1 \neq \text{car}, d, \eta)}{\partial \alpha_{\text{transit}}} = P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \cdot P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta)$$

Therefore:

$$\frac{\partial P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta)}{\partial \alpha_{\text{transit}}} = \\ P_3(a_3 = \text{Uber}|m_3 = \text{transit}, d, \eta) \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \cdot P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta)$$

Step 3: Combining terms. Substituting into the main derivative:

$$\begin{aligned} &\frac{\partial}{\partial \alpha_{\text{transit}}}\Big(P_3(\text{Uber}|d,\eta) + P_3(\text{transit+Uber}|d,\eta)\Big) = \\ &\quad P_1(\text{car}|d,\eta) \cdot \Gamma \cdot [P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) + P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta)] \\ &\quad + (1 - P_1(\text{car}|d,\eta)) \cdot [-P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \\ &\quad + P_3(a_3 = \text{Uber}|m_3 = \text{transit}, d, \eta) \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \cdot P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta)] \end{aligned}$$

Rearranging:

$$\begin{aligned} &\frac{\partial}{\partial \alpha_{\text{transit}}}\Big(P_3(\text{Uber}|d,\eta) + P_3(\text{transit+Uber}|d,\eta)\Big) = \\ &\quad P_1(\text{car}|d,\eta) \cdot \Gamma \cdot [P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) + P_3(\text{transit + Uber}|m_1 \neq \text{car}, d, \eta)] \\ &\quad - (1 - P_1(\text{car}|d,\eta)) \cdot P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \\ &\quad + (1 - P_1(\text{car}|d,\eta)) \cdot P_3(\text{Uber}|m_1 \neq \text{car}, d, \eta) \cdot P_3(\text{transit}|m_1 \neq \text{car}, d, \eta) \cdot P_3(a_3 = \text{Uber}|m_3 = \text{transit}, d, \eta) \end{aligned}$$

■

Proposition 3 *Consider the three-period mode choice model described in the text, where $\beta_{transit} = \beta_{transit}(\alpha_{transit})$ is a differentiable function with $\beta'_{transit}(\alpha_{transit}) \geq 0$. The derivative of the Uber mode share in period 2 (conditional on period 2 occurring) with respect to the transit utility parameter $\alpha_{transit}$ is:*

$$\frac{\partial P_2(Uber|\eta)}{\partial \alpha_{transit}} = \underbrace{P_1(car|\eta) \cdot \Gamma \cdot [P_2(Uber|m_1 \in \{Uber,\ transit\}) - P_2(Uber|car)]}_{\textit{Composition effect } (\geq 0)} \\ - \underbrace{\beta'_{transit}(\alpha_{transit}) \cdot \Omega}_{\textit{Substitution effect } (\geq 0)}$$

where:

- $\Gamma = P_1(transit|\eta) + (1 - P_1(car|\eta)) \cdot P_3(transit|m_1 \in \{Uber,\ transit\}, d, \eta) \geq 0$
- $\Omega = P_1(car|\eta) \cdot P_2(Uber|car) \cdot P_2(transit|car) + [1 - P_1(car|\eta)] \cdot P_2(Uber|m_1 \in \{Uber,\ transit\}) \cdot P_2(transit|m_1 \in \{Uber,\ transit\}) \geq 0$

The derivative is non-negative if and only if:

$$\beta'_{transit}(\alpha_{transit}) \leq \frac{P_1(car|\eta) \cdot \Gamma \cdot [P_2(Uber|m_1 \in \{Uber,\ transit\}) - P_2(Uber|car)]}{\Omega}$$

Proof. The overall Uber mode share in period 2 is:

$$P_2(\text{Uber}|\eta) = P_1(\text{car}|\eta) \cdot P_2(\text{Uber}|\text{car}) + [1 - P_1(\text{car}|\eta)] \cdot P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\})$$

where we use the fact that $P_1(\text{transit}) + P_1(\text{Uber}) = 1 - P_1(\text{car})$ since travelers who don't drive have the same choice set in periods 2 and 3. Also note that conditional on period 1 mode choices, intermediate trip probabilities do not depend on η.

Taking the derivative with respect to α_{transit}:

$$\begin{aligned}\frac{\partial P_2(\text{Uber}|\eta)}{\partial \alpha_{\text{transit}}} &= \frac{\partial P_1(\text{car}|\eta)}{\partial \alpha_{\text{transit}}} \cdot [P_2(\text{Uber}|\text{car}) - P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\})] \\ &\quad + P_1(\text{car}|\eta) \cdot \frac{\partial P_2(\text{Uber}|\text{car})}{\partial \alpha_{\text{transit}}} + [1 - P_1(\text{car}|\eta)] \cdot \frac{\partial P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\})}{\partial \alpha_{\text{transit}}}\end{aligned}$$

Step 1: Period 1 derivatives. *See Step 1 of the proof for Proposition 2.*

Step 2: Period 2 conditional probability derivatives. From the multinomial logit model:

$$\begin{aligned}\frac{\partial P_2(\text{Uber}|\text{car})}{\partial \alpha_{\text{transit}}} &= \frac{\partial P_2(\text{Uber}|\text{car})}{\partial \beta_{\text{transit}}} \cdot \beta'_{\text{transit}}(\alpha_{\text{transit}}) \\ &= -P_2(\text{Uber}|\text{car}) \cdot P_2(\text{transit}|\text{car}) \cdot \beta'_{\text{transit}}(\alpha_{\text{transit}})\end{aligned}$$

Similarly:

$$\begin{aligned}\frac{\partial P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\})}{\partial \alpha_{\text{transit}}} &= -P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\}) \\ &\quad \times P_2(\text{transit}|m_1 \in \{\text{Uber},\ \text{transit}\}) \cdot \beta'_{\text{transit}}(\alpha_{\text{transit}})\end{aligned}$$

Step 3: Combining terms. Substituting into the main derivative:

$$\begin{aligned}\frac{\partial P_2(\text{Uber}|\eta)}{\partial \alpha_{\text{transit}}} &= -P_1(\text{car}|\eta)\cdot\Gamma\cdot[P_2(\text{Uber}|\text{car}) - P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\})] \\ &\quad - \beta'_{\text{transit}}(\alpha_{\text{transit}})\cdot[P_1(\text{car}|\eta)\cdot P_2(\text{Uber}|\text{car})\cdot P_2(\text{transit}|\text{car}) \\ &\quad + [1 - P_1(\text{car}|\eta)]\cdot P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\}) \\ &\quad \cdot P_2(\text{transit}|m_1 \in \{\text{Uber},\ \text{transit}\})]\end{aligned}$$

Since $P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\}) > P_2(\text{Uber}|\text{car})$ when $\beta_{\text{car}} > -\infty$, we can rewrite this as:

$$\begin{aligned}\frac{\partial P_2(\text{Uber}|\eta)}{\partial \alpha_{\text{transit}}} &= P_1(\text{car}|\eta)\cdot\Gamma\cdot[P_2(\text{Uber}|m_1 \in \{\text{Uber},\ \text{transit}\}) \\ &\quad - P_2(\text{Uber}|\text{car})] - \beta'_{\text{transit}}(\alpha_{\text{transit}})\cdot\Omega\end{aligned}$$

The condition for non-negativity follows immediately. ■

Corollary 1 *The derivative of the overall Uber market share in period 2 (conditional on period 2 occurring) with respect to the transit utility parameter $\alpha_{transit}$ is:*

$$\begin{aligned}\frac{\partial P_2(Uber)}{\partial \alpha_{transit}} &= \underbrace{\int P_1(car|\eta)\cdot\Gamma\cdot[P_2(Uber|m_1 \in \{Uber,\ transit\}) - P_2(Uber|car)]dF(\eta)}_{\textit{Aggregate composition effect } (\geq 0)} \\ &\quad - \underbrace{\int \left(\beta'_{transit}(\alpha_{transit})\cdot\Omega\right)dF(\eta)}_{\textit{Aggregate substitution effect } (\geq 0)}\end{aligned}$$

Proof. The unconditional choice probability for Uber in period 2 comes from integrating over η.

$$P_2(\text{Uber}) = \int P_2(\text{Uber}|\eta)\mathrm{d}F(\eta)$$

Because the domain of integration does not depend on α_{transit} and the integrand is sufficiently well-behaved, we can differentiate under the integral sign.

$$\frac{\partial P_2(\text{Uber})}{\partial \alpha_{\text{transit}}} = \int \frac{\partial P_2(\text{Uber}|\eta)}{\partial \alpha_{\text{transit}}}\mathrm{d}F(\eta)$$

The integrand for each of the composition and substitution effects is positive for each η by the prior result, and so integrating over the distribution of η will also be positive. ■

Corollary 2 *When $\beta_{transit}$ is independent of $\alpha_{transit}$ (i.e., $\beta'_{transit}(\alpha_{transit}) = 0$), we have $\frac{\partial P_2(Uber)}{\partial \alpha_{transit}} \geq 0$ with equality if and only if $P_1(car) \in \{0, 1\}$.*

C.6 Mapping the model to the data

It will be simpler to discuss the choice of transit plus an access mode as a single choice, and so note that

$$\begin{gathered}P_1(\text{transit + Uber} \mid d) = P_1(a_1 = \text{Uber} \mid m_1 = \text{transit}, d) \times P_1(m_1 = \text{transit} \mid d),\\ P_3(\text{transit + Uber} \mid m_1 \neq \text{car}, d) = P_3(a_3 = \text{Uber} \mid m_3 = \text{transit}, d) \times P_3(m_3 = \text{transit} \mid m_1 \neq \text{car}, d).\end{gathered} \tag{27}$$

In our data we observe the number of Uber trips near a transit stop and the number of Uber trips in the rest of the city. If N is the total number of travelers within $\bar{d}$ meters of the transit stop, then the number of Uber trips occurring between d_1 and d_2 meters of the transit stop is

$$\begin{aligned}\int_{d=d_1}^{d_2} \Big(P_1(\text{Uber} \mid d) + P_1(\text{transit + Uber} \mid d) \\ + \big[1 - P(\text{car} \mid d)\big] \big[P_3(\text{Uber} \mid m_1 \neq \text{car}, d) \\ + P_3(\text{transit + Uber} \mid m_1 \neq \text{car}, d)\big]\Big) f(d)\, \mathrm{d}d,\end{aligned} \tag{28}$$

where $f(d)$ is the population density function. Assuming uniform population density, $f(d)$ simplifies to $2Nd/\bar{d}^2$.

The number of Uber trips happening in the rest of the city by people living between d_1 and d_2 meters of the transit stop is

$$\begin{aligned}\int_{d=d_1}^{d_2} \Big(P_2(\text{Uber} \mid m_1 = \text{car}) \times P_1(\text{car} \mid d) \\ + P_2(\text{Uber} \mid m_1 \neq \text{car}) \times P_1(m_1 \neq \text{car} \mid d)\Big) f(d)\, \mathrm{d}d\end{aligned} \tag{29}$$

C.7 Model simulations

Table A3: Model parameters for simulation in Figure 2

Parameter	Positive impact on Uber trips	Negative impact on Uber trips	Mixed impact on Uber trips
α_{car}	0	0	0
α_{transit}	-1	0.7	1.5
α_{Uber}	-1.5	-1.4	-2
β_{car}	0	0	0
β_{transit}	0.5	0.2	0.5
β_{Uber}	-1.5	-1.4	-2
f_{Uber}	3	10	3
c_{walk}	0.003	0.007	0.004
c_{Uber}	0.0001	0.004	0.0003
p_s	0.75	0.50	0.50
p_r	1	1	1
$\sigma_{\alpha_{\text{transit}}}$	0	0.5	0.5
$\sigma_{\alpha_{\text{Uber}}}$	1	2	2
n_{sim}	1000	1000	1000

Notes: Each model simulates 1000 individuals with random preferences over modes, and uses distance bands from 0 to 5000 meters for probability calculations. However, all plots are restricted to 0–1200 meters from the station. Parameter definitions are discussed in Section 2.

D Empirics appendix

D.1 Event study for far distances

We adapt our empirical approach slightly to assess the validity of our control group. Because we can no longer use a farther distance band as a control, we instead use untreated observations across all stations and rely on city-by-month fixed effects to absorb secular trends. We implement the estimator of de Chaisemartin and D'Haultfœuille (2024), using a linear model with the inverse hyperbolic sine of trips (in hundreds) as the outcome.[17] The event-study specification for a given distance d^* is:

$$
\begin{aligned}
y_{dit} = \gamma_{c(i)t} + \delta_{di} & \sum_{j \in \{-6,-5,\ldots,6\} \setminus \{-2\}} \alpha_{d^*,j} \times \mathbb{1}\{\tau_{it} = j\} \times \mathbb{1}\{d = d^*\} \\
& + \left(\beta_{1,d^*} \times \mathbb{1}\{\tau_{it} < -6\} + \beta_{2,d^*} \times \mathbb{1}\{\tau_{it} > 6\}\right) \mathbb{1}\{d = d^*\} + \epsilon_{dit}, \quad \text{where } d = d^*
\end{aligned}
$$

D.2 Adjacent distance bins methodology

Total effects by distance calculation

Our main difference-in-differences estimates come from (5), 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 $p \in \{100, \ldots., 1100\}$ 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.

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

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

[17] Log or IHS transformations are not scale-invariant (Chen and Roth, 2023); we normalize trips by 100 so that estimated effects at 0–100 m align with our PPML estimates.

distances. To implement this, we simplify the above regression as follows.

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

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. The average effect for 1-6 months after the transit opening are reported in Figure A10.