

# Optimal Urban Transportation Policy: Evidence from Chicago<sup>†</sup>

Milena Almagro\*, Juan Camilo Castillo<sup>†</sup>, Nathaniel Hickok<sup>‡</sup>, Felipe Kup<sup>§</sup>, Tobias Salz<sup>¶</sup>

SEPTEMBER 1, 2023

**Preliminary and incomplete. Please do not cite or circulate.**

## Abstract

Urban transportation policies have become a focal point in cities' efforts to curb congestion and address environmental and distributional concerns. This paper characterizes the optimal mix of policies and evaluates their welfare and distributional effects. To that end, we present a framework of a municipal government aiming to maximize welfare. The government chooses the prices and frequencies of different modes of transportation, subject to a budget constraint that introduces monopoly-like distortions. We move on to an empirical application of this framework to the city of Chicago. We first construct a novel dataset of all relevant transportation modes. On the demand side, our empirical model captures the rich heterogeneity in travel choices. On the supply side we account for differential congestion and costs of different road-based modes. Our counterfactual results suggest that if the city only intervenes on public transit, it should lower transit prices even further but also lower frequency to meet its budget constraint. On the other hand, introducing a per-kilometer tax on drivers leads to higher welfare gains.

**JEL classification:** L91, L5, L13, H23, R41, R48.

**Keywords:** Urban transportation, Public Transit Subsidy Design, Congestion Pricing, Spatial Equilibrium, Ramsey Pricing.

---

<sup>†</sup> For helpful comments we thank seminar audiences at Booth, Harvard, MIT, Rice, Texas A&M, Yale, CEMFI, EIEF Junior Applied Micro Conference, the Tinos IO Conference, LACEA-LAMES, NBER Market Design 2022 Fall Meeting, Philadelphia Fed, UCLA IO/Spatial Conference, Berkeley, Toronto, Georgetown, Johns Hopkins, Maryland, Wharton Real Estate, UIUC, UNLP, Washington University in St Louis, St Louis Fed, IIOC, LSE, Oxford, Warwick, IFS/LSE/UCL IO workshop, Cowles Models and Measurement, and ERWIT/CURE. We acknowledge financial support from the NSF (Award No. 1559013, Supplement), through the NBER's "Transportation Economics in the 21st Century" initiative. We are grateful to Isis Durrmeyer, Clara Santamaria, and Daniel Sturm for their insightful discussions. We also thank Steve Berry, Gilles Duranton, Jessie Handbury, Gabriel Kreindler, Jing Li, and Chris Severen for helpful comments. We thank Gilles Duranton and Prottoy Akbar for helping us obtain travel time data. All remaining errors are our own.

\* University of Chicago Booth School of Business and NBER, email: [milena.almagro@chicagobooth.edu](mailto:milena.almagro@chicagobooth.edu)

<sup>†</sup> University of Pennsylvania Economics and NBER, email: [jccast@upenn.edu](mailto:jccast@upenn.edu)

<sup>‡</sup> MIT Economics, email: [nhickok@mit.edu](mailto:nhickok@mit.edu)

<sup>§</sup> University of Pennsylvania Economics, email: [kupf@sas.upenn.edu](mailto:kupf@sas.upenn.edu)

<sup>¶</sup> MIT Economics and NBER, email: [tsalz@mit.edu](mailto:tsalz@mit.edu)

# 1 Introduction

Since the 1950s, urban transportation in the US has been characterized by the overwhelming use of personal cars. This heavy reliance on personal vehicles poses a significant challenge for city governments. They are grappling with the task of managing the \$87 billion in annual welfare costs due to congestion, minimizing environmental impact, and ensuring affordable transportation access.<sup>1,2</sup> Moreover, the high usage of personal cars persists despite high subsidies for public transit. These subsidies cover on average of three quarters of marginal trip costs, yet public transit still accounts for less than ten percent of the 1.1 billion trips that Americans undertake every day.<sup>3</sup>

Cities' increased focus on environmental and distributional concerns as well as traffic congestion has led to a renewed discussion about the right mix of urban transportation policies to achieve these goals.<sup>4</sup> Some argue that public transit should be even cheaper, and indeed several municipalities have recently introduced free public transit.<sup>5</sup> Others argue that, instead of lowering fares, cities should provide more frequent, higher-quality public transit.<sup>6</sup> These proposals appeal to different types of riders, and stressed municipal budgets imply that it may not be feasible to do both. Alternatively, other cities have started taxing private modes. For example, London has enacted a £15 congestion surcharge during the daytime and New York City has recently approved a cordon tax below 60th street.<sup>7</sup> This type of policy not only has the power to steer consumers away from private modes in favor of public alternatives, but can also help financially strained municipalities due to tax revenue collection. Importantly, these three different policy levers

---

<sup>1</sup> See [World Economic Forum— US Traffic Congestion Cost in 2018](#).

<sup>2</sup> A private car emits 0.96 pounds of CO<sub>2</sub> per passenger-mile whereas public transit emits 0.45, even with low utilization rates.

<sup>3</sup> See <https://www.bts.dot.gov>. and [newgeography.com](#)

<sup>4</sup> See [Brookings — U.S. Transportation policy](#) and [HKS — Free Public Transit](#)

<sup>5</sup> See [NYT — “Should Public Transit Be Free? More Cities Say, Why Not?”](#)

<sup>6</sup> See [The Conversation — Low-cost, high-quality public transportation](#).

<sup>7</sup> For a comprehensive list of congestion taxes see <https://ops.fhwa.dot.gov>.

cannot be evaluated in isolation — they are linked through mode substitution on the demand side (Anderson, 2014), road congestion, and municipal budget constraints.

Given these interactions, what is the right mix of urban transportation policies? Should cities aim to increase the use of public transit, discourage the use of roads, or some combination of the two? In this paper, we measure the welfare effects of urban transit policies and characterize the optimal mix of policies for a budget-constrained municipal government. Given prevailing difficulties to build new transit infrastructure in the US (Brooks and Liscow, 2021), we focus on three interventions that do not require new infrastructure: changing the fares of public transit, adjusting their service frequency, and road pricing.<sup>8</sup>

In order to measure the welfare effects of these interventions, we must account for several ingredients. First, we need to understand how people substitute across transit modes as prices and travel times change. Second, we need to account for important forces that arise in transportation markets—traffic and environmental externalities as well as scale economies. Third, we have to account for the interaction of travel decisions across different locations, which are linked through traffic congestion.

We first formulate a framework of a municipal government who wants to maximize welfare, accounting for these efficiency and distributional concerns. The government chooses the prices and service levels (frequency) of different modes of transportation, subject to a budget constraint that accounts for operational costs, fare revenue, and taxes. We first find that an unconstrained social planner would set price minus marginal cost equal to the marginal externality—the canonical result about optimal taxes. However, our social planner deviates from this solution because budget considerations introduce two monopoly-like distortions. First, the planner charges markups that downwards-distort quantities. Second, given the ability to choose service levels, the planner is directly affect-

---

<sup>8</sup> We focus on short-run adjustments, keeping residents' and firms' locations as well as infrastructure fixed. We thus measure welfare gains in the short run, before firms and residents readjust.

ing the quality that travelers experience. Budget considerations distort quality towards the marginal consumer, as in Spence (1975).

Next, we move to an empirical application of this framework to the city of Chicago, which is the third most populous city in the US and has the second largest public transit system after New York City.<sup>9</sup> Chicago is an ideal setting for our purposes for several reasons. First, it relies to a meaningful extent on both public and private transportation. We should therefore expect that both changes in public transit provision and the introduction of congestion pricing have sizeable effects. Second, Chicago is a city with large economic disparities, which means that it is a good testing ground to quantify the distributional effects of transportation policies.

Next to economic considerations, Chicago is particularly good in terms of data availability. We combine several data sources to construct a rich dataset of travel flows, travel times, and prices for all relevant modes and markets. First, we have access to the near-universe of public transit trips, ride-share trips, and taxi trips. One challenge we face is that there are no official records of trips done by private cars, the most common mode in the city. To overcome this problem, we determine mobility patterns from individual cellphone location records, which we use to construct the total number of trips in the city. We then recover the number of car trips by subtracting public transit from the total number of trips constructed using our cellphone data. The richness of our data allows us to perform our analysis at a very granular level of aggregation: we define markets at the hourly level from one community area to another and directly measure mode choice across these different markets.<sup>10</sup> Given that we have mode-shares across a multitude of markets, we use across-market and across-mode variation to estimate our structural parameters. This comprehensive approach of measuring mode shares has the advantage that we can use an

---

<sup>9</sup> See US transit agency statistics at [Statista](#).

<sup>10</sup> There is a total of 77 community areas and 801 census tracts in the city of Chicago.

instrumental variable strategy to deal with endogeneity concerns, as opposed to studies that rely on individual variation in travel survey data. Moreover, measuring this granularity is important to capture the distributional effects of our policy counterfactuals and to be able to quantify policies that change across space, such as cordon pricing, or over time, such as peak and non-peak transit fares.

Our data reveals that there is substantial heterogeneity in mode choice across travelers of different income levels. The share of ride-hailing is positively correlated with traveler's income as opposed to the share of buses that presents a negative relationship. The share of car trips, on the other hand, presents an inverted U-shape pattern. We also find that access to transportation varies substantially with income levels. For example, distance to the nearest train station is negatively correlated with choosing to travel by train. These patterns together motivate the need of rich heterogeneity in both locations characteristics, such as access to public transit, but also in traveler's substitution patterns across income levels.

We then use our data to estimate the main parts of our model. To estimate demand we follow Berry *et al.* (1995), while also allowing for rich heterogeneity based on consumer demographics, as in Petrin (2002). This approach has the advantage that we can use standard inversion techniques to address potential endogeneity concerns. At the same time, we allow for heterogeneous substitution patterns across consumer characteristics such as car ownership and income. Our demand estimation results reveal substantial heterogeneity in the value of time across travelers, ranging from \$21 to \$58 for consumers in the lowest and highest income quintiles, respectively.<sup>11</sup>

A key ingredient of our model is a congestion technology that allows us to map traffic flows into speeds. Our estimation accounts for the difference in the marginal congestion

---

<sup>11</sup> For comparison, the average hourly wage in Chicago's metropolitan area is \$30. See [wage statistics from Bureau of Labor Statistics for the Chicago region](#).

effect of a bus and a car. We exploit variation across hours of the day in travel speeds and in the number of vehicles traveling between adjacent community areas, following Akbar and Duranton (2017); Akbar *et al.* (2018); Couture *et al.* (2018); Kreindler (2023). Our analysis reveals that the marginal congestion of a bus is between four to six times the congestion produced by a car. Given that the average bus is utilized only by ten passengers, these results mean that the marginal congestion by passenger is only twice as large for cars compared to buses. Moreover, we find an average elasticity of travel speed to traffic flows between -0.12 to -0.19, comparable to existing estimates in the literature (Akbar and Duranton, 2017; Couture *et al.*, 2018). Finally, we find substantial heterogeneity across different areas of the city: traffic has a larger effect on travel times in central areas than in more peripheral areas.

With our estimates of preferences and the congestion technology, we proceed to quantify the welfare effects of transportation policies. We consider different policy interventions and compare them to the status quo. First, the planner can only intervene on public transit. We find that the planner would like to reduce public transit prices by 90% and service levels by 20-30%. This policy leads to welfare gains of \$2M per week, equivalent to \$1.11 per traveller per week. In a second scenario, the planner only controls prices on private alternatives. We find that optimal congestion prices should be of the order of \$0.25/km, leading to an increase in welfare of \$6M per week. Considering that the average commute trip distance is 8km, this tax on cars imply a daily congestion charge of \$4 for commuters.<sup>12</sup> However, even though optimal congestion prices return the largest efficiency gains relative to the status quo, they cause a large, regressive decrease in consumer surplus. When we put the two policies together, we find similar adjustments in prices and service levels, which lead to a total welfare gain of \$8M per week. Finally, as a benchmark, we let the social planner set all prices including ride-hailing prices. We find

---

<sup>12</sup>For reference, the cordon price currently implemented in London is 15 sterling pounds.

that she only wants to reduce them by 5%. The reason behind such a small change is that market power keeps prices high, implicitly acting as a Pigouvian tax.

**Related Literature** Our work relates to several strands of the literature in transportation economics and industrial organization.

Classical papers in transportation economics develop models that capture the interaction between schedule constraints and congestion (Small, 1982; Arnott *et al.*, 1990, 1993; Small *et al.*, 2005). In this work, we combine a congestion model with the demand approach in industrial organization (Berry, 1994; Berry *et al.*, 1995), which allows us to model rider heterogeneity and account for endogeneity of prices.

Our work contributes to the empirical work that measures the effects of traffic congestion (Akbar and Duranton, 2017; Couture *et al.*, 2018; Akbar *et al.*, 2018). Hall (2018) shows theoretically that pricing some highway lanes (“Lexus lanes”) can lead to Pareto improvements. Yang *et al.* (2020) exploit the variation that is induced by driving restrictions in Beijing to derive the optimal road congestion surcharge. Both of these papers do not account for mode substitution or the interaction between public and private transportation. Like Parry and Small (2009) we derive optimal prices for public transit but our focus lies on the joint effect of prices and quality improvements. Unlike Parry and Small (2009), who calibrate their model, we estimate demand and supply based on rich new data, which allows us to speak to the distributional effects of transit policies.

Some papers analyze transportation markets based on models of spatial equilibrium. These studies are closely linked to theoretical work by Arnott (1996), which shows that taxis should be subsidized because of increasing returns to scale, and Lagos (2003), who formulates a spatial search- and matching model and calibrates it to the New York City taxi market. More recent empirical work includes Frechette *et al.* (2019) and Buchholz (2020) who also both model the New York City taxi market, as well as Brancaccio *et al.*

(2020), who model the dry bulk shipping industry. Castillo (2020) and Rosaia (2020) study ride-hailing platforms. Kreindler (2023) combines a structural approach with experimental evidence to study the effects of congestion on the well-being of travelers. Kreindler *et al.* (2023) characterize the optimal bus network for the city of Jakarta. Finally, Fuchs and Wong (2022) also study a multi-modal transportation model in the context of freight transportation.

Within this strand of literature, Durrmeyer and Martínez (2022) and Kreindler *et al.* (2023) are most closely related to our work. Durrmeyer and Martínez (2022) estimate an equilibrium model of mode substitution, like our study, and they investigate the welfare effects of private car restrictions and congestion prices. We, by contrast, use our model to characterize optimal prices and service levels across *several modes* of transportation that interact through demand, supply and budget considerations. Moreover, we also quantify the distributional implications of public transit policies, for which our granular data is particularly well suited. Similar to Durrmeyer and Martínez (2022), we also estimate a congestion production function at the neighborhood level, allowing for heterogeneity across different areas of the city. We extend their approach by allowing for heterogeneity of the marginal impact of different types of vehicles and by incorporating such mapping of traffic flows on travel times into a planner's problem. Kreindler *et al.* (2023) study optimal transit policies but focus on the optimal network configuration for buses. While our policy simulations are less granular in terms of network planning, we incorporate the trade-off that the social planner faces when setting policies for both public transit and private modes of transportation through congestion surcharges.

There are several papers that investigate the long run effects of urban transportation policy, such as new rail infrastructure and interactions with residential location choices. Severen (2021) measures the effects on commuter welfare and productivity of the Los Angeles Commuter Rail. Tsivanidis (2019) quantifies the equilibrium effects of new public

transit infrastructure in Bogotá. Fajgelbaum and Schaal (2020) and Allen and Arkolakis (2019) characterize the optimal allocation of road infrastructure. Brinkman and Lin (2022) study the welfare implications of highway constructions during the mid-1900s. These papers focus only on one mode of transportation and abstract away from incorporating rich heterogeneity in substitution patterns. We depart from their work by abstracting away from the residential and firm location choice models but allowing for rich demand substitution patterns across modes and heterogeneity, which is a key ingredient to understand distributional effects of transit policies. More recently, Barwick *et al.* (2021) explores the interaction between mode choices and residential location choices and Herzog (2021) quantifies the equilibrium effects of London’s downtown tolls. We depart from their work by focusing on the characterization of optimal prices and service levels through the lens of a social planner’s problem.

## 2 Background and Data

### 2.1 Background

Chicago is the third largest city in the U.S. and its transportation system, which is operated by the Chicago Transit Authority (henceforth CTA), is the second largest after New York’s. It includes a bus network of 152 routes with more than two thousand buses and a train rapid transit system with eight routes and 144 stations, which is known as the “Chicago L”. Prices are per ride and independent of distance. The full fare for bus and the L are \$2.25 and \$2.50 respectively.<sup>13</sup> The Chicago metropolitan area is also served by a commuter rail system, known as the Metra, which is operated by the Regional Transportation Authority (RTA). Both the CTA and the RTA have a history of budget short-

---

<sup>13</sup> Reduced fares exist for students and seniors and a monthly pass for unlimited rides can be obtained at \$75. There are also daily, 3-day, weekly, and monthly passes. See [CTA Fares](#) for additional details.

falls, which suggests that it is important to account for budget considerations. After the pandemic, the CTA has proposed an annual budget of \$1.8 billions to keep fares at the pre-pandemic levels.<sup>14</sup> Lastly, passengers may travel by private for-hire-vehicles (FHV) in the form of taxis and ride hail. Taxis have a regulated rate of \$2.25 per mile or \$0.2 per 36 seconds as well as a \$3.25 base fare.<sup>15</sup> Ride hail companies adjust prices dynamically according to market conditions.

## 2.2 Data description

We define a market as an origin-destination-time tuple. Chicago's community areas serve as our origin and destination locations. In total there are 77 community areas, with an average size of three square miles and an average population of 36,000 people.<sup>16</sup> We define a unit of time as an hour of the day, distinguishing between weekdays and weekends. Thus, we have 48 time periods. Our final goal is to construct a dataset that consists of travel flows, prices, and travel times for every mode in every market. Additionally, we link traveler demographics, such as income and car ownership, to every market.

To construct this dataset, we rely on a variety of raw data sources. First, we use administrative public transit microdata from the CTA for the month of January 2020. We have access to the universe of individual public transit trips for both buses and the Chicago L train. We observe the station or bus stop of origin, the time that the passenger tapped in, and an inferred drop-off L station or bus stop (Zhao *et al.*, 2007).<sup>17</sup>

The second data source is the universe of de-identified taxi and ride hail data at the

---

<sup>14</sup> See [Transit Chicago](#) for the 2023 full budget report.

<sup>15</sup> See [Chicago Taxi Fare Regulation](#).

<sup>16</sup> See [Community Areas in Chicago](#).

<sup>17</sup> Unfortunately, we do not have data on Metra rides as this is managed by the RTA and not the CTA. There were 74 million Metra rides in 2019 in Cook, DuPage, Kane, Kendall, Lake, McHenry, and Will counties combined, corresponding to less than 1% of trips in the region before the pandemic.

See [My Daily Travel](#) and [Annual/Monthly Ridership](#) for details.

trip level for the month of January 2020, which is published by the City of Chicago.<sup>18</sup> Those data include prices, drop-off and pick-up locations, trip length, trip duration, and the number of pooled riders in the case that the trip is a pool ride.

The third data source is based on mobile-phone location records from Veraset for the month of January 2020, which records people's movements over time.<sup>19</sup> This dataset covers about forty percent of active cellphone devices in the U.S. The data record the device ID and a sequence of GPS coordinates with a time stamp for each entry. The frequency with which GPS coordinates are generated depends on the applications installed by the user. We restrict our analysis to devices with very frequent location information. Through a series of steps, we infer trips from the sequence of GPS coordinates that each device generates, ending roughly with 5% coverage of the population. To exclude pedestrian movements, we focus on trips that are faster than 11km/h and cover more than 400 meters. The discarded trips represent 15.6% of our data. See Appendix A.1 for more details. The final output from this process is a dataset with a fraction of the universe of trips that took place in Chicago. We then scale up the number of trips by a factor such that the aggregate share of car trips is consistent with what is reported by the Chicago Metropolitan Agency for Planning (CMAP) Household Travel Survey.<sup>20</sup> For details, see Appendix A.3.1.

To incorporate information about demographics, we rely on census data using the 2016-2020 ACS sample. We use information on income and car ownership rates at the census tract level. We infer a device user's home location based on the modal GPS location during night-time hours. We separate the population of devices into two groups: those who spend at least three nights or more in their modal night location, which we call *residents*, and those who spent at most two nights, which we call *visitors*.<sup>21</sup> Trips made by

---

<sup>18</sup> [Source: Chicago Data Portal, Transportation Network Providers - Trips](#)

<sup>19</sup> [Source: Veraset: Location Data Provider](#)

<sup>20</sup> [Source: My Daily Travel survey \(website\)](#)

<sup>21</sup> Figure 14 in Appendix A.1.3 shows the average share of visitors by origin location. As expected, visitors are highly concentrated in the city center and in the airports.

residents represent 93.3% of all cellphone trips in our sample. For residents, we impute their income and car ownership probability as the median income and car ownership rates of their home census tract. Given that we estimate income and car-ownership for each device and can track all trips made by cellphone devices, we are able to construct a distribution of travelers' demographics for every market.

Since the cellphone data only covers forty percent of devices, and among those we select the ones that provide sufficient location data, one may wonder whether some demographic groups are over- or underrepresented. We test the representativeness of our data based on census covariates assigned to devices by our inferred home location. The results from this test show that income groups are equally represented in the data as shown in Figure 15 in the Appendix. Further, we also test whether the trips inferred from our cell-phone are representative of travel patterns. To do so, we plot the time and distance travelled for both the trips in the cell-phone data and in the survey data. Figure 16 in the Appendix show that both aggregate distributions align quite well. Given their representativeness, we can multiply the cell-phone trips by a common inflation factor to arrive at the total number of trips (see Appendix A.3 for details).

We combine the previous data sources to construct the near-universe of market flows across all modes: car, taxi and ride-hail, and public transit trips. While the CTA data allow us to observe the total number of trips for buses, trains, taxis, and ride-hailing apps, we do not have official records to directly measure car trips. Given that the cell-phone data covers all trips, excluding walking trips, we can recover car trips by subtracting public transit, taxi, and ride-share trips from the cell-phone trips. Finally, we measure the overall size of the market as twice the maximum of the number of trips that we observe in the cellphone data for that specific market. We believe this margin of adjustment is important for our counterfactual simulations as some policies may induce travelers not

only to switch across modes but also to stop or start traveling.<sup>22</sup> In what follows, we define the option of staying put or walking as the traveler’s outside option.

Our last source of data are Google Maps queries. We query 30,796,848 counterfactual trips, one for each (origin census tract, destination census tract, hour of the week) triple. We combine our flows data with those queries to obtain travel times and routes across origin-destination pairs by mode. We sample real-time traffic for origin-destination pairs for weekends and weekdays.

Our comprehensive dataset has several advantages over survey data. First, survey data is constructed to be representative at the city level but often lacks representativeness at finer resolution, such as conditional on the hour or a community district. Even worse, at finer resolution surveys often do not provide any data, as can be seen in Figure 17 (Appendix G.1), which shows two matrices whose cells contain the total of number of travellers by origin and destination aggregated at the community area level. In this representation of flows, we see that survey data is very sparse. At this level of aggregation, 60.7% of the origin-destination pairs have zero trips. This sparsity means that, although ridership surveys are useful to understand aggregate movement patterns, they are not very useful if one wants to understand fine-grained movement patterns or policy interventions that are heterogeneous across space and across time. Economically, this is relevant because results from survey data may not adequately capture the effect on low income travelers of policies meant to provide a minimum level of mobility. Our detailed data also allows us to measure how different types of vehicles—cars or buses—contribute to traffic congestion throughout the city. Lastly, the richness of our data has econometric advantages. We can invert market shares (Berry (1994), Berry *et al.* (1995)) at a very granular level and allows us to construct moment conditions to address the endogeneity of

---

<sup>22</sup>For example, Rosenblum *et al.* (2020) shows that a \$0.50 bus subsidy in Cambridge, MA increased ridership by 30%, without decreasing the shares of other modes for their surveyed population.

prices and travel times.

## 2.3 Summary Statistics and Descriptive Results

In this section, we provide descriptive evidence based on our data set. Table 1 below displays the overall market share for each mode, conditional on traveling using motorized transportation.<sup>23</sup> Although Chicago has one of the best public transit systems in the US, about 70% of trips are taken by car.

Table 1: Overall market shares

Mode share	
Car	66.20
Bus	16.05
Train	13.19
Ride Hail	4.08
Taxi	0.47

Source: Chicago Metropolitan Agency for Planning (CMAP) Household Travel Survey (January 2019)

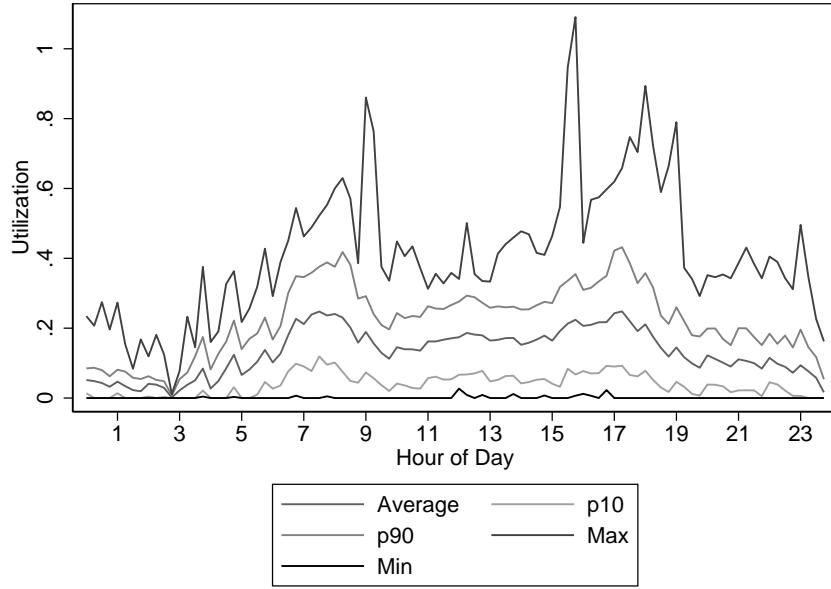
Figure 1 shows that bus utilization rates are strikingly low. Even during the morning and afternoon rush hours, average utilization rates stay below 25%. If we look at the distribution within hour, we see that 90% of buses are at a utilization below 40% during the morning rush hour. These low utilization rates indicate that resources are not optimally allocated.

Moreover, this graph also shows that, while certain buses can be traveling at full capacity at given times, those events are extremely rare. Only one bus route ever reaches full capacity, and that only occurs during a very short period of time during the average day.

---

<sup>23</sup>We exclude biking and multi-modal trips, but these modes only represents 1.8% and 2% of the overall trips.

Figure 1: Bus utilization rates



*Notes:* This figure shows the minimum, 10th, 50th, 90th percentile, and maximum utilization rates across bus routes for each hour of the day. To compute these percentiles we first measure utilization for each bus every fifteen minutes by taking the number of riders on the bus divided by the capacity of the bus. We conservatively assume each bus has a capacity of 53, which is the smaller of the two bus sizes used by the CTA. The percentiles of utilization rates for a given bus route in a certain hour are then computed across each bus on the route over that hour.

*Source:* Authors' calculation using CTA bus trip-level data.

In what follows, we describe how market characteristics and mode choices vary with traveler's demographics. The upper left panel of Figure 2 shows that car market shares follow an inverted U-shape. We also observe a nonmonotonic relationship for trains in the upper right panel.<sup>24</sup> For bus (lower left panel), we observe a negative relationship between market shares and income: higher-income travelers tend to travel less frequently using public alternative compared to low-income travelers. On the other hand, we observe the opposite pattern for ride hailing (lower right panel). We take these graphs as evidence of heterogeneity across income levels for different modes of transportation. This

<sup>24</sup>This non-monotonicity likely relates to residential sorting, since high-income travelers choose to live in areas with good access to trains.

heterogeneity could be driven by preferences, with higher income travellers preferring faster modes of transportation, but also by accessibility, as they could be more likely to own a car.

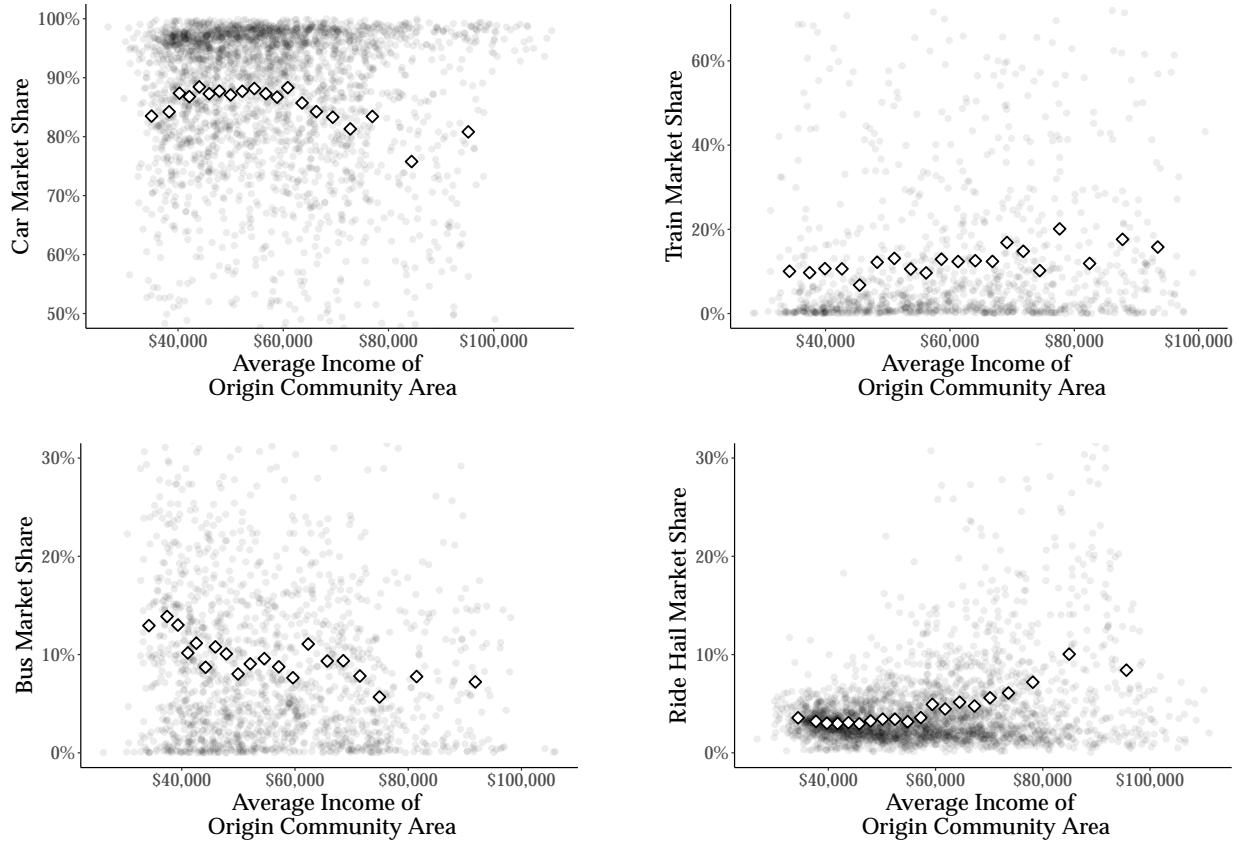


Figure 2: Mode market shares by traveler's income

*Notes:* This figure shows both a scatter and a binscatter of mode market shares against average income. Our level of observation is a census tract. We exclude the share of those who do not travel for the ease of exposition.

*Source:* Authors' calculations using data from the Chicago Transportation Authority, City of Chicago Data Portal, mobile phone location records, and U.S. Census Bureau.

Next, we show that indeed there are important differences in the share of car ownership across the income distribution. Figure 3 displays car ownership by income level. We observe that for low-income levels, car ownership increases with income, then slightly decreases and finally flattens out.<sup>25</sup> This pattern implies that it is important to account

<sup>25</sup>We also believe that this non-monotonicity is probably driven by residential sorting.

for differences in car ownership in our estimation: Ignoring this margin would conflate strong preferences for other modes of transportation with the actual possibility of traveling by car.

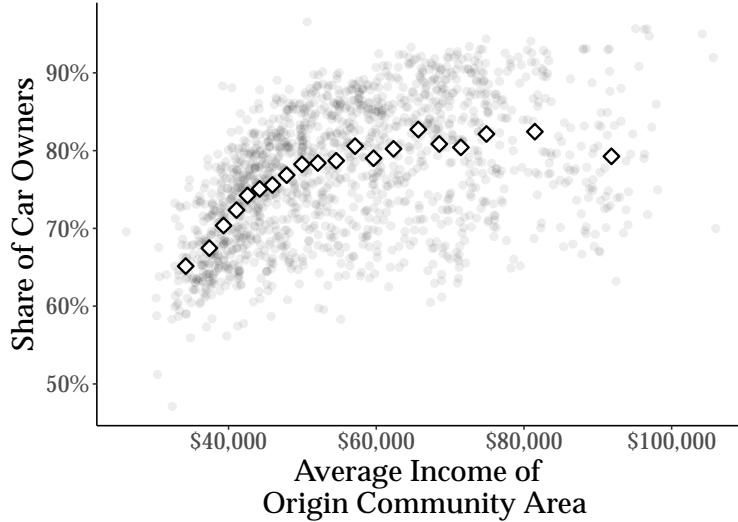


Figure 3: Car ownership by traveler's income

*Notes:* This figure plots a binscatter of car ownership against median income. Our level of observation is a census tract.

*Source:* Authors' calculations using data from the U.S. Census Bureau.

Figure 4a presents the distribution of transportation expenses across markets. Most trips in our dataset cost between \$1 and \$10, with a long tail of expensive trips. In Figure 4b, we present the distribution for high- and low-income markets. As anticipated, markets predominantly composed of lower-income travelers exhibit markedly lower travel expenses when compared to those consisting of higher-income travelers.

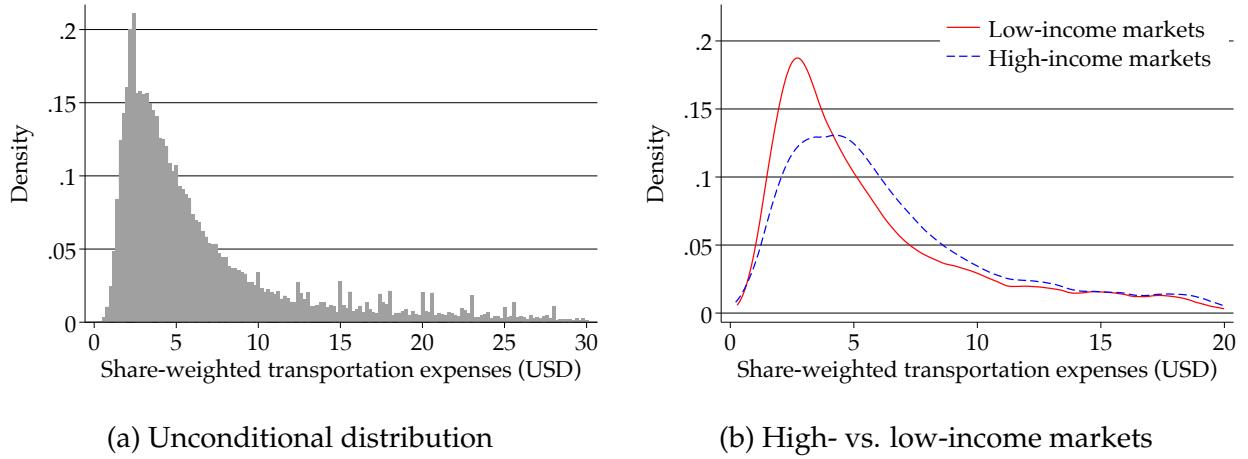


Figure 4: Distribution of transportation expenses across markets

*Notes:* Subfigure (a) presents the distribution of transportation expenses. Every observation represents the average expenses per trip within a market. The peak at \$2.25 consists of markets where every trip in the market is a bus trip. Subfigure (b) presents the distribution for high- and low-income markets, which we define as markets in which the median traveler belongs to the top or bottom quintile of the income distribution of the city.

Similarly, figure 5a shows the distribution of mode-share trip travel time across markets, and figure 5b breaks it down by top and bottom income quintiles. We see similar patterns here as in the transportation expense figures: markets with low-income median travelers are characterized by lower travel times, and markets with high-income median travelers show higher travel times.

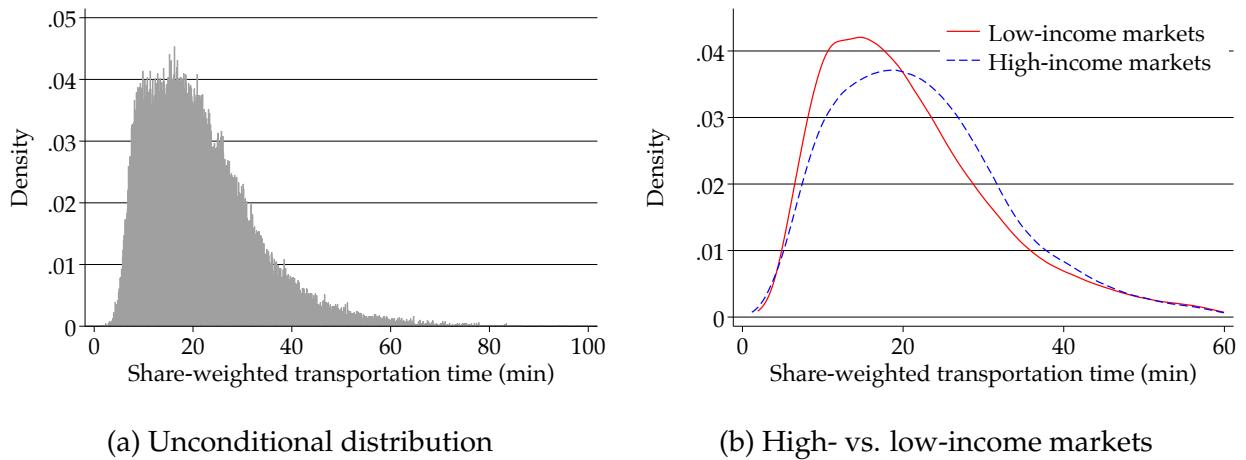


Figure 5: Distribution of transportation time across markets

*Notes:* Subfigure (a) presents the distribution of transportation time. Every observation represents the average times for all trips within a market. Subfigure (b) presents the distribution for high- and low-income markets, which we define as markets in which the median traveler belongs to the top or bottom quintile of the income distribution of the city.

To summarize, we take the last facts as evidence that travelers across locations and income levels have different access to transit and face different trade-offs, translating into different mode choices. Therefore, in our estimation and simulations it is important to account for such heterogeneity across locations as well as travelers. Moreover, such heterogeneity implies that different transportation policies may have unequal effects across demographic groups, and thus important distributional implications.

3 Model

### 3.1 Setup and Equilibrium Definition

The model has three parts. First, there are travelers with fixed origin and destination who choose either one of the available modes or not to travel at all. Second, there is a transportation technology that captures the relationship between the number of people who use a mode and its travel time. Finally, there is a social planner that maximizes welfare

subject to a budget constraint. In this section, we first present a theoretical result that highlights the general forces in the social planner's objective function. To keep the exposition simple we focus on only one market and assume that the social planner controls all modes. Then, in section 3.3, we present an empirical version of our model. The empirical version of our model accounts for inter-temporal and spatial variation in supply and demand and for the linkages across space due to congestion and driver movements. In the empirical version of the model, the social planner can tax road based modes of transportation and directly control the prices of public transit.

Travelers differ across three observable dimensions: their income, their access to public transit, and whether they own a car. We denote a traveler's type by  $\theta \in \mathbb{R}^n$ , with density  $f(\cdot)$ , which captures observable characteristics as well as unobservable preferences for modes.

A traveler of type  $\theta$  decides which transportation mode  $j$  to take to her destination. She can choose among the set  $\mathcal{J}(\theta)$ , which varies depending on whether public transit is easily accessible and on whether she owns a car. She can also choose the outside option of not taking a trip, which we denote by 0. Thus, her choice set is given by  $\mathcal{J}(\theta) \cup \{0\}$ . She gets utility  $u_j(t_j, \theta) - p_j$  if she takes transportation mode  $j$ , where  $p_j$  is the price and  $t_j$  is the total trip time—the sum of the travel time and the waiting time before the trip starts.<sup>26</sup> We normalize the utility of the outside option to zero. Therefore,  $u_j(t_j, \theta) - p_j$  is the utility measured relative to not taking a trip.

The traveler chooses the mode that maximizes her utility among her choice set:

$$j^*(\theta) = \underset{j \in \mathcal{J}(\theta) \cup \{0\}}{\operatorname{argmax}} u_j(t_j, \theta) - p_j \quad (1)$$

---

<sup>26</sup>Observe that we can also incorporate heterogeneity in traveler's sensitivity to prices  $u_j(t_j, \theta) - \theta_p \cdot p_j$ . Note that a re-scaled version of utility, namely  $(u_j(t_j, \theta) - \theta_p \cdot p_j)/\theta_p$  leads to the same optimal choices. Using  $u_j(t_j, \theta) - p_j$  has the advantage of measuring utility directly in monetary terms for all consumers  $\theta \in \Theta$ , where the Value of Time (VOT) can be computed as  $\partial u_j(t_j, \theta)/\partial t_j$ .

Given vectors of prices  $\mathbf{p}$  and total trip times  $\mathbf{t}$  for all modes, demand for mode  $j$  is given by

$$q_j = q_j(\mathbf{p}, \mathbf{t}) = \int_{\Theta_j(\mathbf{p}, \mathbf{t})} f(\theta) d\theta, \quad (2)$$

where  $\Theta_j(\mathbf{p}, \mathbf{t})$  is the set of consumer choosing mode  $j$  at  $(\mathbf{p}, \mathbf{t})$ .

We refer to the vector  $\mathbf{q}$  as trips. Gross utility and consumer surplus are given by

$$U(\mathbf{p}, \mathbf{t}) = \sum_j \int_{\Theta_j(\mathbf{p}, \mathbf{t})} u(t_j, \theta) f(\theta) d\theta \quad \text{and} \quad CS(\mathbf{p}, \mathbf{t}) = \sum_j \int_{\Theta_j(\mathbf{p}, \mathbf{t})} (u(t_j, \theta) - p_j) f(\theta) d\theta. \quad (3)$$

Total times are determined by a transportation technology that captures the dependence of travel times on the overall capacity of the fleet for each mode as well as the number of traveling passengers. The fleet size for public transit is a policy choice and determines the frequency at which buses run.<sup>27</sup> For ride-hailing, the fleet size is determined by the number of drivers. The transportation technology also captures the fact that road-based modes of transportation are subject to congestion.

Accounting for all these considerations, we can compactly write the vector  $\mathbf{t}$  of travel times for all modes is given by

$$\mathbf{t} = T(\mathbf{q}, \mathbf{k}), \quad (4)$$

where  $\mathbf{k}$  is the vector of fleet sizes for all modes.

The transportation technology also determines the cost  $C(\mathbf{q}, \mathbf{k})$  of supplying  $\mathbf{q}$  trips with fleet size  $\mathbf{k}$ . This cost function includes both labor costs and physical costs, such as fuel and vehicle depreciation. Additionally, there is an environmental externality  $E(\mathbf{q}, \mathbf{k})$  that is borne by society. We assume that both functions are increasing and convex in both

---

<sup>27</sup>We are accounting for trains in the demand estimation but do not consider train frequency and pricing a policy choice because adjustments in those may be subject to technical constraints that we are not capturing here.

arguments. With this notation we can now define an equilibrium.

**Definition 1** (Transportation Equilibrium). *Given a set of prices  $\mathbf{p}$  and fleet sizes  $\mathbf{k}$ , a market equilibrium is a set of trips  $\mathbf{q}$  and travel times  $\mathbf{t}$ ,  $(q^*(\mathbf{p}, \mathbf{k}), t^*(\mathbf{q}, \mathbf{k}))$  such that (2) and (4) are satisfied.*

Note that in such definition times and quantities are the equilibrium objects, while prices and fleet size are fixed. Therefore, for any given set of fleet size and prices, travel times adjust to bring the market into equilibrium. There might exist multiple equilibria given  $(\mathbf{p}, \mathbf{k})$ —although empirically we have not found this to be an issue.

## 3.2 The Social Planner's Problem

The city government chooses prices  $\mathbf{p}$  and capacities  $\mathbf{k}$ . Its goal is to maximize welfare subject to a budget constraint. However, to characterize a solution it will be easier to think of the government as choosing an allocation  $(\mathbf{q}, \mathbf{k})$  to maximize its objective function. We can do this change of variables because for any  $(\mathbf{q}, \mathbf{k})$  there exists a unique set  $(\mathbf{p}, \mathbf{t}) = (p(\mathbf{q}, T(\mathbf{q}, \mathbf{k})), T(\mathbf{q}, \mathbf{k}))$  that is consistent with market equilibrium. This alternative representation of the government's problem has two advantages over the original problem. First, it allows us to sidestep the problems that could arise from multiple equilibria.<sup>28</sup> Second, as we will see below, the first order conditions yield easily interpretable expressions.

The government's revenue is equal to the payments it obtains from travelers minus its costs:

$$\Pi(\mathbf{q}, \mathbf{k}) = \sum_j p_j(\mathbf{q}, T(\mathbf{q}, \mathbf{k})) q_j - C(\mathbf{q}, \mathbf{k}).$$

---

<sup>28</sup>The only concern is that the market could end up in a wrong equilibrium. However, the government can ensure the right equilibrium plays out in the short run: if agents expect a certain equilibrium, the government can set prices until agents' expectations adjust. It can then revert to the equilibrium prices for the desired allocation. This follows the idea of an *insulating tariff* from Weyl (2010).

This revenue cannot be lower than  $-B$ , where  $B$  is the transportation budget. Welfare is equal to consumer surplus plus the government's revenue minus externalities, which can be written as

$$W(\mathbf{q}, \mathbf{k}) = U(\mathbf{q}, T(\mathbf{q}, \mathbf{k})) - C(\mathbf{q}, \mathbf{k}) - E(\mathbf{q}, \mathbf{k}).^{29}$$

The social planner's objective function is thus:

$$\begin{aligned} & \max_{\mathbf{q}, \mathbf{k}} \quad U(\mathbf{q}, T(\mathbf{q}, \mathbf{k})) - C(\mathbf{q}, \mathbf{k}) - E(\mathbf{q}, \mathbf{k}) \\ \text{s.t.} \quad & \sum_j p_j(\mathbf{q}, T(\mathbf{q}, \mathbf{k})) \cdot q_j - C(\mathbf{q}, \mathbf{k}) \geq -B \end{aligned} \tag{5}$$

In order to derive optimality conditions for this problem, we first define some useful notation. Let

$$\tilde{C}_j^q = \frac{\partial C}{\partial q_j} + \frac{\partial C}{\partial k_j} \cdot \frac{k_j}{q_j} \quad \text{and} \quad \tilde{E}_j^q = \frac{\partial E}{\partial q_j} + \frac{\partial E}{\partial k_j} \cdot \frac{k_j}{q_j}$$

represent the marginal cost and externality of a trip when holding the fleet size per trip fixed—i.e., if the number of trips using mode  $j$  increase by 1%, so does fleet size  $k_j$ . Similarly, let

$$\tilde{T}_{jk} = \frac{\partial T_j}{\partial q_k} + \frac{\partial T_j}{\partial k_k} \cdot \frac{k_k}{q_k}$$

be the derivative of mode- $j$  time with respect of the number of trips using mode  $k$ , holding the fleet size per trip fixed. Also let  $\bar{u}_j^T = \partial U / \partial T_j$  be the derivative of gross utility per traveler with respect to the total time of mode  $j$ . Finally,  $\Omega_{kj}$  represents elements of the inverse Jacobian of  $q(p, T)$  with respect to  $p$ .

**Proposition 1.** *The first order conditions for the social planner's problem (5) can be written as*

---

<sup>29</sup>This expression can be derived by noting that  $W = CS + \Pi - E = U - \sum_j p_j q_j + \sum_j p_j q_j - C - E$ .

$$\begin{aligned}
p_j = & \underbrace{\tilde{C}_j}_{\text{Marginal cost}} + \underbrace{\tilde{E}_j}_{\text{Marginal env. externality}} - \sum_k \underbrace{\bar{u}_k^T}_{\text{Mg. utility of time}} \cdot \underbrace{\tilde{T}_{kj}}_{\text{Network effects}} \\
& + \frac{\lambda}{1+\lambda} \cdot \left\{ \underbrace{\sum_{k \in J} q_k \cdot \Omega_{kj} - \tilde{E}_j}_{\text{Market power markup}} - \underbrace{\sum_k (\tilde{u}_k^T - \bar{u}_k^T) \cdot \tilde{T}_{kj}}_{\text{Spence distortion}} \right\}
\end{aligned} \quad (6)$$

$$\begin{aligned}
\frac{\partial C}{\partial k_j} = & - \underbrace{\frac{\partial E}{\partial k_j}}_{\text{Mg. env. externality of fleet size}} + \underbrace{\sum_k \bar{u}_k^T \cdot \frac{\partial T_k}{\partial k_j}}_{\text{Mg. gross utility from fleet size}} + \frac{\lambda}{1+\lambda} \cdot \underbrace{\left\{ \frac{\partial E}{\partial k_j} + \sum_k (\tilde{u}_k^T - \bar{u}_k^T) \cdot \frac{\partial T_k}{\partial k_j} \right\}}_{\text{Spence distortion}}
\end{aligned} \quad (7)$$

where  $\lambda$  is the Lagrange multiplier for the budget constraint and  $\tilde{u}_j^T$  is a weighted sum of the derivative of gross utility among marginal travelers with respect of the total time of if pickup times for mode  $j$  increase by 1%.

*Proof.* See appendix B.1 □

To understand these expressions, think first of an unconstrained social planner, in which case  $\lambda = 0$  and the final terms drop out. Both expressions take a Pigouvian form. The price of a trip,  $p_j = \tilde{C}_j + \tilde{E}_j - \sum_k \bar{u}_k^T \cdot \tilde{T}_{kj}$ , is equal to the marginal cost—accounting also for the marginal cost of additional fleet size—plus two terms to correct for externalities. The first one accounts for environmental externalities. The second one accounts for the change in other travelers' utility. It is equal to the sum over modes of the product of  $\bar{u}_k^T$ , the derivative of gross utility with respect to time, times  $\tilde{T}_{kj}$ , the change in total time given an additional trip using mode  $j$ . For  $j \neq k$ ,  $\tilde{T}_{kj}$  is positive due to congestion (or zero if the two modes are not interrelated), so those terms lead to a Pigouvian tax. But  $\tilde{T}_{jj}$  can be negative. For taxis and ride-hailing, for instance, increasing the number of trips and fleet

size by the same factor results in a reduction in waiting times due to returns to scale (i.e., economies of density) in the matching process. In that case there should be a Pigouvian subsidy, as noted by Arnott (1996).

Equation (7) also takes a Pigouvian form with  $\lambda = 0$ . The marginal cost is the price of fleet size if it is supplied in a competitive market. That is the case, for instance, if labor is supplied competitively in taxi or ride-hailing markets. The price of fleet size is minus the marginal environmental externality plus the change in gross utility that a higher fleet size induces due to times.

With a budget constraint, however, the social planner behaves qualitatively like a monopolist. To meet the budget the social planner needs to raise revenue. This introduces a market power markup in equation (6). The social planner now also under-weights environmental externalities. Finally, there is a Spence distortion. The government internalizes effects on other travelers' utility, but imperfectly: it accounts for changes in the utility of marginal travelers, rather than that of all travelers.

As  $\lambda \rightarrow \infty$ , the social planner no longer cares about welfare and it only cares about its budget. These two expressions then become the first order conditions for profit maximization: terms related to environmental externalities cancel out, and there is a full markup and a full Spence distortion as  $\lambda/(1 + \lambda) \rightarrow 1$ .

### 3.3 Empirical Model

Here we describe the empirical version of our demand model and of the transportation technology. The city is composed of community areas  $a \in \mathcal{A}$ , the main level of spatial aggregation we consider. We also refer to them as *locations*. Every community area  $a$  is composed of several census tracts  $c \in a$ , the smallest geographic area that we consider. We denote by  $a(c)$  the community area which  $c$  belongs to. Time is divided into hours

$h \in \mathcal{H}$ .<sup>30</sup>

### 3.3.1 Demand

First, we define a market  $m = (a, \tilde{a}, h) \in \mathcal{M}$  as the collection of people who make travel decisions from community area  $a$  to community area  $\tilde{a}$  at a particular time  $h$ .<sup>31</sup> Then, for each sub-market  $m$ , people  $i \in \mathcal{I}_m$  arrive at an exogenous arrival rate  $\mu_m$ . They decide which mode  $j \in \mathcal{J}_m^i \cup \{0\}$  to use, where  $j = 0$  denotes the outside option of staying put, by solving the following problem:

$$V_m^i = \max_{j \in \mathcal{J}_m^i \cup \{0\}} \delta_{mj}^i + \varsigma_{mg(j)}^i + (1 - \rho)\epsilon_{mj}^i = \max_{j \in \mathcal{J}_m^i \cup \{0\}} \xi_{mj} + \alpha_T \cdot T_{mj} + \alpha_p^i \cdot p_{mj} + \varsigma_{mg(j)}^i + (1 - \rho)\epsilon_{mj}^i \quad (8)$$

where  $T_{mj}$  denotes the sum of the waiting and travel times for mode  $j$ ,  $p_{mj}$  is the price for mode  $j$ ,  $\alpha_T$  is the preference parameter over travel times, and  $\alpha_p^i$  are person  $i$ -specific price coefficients.<sup>32</sup> Finally,  $\varsigma_{mg(j)}^i$  and  $\epsilon_{mj}^i$  are idiosyncratic unobserved taste shocks. The taste shock  $\varsigma_{mg(j)}^i$  is common to all goods within the group  $g(j) \equiv \mathcal{J}_m^i$ , defined as all modes of transportation excluding staying put, which aims to capture uncertainty in the option to travel relative to the outside option across markets. For example, one can imagine that choosing to travel is more uncertain outside commuting hours. The taste shock  $\epsilon_{mj}^i$  is specific to mode  $j$  and captures things such as low weather temperatures that make public transit less attractive. This model can be thought of as a sequential choice problem: First, travellers choose whether they want to travel or not, and second, conditional on

---

<sup>30</sup> Specifically, we define  $h$  as hour-of-the-week. We consider  $48 = 2 \times 24$  hours-of-the-week, corresponding to an average weekday day and an average weekend day. Daily variation is aggregated by taking averages across dates for the same hour-of-the-week.

<sup>31</sup> Given that our level of temporal variation averages across dates, traveling decisions should be thought as the choice over average trips at a given hour-of-the-week  $h$  rather than stemming from very short-run temporal variation coming from shocks or special occasions.

<sup>32</sup> In this model, the VOT of consumer  $i$  is computed as  $\alpha_T / \alpha_p^i$ .

travelling, travellers choose which mode they use.

Importantly, we allow the choice sets  $\mathcal{J}_m^i$  to vary across markets and across consumers within the same market, capturing the fact that certain markets  $m$  and certain consumers  $i$  within  $m$  may not have access to certain modes of transportation. Some community areas cannot be reached by train, for instance, and some consumers do not own a car. We also allow for preference heterogeneity in the price coefficient. This heterogeneity captures potential differences across consumers in the trade-off between price and travel times. For example, travelers of different income groups may have different opportunity costs in terms of time due to wage differentials.

We assume that  $\epsilon_{mj}^i$  follows a Type I Extreme Value distribution and that the *iid* shock for group  $g(j)$  follows the unique distribution such that  $\varsigma_{mg(j)}^i + (1-\rho)\epsilon_{mj}^i$  is also an extreme value random variable. The parameter  $\rho \in [0, 1]$  governs the within group correlation; a larger  $\rho$  implies modes within group  $g$  are closer to perfect substitutes. With these assumptions, we have that person  $i$ , conditional on choosing from modes in  $g(j)$ , chooses mode  $j$  with probability

$$\mathbb{P}_{mj}^i | g = \frac{\exp\left(\frac{\delta_{mj}^i}{1-\rho}\right)}{D_g}, \quad (9)$$

where

$$D_g = \sum_{j \in g} \exp\left(\frac{\delta_{mj}^i}{1-\rho}\right). \quad (10)$$

The unconditional probability of choosing from group  $g$  is given by

$$\mathbb{P}_g^i = \frac{D_g^{(1-\rho)}}{\sum_g D_g^{(1-\rho)}}. \quad (11)$$

The unconditional choice probability of mode  $j$  for person  $i$  is therefore given by

$$\mathbb{P}_j^i = \mathbb{P}_{mj}^i | g \cdot \mathbb{P}_g^i = \frac{\exp\left(\frac{\delta_{mj}^i}{1-\rho}\right)}{D_g^\rho \cdot \left[\sum_g D_g^{(1-\rho)}\right]} \quad (12)$$

Integrating over  $\alpha_p^i$ , which follows a market-specific distribution  $F_m$ , we obtain that market shares and trips for mode  $j$  in market  $m$  are given by:

$$\mathbb{P}_{mj} = \int \mathbb{P}_{mj}^i dF_m(\alpha_p^i) \quad q_{mj} = \lambda_m \cdot \mathbb{P}_{mj}.$$

### 3.3.2 Transportation Technology

In this section, we describe the transportation technology, which determines times as a function of trips and fleet sizes (frequencies). We first describe how we model traffic congestion, which affects most modes, and we then describe how we model each individual mode.

**Traffic congestion** To account for traffic congestion, we model the city as a directed graph in which nodes are community areas  $a$ , and edges  $e = (a, \tilde{a})$  connect community areas that are spatially adjacent. Let  $\mathcal{E}$  be the set of all edges. We assume that routes are exogenous and pre-determined. If a traveler uses mode  $j$  in market  $m = (a, \tilde{a}, h)$ , she follows a directed path

$$r_{a,\tilde{a},j} = (e_{a,a_1,j}, e_{a_1,a_2,j}, \dots, e_{a_{N_m-1},\tilde{a},j})$$

over edges that connects  $a$  with  $\tilde{a}$ . During hour  $h$ , the congestion on edge  $e$  is

$$Q_{eh} = \left( \sum_j w_j \cdot q_{ehj} \right)^\beta, \quad (13)$$

where  $q_{ehj}$  is the total number of vehicles for mode  $j$  that go through edge  $e$  and can be defined as the sum across all routes that go through edge  $e$ ,  $\mathcal{R}_{hj}^e$ :

$$q_{ehj} \equiv \sum_{r \in \mathcal{R}_{hj}^e} q_{rhj}.$$

The weight  $w_j$  on vehicle flows captures the fact that some modes congest more than others—buses congest more than cars, and trains do not congest at all. We assume that the travel time over edge  $e$  at time  $h$  is given by

$$\tau_{eh} = A_{eh} \cdot Q_{eh} = A_{eh} \cdot \left( \sum_j w_j \cdot q_{ehj} \right)^\beta, \quad (14)$$

where  $A_{eh}$  is an edge-specific term that measures the distance between the nodes  $e$  connects and the quality of infrastructure. Finally, the travel time in market  $m = (a, \tilde{a}, h)$  is given by

$$T_{a,\tilde{a}}^T = \sum_{e \in r_{m,j}} \tau_{eh}, \quad (15)$$

the sum of the travel times over all edges in the path  $r_{m,j}$ .

**Public transit** Travel times via public transit depend on the time in the vehicle, the wait time, and the walking times to the pickup location and from the drop-off location. Suppose a traveler takes public transit mode  $j$  (either buses or trains) from community area  $a$  to community area  $\tilde{a}$  during hour  $h$ . She uses some public transit route  $r$  that depends on her origin and destination. Her total travel time is:

$$T_{a,\tilde{a},h}^j = \underbrace{T_{ar}^{j,1}}_{\text{Walk to stop}} + \underbrace{T_{r,h}^{j,2}}_{\text{Waiting time}} + \underbrace{T_{a,\tilde{a}}^T}_{\text{Time in vehicle}} + \underbrace{T_{r\tilde{a}}^{j,3}}_{\text{Walk to destination}}.$$

In this expression,  $T_{ar}^{j,1}$  denotes walking times to the pickup location and  $T_{r\tilde{a}}^{j,3}$  denotes walking times from the drop-off location to the final destination.  $T_{r,h}^{j,2}$  is the time travelers wait before getting on the bus or train. We assume that travelers expect buses and trains to arrive at some uniform rate, which leads to an expected route-specific wait time of  $T_{r,h}^{j,2} = 1/(2 \cdot k_{rhj})$ . After getting on the bus or train, the travel time is  $T_{a,\tilde{a}}^T$  where  $a$  and  $\tilde{a}$  are the community areas where the origin and destination stops are located. In the case of buses these are determined by the congestion equation 15.

**Ride Hailing** The total time for ride hailing passengers is given by

$$T_{a\tilde{a}h}^j = \underbrace{T_{a\tilde{a}h}^W}_{\text{Waiting time}} + \underbrace{T_{a,\tilde{a}}^T}_{\text{Time in vehicle}}, \quad (16)$$

where  $T_{a\tilde{a}h}^W$  represents the time they need to wait from the moment they request a trip until a driver picks them up, and  $T_{a,\tilde{a}}^T$  represents the time they spend in the vehicle, which is given by equation (14) from the traffic congestion model.

The waiting time  $T_{a\tilde{a}h}^W$  arises from a simple model of matching and of driver movements that captures three main forces. First, waiting times are lower when many drivers are working: there will be a large number of idle drivers and, thus, the nearest driver will be close to the location where the rider requested a trip. Second, waiting times are higher when many travelers demand ride-hailing trips, depleting the market of available drivers. Third, at a given point in time, waiting times are lowest in those areas where more idle drivers are located. We set up a model of driver movements that parsimoniously accounts for the fact that there will mechanically be more idle drivers in neighborhoods with a net inflows of trips, as well as for the fact that drivers tend to relocate towards areas with higher earnings opportunities. Appendix C presents further details, and explains how we estimate our model using data on Uber waiting times.

**Costs and environmental externalities** We assume that costs and environmental externalities are proportional to the number of miles driven by the vehicles involved in each mode. For cars and ride hailing, the number of miles driven depends on how many passengers choose to travel using these modes. For buses and trains, on the other hand, the number of miles driven depends on the frequency with which routes run; hence, the marginal cost and externality of an additional passenger is zero.

For all modes, the cost per mile accounts for fuel or energy, vehicle depreciation, and maintenance. For buses and trains, it also includes labor costs.<sup>33</sup> Environmental externalities account for the social cost of carbon, for which we use latest EPA proposal of \$190 per tonne as the baseline number, as well as for the social cost of local pollutants, which we obtain from Holland *et al.* (2016).<sup>34</sup> Appendix F describes in detail the numbers that we use for all costs and externalities.

## 4 Estimation and Computation

### 4.1 Demand model

We estimate several specification of our demand model, gradually building up to the main specification that we outline in section 3.3.1. Our first specification shows estimates from a model with no income heterogeneity ( $\alpha_p^i = \alpha_p$ ), no heterogeneity in car-ownership, and no nest ( $\varsigma_{mg}^i = 0$ ). The model then reduces to a standard logit demand model, allowing us to recover mode-market mean utilities directly from the data ( $\delta_{mj} = \ln(s_{mj}) - \ln(s_{m0})$ ) and estimate the parameters of the model by regressing  $\delta_{mj}$  on prices and travel times.

---

<sup>33</sup>For ride hailing, labor costs depend on the number of drivers that are working, which is an exogenous quantity that is independent of the number of people that request ride-hailing trips.

<sup>34</sup>See [EPA Issues Supplemental Proposal to Reduce Methane and Other Harmful Pollution from Oil and Natural Gas Operations](#).

However, while public transit prices are regulated and can be treated as exogenous, one may be concerned about endogeneity in ride-hailing prices. The reason is that ride-hail companies are adjusting prices to demand conditions, which could lead to a correlation between prices and  $\xi_{mj}$ . To address this concern, we use differentiation instruments, as suggested in Gandhi and Houde (2017). Our instrument exploit the idea that ride-hail prices must be lower in markets where other modes offer lower travel times.

In particular, we construct the following two instruments:

$$Z_{m,ridehail}^{local} = \sum_{j \neq ridehail} \mathbb{1} \cdot \{|T_{m,ridehail} - T_{mj}| < SD_T\}$$

$$Z_{m,ridehail}^{quad} = \sum_{j \neq ridehail} (T_{m,ridehail} - T_{mj})^2$$

where  $SD_T$  is the standard deviation of the pairwise difference  $|T_{m,ridehail} - T_{mj}|$  across markets.

Motivated by the disparities in mode choice across the income distribution from Section 2.3, we let the price sensitivity of our travellers be a function of their income. In particular, we take  $\alpha_p^i = \alpha_p + \alpha_{py} \cdot y^i$ , where  $y^i$  is the income of traveler  $i$ .<sup>35</sup> Because allowing for heterogeneity by income precludes using linear regression to estimate the parameters, we instead estimate this model using two-step GMM, with moment conditions in the form of:

$$\mathbb{E}[\xi_{mj} \cdot \mathbf{Z}_{mj}] = 0,$$

for a given vector of instruments  $\mathbf{Z}_{mj}$ . Concretely, we use as instruments trip times, non-ridehail prices, the differentiation instruments, and the share  $\pi^y$  of consumers in each

---

<sup>35</sup>For computational simplicity, we divide the population into five income bins.

income quintile:

$$\mathbf{Z}_{mj} = (T_{mj}, p_{mj} \cdot \mathbb{1}\{j \neq ridehail\}, (Z_{m,ridehail}^{local}, Z_{m,ridehail}^{quad}) \cdot \mathbb{1}\{j = ridehail\}, \pi_m^y \cdot T_{mj}).$$

To compute the GMM objective function, we follow the procedure outlined in Berry *et al.* (1995). First, we guess values for the parameter vector  $\theta \equiv (\alpha_p, \alpha_T, \alpha_{py}, \rho)$ . We can then recover mode-market mean utilities  $\hat{\delta}_{mj}(\theta) = \xi_{mj} + \alpha_T \cdot T_{mj} + \alpha_p \cdot p_{mj}$  using an iterative contraction mapping. Second, we construct the residuals  $\hat{\xi}_{mj}(\theta) = \hat{\delta}_{mj}(\theta) - \alpha_T \cdot T_{mj} - \alpha_p \cdot p_{mj}$ .<sup>36</sup> Finally, we compute the value of the GMM objective function

$$J(\theta) = (\mathbf{Z} \cdot \hat{\xi} \cdot (\theta))' \cdot W \cdot (\mathbf{Z} \cdot \hat{\xi}(\theta)),$$

where  $\mathbf{Z}$  is the matrix whose columns are  $\mathbf{Z}_{mj}$  and  $\hat{\xi}(\theta)$  is the vector whose elements are  $\hat{\xi}_{mj}(\theta)$ .

Our next specification also allows for heterogeneity in car ownership by allowing for random choice set variation. In particular, for each consumer type  $i$  in a given market  $m$  we compute the probability  $p_{car,m}^i$  of owning a car. Then, if cars are an available mode in market  $m$ , we assume that cars are in the choice set  $\mathcal{J}_m^i$  of consumers of type  $i$  only with probability  $p_{car,m}^i$ . This heterogeneity component is important, because a model where everyone can choose to travel by car would lead to biased coefficients. The reason is that, without imposing constrained choice sets, we would infer stronger preferences for alternative modes of transportation, when in reality, the choice set of a significant share of the population is limited.

In some specifications we constrain the price coefficient by taking  $\alpha_p^i = s \cdot (\alpha_p + \alpha_{py} \cdot y^i)$ , where  $s(x) = -(1/a) \cdot (\ln(\exp(-a \cdot b) + \exp(-a \cdot x)))$ . The function  $s(\cdot)$  is close to linear before asymptoting to  $b$  at a rate controlled by  $a$ . In all specifications with constrained

---

<sup>36</sup>In specifications with fixed effects we also concentrate out any fixed effects.

price coefficients we set  $a = 100$  and  $b = -0.001$ , which largely maintains the linear relationship between income and price sensitivity while preventing the estimated price coefficient from becoming positive by constraining it to be below  $-0.001$ . In our preferred specification that includes price sensitivity by income, heterogeneity in car ownership, and a nest for all inside options this constraint is not binding.

Finally, we also allow for an idiosyncratic shock of travelling relative to staying put. We model this shock as a nesting structure as specified in Section 3.3.

Table 2: Demand Estimation Results

	Dependent variable: Log Market Shares					
	(1) Logit, OLS	(2) Logit, IV	(3) BLP, GMM	(4) BLP, GMM	(5) BLP, GMM	(6) BLP, GMM
$\alpha_T$	-2.46 (0.019)	-2.87 (0.022)	-2.99 (0.022)	-3.07 (0.020)	-3.11 (0.020)	-2.66 (0.017)
$\alpha_p$	-0.03 (0.001)	-0.09 (0.001)	-0.11 (0.001)	-0.097 (0.002)	-0.104 (0.001)	-0.100 (0.001)
$\alpha_{py}$	.	.	0.02 (0.0004)	.09 (0.0003)	0.02 (0.0004)	0.01 (0.0005)
$\rho$	.	.	.	.	.	0.24 (0.015)
Mode FE	✓	✓	✓	✓	✓	✓
Market FE	✓	✓	✓	✓	✓	✓
Constrained $\alpha_p^i$				✓	✓	✓
Car Ownership					✓	✓
Nest						✓
Avg. VOT	89.35	33.09	-0.61	1,280.32	92.66	34.80
VOT (Top Quintile)	.	.	-90.82	3,071.81	261.87	58.11
VOT (Bot. Quintile)	.	.	18.67	8.73	20.63	20.97
Avg. Price Elast.	-0.175	-0.588	-0.232	-0.291	-0.359	-0.506
Avg. Time Elast.	-0.738	-0.83	-0.718	-0.738	-0.721	-0.751
Num. Markets	189,293	189,293	152,725	152,725	152,317	152,317
N	383,990	383,990	339,876	339,876	339,017	339,017

Notes: This table presents demand estimation results from the specifications outlined in section 4.1. The average VOT is computed by first computing the within market average VOT as the weighted average of  $\alpha_T/\alpha_p^i$  and then averaging across markets, with weights given by market size. Similarly, the average elasticities are computed as the weighted average of own-price and own-time elasticities across all mode-market observations, with weights given by market size.

Source: Authors' calculations.

Table 2 shows the results from all those specifications. We see first that the OLS estimates exhibit the typical bias towards zero, which results from the positive correlation of prices with demand shocks. Since consumers are not very price elastic, we obtain a value of time of \$90, which seems too high. Instrumenting for prices more than doubles the price coefficient and therefore reduces the value of time to a more sensible value of \$33.1, which is approximately equal to the average hourly wage in Chicago for 2020. This estimated VOT is consistent with the literature, which typically estimates VOT to be between 0.5 and 1 times the average hourly wage.<sup>37</sup> In the third and fourth specifications we estimate that high income consumers are extremely price inelastic. However, we can see from the fifth specification that this result is an artifact of not accounting for car ownership. Once we account for car ownership those with income in the highest quintile have a VOT of \$262 while those with income in the lowest quintile have a VOT of \$21. Finally, in the sixth specification, we allow for a nest of all inside options. This empirical model leads to more reasonable substitution patterns to the outside option. We estimate a VOT of \$58 and \$21 for high and low income consumers, respectively. Hence, there is considerable heterogeneity across income groups. We take specification in column 6 as our preferred specification.

---

<sup>37</sup>In his review, Small (2012) shows that the value of personal travel time is roughly half of the gross wage rate and that the value of travel time does not increase proportionally with income or hourly wage, with elasticity estimates ranging between 0.5 and 0.9.

Figure 6 shows the value of time that we infer for Chicago's community areas. For the North Side, characterized by higher incomes, we infer values of time between \$30 and \$50, on average. On the South Side we also see a few neighborhoods with a relatively high VOT. One is Hyde Park, the neighborhood that is home to the University of Chicago, Midway airport (center left) as well as the neighborhoods of Beverly, Mount Greenwood, and Morgan Park (bottom left), that were popular white-flight destinations during the 1950s and 1960s.

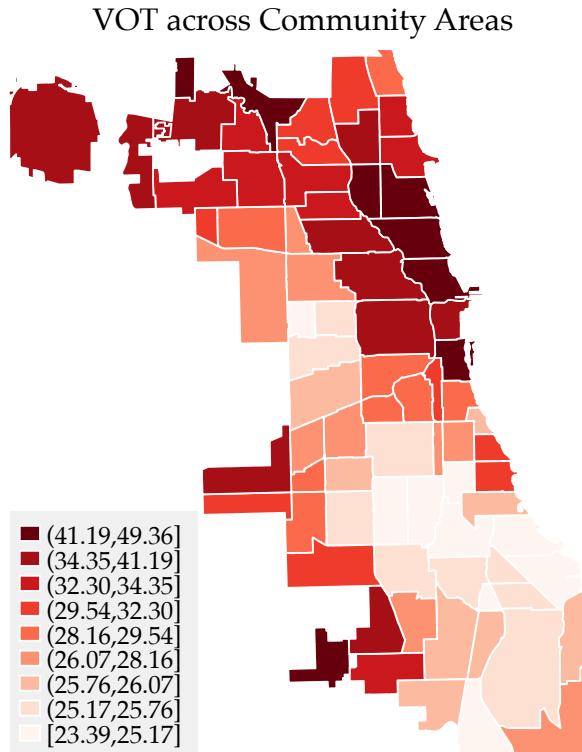


Figure 6: Value of Time across Space

*Notes:* This figure shows the VOT estimates across different regions in the city of Chicago under specification (6) of Table 2, which includes income heterogeneity, car ownership rates, and a nest for all inside goods. Income heterogeneity is specified as  $\alpha_p^i = s(\alpha_p + \alpha_{py}y^i)$  in the utility function given by equation 8. We instrument ride hail prices using two Gandhi-Houde instruments and their interaction with the share of consumers across five different income quintiles.

*Source:* Authors' calculation using mobile phone geo-location records, public transit trip-level data from the Chicago Transportation Authority, and Google Maps trips queries.

Table 3 presents substitution patterns in the form of diversion ratios. We can see that

Table 3: Diversion Ratios

(a) Overall

From \ To	<b>Bus</b>	<b>Car</b>	<b>Ridehail</b>	<b>Train</b>	<b>Outside</b>
<b>Bus</b>	.	0.30	0.08	0.10	0.52
<b>Car</b>	0.10	.	0.05	0.1	0.76
<b>Ridehail</b>	0.13	0.28	.	0.11	0.49
<b>Train</b>	0.12	0.27	0.10	.	0.51

(b) Own Car

From \ To	<b>Bus</b>	<b>Car</b>	<b>Ridehail</b>	<b>Train</b>	<b>Outside</b>
<b>Bus</b>	.	0.57	0.03	0.07	0.33
<b>Car</b>	0.10	.	0.05	0.10	0.76
<b>Ridehail</b>	0.05	0.56	.	0.07	0.31
<b>Train</b>	0.06	0.53	0.05	.	0.36

(c) No Car

From \ To	<b>Bus</b>	<b>Car</b>	<b>Ridehail</b>	<b>Train</b>	<b>Outside</b>
<b>Bus</b>	.	.	0.13	0.15	0.72
<b>Car</b>	.	.	.	.	.
<b>Ridehail</b>	0.20	.	.	0.15	0.65
<b>Train</b>	0.18	.	0.16	.	0.66

(d) Top Income Quintile

From \ To	<b>Bus</b>	<b>Car</b>	<b>Ridehail</b>	<b>Train</b>	<b>Outside</b>
<b>Bus</b>	.	0.41	0.09	0.09	0.41
<b>Car</b>	0.10	.	0.09	0.09	0.72
<b>Ridehail</b>	0.09	0.41	.	0.10	0.40
<b>Train</b>	0.10	0.34	0.13	.	0.43

(e) Bottom Income Quintile

From \ To	<b>Bus</b>	<b>Car</b>	<b>Ridehail</b>	<b>Train</b>	<b>Outside</b>
<b>Bus</b>	.	0.31	0.06	0.11	0.53
<b>Car</b>	0.09	.	0.03	0.10	0.78
<b>Ridehail</b>	0.13	0.30	.	0.10	0.47
<b>Train</b>	0.13	0.29	0.07	.	0.52

*Notes:* These tables present average diversion ratios for various consumer types for the demand estimates in specification (6). To construct them, the diversion ratios are first averaged across markets, weighted by market size. Then, to account for the fact that not every mode is present in every market, they are normalized so that each row sums to 1. Element  $(m, m')$  of each table gives the diversion ratio from mode  $m$  to mode  $m'$ .

*Source:* Authors' calculations.

there is substantially more substitution to the outside option, which is staying put, among those without a car compared to those with a car. Furthermore, those in the top income quintile have significantly more substitution towards ridehail and cars modes than those in the bottom income quintile. By contrast, those in the lowest income quintile are more likely to substitute towards buses or the outside option. Among people without cars, we find that more people who take buses substitute to the outside option than people who use ride hail. Among cars owners, car trips are twice as likely to substitute to the outside option than the remaining modes.

Additionally, we run a few additional specifications to asses alternative specifications. Results are shown in Table 10. First, we allowed for travellers not only to care about the average travel time, but also about reliability. To do so, we also incorporate the standard deviation of travel time for each mode. We find that increasing the average travel time by one minute is 2.5 times more important than increasing the standard deviation of travel time by one minute, which suggests that preferences for reliability are secondary relative to the travel time itself. Moreover, this empirical model exhibits extreme heterogeneity with a very large VOT that is hard to reconcile with the existing literature. For that reason, we still choose specification column 6 on Table 2 as our preferred specification. Finally, we also run a model that includes a quadratic term on time. The results of this model show less heterogeneity with VOTs on the low end compared to the literature. As a consequence, counterfactual simulations with this alternative model would imply lower values of consumer surplus relative to costs and environmental externalities. It follows from equation 6, that counterfactuals with this alternative specification would lead to higher cost reductions and higher congestion prices to a first order. Therefore, our results should be interpreted as conservative lower bounds on cost reductions and congestion prices.

## 4.2 Congestion

Taking the logarithm of equation 14 leads to the following relationship between traffic flows and travel times:

$$\log \tau_{eh} = \log A_{eh} + \beta \cdot \log \left( \sum_j w_j \cdot q_{ehj} \right)$$

We also assume that congestion at an edge can be decomposed into a permanent and a time varying component:  $\log A_{eh} = a_e + \epsilon_{eh}$ . Our final empirical equation is therefore given by

$$\log \tau_{eh} = a_e + \beta \log \left( \sum_j w_j q_{ehj} \right) + \epsilon_{eh}. \quad (17)$$

We cannot separately identify all  $w_j$  from  $a_e$  without a normalization, so we set  $w_{\text{car}} = 1$ .<sup>38</sup>

Under the assumption that  $\mathbb{E}[\epsilon_{eh}|a_e, a_h, q_{ehj}] = 0$ , we can estimate the remaining  $w_j$  and  $\beta$  via non-linear least squares. We exploit within-edge variation, e.g. the extent to which travel times correlate with the number of cars on the road across different hours of the week. This orthogonality assumption does not hold if there are shocks to the traffic equation (17) that affect travelers' mode choices. In other words, if people expect travel times to be high for reasons other than congestion (such as bad weather or road closures) and react accordingly. Since we aggregate data at the week hour level, one-off events like storms or road closures are not a threat to our identification. Expected weather and visibility conditions (such as the difference between day and night) could be a problem, so we include day/night dummies, weekend dummies, as well as weather controls. In the estimation, we distinguish only between cars and buses and treat ride hailing cars the same as private cars.

---

<sup>38</sup> A model with  $\tilde{b} = c \cdot b$  and  $\tilde{a}_e = a_e - \log c$  gives observationally equivalent travel times.

Table 4 presents the baseline estimates of the congestion technology. Column (1) only includes edge fixed effects, column (2) includes a control for weekdays, column (3) includes a control for day versus night, and column (4) includes weather controls such as average visibility or precipitation at the hour-of-the-week level. Column (5) includes all controls at once. The second row showing the estimates of  $w_{bus}$  reveals that the marginal congestion of one extra bus is equivalent to 4.0-5.9 cars. The estimates of  $\beta$ , which represent the elasticity of vehicle flow on travel times, is in line with previous estimates of the literature (Akbar and Duranton, 2017; Akbar *et al.*, 2018; Couture *et al.*, 2018).

Table 4: Estimates of the congestion function

	Dep. var.: Log Speed				
	Baseline	Weekday	Day/Night	Weather	All
$\beta$	-0.184 (0.001)	-0.189 (0.001)	-0.123 (0.001)	-0.134 (0.001)	-0.11 (0.001)
$w_{bus}$	5.162 (0.105)	5.36 (0.106)	3.994 (0.125)	5.878 (0.147)	4.711 (0.153)
Edge FE	✓	✓	✓	✓	✓
Weekday		✓			✓
Day/Night			✓		✓
Weather FE				✓	✓
N	14408	14408	14408	14408	14408

*Notes:* This figure shows the estimates for regression equation 17 across different specifications. The level of observation is two adjacent community areas (edges) at 48 different time intervals (24 hours for weekdays and 24 hours for weekends). The dependent variable variable is Log Speed as measured by real time Google Maps queries. The explanatory variables contain the number of cars as well as buses that travel across edges at a given hour.

*Source:* Authors' calculation using mobile phone geo-location records, public transit trip-level data from the Chicago Transportation Authority, and Google Maps trips queries.

Our estimates imply effects that are non-linear across different types of vehicles. Figure 7 shows how buses and cars interact in how travel speeds are determined, and the heterogeneous effects of different types of vehicles. Travel speed decreases much more

steeply with the number of cars than with the number of buses: the marginal congestion of a bus is much larger than that of a car.

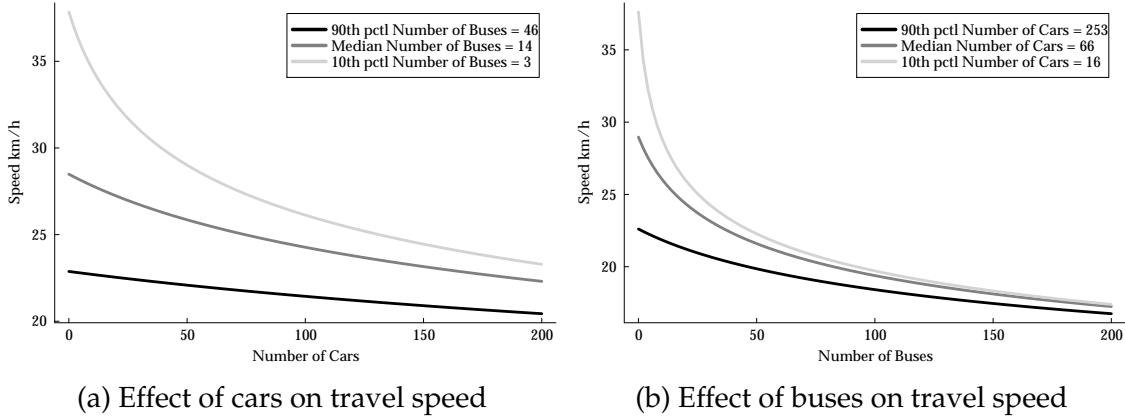


Figure 7: Predicted travel speeds across different points of vehicle flows

*Notes:* These figures plots predicted travel speed in km/h for different values of vehicle flows for both cars and buses.

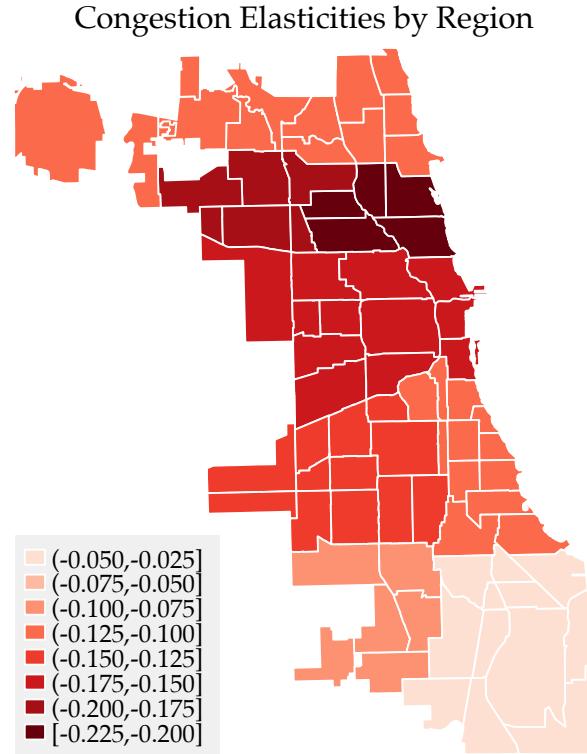
*Source:* Authors' calculations using data from the Chicago Transportation Authority, City of Chicago Data Portal, mobile phone location records, and U.S. Census Bureau.

We also allow for the elasticity of congestion to vary across regions. We therefore divide Chicago into nine larger areas. Figure 8 shows the congestion elasticity estimates from specification 17, allowing for region-specific  $\beta_r$ . We see that there is substantial heterogeneity across different areas of Chicago, with central areas showing a larger congestion elasticity compared to more peripheral areas.

### 4.3 Solving for Equilibrium and the Planner's Problem

Here we explain how we numerically solve the planners' problem and the nested transport equilibria for a given policy guess. Suppose that the government sets prices and capacities that are described by vectors  $p$  and  $k$ . Following section 3.1, an equilibrium is given by vectors of quantities  $q$  and times  $t$  such that: The number of trips is consistent with demand,  $q = q(p, t)$ , where  $q(p, t)$  is given by our estimated demand model,

Figure 8: Estimates of congestion elasticities by region



*Notes:* This figure shows the congestion elasticities,  $\beta$ , for regression equation 17 across different regions in the city of Chicago. We follow the geographical division from [Community Areas in Chicago](#).

*Source:* Authors' calculation using mobile phone geo-location records, public transit trip-level data from the Chicago Transportation Authority, and Google Maps trips queries.

and times are consistent with the transportation technology,  $t = T(p, t)$ , where  $T(p, t)$  is given by the congestion and driver movement models that we estimate. What does this mean concretely in our case? The planner proposes public transit prices and a fleet size. Given those, travellers take as given travel times, prices, and wait times for buses and for hire vehicle when making their travel decisions. These decisions generate flows across community areas and therefore congestion at a given edge, which is a function of all the flows that pass through an edge. Road-based travel times have to be consistent with edge level congestion. Wait times for public transit have to be consistent with the fleet size of public transit and wait times for for hire vehicles with the driver movements. Driver movements in turn have to be consistent with consumer demand and road based travel

times. Given the linkages of markets through the congestion technology and driver movements, an equilibrium therefore requires that the plans of agents in thousands of markets have to be consistent with each other.

To find an equilibrium, we write the previous two conditions as a fixed point problem with prices  $\mathbf{p}$  and capacities  $\mathbf{k}$ . Let  $f^{\mathbf{p}, \mathbf{k}}(\mathbf{q}) = q(\mathbf{p}, T(\mathbf{q}, \mathbf{k}))$ . The problem of finding an equilibrium is equivalent to finding a fixed point of  $f^{\mathbf{p}, \mathbf{k}}$ . In words,  $q^*$  is an equilibrium if people would demand  $q^*$  trips when (wait and travel) times are those that arise when  $q^*$  trips take place. Naive algorithms to find fixed points—such as simple or damped fixed point iteration—often diverge. We thus rewrite the fixed point problem as a root-finding problem (i.e., finding a root  $f^{\mathbf{p}, \mathbf{k}}(\mathbf{q}) - \mathbf{q} = 0$ ) and we use a limited-memory version of Broyden’s method to solve it. Appendix D describes in detail the algorithm that we use.

Once we find an equilibrium, we can compute all quantities that go into the city government’s objective function. To solve the social planner’s problem—which involves a budget constraint—we follow the augmented Lagrangian method (Wright, 2006), where we iteratively maximize problems that approximate the Lagrangian of the main problem until convergence. We provide further details in Appendix E.

## 5 Preliminary counterfactuals

We now present several counterfactual exercises. They are based on a previous version of our model that differs in several ways from the main model described above. First, it is based on a simpler demand model without variation in car ownership and without a nest for the outside option. Although that model captures well price elasticities and the value of time, it leads to unrealistic substitution patterns. Second, the model is based on a calibrated version of the congestion model. Third, the routes taken by cars and for-hire vehicles do not follow the empirical routes we obtain from Google maps; instead, they

are based on a minimum-distance approximation that does not capture detours to avoid traffic. Our goal for the next few months is to run a version of these counterfactuals that uses our final model.

In what follows, we characterize the optimal prices and service levels of a budget-constrained planner following a city-wide version of problem 5 that has access to different policy levers. We focus on the following counterfactual scenarios:

1. *Transit*: The planner can only set optimal public transit prices and service levels.
2. *Congestion*: The planner sets congestion prices on cars.
3. *Transit + congestion*: The planner sets both policy instruments simultaneously.
4. *Social planner*: The planner can choose all policy levers: transit prices and service levels, congestion prices, and the price of taxis and ride-hailing.

The results obtained under the optimal schemes across different policy interventions can be found on Table 5. Results are reported relative to the status quo (the first column).

Observe that under all scenarios, the planner would like to decrease public transit prices by roughly 90% and service levels by 20%-30%. Setting this type of policy alone, column (2) of Table 5, leads to a welfare gain of \$2M per week, most of which comes from consumer surplus as now travelers face substantially lower prices. The budget constraint is binding in this scenario.

On the other hand, we find that optimal congestion prices are of the order of \$0.25/km (column (3)). When the planner uses this policy instrument alone, we see welfare gains of \$6M per week, which mostly come from reductions in externalities. This tax collection leads to a budget surplus. However, consumer surplus decreases by roughly \$39M per week. Even after redistributing tax revenue back to consumers, the loss in consumer surplus still is of \$1M per week.

Table 5: Counterfactual Results

		Status Quo	Transit	Congestion	Transit + Congestion	Social Planner
$\Delta$ Price	Uber	0%	0%	0%	0%	-5.19%
	Bus	0%	-88.4%	0%	-83.1%	-93.4%
	Train	0%	-91%	0%	-86.6%	-96.4%
Car Surcharge (\$/km)		0	0	0.264	0.261	0.260
$\Delta$ Fleet	Bus	0%	-25.8%	0%	-24.3%	-23.3%
	Train	0%	-30.1%	0%	-29.1%	-27.8%
$\Delta$ Welfare (\$M/week)		0	2.03	6.08	7.95	8.29
$\Delta$ CS (\$M/week)		0	1.46	-38.7	-36.7	-34.9
City Profit (\$M/week)		-13.0	-13.1	-12.8	-12.9	-13.9
Taxes (\$M/week)		0	0	37.8	37.4	37.2
Externality (\$M/week)		26.1	25.5	21.1	20.5	20.6

*Notes:* This table presents the changes in prices, service levels, and welfare relative to the status quo (column 1) across different counterfactual scenarios. Column 2 only changes public transit prices and service levels. Column 3 changes congestion prices. Column 4 combines both. Column 5 also set optima ride-hailing prices.

*Sources:* Authors' calculations.

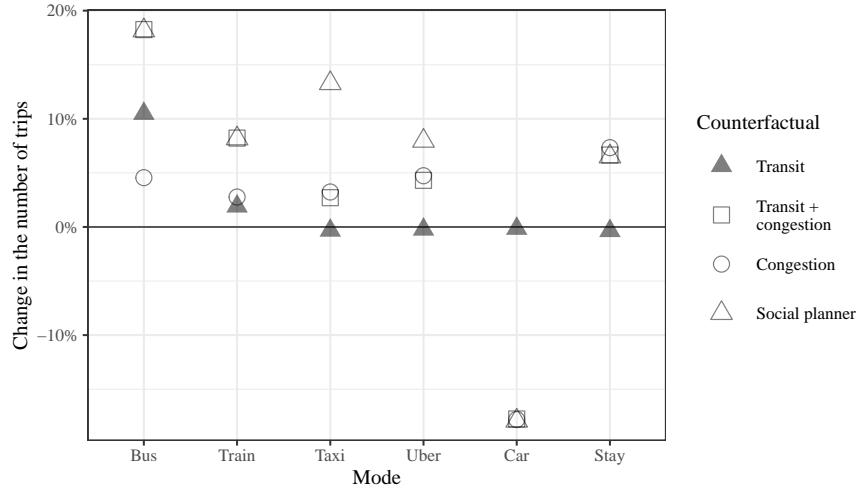
In the third scenario, where the planner combines both policies, column (4), we obtain similar qualitative results in terms of prices and service levels. In this case, the gains in consumer surplus add up to \$8M per week.

Finally, we also allow the planner to set ride-hailing prices, column (5). In such case, on top of similar adjustments compared to the previous counterfactuals, the planner would like to decrease ride-hailing prices by 5%. One might have expected the optimal price to be higher than in the status quo, in order to correct for the externalities caused by ride-hailing. However, platforms charge a substantial markup on these services, which acts like a Pigouvian tax.

## 5.1 Substitution patterns

We can also measure how travelers adjust their behavior in response to the policy interventions, as reported on Figure 9. Not surprisingly, when public transit prices decrease, travelers switch to these alternatives. Similarly, when the planner imposes congestion prices, there are fewer car trips and travelers start substituting toward staying put. Finally, when the planner also decreases ride-hailing prices, we see an increase in the market share of this mode.

Figure 9: Substitution patterns across different policies



*Notes:* This figure presents percentual changes in mode market shares relative to the status quo under optimal policies across different counterfactual scenarios.

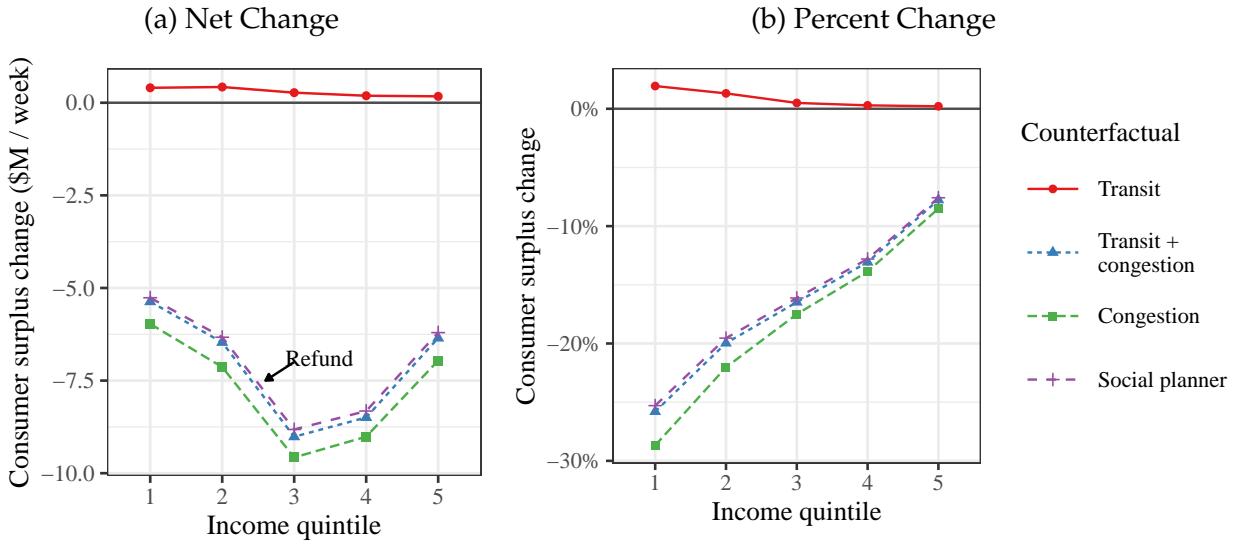
*Sources:* Authors' calculations.

## 5.2 Distributional Effects

We can also quantify how consumer surplus changes across different regions of the income distribution. To do so, we compute consumer surplus across income quintiles. Results can be seen on Figure 10. Observe that intervening only in public transit lead to gains in consumer surplus and that these gains are progressively distributed. On the other hand, as soon as the planner sets optimal congestion prices, we observe large losses

in consumer surplus. The pattern of net losses is U-shaped, which is related to how car market shares change by income as shown in panel (a) of Figure 2. However, when we measure those losses as a percentage of consumer surplus, we find that these policies are highly regressive. To summarize, our counterfactual analysis reveals that the policies that deliver the largest efficiency gains are also the ones that are the most regressive in terms of consumer surplus.

Figure 10: Change in Consumer Surplus Across Income Quintiles



*Notes:* This figure presents changes in consumer surplus relative to the status quo under optimal policies across different counterfactual scenarios. We calculate consumer surplus for five groups according to traveler's income quintiles. Panel (a) displays net changes in dollar values across the five quintiles of the income distribution. Panel (b) displays changes as a percent.

*Sources:* Authors' calculations.

### 5.3 Decomposition

In this section, we decompose the change in consumer surplus and in environmental externalities attributed to different channels. The change in consumer surplus is a product of two forces: the direct change in prices and the indirect effect in time due to changes in mode choices. The change in the environmental externalities is also due to two channels: the change in services levels (capacity) and the change in traveler's mode choices (sub-

stitution). Table 6 shows how each of these channels contribute to the overall aggregate effects across different scenarios.

Table 6: Decomposition of Consumer Surplus and Environmental Externalities

		Status quo	Transit	Congestion	Transit + Congestion	Social Planner
	Total	0	1.428	-38.654	-35.758	-32.658
$\Delta$ CS (\$M/week)	Price	0	5.850	-43.532	-36.512	-32.806
	Time	0	-4.422	4.878	0.753	0.148
	Capacity	0	-4.446	0	-4.124	-3.662
	Substitution	0	0.024	4.878	4.877	3.809
	Total	0	-0.627	-4.994	-5.516	-5.273
$\Delta$ Externality (\$M/week)	Capacity	0	-0.531	0	-0.504	-0.435
	Substitution	0	-0.095	-4.994	-5.012	-4.839
$\Delta$ Average speed (km/h)		0%	0.001%	0.029%	0.029%	0.028%

*Notes:* This table represent the change in consumer surplus and environmental externalities attributed to different channels. Changes in consumer surplus (first row) are divided into changes in prices (second row) and times (third row). Changes in times are a product in changes in fleet size (fourth row) and substitution of consumers across modes (fifth row). Total changes in externalities (sixth row) are decomposed into changes in fleet size (seventh row) and substitution across consumer (eighth row).

*Sources:* Authors' calculations.

Focusing on the counterfactual where the planner only sets public transit prices and service levels, we see that consumers face two opposing effects. On the one hand, lower prices means an increase in consumer surplus of \$7M per week. On the other hand, lower service levels increases the overall travel times and, in turn, decreases consumer surplus by \$4M per week. In terms of externalities, most of the reduction accrues through the reduction in service levels and fewer vehicles running throughout the city.

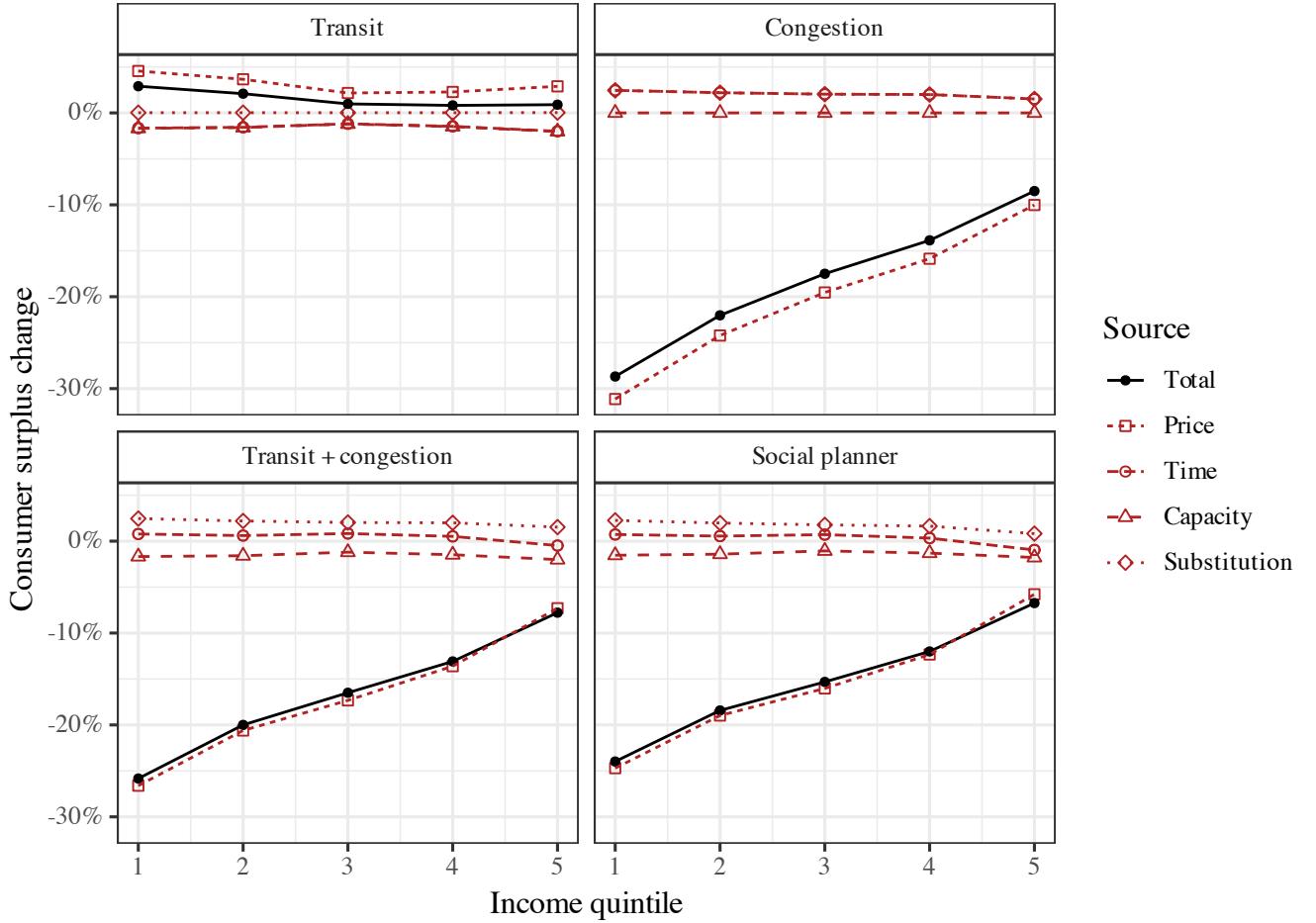
When the planner only set congestion prices, we see a large reduction in consumer surplus of \$43M per week. The reason is that consumers face an increase of prices of 25% for the most common mode of transportation, namely private cars. Because due to this increase in prices, consumers stop traveling by car, traveling speed goes up by 3%, which translates into lower overall travel times and an increase in consumer surplus of \$5M per week.

Simultaneously setting public transit prices and levels as well as congestion prices can be viewed as the combination of the previous two cases. However, in this case we have two opposing effects for both prices and travel times that net each other out in the aggregate overall results.

Finally, when the planner sets all prices and reduces ride-hailing prices by 5%, we see some interactions of these policies accruing through to channels. First, consumer surplus increases by \$3.5M per week relative to the previous scenario. However, as travelers start substituting toward ride-hail, travel speeds decrease and overall travel time increases, which partially undoes the effect of congestion of car surcharges.

Next, we zoom in on how consumer surplus changes across the income distribution. The results, in percentage terms, can be seen in Figure 11. Observe that the absolute effect of prices is larger for lower income consumers, as they are the ones who are the most sensitive to prices.

Figure 11: Caption



*Notes:* These graphs presents changes in consumer surplus across income quintiles for four different counterfactual scenarios scenarios. Each of the lines represent the change in consumer surplus from each of the channels that affect traveler's utility.

*Sources:* Authors' calculations.

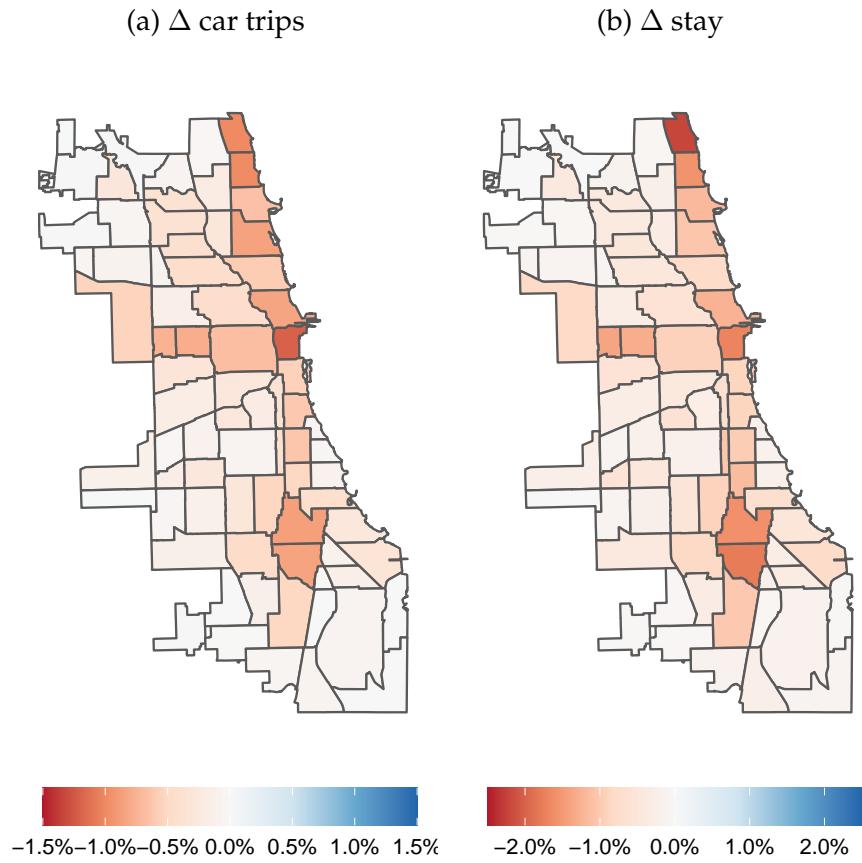
## 5.4 Heterogeneity Across Space

We can also quantify how the behavior of traveler's is changing across space due to the introduction of these policies.

The results for optimal public transit prices and service levels are shown on Figure 12. As the planner reduces public transit prices by 90%, travelers start switching to public transit alternatives and away from car trips. Moreover, this policy also induces consumers

to travel more as the share in the number of consumers who stay put goes down. Not surprisingly, these changes are more pronounced along the train lines.

Figure 12: Change in Market Shares Across Space for Public Transit Policies



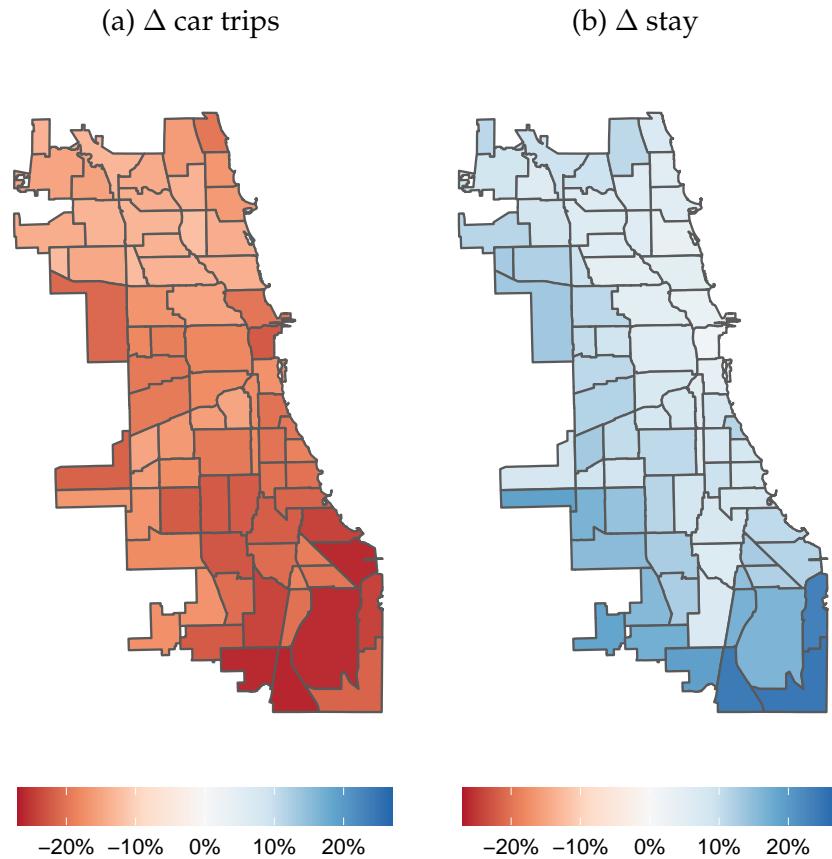
*Notes:* This map presents changes in modes shares under optimal public transit prices and service levels relative to the status quo. Panel (a) displays changes in the share of car trips. Panel (b) displays changes in the share of traveler's who stay put.

*Sources:* Authors' calculations.

Figure 13 shows changes in car trips and staying put for congestion prices. In this case, we observe that the number of car trips decreases substantially across all areas in Chicago, but especially so for the South Side of Chicago, where low-income residents are concentrated, since those are the travelers that are most affected by the increase in car trip prices. Finally, we observe that most of the substitution is toward staying put—and it is especially large in the South Side. Therefore, this policy induces less travel in the city as

a whole.

Figure 13: Change in Market Shares Across Space for Congestion Prices



Notes: This map presents changes in modes shares under optimal congestion prices relative to the status quo. Panel (a) displays changes in the share of car trips. Panel (b) displays changes in the share of traveler's who stay put.

Sources: Authors' calculations.

## 5.5 Robustness: social cost of carbon

For our main counterfactual analysis, we account for externalities that arise from a social cost of carbon dioxide of \$200 per ton. This value is roughly in line with the latest value of \$190 reported by the United States Environmental Protection Agency as of November

2022 plus an upward adjustment due to the social cost of local pollutants.<sup>39,40</sup>

Unfortunately, there is no consensus on the right value of the social cost of carbon, as it is very sensitive to climate model parameters as well as the discount factor. To deal with this issue, we evaluate optimal policies under different values of the social cost of carbon. We do so for a high social cost of carbon of \$300 and \$100 per ton of carbon dioxide. We report results in Table 7 and Table 8, respectively. While optimal public transit levels and services are of a similar magnitude as in our baseline scenario, optimal congestion prices are more sensitive to the social cost of carbon. In particular, in the high cost scenario congestion prices increase by more than 20% to roughly \$0.32/km, and in the low cost scenario they go down by 20%, to roughly \$0.20/km. These relatively small responses arise from the fact that a large part of the costs of private transit to society is traffic congestion.

## 6 Conclusions

In this paper, we measure the welfare effects of urban transportation policies, and we explore how a budget-constrained planner should set optimal prices and public transit service levels. We first start by showing in a theoretical model that on top of congestion and environmental externalities, budget considerations introduce additional sources of inefficiency as the planner starts behaving somewhat like a monopolist.

Then, we move on to empirically quantify welfare effects. We focus on the city of Chicago due to data quality and the importance of its public transit system, the second largest in the nation. To do so, we construct a novel data set that contains the universe of trip flows across locations, hours, and modes. Using this dataset, we estimate a model of

---

<sup>39</sup> See [The Social Cost of Greenhouse Gases](#).

<sup>40</sup> The social cost of local pollutants is only a small fraction of the total environmental externalities. Holland *et al.* (2016) estimate a social cost of local pollutants of \$0.48 per gallon of gasoline for Chicago. The latest number by the EPA that we cite above implies a social cost of carbon of \$1.74 per gallon of gasoline.

Table 7: Counterfactual Results: High Social Cost of Carbon

		Status Quo	Transit	Congestion	Transit + Congestion	Social Planner
$\Delta$ Price	Uber	0%	0%	0%	0%	-0.827%
	Bus	0%	-92.9%	0%	-98.6%	-93.9%
	Train	0%	-95.6%	0%	-96.2%	-95.8%
Car Surcharge (\$/km)		0	0	0.321	0.316	0.316
$\Delta$ Frequency	Bus	0%	-25.2%	0%	-22.9%	-23.7%
	Train	0%	-32.6%	0%	-31%	-29.7%
$\Delta$ Welfare (\$M/week)		0	2.36	8.81	11.0	11.2
$\Delta$ CS (\$M/week)		0	1.37	-45.6	-43.0	-42.6
City Profit (\$M/week)		-13.0	-13.0	-12.8	-13.6	-13.5
Taxes (\$M/week)		0	0	43.3	42.8	42.7
Externality (\$M/week)		39.1	38.1	30.3	29.4	29.5

*Notes:* This table presents the changes in prices, service levels, and welfare relative to the status quo (column 1) across different counterfactual scenarios for a social cost of carbon of \$300 per ton of  $CO_2$ . Column 2 only changes public transit prices and service levels. Column 3 changes congestion prices. Column 4 combines both. Column 5 also set optima ride-hailing prices.

*Sources:* Authors' calculations.

mode demand that allows us to quantify how travelers substitute across modes. We also estimate a congestion technology that allows us to measure how vehicle flows map onto travel times.

Finally, we use our estimates to run counterfactual simulations and characterize optimal policies for a battery of counterfactual scenarios. We find that congestion prices deliver the largest welfare gains but they also come at the cost of being highly regressive.

Table 8: Counterfactual Results: Low Social Cost of Carbon

		Status Quo	Transit	Congestion	Transit + Congestion	Social Planner
$\Delta$ Price	Uber	0%	0%	0%	0%	-4.1%
	Bus	0%	-82.2%	0%	-89%	-101%
	Train	0%	-83.7%	0%	-96.8%	-106%
Car Surcharge (\$/km)		0	0	0.208	0.204	0.202
$\Delta$ Frequency	Bus	0%	-25.7%	0%	-23.4%	-24.5%
	Train	0%	-27.2%	0%	-24.8%	-23.5%
$\Delta$ Welfare (\$M/week)		0	1.73	3.83	5.45	5.88
$\Delta$ CS (\$M/week)		0	1.47	-31.3	-28.0	-26.1
City Profit (\$M/week)		-13.0	-13.0	-12.9	-14.2	-15.1
Taxes (\$M/week)		0	0	31.6	31.1	30.8
Externality (\$M/week)		13.0	12.8	11.0	10.8	10.9

*Notes:* This table presents the changes in prices, service levels, and welfare relative to the status quo (column 1) across different counterfactual scenarios for a social cost of carbon of \$100 per ton of  $CO_2$ . Column 2 only changes public transit prices and service levels. Column 3 changes congestion prices. Column 4 combines both. Column 5 also set optima ride-hailing prices.

*Sources:* Authors' calculations.

## References

- AKBAR, P. and DURANTON, G. (2017). Measuring the cost of congestion in highly congested city: Bogotá. 5, 6, 40
- AKBAR, P. A., COUTURE, V., DURANTON, G. and STOREYGARD, A. (2018). *Mobility and congestion in urban India*. Tech. rep., National Bureau of Economic Research. 5, 6, 40, 66
- ALLEN, T. and ARKOLAKIS, C. (2019). *The welfare effects of transportation infrastructure improvements*. Tech. rep., National Bureau of Economic Research. 8
- ANDERSON, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *American Economic Review*, 104 (9), 2763–2796. 2
- ARNOTT, R. (1996). Taxi travel should be subsidized. *Journal of Urban Economics*, 40 (3), 316–333. 6, 24

- , DE PALMA, A. and LINDSEY, R. (1990). Economics of a bottleneck. *Journal of urban economics*, **27** (1), 111–130. 6
- , — and — (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, pp. 161–179. 6
- BARWICK, P. J., LI, S., WAXMAN, A. R., WU, J. and XIA, T. (2021). *Efficiency and equity impacts of urban transportation policies with equilibrium sorting*. Tech. rep., National Bureau of Economic Research. 8
- BERRY, S., LEVINSOHN, J. and PAKES, A. (1995). Automobile prices in market equilibrium. *Econometrica*, **63** (4), 841–890. 4, 6, 12, 32, 76
- BERRY, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pp. 242–262. 6, 12
- BRANCACCIO, G., KALOUPTSIDI, M. and PAPAGEORGIOU, T. (2020). Geography, transportation, and endogenous trade costs. *Econometrica*, **88** (2), 657–691. 6
- BRINKMAN, J. and LIN, J. (2022). Freeway revolts! the quality of life effects of highways. *The Review of Economics and Statistics*, pp. 1–45. 8
- BROOKS, L. and LISCOW, Z. D. (2021). Infrastructure costs. Available at SSRN 3428675. 2
- BUCHHOLZ, N. (2020). Spatial equilibrium, search frictions and efficient regulation in the taxi industry. *Working paper*. 6
- BYRD, R. H., NOCEDAL, J. and SCHNABEL, R. B. (1994). Representations of quasi-newton matrices and their use in limited memory methods. *Mathematical Programming*, **63** (1), 129–156. 77
- CASTILLO, J. C. (2020). Who benefits from surge pricing? *Working paper*. 7
- , KNOEPFLE, D. and WEYL, G. (2018). Surge pricing solves the wild goose chase. *Working Paper*. 74
- COUTURE, V., DURANTON, G. and TURNER, M. A. (2018). Speed. *Review of Economics and Statistics*, **100** (4), 725–739. 5, 6, 40

- DURRMEYER, I. and MARTÍNEZ, N. (2022). The welfare consequences of urban traffic regulations. 7
- FAJGELBAUM, P. D. and SCHAAL, E. (2020). Optimal transport networks in spatial equilibrium. *Econometrica*, **88** (4), 1411–1452. 8
- FRECHETTE, G. R., LIZZERI, A. and SALZ, T. (2019). Frictions in a competitive, regulated market: Evidence from taxis. *American Economic Review*, **109** (8), 2954–92. 6
- FUCHS, S. and WONG, W. F. (2022). *Multimodal Transport Networks*. Federal Reserve Bank of Boston. 7
- HALL, J. D. (2018). Pareto improvements from lexus lanes: The effects of pricing a portion of the lanes on congested highways. *Journal of Public Economics*, **158**, 113–125. 6
- HERZOG, I. (2021). *The city-wide effects of tolling downtown drivers: Evidence from london's congestion charge*. Tech. rep., Working Paper. 8
- HOLLAND, S. P., MANSUR, E. T., MULLER, N. Z. and YATES, A. J. (2016). Are there environmental benefits from driving electric vehicles? the importance of local factors. *American Economic Review*, **106** (12), 3700–3729. 30, 54, 79
- KREINDLER, G. (2023). *Peak-hour road congestion pricing: Experimental evidence and equilibrium implications*. Tech. rep., National Bureau of Economic Research. 5, 7
- , GADUH, A., GRAFF, T., HANNA, R. and OLKEN, B. A. (2023). *Optimal Public Transportation Networks: Evidence from the World's Largest Bus Rapid Transit System in Jakarta*. Tech. rep., National Bureau of Economic Research. 7
- LAGOS, R. (2003). An analysis of the market for taxicab rides in new york city. *International Economic Review*, **44** (2), 423–434. 6, 74
- PARRY, I. W. H. and SMALL, K. A. (2009). Should urban transit subsidies be reduced? *American Economic Review*, **99** (3), 700–724. 6
- PETRIN, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of political Economy*, **110** (4), 705–729. 4

- ROSAIA, N. (2020). *Competing platforms and transport equilibrium: Evidence from New York City*. Tech. rep., mimeo, Harvard University. 7
- ROSENBLUM, J. L., ZHAO, J., ARCAYA, M., STEIL, J. and ZEGRAS, P. C. (2020). How low-income transit riders in boston respond to discounted fares: A randomized controlled evaluation. In *2020 APPAM Fall Research Conference*, APPAM. 12
- SEVEREN, C. (2021). Commuting, labor, and housing market effects of mass transportation: Welfare and identification. *Review of Economics and Statistics*, pp. 1–99. 7
- SMALL, K. A. (1982). The scheduling of consumer activities: work trips. *The American Economic Review*, **72** (3), 467–479. 6
- , WINSTON, C. and YAN, J. (2005). Uncovering the distribution of motorists' preferences for travel time and reliability. *Econometrica*, **73** (4), 1367–1382. 6
- SPENCE, A. M. (1975). Monopoly, quality, and regulation. *The Bell Journal of Economics*, pp. 417–429. 3
- TSIVANIDIS, N. (2019). Evaluating the impact of urban transit infrastructure: Evidence from bogota's transmilenio. 7
- WEYL, E. G. (2010). A Price Theory of Multi-Sided Platforms. *American Economic Review*, **100** (4), 1642–1672. 21
- WRIGHT, J. N. S. J. (2006). Numerical optimization. 43, 78, 79
- YANG, J., PUREVJAV, A.-O. and LI, S. (2020). The marginal cost of traffic congestion and road pricing: evidence from a natural experiment in beijing. *American Economic Journal: Economic Policy*, **12** (1), 418–53. 6
- ZHAO, J., RAHBEE, A. and WILSON, N. H. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. *Computer-Aided Civil and Infrastructure Engineering*, **22** (5), 376–387. 9, 68

# A Data Appendix

## A.1 Cellphone location records

This subsection details how we construct our sample of trips based on the raw cellphone data. We define a “ping” as one individual entry of the raw data. Each ping contains a timestamp, latitude, longitude, and a device identifier.

### A.1.1 Data filtering

We start by subsetting cellphone pings to a rectangle around the city of Chicago (i.e., latitude between 41.11512 and 42.494693, longitude between -88.706994 and -87.527174) for the months of July 2019 - March 2020 and for the months of June 2020 - August 2020.

Next, using the cellphone device identifier, the timestamp and geolocation of each ping, we calculate the time between two consecutive pings as well as the geodesic distance. These distances allow us to obtain the speed between consecutive pings. We then filter out “noisy” pings by using distance, time, and speed variables. In particular, we remove pings that are moving at an excessive speed since these pings are likely to be GPS “jumps” resulting from noise in the measurement of the GPS coordinates of the device.<sup>41</sup> We also drop “isolated” pings since they are not helpful for identifying whether people are moving. Additionally, we only keep pings belonging to a “stream” of pings.<sup>42</sup> We define a stream of pings as a sequence of pings for the same cellphone identifier such that a ping always has another ping within the next 15 minutes and within 1,000 meters. We drop streams with less than 3 pings. Finally, we aggregate pings to the minute of the day by taking the average location and timestamp across pings within each minute for a given

---

<sup>41</sup> 40 meters per second, i.e. about 145 kilometers per hour

<sup>42</sup> In particular, we only keep pings that satisfy the following two conditions: (i) no more than ten minutes to either the next or the previous ping, (ii) no more than 5,000 meters to either the next or the previous ping.

cellphone identifier. In what follows, we focus on the remaining filtered pings aggregated at the minute level.

### A.1.2 Defining movements, stays, and trips

We identify two consecutive (aggregated) pings as a “movement” for a given cellphone identifier if their distance is at least 50 meters or if their implied speed is at least 3 meters per second (6.7 miles per hour or 10.8 kilometers per hour). We then define a “stay” as a sequence of two or more successive pings with no movement.

Finally, we take all streams of pings and define trips as being a stream (i) with movement, (ii) that starts with a stay, and (iii) that ends with a stay. We remove all trips with a total geodesic trip distance between the starting and ending point below 0.25 miles (about 400 meters).

### A.1.3 Estimation of home locations and traveler’s income

This subsection details how we assign a home location and an income level to each individual cellphone identifier.

**Home locations** We start by assigning all cellphone pings to census block polygons for the subset of pings within Chicago during our sample period.<sup>43</sup> Next, we focus on pings during night hours, defined as between 10pm and 8am, when individuals are more likely to be at home.

Using this subset of pings, we attribute a score system for each hour between 10pm and 8am. Specifically, regardless of the number of pings, scores are assigned as follows:

- A value of 10 to all census blocks that were pinged between 1 am and 5 am.

---

<sup>43</sup> See Appendix A.1 for the sample restrictions.

- A value of 5 to all census blocks that were pinged between 11 pm and 1am or between 5 am and 7 am.
- A value of 2 to all census blocks that were pinged between 10pm and 11pm, or between 7am and 8am.

The basic idea is to assign a higher score to blocks where the cellphone owner is more likely to be at home. Finally, we sum the scores across all census blocks for each cellphone ID - month combination and keep the census block with highest score. If this highest-score census block appears on at least 3 or more separate nights during the month, we assign it as the cellphone's home census block for that month. Otherwise, we consider the cellphone as having an unknown home location, which we believe captures occasional Chicago visitors such as tourists. Throughout the text, we refer to these devices as *visitors*. Figure 14 plots the share of visitors by origin locations. We see that for trips done by visitors, the most common origin locations are the city center (center right), both airports (top left and center left), as well as Hyde Park the neighborhood home to the University of Chicago (right, south of the center).

**Individual cellphone's income** For all cellphones with an assigned home location, we impute their income by using the census tract median household income.<sup>44</sup>

**Market-level income quintile shares** Next, for each market, we estimate traveler's income distribution.<sup>45</sup> First, we take median income by tracts and divide tracts according to quintiles.<sup>46</sup> Next, assign an income quintile to each device according to their home location. Since we can follow how devices travel across space and over time, for each market

---

<sup>44</sup>We compute the census-tract median income percentile using the 2010 Census data.

<sup>45</sup>Recall, a market is defined as an (origin Community Area, destination Community Area, hour of the week)-tuple.

<sup>46</sup>For 2010, income quintiles are defined using the following cut-offs: 34,875, 46,261, 60,590 and 85,762 U.S. Dollars.

Share of trips made by visitors, by origin

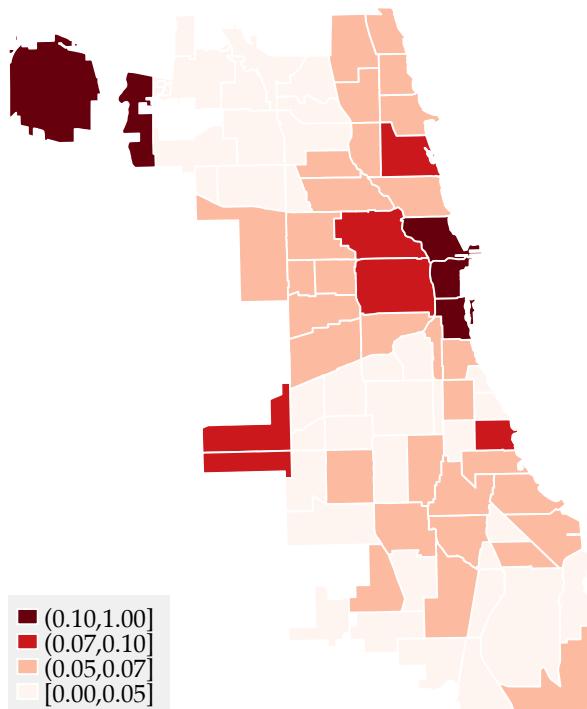


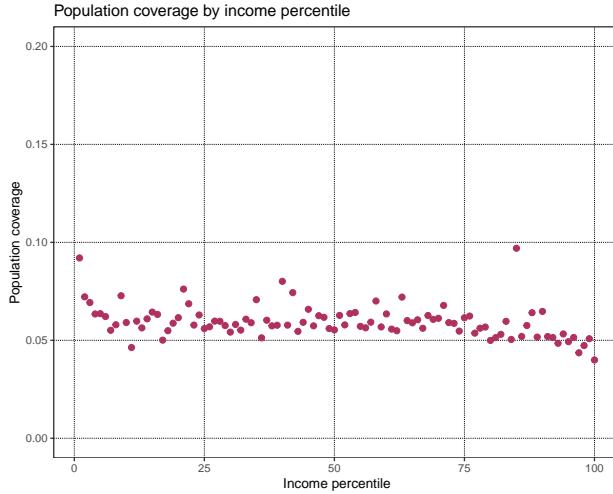
Figure 14: Share of visitors by origin location

*Notes:* This figure shows the share of trips at the origin Community Area level made by visitors. In our cellphone trips data, each market (origin-destination-hour triple) has a share of trips made by visitors. To construct the shares displayed in the figure, we take the weighted average of the share of trips made by visitors across destinations and hours of the week, for each origin Community Area, using inside market size (number of cellphone trips per market) as weight.

*Source:* Authors' calculations using mobile phone location records.

we can measure the quintile from each traveler departing from its destination. Finally, for each market, we construct shares of traveler's income quintile. For markets with less than 5 trips, we impute market-level income shares using the underlying distribution of census tract-level income for the origin community area of that market.

Figure 15: Representativeness across income groups



*Notes:* This figure plots a binscatter of the fraction of the population in each income percentile covered by the mobile phone data. We define the census-tract specific population coverage as the ratio between (i) the number of cellphones whose home location is assigned to that specific census tract, and (ii) the number of inhabitants of the census tract according to the 2010 Census data. Income percentiles are defined by the census tract median household income.

*Source:* Authors' calculation using mobile phone geo-location records and 2010 Census data.

#### A.1.4 Cell-phone data validation

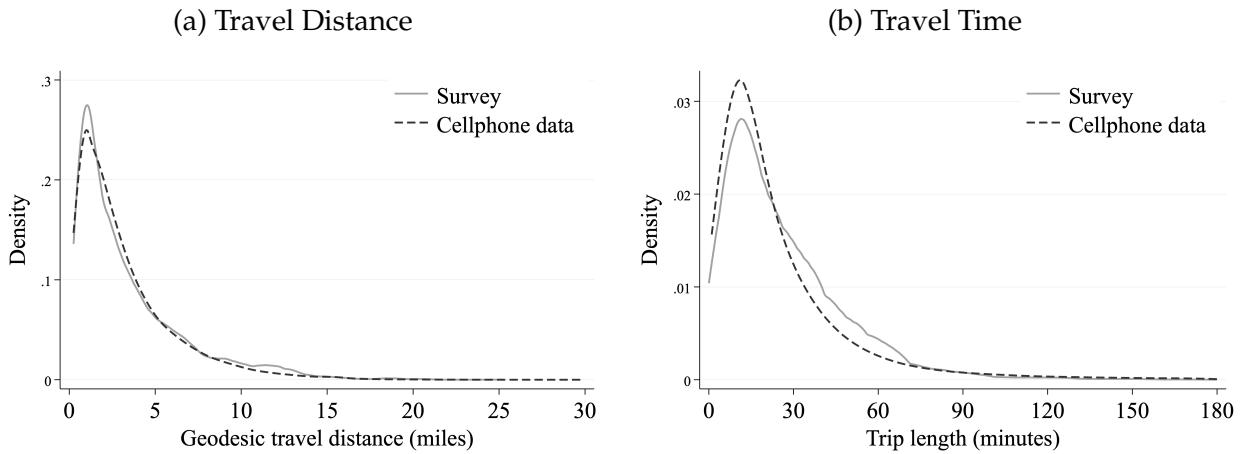
In this section, we provide evidence that the cell-phone geo-location data is representative of the population and that it measures traveling patterns with precision.

First, we measure the coverage of our cell-phone devices across the income distribution. To do so, for each census tract, we count the number of devices whose home location falls within the tract boundaries. We divide those counts by the census population. Figure 15 displays the share of the tract population covered by the cell-phone data. We order tracts by income percentiles. We see that our coverage is fairly constant and hovers 5%

across all percentiles of the income distribution.<sup>47</sup> We take this as evidence that our cell-phone location records cover a representative sample of the population.

Next, we show that the cell-phone location records can accurately represent travelling patterns. To do so, we plot the travel (geodesic) distance distribution for the universe of trips for both the cell-phone data and the survey data. The two distributions, plotted in Figure 16, present a high degree of overlap and similarity. We conclude that movements measured using cell-phone location records with the procedure described in Section A.1 are a good representations of the travel patterns in the city of Chicago.

Figure 16: Travel Distance Histograms



*Notes:* This figure plots kernel densities of the distribution of travel distances (Panel a) and travel times (Panel b) using trips in the survey data as well as in the cell-phone data. Our level of observation is a trip. Trips in the cell-phone data are constructed following the steps in Appendix A.1. Trips in the survey data do not include walking, biking or multi-modal trips.

*Source:* Authors' calculations using survey data from the Chicago Transportation Authority and mobile phone location records.

---

<sup>47</sup> A perfect representative sample should show a straight horizontal line.

## A.2 Travel times, routes, and schedules

### A.2.1 Travel times and routes

Similar to Akbar *et al.* (2018), we query and geocode trips using Google Maps. For each mode of transportation, we query 30,796,848 counterfactual trips and obtain their distance, duration and path.<sup>48</sup> Importantly, we can measure trip duration for the same origin-destination tuple over the time of the weekday (or weekend) and how this varies with traffic conditions. Moreover, using the detailed “steps” of the public transit Google Maps queries, we obtain walk times from the origin latitude/longitude to the “best” train or bus station.<sup>49</sup>

We also obtain Google Maps data on train trip times by querying Google Maps three times for each pair of train stations in Chicago. These times represented three broad time categories: weekday peak, weekday non-peak, and weekend. In particular, the first query requested a trip time of 8am on Wednesday July 6th, 2022, the second query requested a trip time of 11am on Wednesday July 6th, 2022, and the third query requested a trip time of 11am on Saturday July 9th 2022.

### A.2.2 Public transit schedules

We obtain historical GTFS data from [Open Mobility Data](#). These data contain bus and train schedules for December 2019 through February 2020.

---

<sup>48</sup>One trip for each (origin census tract, destination census tract, hour of day, weekend dummy) combination. We use all the 801 Chicago census tracts boundaries for the year 2010 from the [Chicago Data City portal website](#).

<sup>49</sup>The “best” bus or train station is not necessarily the closest one, depending on the destination and/or the time of the day.

## A.3 Constructing Market Shares

The market shares are constructed using five main sources: (1) Taxi and TNP trips data from the City of Chicago, (2) Google Maps data, (3) cellphone trips data, (4) historical GTFS data containing public transit route schedules, and (5) Chicago public transit data from the MIT Transit Lab and the CTA.

### A.3.1 Raw data processing

#### Taxi and Transportation Network Provider (TNP) data

We obtain trip times, distances, and origin-destination census tracts for both Taxi and Transportation Network Provider (TNP) trips from the [City of Chicago's Data Portal](#).<sup>50</sup>

#### Cellphone trips data

We construct cellphone trips from cellphone pings using the procedure detailed in Appendix A.1. This procedure results in a trip-level dataset. Since our cellphone data only captures a portion of the total trips, we adjust for this by assigning an inflation factor to each trip. To account for varying rates of unobserved trips across different city areas, we allow inflation factors to vary by the neighborhood of the trip's origin.<sup>51</sup> Specifically, we calibrate these factors to ensure that the proportion of car trips beginning in each neighborhood in our dataset matches the corresponding proportion in the Chicago Metropolitan Agency for Planning (CMAP) Household Travel Survey.<sup>52</sup>

#### Public transit data

We obtain individual public transit trips via a partnership between the MIT Transit Lab

---

<sup>50</sup>For privacy reasons, during periods of the day and for locations with very few trips, only the origin and/or destination Community Area of a trip is reported. See [this page](#) for a discussion of the approach to privacy in this data set.

<sup>51</sup>Each neighborhood is a group of about 8-9 community areas. The exact make-up of neighborhoods can be found on [Wikipedia](#).

<sup>52</sup>Source: [My Daily Travel survey \(website\)](#)

and the CTA. Specifically, we use the near-universe of CTA bus and train trips within the city of Chicago. This data notably excludes trips taken via the Metra, which is a suburban rail system operating in and around Chicago. Metra is managed by a different agency, the Regional Transportation Authority. Each observations corresponds to a passenger swiping in to access the bus or the train station. For buses, we observe the specific bus stop, bus line, and boarding time. For trains, we observe the station and swiping time. Drop-off locations are given to us and imputed following Zhao *et al.* (2007). Unfortunately, these data are missing travel times for train trips, and so we are forced to impute these travel times. To do so, we first match each train trip to the historical GTFS data. To compute the match for a given train trip we first find all scheduled trips between the origin and destination stops of that trip. We then take the match to be the scheduled trip whose boarding time is closest to the observed boarding time. We then take the scheduled travel time as the travel time. This matching process enables us to compute travel times for close to 90% of train trips.

For trips that have no matches in the schedule data, we impute travel times using Google Maps data.<sup>53</sup> In particular, we first assign each trip one of three time categorizations: weekend (if Saturday or Sunday), peak weekday (if between 5-9:59am or 2-6:59pm on a weekday), or non-peak weekday (otherwise). We then take the time to be the travel time of the matching train trip from the Google Maps data.

As an additional cleaning step, we drop any duplicate trips along with any trips that have missing coordinates, mode information, or time (for bus trips). We also compute travel distances for each trip. We use the Harversine formula to compute distances, with radius equal to 6371.0088, which is the mean radius of Earth in km. For bus trips, we compute the travel distances as the Manhattan distance between the boarding and alighting coordinates, while for train trips we compute the travel distances as the Euclidean

---

<sup>53</sup>Manual inspection suggests these trips typically involve an unobserved transfer between two lines.

distance between the boarding and alighting coordinates. Finally, since there could be measurement error, we account for trips that may be missing in our dataset by assigning each trip an inflation factor. This inflation factor is computed at the day-mode level such that

$$infl_{dm} T_{dm} = R_{dm},$$

where  $dm$  indexes the day-mode,  $T$  is the total number of observed trips, and  $R$  is the observed ridership, which we obtain from the [City of Chicago's Data Portal](#).

## A.4 Market Share Calculations

We first append together the transit, TNP, taxi, and cellphone trips data. We incorporate walk times to bus/train stations from the Google Maps data. We drop any trips that have a negative trip time, trip time exceeding 6 hours, negative prices, or missing values for origin, destination, distance, duration, mode, trip time, or price.

We calculate market shares at the (origin Community Area, destination Community Area, hour-of-the-week) level using a two-step process. First, we aggregate trips at the (origin Community Area, destination Community Area, hour-of-the-week) level. In particular, we compute trip counts by mode as

$$ntrips_{odtj} = \sum_{k \in \mathcal{I}(odtj)} infl_k,$$

where  $\mathcal{I}(\cdot)$  gives the observations corresponding to the given origin, destination, time, mode tuple and  $infl_k$  gives the inflation factor for observation  $k$ . Travel times, travel distances, and prices were computed analogously as weighted averages (with inflation factors as weights) of the times and prices for each individual trip in  $\mathcal{I}(odtj)$ . We let the number of car trips be the residual after subtracting public transit, taxi, TNP, and shared

trips from the cellphone trips.<sup>54</sup> Car prices are computed as 0.6374 U.S. Dollars per trip mile, which is AAA's estimate of per mile driving costs for an average 2020 model.<sup>55</sup>

We then get the trip counts, prices, times, and distances for a given mode  $j$  for each origin Community Area, destination Community Area, and hour-of-the-week by averaging across the corresponding (origin Community Area, destination Community Area, hour-of-the-week)-level observations. Finally, the market size is taken to be twice the maximum number of total trips, where the max is taken over the (origin Community Area, destination Community Area, hour-of-the-week)-level observations.

To compute market shares, we need to take a stance on the size of the market, which captured how many people could be traveling at a given moment in time. We set the market size to two times the total number of trips observed in a given market. We then probe how sensitive our counterfactuals are to this assumption.

## B Proofs

### B.1 Proof of proposition 1

*Proof.* The Lagrangian for the social planner's problem is:

$$U(q, T(q, k)) - C(q, k) - E(q, k) - \lambda \left( C(q, k) - \sum_j p_j(q, T(q, k))q_j - B \right)$$

The first order condition for  $q_j$  is:

$$\frac{\partial U}{\partial q_j} + \sum_k \frac{\partial U}{\partial t_k} \frac{\partial T_k}{\partial q_j} - \frac{\partial C}{\partial q_j} - \frac{\partial E}{\partial q_j} + \lambda \left( p_j + \sum_k q_k \frac{dp_k}{dq_j} - \frac{\partial C}{\partial q_j} \right) = 0$$

---

<sup>54</sup>If the residual is negative we assume that there are no car trips.

<sup>55</sup>Source: [AAA brochure "Your driving costs"](#).

The first order condition for  $k_j$  is:

$$\sum_k \frac{\partial U}{\partial t_k} \frac{\partial T_k}{\partial k_j} - \frac{\partial C}{\partial k_j} - \frac{\partial E}{\partial k_j} + \lambda \left( \sum_k q_k \frac{dp_k}{dk_j} - \frac{\partial C}{\partial k_j} \right) = 0$$

We now show that  $\partial U / \partial q_j = p_j$ . Let  $\partial\Theta_j(p, t)$  be the boundary between  $\Theta_j(p, t)$  and  $\Theta_0(p, t)$ , and let  $\partial\Theta_{jk}(p, t)$  be the boundary between  $\Theta_j(p, t)$  and  $\Theta_k(p, t)$ . Gross utility can be written as

$$U(q, t) = \int_{\partial\Theta_j(q, t)} u(t_j, \theta) f(\theta) d\theta$$

Using the Leibniz integral rule, we get that

$$\frac{\partial}{\partial q_j} U(q, t) = \sum_k \int_{\partial\Theta_k(q, t)} u_k(t_k, \theta) e_k(\theta) f(\theta) d\theta + \sum_{kl} \int_{\partial\Theta_{kl}(q, t)} u_k(t_j, \theta) e_k(\theta) f(\theta) d\theta$$

(the interior term from the integral rule is zero because  $t$  is fixed), where  $e_k(\theta)$  denotes by how much  $\Theta_k(q, t)$  expands at  $\theta$  as  $q_j$  increases. This can also be written as:

$$\sum_k \int_{\partial\Theta_k(q, t)} u_k(t_k, \theta) e_k(\theta) f(\theta) d\theta + \sum_{k, l > k} \int_{\partial\Theta_{kl}(q, t)} (u_k(t_k, \theta) - u_l(t_l, \theta)) e_l(\theta) f(\theta) d\theta$$

Since agents in the boundaries are indifferent between two choices,  $u_k(t_k, \theta) = p_k$  for the first sum and  $u_k(t_k, \theta) - u_l(t_l, \theta) = p_k - p_l$  for the second sum. After substituting and rearranging terms, our main expression can be written as:

$$\sum_k p_k \left( \int_{\partial\Theta_k(q, t)} e_k(\theta) f(\theta) d\theta + \sum_l \int_{\partial\Theta_{kl}(q, t)} e_k(\theta) f(\theta) d\theta \right)$$

The term in parentheses is how much  $\Theta_k(p, t)$  expands in total into all other regions, so it is equal to  $\partial q_k / \partial q_j$ . It is thus equal to 1 for  $j$  and 0 for  $k \neq j$ . We can thus conclude that  $\partial / \partial q_j U(q, t) = p_j$ . Substituting  $\partial U / \partial q_j = p_j$ ,  $dp_k / dq_j = \partial p_k / \partial q_j + \partial p_k / \partial t_j \partial T_k / \partial q_j$ , and

$dp_k/dk_j = \partial p_k/\partial t_j \partial T_k/\partial k_j$  on the first order conditions and a few steps of algebra yield

$$p_j = \frac{\partial C}{\partial q_j} - \frac{1}{1+\lambda} \sum_k \frac{\partial U}{\partial T_k} \frac{\partial T_k}{\partial q_j} + \frac{1}{1+\lambda} \frac{\partial E}{\partial q_j} - \frac{\lambda}{1+\lambda} \sum_k \left( q_k \frac{\partial p_k}{\partial q_j} + q_k \sum_{k'} \frac{\partial p_k}{\partial T_{k'}} \frac{\partial T_{k'}}{\partial q_j} \right), \quad (18)$$

$$\frac{\partial C}{\partial k_j} = \frac{1}{1+\lambda} \sum_k \frac{\partial U}{\partial T_k} \frac{\partial T_k}{\partial k_j} - \frac{1}{1+\lambda} \frac{\partial E}{\partial k_j} + \frac{\lambda}{1+\lambda} \sum_k q_k \sum_{k'} \frac{\partial p_k}{\partial T_{k'}} \frac{\partial T_{k'}}{\partial k_j}. \quad (19)$$

The final term in equation (18) can be written as

$$\sum_k q_k \sum_{k'} \frac{\partial p_k}{\partial T_{k'}} \frac{\partial T_{k'}}{\partial q_j} = \sum_k \left( \sum_{k'} q_{k'} \frac{\partial p_{k'}}{\partial T_k} \right) \frac{\partial T_k}{\partial q_j}$$

We now show that  $\sum_{k'} q_{k'} \left( \frac{\partial p_{k'}}{\partial T_k} \right)$  can be written as a weighted average of the change in gross utility among marginal travelers. By Leibniz's integration rule,

$$\frac{\partial q_j}{\partial p_j} = -W_j(p, t) - \sum_{k \neq j} W_{jk}(p, t)$$

where

$$W_j(p, t) = \int_{\partial\Theta_j(p, t)} v_j(\theta) \cdot \hat{n}_j(p, t, \theta) f(\theta) d\theta$$

and

$$W_{jk}(p, t) = \int_{\partial\Theta_{jk}(p, t)} v_{jk}(\theta) \cdot \hat{n}_{jk}(p, t, \theta) f(\theta) d\theta$$

are integrals over boundaries  $\partial\Theta_j(p, t)$  and  $\partial\Theta_{jk}(p, t)$ , where the integrand is the density of riders that are willing to switch modes in response to an increase in utility. That density is given by the dot product of  $v_{jk}(\theta)$ , the vector whose elements are the inverse of  $\partial u_j/\partial\theta - \partial u_k/\partial\theta$  (and the inverse of  $\partial u_j/\partial\theta$  for  $v_j(\theta)$ ), and  $\hat{n}_x(p, t, \theta)$ , the unit normal component of the boundary  $\partial\Theta_x(p, t)$  at  $\theta$ . Also by Leibniz's integration rule,

$$\frac{\partial q_j}{\partial t_j} = V_j(p, t) + \sum_{k \neq j} V_{jk}(p, t)$$

where

$$V_j(p, t) = \int_{\partial\Theta_j(p, t)} \frac{\partial u_j(t, \theta)}{\partial t} v_j(\theta) \cdot \hat{n}_j(p, t, \theta) f(\theta) d\theta$$

and

$$V_{jk}(p, t) = \int_{\partial\Theta_{jk}(p, t)} \frac{\partial u_j(t, \theta)}{\partial t} v_{jk}(\theta) \cdot \hat{n}_{jk}(p, t, \theta) f(\theta) d\theta$$

These are similar integrals as before, only that the integrand is the density of riders that are willing to switch modes in response to an increase in pickup times.

Let  $\Lambda(p, t)$  be the matrix whose  $j$ -th diagonal element is  $W_j(p, t) + \sum_k W_{jk}(p, t)$ , and whose non-diagonal element  $(j, k)$  is  $V_{jk}(p, t)$ . Let  $\Sigma(p, t)$  be a matrix that is defined similarly, but whose elements arise from  $V_{jk}(p, t)$  instead of  $W_{jk}(p, t)$ . Then, by the implicit function theorem, the matrix of derivatives  $\partial p_j / \partial t_k$  is given by

$$\Psi(p, t) = \Lambda^{-1}(p, t) \Sigma(p, t)$$

From the definition of  $W$  and  $V$ , it is clear that this is a weighted average of  $\partial u_j(t, \theta) / \partial t$  over sets of marginal agents. We define

$$\tilde{u}_j^T \equiv \sum_k q_k \frac{\partial p_k}{\partial t_j} = \sum_k q_k \Psi_{kj}$$

the sum of such weighted averages, weighted by the number of agents in each market.

Substituting  $\tilde{u}_j^T \equiv \sum_k q_k \partial p_k / \partial t_j$  and the definitions of  $\epsilon_{jk}^{T,q}, \epsilon_{jk}^{T,k}, \bar{u}_j^T, \Omega_{kj}, \tilde{C}_j^q$ , and  $\tilde{E}_j^q$  into expressions (18) and (19) yields expression (7) and

$$p_j = \underbrace{\frac{\partial C}{\partial q_j}}_{\text{Mg. cost of a trip}} + \underbrace{\frac{\partial E}{\partial q_j}}_{\text{Mg. env. ext. of a trip}} - \sum_k \underbrace{\bar{u}_k^T}_{\text{Mg. utility of time}} \underbrace{\frac{\partial T_k}{\partial q_j}}_{\text{Congestion effect}} + \frac{\lambda}{1+\lambda} \left\{ \underbrace{\sum_{k \in J} q_k \Omega_{kj}}_{\text{"Monopolist" markup}} - \underbrace{\frac{\partial E}{\partial q_j}}_{\text{Spence distortion}} + \underbrace{\sum_k (\tilde{u}_k^T - \bar{u}_k^T) \frac{\partial T_k}{\partial q_j}}_{\text{Spence distortion}} \right\}. \quad (20)$$

To obtain expression (6), multiply (7) by  $k_j/q_j$  and subtract it from (20).  $\square$

## C Model of Waiting Times for Ride-hailing and Taxis

Consider mode  $j$  (either taxi or ride hailing). Let  $q_{ahj}$  be the number of mode- $j$  trips with origin  $a$  during hour  $h$ , and let  $I_{ahj}$  be the number of drivers working for mode  $j$  that are idle in location  $a$  during this time. We assume that there is a matching technology such that the expected waiting time for riders before their trip starts is given by

$$T_{ahj}^W = A_{aj}^W I_{ahj}^{-\phi_j} \quad (21)$$

$A_{aj}^W$  is a scale factor that measures the overall matching inefficiency for mode  $j$  in location  $a$ . The parameter  $\phi_j$  is an elasticity that determines how quickly waiting times decrease with the number of idle drivers. This flexible specification nests simple models of matching in taxi and ride hailing markets.<sup>56</sup>

To determine the number of idle drivers in every location, we assume that the distribution of drivers across the city arises from a parsimonious model that captures the spatial dynamics of the market. Let  $L_{hj}$  be the total number of drivers working for mode  $j$  during hour  $h$ . The number of drivers that are busy is given by  $B_{hj} = \sum_{od} T_{odh}^T q_{odhj}$ , where  $T_{odh}^T$  are the travel times from the traffic congestion model, and  $q_{odhj}$  is the number of people taking mode  $j$  from  $o$  to  $d$ . Thus, the total number of idle drivers is given by  $I_{hj} = L_{hj} - B_{hj}$ .

The probability that an idle driver is in location  $a$  during hour  $h$  is given by

$$\frac{\exp(\mu_a + \sum_b B_{ab} F_{hb})}{\sum_{a'} \exp(\mu_{a'} + \sum_b B_{a'b} F_{hb})}, \quad (22)$$

where  $F_{ha} = \sum_b (q_{bahj} - q_{abhj})$  represents the net inflow of mode- $j$  trips into location  $a$ ,  $B_{ab} = \lambda r_{ab}^{-\rho}$  is a factor for each pair of locations  $a$  and  $b$  that decays with the distance  $r_{ab}$

---

<sup>56</sup>In the taxi model in Lagos (2003), for instance,  $\phi_j = 1$ . In the simplest ride-hailing model described by Castillo *et al.* (2018),  $\phi_j = 1/n$  in  $n$ -dimensional space.

between them. This probability takes the form of a multinomial logit model that depends on two terms. First,  $\mu_a$ , which are fixed effects that capture the fact that drivers tend to work in certain locations of the city. Second,  $\sum_b B_{ab}F_b$ , which models the extent to which idle drivers are more likely to be located near areas where net inflows are high. The latter term is driven by two opposing forces: areas with high net inflows of trips are also areas with a high net inflow of drivers, so they tend to have many idle drivers; however, these areas have an oversupply of drivers so earnings go down, and drivers will try to move away from them.

Putting all these pieces together, the number of idle drivers in every location is given by

$$I_{ahj} = (L_{hj} - B_{hj}) \frac{\exp(\mu_a + \sum_b B_{ab}F_{hb})}{\sum_{a'} \exp(\mu_{a'} + \sum_b B_{a'b}F_{hb})}. \quad (23)$$

This expression, coupled with equation (21), determines the waiting times for taxis and ride hailing.

## C.1 Estimation

We first estimate the parameters  $A_{ahj}^W$  and  $\phi_j$  that map the number of idle drivers into waiting times. Consider community area  $a$ . We make the simple assumption that the  $I_{ahj}$  available drivers are distributed homogeneously across  $a$  and that the pickup time conditional on distance is  $t(x) = M_{aj}x^{c_j}$ . That implies that the pickup time has a distribution whose expectation is<sup>57</sup>

$$T_{ahj}^W = M_{aj}\Gamma\left(1 + \frac{c_j}{2}\right)\left(\frac{1}{\pi I_{ahj}}\right)^{\frac{c_j}{2}}. \quad (24)$$

This takes the desired form  $A_{aj}^W I_{ahj}^{-\phi_j}$ , where  $A_{aj}^W = M_{aj}\Gamma\left(1 + \frac{c_j}{2}\right)\left(\frac{1}{\pi}\right)^{\frac{c_j}{2}}$  and  $\phi_j = \frac{c_j}{2}$ .

---

<sup>57</sup>With a density of idle drivers  $I_{ahj}$ , the pdf of the distance to the nearest driver is given by  $2\pi x I_{ahj} e^{-\pi I_{ahj} x^2}$ , a Weibull distribution with parameters  $k = 2$  and  $\lambda = 1/\pi L$ . We integrate the travel time over this density to obtain equation (24).

We obtain  $M_{aj}$  and  $c_j$  from a regression of the log of the travel time on the log of the travel distance for all car trips in our Google Maps dataset originating and ending within the same community area, where we include community area fixed effects. The main coefficient from this regression is  $c_j = 0.73$  (s.e.=), and  $M_{aj}$  are the fixed effects that we estimate. Based on those results, we can conclude that  $\phi_j = \frac{c_j}{2} = 0.365$ , and we back out  $A_{ahj}^W$  from the expression above.

We then move on to estimate the parameters of the driver location model ( $\mu_a$ ,  $\lambda$ , and  $\rho$ ). We do not observe drivers directly, but we use Uber data for the average waiting time at the community area by hour of the week level—i.e.,  $T_{ahj}^W$ . Inverting equation (24) allows us to compute all values of  $I_{ahj}$ . We can then estimate ( $\mu_a$ ,  $\lambda$ , and  $\rho$ ) by maximum likelihood, based on equation (22). Maximizing this likelihood is not a simple problem since the vector of  $\mu_a$  has 77 elements. We simplify the task by splitting the problem into an inner loop that computes the optimal vector of  $\mu_a$  given  $\lambda$  and  $\rho$  using a contraction mapping, as in Berry *et al.* (1995), and an outer loop that maximizes over  $\lambda$  and  $\rho$ . Table 9 presents our main estimates.

Table 9: Driver Movement Estimates

	Coefficient	Standard Error
$\lambda$	0.04185	0.00007
$\rho$	-0.13119	0.01013

*Notes:* Standard errors are computed using a sandwich estimator.

## D Equilibrium computation

Given prices and capacities  $(\mathbf{p}, \mathbf{k})$ , an equilibrium is a set  $(\mathbf{q}, \mathbf{t})$  that satisfies  $\mathbf{q} = q(\mathbf{p}, \mathbf{t})$  and  $\mathbf{t} = T(\mathbf{q}, \mathbf{k})$ , the demand and transportation technology equations. By plugging in the technology equation in the demand equation, the equilibrium condition can alternatively

be written as  $\mathbf{q} = q(\mathbf{p}, T(\mathbf{q}, \mathbf{k}))$ . Thus, if we define the function  $f^{\mathbf{p}, \mathbf{k}}(\mathbf{q}) = q(\mathbf{p}, T(\mathbf{q}, \mathbf{k}))$ , an equilibrium is characterized by a vector of flows  $\mathbf{q}^{\mathbf{p}, \mathbf{k}}$  that is a fixed point of  $f^{\mathbf{p}, \mathbf{k}}$ . After finding a fixed point, the equilibrium vector of travel times can then be computed as  $\mathbf{t}^{\mathbf{p}, \mathbf{k}} = T(\mathbf{q}^{\mathbf{p}, \mathbf{k}}, \mathbf{k})$ .

One naive way to search for an equilibrium is by fixed point iteration. However, this procedure typically diverges. We, instead, find a root of  $f^{\mathbf{p}, \mathbf{k}}(\mathbf{q}) - \mathbf{q} = 0$  using a limited-memory version of Broyden's method. We use the actual vector of trips in the data as the initial point, and we use an identity matrix as the initial guess for the Jacobian. The full Broyden algorithm is:

---

**Algorithm 1** Equilibrium computation using Broyden's method

---

Set initial value of trips  $\mathbf{q}$ .

Compute initial times  $\mathbf{t} = T(\mathbf{q}, \mathbf{k})$ .

Compute deviation  $\mathbf{d} = q(\mathbf{p}, \mathbf{t}) - \mathbf{q}$ .

Set new vector of trips  $\mathbf{q}' = \mathbf{q} + \gamma \mathbf{d}$  for a small step size  $\gamma > 0$ .

Compute new vector of times  $\mathbf{t}' = T(\mathbf{q}', \mathbf{k})$ .

Compute deviation  $\mathbf{d}' = q(\mathbf{p}, \mathbf{t}') - \mathbf{q}'$ .

Set initial approximation to inverse Jacobian  $\mathbf{A} = \mathbf{I}$ .

**while**  $\|\mathbf{d}'\| > \text{tolerance}$  **do**

Define differences  $\Delta \mathbf{q} = \mathbf{q}' - \mathbf{q}$  and  $\Delta \mathbf{d} = \mathbf{d}' - \mathbf{d}$ .

Update vectors of trips  $\mathbf{q} = \mathbf{q}'$  and deviation  $\mathbf{d} = \mathbf{d}'$ .

Compute new approximation to inverse Jacobian  $\mathbf{A} = \mathbf{A} + \frac{\Delta \mathbf{q} - \mathbf{A} \Delta \mathbf{d}}{\Delta \mathbf{q}^T \mathbf{A} \Delta \mathbf{d}} \Delta \mathbf{q}^T \mathbf{A}$ .

Compute new vector of trips  $\mathbf{q}' = \mathbf{q} - \mathbf{A} \mathbf{d}$ .

Compute new vector of times  $\mathbf{t}' = T(\mathbf{q}', \mathbf{k})$ .

Compute new deviation  $\mathbf{d}' = q(\mathbf{p}, \mathbf{t}') - \mathbf{q}'$ .

**end**

---

We make two adjustments to the above algorithm. First, we compute the approximation to the inverse Jacobian  $\mathbf{A}$  with the limited-memory approach in Byrd *et al.* (1994). Second, when we compute the new vector  $\mathbf{q}'$  we often obtain an infeasible vector of trips (the number of Uber or taxi drivers is not enough to satisfy demand). Whenever that is the case, we iteratively update  $\mathbf{q}' = \mathbf{q} + 1/2(\mathbf{q}' - \mathbf{q})$  until we get back to a feasible value.

## E Optimization

Based on the procedure we describe in Appendix D, which we use to compute an equilibrium, we can compute welfare  $W(\mathbf{p}, \mathbf{t})$  and the net revenue of the city  $\Pi(\mathbf{p}, \mathbf{t})$ . The welfare maximization problem is

$$\max_{\mathbf{p}, \mathbf{t}} W(\mathbf{p}, \mathbf{t}). \quad (25)$$

We solve this problem in two steps. First, we approximate the solution with a Nelder-Mead optimizer, starting from the true prices and capacities and stopping after 100 iterations. Second, we run a quasi-Newton method starting from the Nelder-Mead optimum. This method differs from Newton's method in two ways, both of which greatly reduce the computational cost of our procedure. First, to avoid computing the Hessian of the objective function, we use the BFGS approximation (Wright, 2006), which only requires computing the gradient. Second, we approximate the gradient with central differences. Every time we compute a finite difference, instead of fully running Broyden's method until convergence to an equilibrium, we only take a few steps (typically three) starting from the central point, which allows us to obtain a good approximation to the gradient at a small fraction of the computational cost.

The social planner's problem is

$$\max_{\mathbf{p}, \mathbf{t}} W(\mathbf{p}, \mathbf{t}) \quad \text{s.t.} \quad \Pi(\mathbf{p}, \mathbf{t}) = -B, \quad (26)$$

where  $B$  is the city's transportation budget. To solve this problem, we use the augmented Lagrangian method. We iteratively solve the following approximation to the Lagrangian:

$$\max_{\mathbf{p}, \mathbf{t}} W(\mathbf{p}, \mathbf{t}) - \lambda_n (\Pi(\mathbf{p}, \mathbf{t}) + B) + \mu_n (\Pi(\mathbf{p}, \mathbf{t}) + B)^2. \quad (27)$$

We initialize this iterative procedure by setting  $\mu_0 = 10^{-6}$  and  $\lambda_0 = 0$ . In every step

$n$  we use the method we described above to maximize the objective function, and we set  $\mu_{n+1} = 2\mu_n$  and  $\lambda_{n+1} = \lambda_n + \mu_n(\Pi^n + B)$ , where  $\Pi^n$  is the net revenue at the  $n$ -th step optimum. In this algorithm  $\lambda_n$  converges to the Lagrange multiplier that results in the budget constraint being satisfied with equality (Wright, 2006). This means that (27) converges to the true Lagrangian plus an extra penalty for deviations from the budget constraint—and thus, the sequence of solutions converge to the solution of (26).

## F Additional Parameters and Assumptions

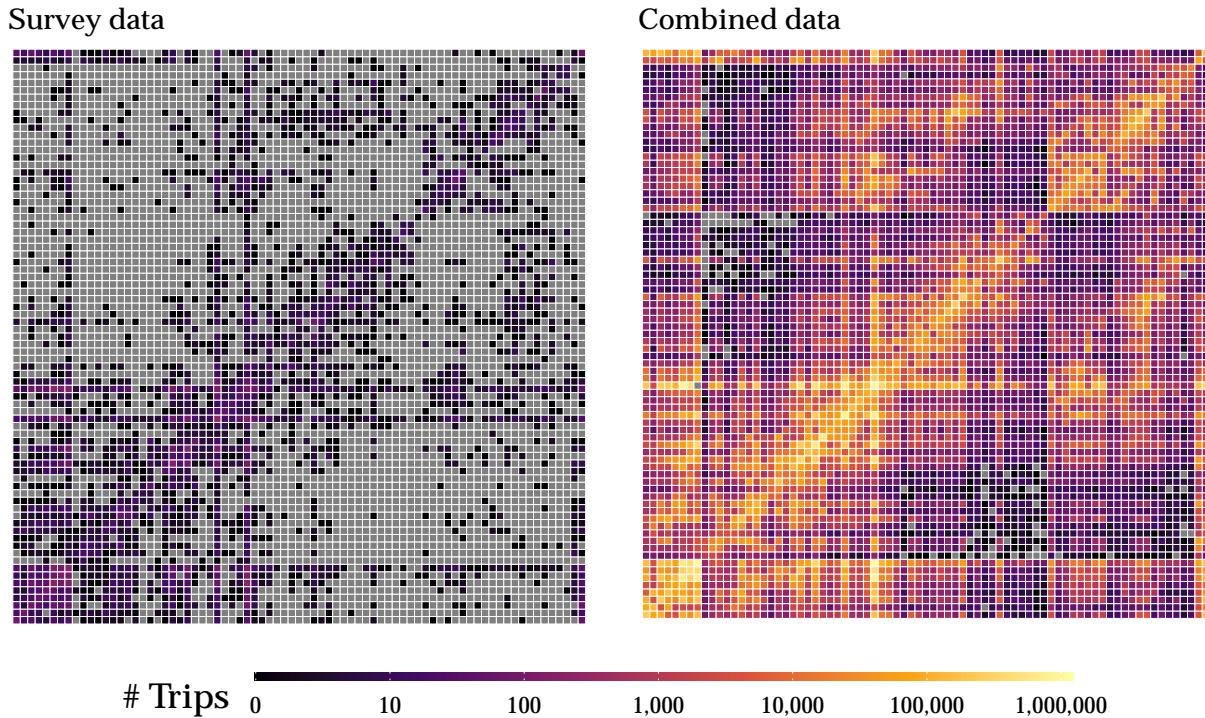
- Social cost of carbon: \$190 per ton
- Cost of local pollutants: 44.93 cents per gallon for gasoline, 41.32 cents per gallon for diesel, based on Holland *et al.* (2016).
- Mg. cost of cars: \$0.35 / km
- Uber/taxi wages: \$10 / h
- Mg. cost of buses (including labor): \$7.60 / km. This is the sum of
  - Capital costs of \$900,000 per bus, each one of which lasts 250,000 miles
  - Fuel costs of \$3.5 per gallon with a fuel efficiency of 3.38 mpg
  - Wages of \$33 per hour at 20 km/h, multiplied by a factor of 2 to account for benefits and the wages of supervisors, schedulers, etc.
  - \$2.51 per km of maintenance costs.
- Mg. cost of trains (including labor): \$12.90 / km. This is the sum of:
  - Capital costs of \$11M per train that lasts 2M miles
  - Energy costs twice the fuel costs of a bus

- Wages of \$33 per hour at 20 km/h, multiplied by a factor of 2 to account for benefits and the wages of supervisors, schedulers, etc.
  - \$5 per km of maintenance costs.
- Number of passengers per private car: 1.5
  - Number of ride-hailing passengers per trip : 1.3

## G Additional Graphs and Figures

### G.1 Sparsity of Our Data versus Survey

Figure 17: Combined vs. Survey Data: Flows Across Community Areas



*Notes:* These figures show the number of trips from every origin community area to every destination community area in our combined data (left panel) and in the survey data (right panel). Each row represents an origin community area and each column represents a destination community area. Grey points represent empty cells.

*Source:* Authors' calculation using mobile phone geo-location records and 2019 Chicago Transit Survey data.

## H Demand Robustness

Table 10: Demand Estimation Robustness

	Dependent variable: Log Market Shares		
	(1) BLP, GMM	(2) BLP, GMM	(3) BLP, GMM
$\alpha_T$	-2.66 (0.017)	-2.53 (0.022)	-0.37 (0.055)
$\alpha_{std(T)}$	.	-1.08 (0.058)	.
$\alpha_{T^2}$	.	.	-1.47 (0.038)
$\alpha_p$	-0.100 (0.001)	-0.099 (0.002)	-0.103 (0.001)
$\alpha_{py}$	0.01 (0.0005)	0.02 (0.0009)	0.01 (0.0006)
$\rho$	0.24 (0.015)	0.28 (0.015)	0.32 (0.019)
Mode FE	✓	✓	✓
Market FE	✓	✓	✓
Car Ownership	✓	✓	✓
Constrained $\alpha_p^i$	✓	✓	✓
Nest	✓	✓	✓
Avg. VOT	34.801	120.547	17.472
VOT (Top Quintile)	58.11	363.63	24.82
VOT (Bot. Quintile)	20.97	17.11	12.51
Avg. Price Elast.	-0.506	-0.390	-0.597
Avg. Time Elast.	-0.751	-0.715	-0.472
Num. Markets	152,317	87,244	152,317
N	339,017	202,083	339,017